Distortion Prediction for Video Quality Optimization over Packet Switched Networks

Andrea Vesco, Enrico Masala, Carlo Novara
Control and Computer Engineering Department
Politecnico di Torino — 10129 Torino, Italy
Email: andrea.vesco|masala|carlo.novara@polito.it

***Abstract*—Scheduling techniques are often deployed at the network edge to maximize the quality of the video communication while satisfying a given constraint on the maximum high priority usable bandwidth. For the case of video communications, the importance of each packet, in terms of the distortion that would be caused by its loss, can be used to decide which packets should be prioritized in order to maximize the expected video quality. However, the resulting performance strictly depends on the length of the video stream segment considered, at each time instant, by the distortion-aware scheduling algorithms. This work focuses on improving the performance of those algorithms by predicting the characteristics of the near-future part of video streams, exploiting the short and long term dependencies in packet distortion values. This approach could be particularly valuable in live video scenarios where accessing the future part of the video would imply the insertion of an additional delay. Several distortion prediction models are derived and their performance evaluated with actual scheduling algorithms. Results show that the best predictions are provided by neural network models, yielding significant improvements of the video communication quality.**

***Index Terms*—Multimedia Communication, Scheduling, Distortion Prediction, Neural Networks.**

I. Introduction

Multimedia traffic is an increasingly important share of today's Internet traffic. Many techniques have been proposed to provide successful multimedia communications over the Internet despite the lack of quality of service guarantees. Some of them focus on introducing error resilience in the compressed bitstream at the application level, whereas others address the issue of improving the performance of the transmission channel by using, *e.g.*, error correcting codes or retransmissions. In both cases, error concealment techniques can be further employed at the decoder to recover damaged or missing areas of the video stream.

In the past years, the approach to optimize the performance of video communications employing error control techniques was mainly heuristic and generally based on *a priori* characteristics of the video stream, such as video coding dependencies between frames. Recently, however, a new class of algorithms considering the short-term characteristics of the multimedia data, and in particular the rate and the distortion of each coding unit, have been proposed. The rate-distortion optimization [1] approach has been shown to provide significant improvements in the video communication. In this context, the current research direction is to include as many aspects of the communication system as possible in the rate-distortion optimization framework, ranging from physical, link and network level parameters, *e.g.*, in [2], to application level parameters, sometimes yielding to high-complexity optimization algorithms.

However, a common characteristic of all the rate-distortion optimization strategies, regardless of their complexity, is their need to compute the expected distortion at the receiver for each of the possible transmission policy in order to choose the most promising one to improve the quality of the communication as perceived by the end users. Thus, efficient methods are needed to quickly and reliably estimate the expected distortion for a given transmission policy. Often, the expected distortion is estimated starting from distortion values of single video data computed and stored as side information during compression. The techniques proposed in [3], [4], for instance, are able to precompute distortion values and use them in order to optimize transmission policy. The efficiency of communication optimization algorithms relies on the knowledge of the characteristics of a relatively large portion of the video stream. However in the case of live video, the distortion information is available only for a small part of the video stream. In this scenario, knowledge about the distortion of the future part of the video stream would, of course, be useful in optimizing the transmission strategy for the packets being transmitted. Such information would allow to optimize over a larger portion of the video stream, potentially achieving better performance than a locally optimal solution. This work focuses on the difficult task to estimate the distortion of the near-future part of the video by exploiting the short and long term dependencies in packet distortion values usually present in an encoded video stream, due to the heavy use of differential encoding techniques. In this work, several linear auto regressive with external input (ARX) models and neural network (NN) models for distortion prediction are developed to take into account the short and long term dependencies between packets. Note that these kinds of models are widely used in the literature on time series prediction, *e.g.*, [5], [6], [7], [8].

The distortion models are identified using three video sequences and then validated on five video sequences not used for identification. It is shown that the NN models provide better prediction performance, indicating a nonlinear behavior of the distortion process. Thus, three NN models are finally selected for evaluating the communication performance by

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

simulation. This work addresses the common scenario of video communication over packet switched networks providing different classes of service. The distortion prediction is exploited to optimize the scheduling decisions at the sender side given a constraint on the maximum bandwidth available in the highest quality service class. The performance of the proposed distortion prediction technique is compared to simpler prediction techniques showing its effectiveness in optimizing scheduling decision at the sender side and consequently in improving the quality of the video communication.

The paper is organized as follows. Section II presents the network model whereas the analysis-by-synthesis technique, employed to estimate the packet distortion, and the distortion-aware sender scheduling algorithm are discussed in Section III. The distortion prediction models and simulation results are presented in Section IV and Section V respectively. Finally, the proposed approach and research directions are discussed in Section VI.

II. Network Model

The network model assumes a backbone of a generic topology. Each network node implements the Static Priority scheduling. The scheduler in each node, consists of a number of n prioritized FIFO queues. Each packet entering the network is marked in accordance with a given priority level i and, therefore, is assigned to queue i inside the scheduler. The scheduler always serves the packet at the head of the highest priority non-empty queue.

In the simple, but commonly deployed, case of two classes of service packets marked as high priority receive the *premium* service while low priority packets experience the *best effort* service. The *premium* service ensures no packet loss and upper-bounded end-to-end delay such that the loss/late probability (P_{ll}), *i.e.*, the probability of a packet either being lost or not arriving at the receiver on time for playback, is equal to zero. On the contrary, a low priority packet sent at time t experiences a loss probability P^L_{lost}, independent of t, and a variable end-to-end delay whose distribution is $f_r\left(t \mid not\,lost\right)$. Let t_0 and T_{dd} be the forwarding time of a packet and its delivery deadline at the receiver, respectively. Let also define $\theta = (t_{dd} - t_0)$ as the maximum time the packet can spend in the network. Thus the loss/late probability is given by

$$P^L_{ll} = P^L_{lost} + \left(1 - P^L_{lost}\right) \int_{\theta}^{\infty} f_r\left(t \mid not\,lost\right)\, dt \qquad (1)$$

III. Quality-Oriented Video Transmission

A. Analysis-by-Synthesis Technique

The quality of multimedia communications over packet networks is affected by packet losses. The amount of quality degradation strongly differs depending on the perceptual importance of the lost data. Such quality degradation could be estimated by means of the distortion introduced by the concealment algorithm in case of losses. The potential distortion introduced in the reconstructed stream by the loss of each element could be computed using the analysis-by-synthesis technique, presented in [4]. A possible method to compute the distortion caused by the loss of a certain data unit, *e.g.*, a video packet, referred to as the *distortion of the packet* in the following, is composed of the following steps:

1) Decoding, including concealment, of the bit stream simulating the loss of the element being analyzed (synthesis stage).
2) Quality evaluation, that is, computation of the distortion caused by the loss of the element; the original and the reconstructed picture after concealment are compared using, *e.g.*, Mean Squared Error (MSE).
3) Storage of the distortion value as an indication of the perceptual importance of the analyzed video element.

The previous operations can be implemented by small modifications of the standard encoding process, which already simulates decoding operations since this is needed for motion-compensated prediction. Thus complexity is only due to the concealment algorithm. Moreover, the analysis-by-synthesis technique, as a principle, can be applied to any video coding standard. The values computed with the analysis-by-synthesis algorithm are, of course, dependent on the particular encoder used. Due to the inter-dependencies which typically exist between data units, the simulation of an isolated loss is not completely realistic, particularly for high packet loss rates. Every possible combination of events should ideally be considered, weighted by its probability, and its distortion computed by the analysis-by-synthesis technique, obtaining the expected distortion value. For simplicity, however, we assume that all preceding data units have been correctly received and decoded. Nevertheless, this leads to a useful approximation as demonstrated by some applications of the analysis-by-synthesis approach to MPEG coded video [9], [10]. Moreover, the effectiveness of this approximation is demonstrated by the results provided in Section V. Finally, note that a low-complexity model-based approach, first presented in [11], is employed in this work to estimate the distortion caused by packet losses in future frames due to error propagation. This model allows to estimate the distortion propagation in future frames even in case of low delay video communications scenarios.

B. Distortion-Aware Sender Scheduling

When a video sequence is transmitted throughout the network errors and packet losses contribute to increase the distortion d_0 introduced by the video encoding process. Such distortion is evaluated by computing the Mean Squared Error (MSE), for each frame, between the distorted sequence and the original uncompressed one. The exact expected distortion $E\left[d\right]$ value at the receiver for a given video sequence could be computed as the weighted average of the distortions corresponding to all the possible realizations of the network channel where the weights are the probability of a specific channel realization, as formulated in [12]. However, this procedure is impractical due to its computational complexity, hence a linear approximation is commonly used [3], [4], [12], [13].

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

In other words, it is assumed that if two packets p_1 and p_2 have distortion d_1 and d_2, respectively, their loss causes an overall distortion $d_1 + d_2$. Let Ω be the set of packets in which the video sequence is packetized, α the subset of packets sent as high priority and β the subset of packets sent as low priority such that $\Omega = \alpha \cup \beta$ and $\alpha \cap \beta = 0$. Since the $p_{ll}^{H} = 0$ the expected distortion as seen by the receiver is given by

$$E[d] = d_0 + \sum_{i \in \beta} d_i \cdot p_{ll}^{L}. \tag{2}$$

In order to minimize the expected distortion, *i.e.*, maximize the quality of the video communication, the transmission of a video flow should be performed by sending all packets as high priority in the network. However, a video stream is inherently variable as the amount of bits required to encode each video frame changes significantly. Thus the maximum frame size should be used to determine the allocation, but this might yield to inefficient allocation of premium bandwidth. Therefore, more efficient techniques could be implemented by reducing the allocation of the premium bandwidth and sending video packets in excess as low priority. As a result the expected distortion can be efficiently minimized by sending the packets which would cause the highest distortion in case of loss as high priority, and by sending, at the earliest transmission opportunity, low priority packets in order to minimize their late probability, see Eq. (2). To find a global optimum which minimizes the expected distortion, the scheduling algorithms should run on the entire video sequence, which is obviously not possible. Therefore the maximum length of the video sequence on which the algorithm is run can be, for example, determined on the basis of the maximum delay high priority packets can experience in the network, or a percentile thereof.

The capability of predicting the distortion of packets in future frames would allow to optimize the scheduling without waiting for the future packets to be available. This scheduling scheme, reducing the sender side delay, *i.e.*, the delay introduced by the sender scheduler, is suitable for live video scenarios where minimization of the end-to-end delay is desirable. Assuming that all packets of video frame are available at the sender scheduler every $1/F_r$, where F_r is the video frame rate, the scheduler operates as follows. At time t packets of the video frame i are available and the scheduler immediately sends the most important as high priority. The total amount of packets sent as high priority is controlled such that the sending rate is less than or equal to the dedicated premium bandwidth over a given time interval. Then the scheduler estimates, using a prediction algorithm, the distortion d^{t+1} and the size b^{t+1} of packets of video frame $i+1$. The packet size prediction is based on a "persistent model" aware of the group of picture (GOP) structure. Video coding is, in fact, often characterized by a GOP structure which is repeated periodically. Given the GOP structure, the model works as follows: either $\widehat{b}^{t+1} = b^{t+1-GOP}$ if the next frame is the first of a new GOP or $\widehat{b}^{t+1} = b^{t}$ otherwise. At this time the scheduler processes the remaining packets of video frame i and the predicted packets of frame $i+1$, choosing which packets of the video frame i have to be delayed because they will be sent as high priority after scheduling at time $t+1$ and which packets have to be immediately sent as low priority. This increases the time low priority packets can spend in the network, *i.e.*, decreases the p_{late}^{L}, and consequently further reduces the expected distortion, see Eq. (2). At time $t+1$ packets of the video frame $i+1$ and delayed packets of frame i are available, thus the scheduler immediately sends the most important ones as high priority and the remaining packets of video frame i as low priority. The remaining packets of frame $i+1$ are processed, after the prediction phase, as described above. The expected distortion is negatively affected by the number of packets delayed at time t and not sent as high priority at time $t+1$ as they have less probability to reach the destination on time for decoding than packets immediately sent at time t. This means that the efficiency of the proposed scheduling algorithm is strictly related to the accuracy of the distortion prediction model.

IV. Distortion Prediction Models

Consider a dynamic process described by a discrete-time variable $y^t \in \mathbb{R}$, $t = 1, 2, ...$, where t indicates the time. Suppose that, at each time t, it is of interest to have a prediction of this variable at time $t+1$. Then, a prediction model is a function mapping all the *experimental* and *prior* information on the process available until time t into a prediction $\widehat{y}^{t+1}$ of y^{t+1}, *e.g.*, [5], [6], [7], [8], [14].

In the application considered in this paper, the variable to predict is the distortion d^{t+1}. The sampling time is $1/F_r$ s. The available experimental information at each time t is given by measurements of the distortion d^{τ} itself and of the packet size b^{τ} for $\tau \le t$. As prior information on the process, it is known that d^t and b^t have an approximately periodic behavior with a period of 12 samples given by the GOP structure (one I-type frame followed by eleven P-type frames).

Several distortion prediction models have been developed, aimed at optimizing the performances of the scheduling algorithm presented in the previous section. The models are now described.

A. *Omniscient Model (OM)*

The OM model gives the exact prediction:

$$\widehat{d}^{t+1} = d^{t+1}.$$

Clearly, the OM model cannot be used in on-line applications, since d^{t+1} is not known at time t. However, the OM model gives the best possible performance in both prediction and scheduling, it thus provides an upper bound on the performances which can be achieved by any model. In this work, the OM model is used as a term of comparison for the other models.

B. *All-No-Delayed Model (AN)*

The AN model corresponds to what is called "persistent model" in the literature on time series prediction. It gives a prediction equal to the value of the variable measured at the current time t:

$$\widehat{d}^{t+1} = d^t.$$

Such a trivial model is often used to have a lower bound on the performance which should be achieved by any model. The AN model is interesting for video transmission applications since it leads to a scheduling algorithm such that no packets are delayed.

C. All Delayed Model (AD)

The AD model gives a prediction greater than the value of the variable measured at the current time t:

$$\widehat{d}^{t+1} > d^t.$$

This model is interesting for video transmission applications because, independently of the exact value of $\widehat{d}^{t+1}$, it leads to a scheduling algorithm such that all packets are delayed.

D. Linear Models (ARX)

Several linear ARX models have been identified. These models are of the form

$$\begin{aligned} \widehat{d}^{t+1} &= \beta w^t + \lambda \\ w^t &= [d^t; b^t; d^{t-11}; b^{t-11}] \end{aligned}$$

where $\theta = \{\lambda \in \mathbb{R},\, \beta \in \mathbb{R}^4\}$ is a set of parameters.

The particular choice of the lagged values d^t, b^t, d^{t-11} and b^{t-11} in the regressor w^t has been suggested by the fact that the distortion process has an approximately periodic behavior with a period of 12 samples.

The linear models have been identified on three video sequences called "akiyo", "city" and "coastguard". The identification has been performed by minimizing the following cost function:

$$\begin{aligned} J(\theta) &= -aD(\theta) + bE(\theta) \\ D(\theta) &= \tfrac{1}{N}\sum_{t=1}^{N} n_D^t(\theta) \\ E(\theta) &= \tfrac{1}{N}\sum_{t=1}^{N} n_E^t(\theta) \end{aligned} \tag{3}$$

where:

- N is the length of the video sequence obtained by joining in series the three video sequences.

- $n_D^t(\theta)$ is the total number of packets delayed at time t by the scheduler, based on the distortion prediction.

- $n_E^t(\theta)$ is the number of wrong scheduling decisions at time t, *i.e.*, the number of packets delayed at time t and not sent as high priority at time $t+1$.

- a and b are weights to be chosen according to: $a >> b \Longrightarrow$ large number of right scheduling decisions, $a << b \Longrightarrow$ small number of wrong scheduling decisions.

The motivation for using this cost function is that a prediction model "close" to the OM model is looked for. Indeed, the OM model maximizes the number of delayed packets and minimizes the number of wrong scheduling decisions as well.

The estimate of the parameters has been obtained as

$$\widehat{\theta} = \arg\min_{\theta} J(\theta).$$

Many values of (a, b) have been considered in order to characterize the trade-off between number of delayed packets $D(\theta)$ and number of wrongly delayed packets $E(\theta)$. For each value of (a, b), an ARX model has been obtained. Then, all the ARX models have been used to construct a Pareto optimality curve in the plane (D, E), *e.g.*, [15]. This curve has been constructed by simulating the ARX models and the corresponding scheduling algorithms on a video sequence obtained by joining in series five video sequences called "pamphlet", "deadline", "paris", "crew" and "mobile". Note that these sequences have not been used for identification, and thus they provide a reliable data set for the validation of the prediction models [6]. The Pareto optimality curve is shown in Figure 1.

E. Neural Network Nonlinear Models (NN)

The NN models are of the form

$$\begin{aligned} \widehat{d}^{t+1} &= f(w^t) \\ w^t &= [d^t; b^t; d^{t-11}; b^{t-11}] \end{aligned}$$

where the f is a one hidden layer perceptron [16] composed by r neurons:

$$f(w) = \sum_{i=1}^{r} \alpha_i \sigma(\beta_i w - \lambda_i) + \zeta. \tag{4}$$

Here $\theta = \{\alpha_i, \lambda_i, \zeta \in \mathbb{R},\, \beta_i \in \mathbb{R}^4\}$ is a set of parameters and $\sigma(x) = 2/(1 + e^{-2x}) - 1$ is a sigmoidal function.

Several neural networks of the form Eq. (4) have been identified (trained, according to the neural networks terminology) on the three video sequences "akiyo", "city" and "coastguard". The training has been performed by minimizing the cost function in Eq. (3). Many values of (a, b) have been considered in order to characterize the trade-off between number of delayed packets $D(\theta)$ and number of wrongly delayed packets $E(\theta)$. For each value of (a, b), an NN model has been obtained. Then, all the NN models have been used to construct a Pareto optimality curve in the plane (D, E). As for the ARX models, this curve has been obtained using the five video sequences, not used for identification, "pamphlet", "deadline", "paris", "crew" and "mobile". The Pareto optimality curve is shown in Figure 1, together with Pareto curve of the ARX models, and the values of (D, E) provided by the OM, AN and AD models. It can be observed that the trade-off between $D(\theta)$ and $E(\theta)$ is well described by the curve, even if the curve has been obtained from a data set different from the one used for identification. This suggests that some features of the distortion process are common to all video sequences.

Another observation stemming from Figure 1 is that the ARX optimality curve is farther than the NN optimality curve from the point corresponding to the omniscient prediction model OM. This indicates that distortion is a nonlinear process and motivates the use of nonlinear prediction models.

Then, three NN models, called NN1, NN2 and NN3, have been selected on the optimality curve and used in the next section to assess the communication performance by simulation.

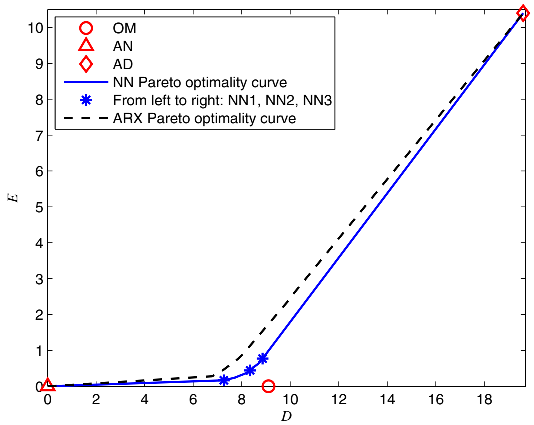

Fig. 1. Prediction models performance on the validation set: Pareto optimality curves.

	OM	NN1		NN2		NN3	
Sequence	D	D	E	D	E	D	E
Foreman	10.74	7.70	0.10	10.26	0.60	11.41	1.20
Mad	9.98	6.76	0.00	9.22	0.04	10.02	0.42
Lts	9.20	7.40	0.02	9.11	0.33	9.81	0.85
Harbour	8.85	6.00	0.42	7.65	0.70	8.81	1.11

TABLE I
DISTORTION PREDICTION PERFORMANCE (PERCENTAGE).

V. MODEL-BASED SIMULATIONS

The performance evaluation focuses on evaluating the effectiveness of the proposed scheduling algorithm exploiting the distortion prediction technique. A model implementing the entire sender scheduler was developed in Matlab. The video sequences in the simulation set known as “foreman”, “mad”, “lts” and “harbour”, are encoded at CIF resolution (352x288), 30 fps. The H.264 encoder v. JM11 [17] is configured to produce packets containing a fixed number of macroblocks, hence producing a fixed number of packets per frame. In more details, 22 macroblocks (each one is 16x16 pixels) are grouped into a packet, then each video frame is packetized into 18 packets each containing an entire row (16x352 pixels) of the encoded picture. Thus, the total number of packets per video sequence is 5400. The quantization parameter is fixed, yielding an approximately constant video quality. Moreover, the amount of premium bandwidth reserved to each video flow is chosen on the basis of its average bitrate at network level in all the experiments. The average bitrates are 862Kb/s, 511Kb/s, 927Kb/s and 1.85Mb/s for the “foreman”, “mad”, “lts” and “harbour” video sequences respectively.

First, the performance of the proposed scheduling algorithm has been assessed considering the relative indices D and E defined in Eq. (3).

The values of these indices obtained on the simulation set are reported in Table I. It is worth noting that these values are consistent with the values obtained on the validation set, represented in Figure 1. In fact, the values of D and E increase as the scheduler uses NN1, NN2 and NN3 on both the data sets. Since the simulation set is different from the training and validation sets used in the previous section, this is a further confirmation that some features of the distortion process are common to all video sequences, which allows to design reliable prediction methods.

Then, the performance of the scheduling algorithm has been assessed by means of a network model. The high priority packets are always considered correctly decoded whereas low priority packets are subject to uniform random losses. In accordance to Eq. (1) the loss/late probability suffered by packets delayed by the scheduler and then handled as low priority is greater then loss/late probability suffered by low priority packets immediately scheduled for transmission by the sender scheduler. The results are evaluated in terms of video quality using the Peak Signal-to-Noise Ratio (PSNR), which is commonly used in the multimedia communications community. The increment in late probability due to the further delay introduced by the sender scheduler depends on many network-dependent factors, *e.g.*, packet delay distribution. Thus, Figures 2, 3, 4 and 5 show the PSNR performance for the video sequences in the simulation set as a function of the late probability suffered by packets delayed by the sender scheduler and then handled as low priority given a loss probability experienced by all low priority packets. Note that this allow to assess the end-to-end performance of the video communication independently of a particular network scenario.

The results show that the proposed prediction technique based on neural networks is able to achieve a PSNR performance very close to the omniscient algorithm, which has perfect knowledge of future packet distortion values. The all-delayed (AD) packet technique provides, as expected, the worst performance, since delaying packets at the transmitter increase their PLR, thus reducing the quality of the video communication. The all-no-delayed (AN) technique, instead, provides a better performance, which however can be improved if some knowledge of future distortion packets is available, as with the proposed prediction techniques. In that case, in fact, less important packets are delayed hence increasing their PLR. This fact is, however, more than counterbalanced by reserving the better service for future packets which are predicted to have higher distortion, *i.e.*, to be more important, than the delayed ones. The general performance trend of the proposed scheduling algorithm is maintained even when the loss probability suffered by low-priority packets increases. Similar experiments on other video sequences, not shown due to space constraints, further confirm the trend.

VI. CONCLUSIONS AND FUTURE WORK

This work focused on how to efficiently and reliably estimate the distortion of the near-future part of video streams by exploiting the short and long term dependencies in packet distortion values. The aim is to run the optimization over a larger portion of the video stream, potentially achieving better performance than a locally optimal solution. In this work, several linear auto regressive with external input (ARX) models

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

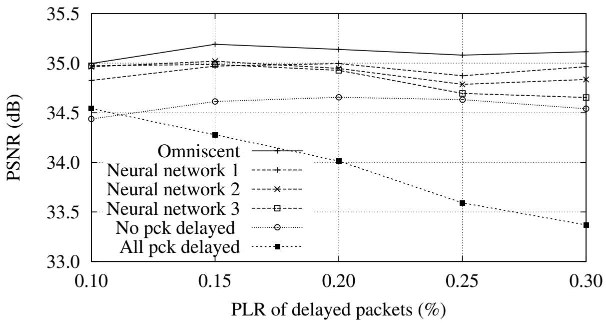

Fig. 2. PSNR performance, for the Foreman sequence, as a function of the PLR of the delayed packets (PLR of non-delayed packets equal to 5%).

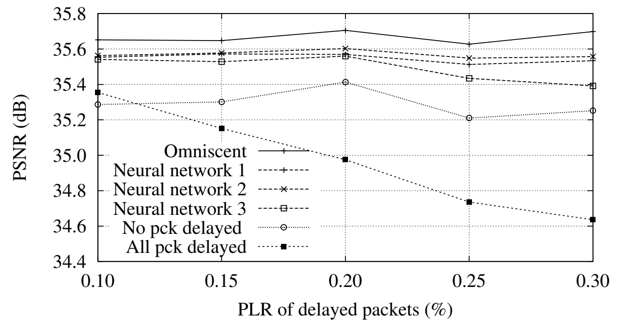

Fig. 4. PSNR performance, for the Lts sequence, as a function of the PLR of the delayed packets (PLR of non-delayed packets equal to 5%).

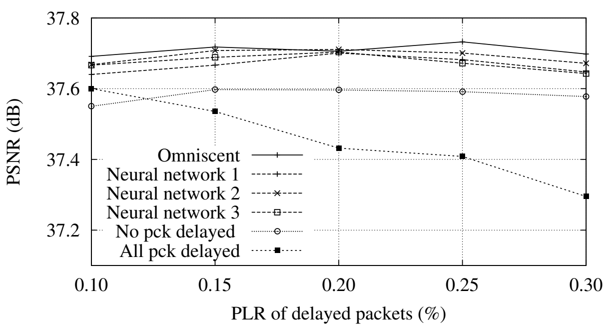

Fig. 3. PSNR performance, for the Mad sequence, as a function of the PLR of the delayed packets (PLR of non-delayed packets equal to 5%).

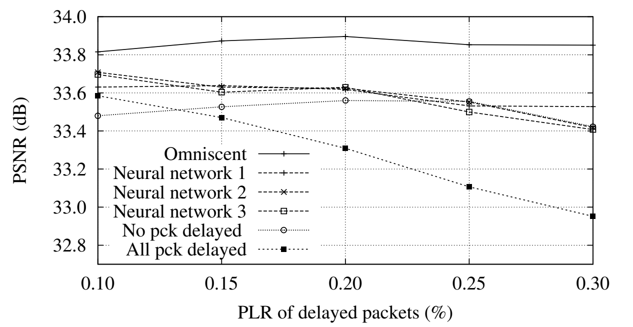

Fig. 5. PSNR performance, for the Harbour sequence, as a function of the PLR of the delayed packets (PLR of non-delayed packets equal to 5%).

and neural network (NN) models for distortion prediction have been developed to take into account the short and long term dependencies between packets. It is shown that the NN models provide better prediction performance, indicating a nonlinear behavior of the distortion process. Thus, three NN models have been selected to simulate the video quality communication performance over a QoS-enabled network while employing a standard rate-distortion optimization framework. Results show that the models allow the achievement of good video quality performance, very close to the one of the reference prescient algorithm. Future work will be devoted to generalize the distortion prediction models in order to support different coding structures and packetization schemes. Moreover, a low complexity implementation of the distortion prediction models will be developed, to allow efficient on-line utilization of these models by the schedulers.

References

[1] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[2] Q. Li and M. V. D. Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 278–290, April 2004.

[3] R. Zhang, S.L. Regunathan, and K. Rose, "End-to-end distortion estimation for RD-based robust delivery of pre-compressed video," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov 2001, pp. 210–214.

[4] E. Masala and J. D. Martin, "Analysis-by-synthesis distortion computation for rate-distortion optimized multimedia streaming," in *Proc. IEEE Int. Conf. on Multimedia & Expo*, vol. 3, Baltimore, MD, July 2003, pp. 345–348.

[5] G. Box and G. Jenkins, *Time series analysis: Forecasting and control*. San Francisco, CA: Holden Day, 1976.

[6] L. Ljung, *System identification: theory for the user*. Upper Saddle River, N.J.: Prentice Hall, 1999.

[7] K. R. Muller, A. Smola, G. Rtsch, B. Schlkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *Advances in Kernel Methods — Support Vector Learning*, Cambridge, MA, 1999, pp. 243–254.

[8] J. Connor, R. Martin, and L. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Transaction on Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994.

[9] P. Bucciol, E. Masala, E. Filippi, and J.C. De Martin, "Cross-layer perceptual ARQ for video communications over 802.11e wireless networks," *Hindawi Journal of Advances in Multimedia*, Article ID 13969, 2007.

[10] F. De Vito, L. Farinetti, J.C. De Martin, "Perceptual classification of MPEG video for Differentiated-Services communications," in *Proc. IEEE Int. Conf. on Multimedia & Expo*, vol. 1, Lausanne, Switzerland, August 2002, pp. 141–144.

[11] F. De Vito, D. Quaglia, and J.C. De Martin, "Model-based distortion estimation for perceptual classification of video packets," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, Sep 2004, pp. 79–82.

[12] P.A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, Apr 2006.

[13] E. Masala, D. Quaglia, and J.C. De Martin, "Adaptive picture slicing for distortion-based classification of video packets," in *Proc. IEEE Workshop on Multimedia Signal Processing*, Cannes, France, October 2001, pp. 111–116.

[14] D. S. Modha and E. Masry, "Memory-universal prediction of stationary random processes," *IEEE Transaction on Information Theory*, vol. 44, pp. 117–133, 1998.

[15] M. J. Osborne and A. Rubenstein, *A Course in Game Theory*. MIT Press, 1994.

[16] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[17] ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," *ITU-T*, May 2003.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.