Distortion-Aware Video Communication with Pipeline Forwarding

Mario Baldi Juan Carlos De Martin Enrico Masala Andrea Vesco
mario.baldi@polito.it demartin@polito.it masala@polito.it andrea.vesco@polito.it
Dipartimento di Automatica e Informatica — Politecnico di Torino
Corso Duca degli Abruzzi, 24 - 10129 Torino (Italy)

ABSTRACT

This paper tackles the issue of optimizing the transport of video over packet networks with respect to both resource utilization and user perceived quality. Previous work showed that the quality of service requirements of multimedia applications can be satisfied by pipeline forwarding of packets. However, the current Internet is not based on such technology and its incremental introduction raises questions on how to handle video packets generated by pipeline forwarding unaware sources at the interface between a subnetwork deploying conventional packet scheduling techniques and one implementing pipeline forwarding. This work proposes to use the perceptual importance of the carried video samples to determine which packets shall be transferred with pipeline forwarding — thus receiving deterministic service — and which with a traditional, e.g., best effort or differentiated, service. Simulation results with the first implemented variants of this solution are presented.

Categories and Subject Descriptors

C.2.1 [**Computer-communications networks**]: Network Architecture and Design – packet switching networks

General Terms

Design, Experimentation, Algorithms.

Keywords

Video streaming, Quality of service, Multimedia networking.

1. INTRODUCTION

Various multimedia applications, such as telephony over IP, voice over IP (VoIP), video broadcasting (e.g., IPTV), and video on demand, are becoming more widely available. These applications are often referred to as real-time to juxtapose them to traditional data applications as timely packet delivery is important for them to work properly. Packet networks, originally designed for data applications are not engineered to tightly control the delay packets experience in routers where they might contend for resources (e.g., transmission capacity), consequently be queued for a variable time, and possibly be dropped. Moreover, multimedia applications are usually of a streaming nature as they generate a more or less continuous flow of data and not elastic, i.e., they need a minimum fraction of their data to reach the destination and do not adapt to particularly poor network service.

Right now the requirements of multimedia applications are commonly satisfied through overprovisioning, i.e., by keeping the network lightly loaded so that contention for network resources is low and queuing time consequently small. This approach is not feasible if multimedia traffic grows faster than technology enables proportionally more powerful network infrastructures. And this might be the case not only because a larger fraction of broadband users might subscribe current multimedia services, but especially because even more This paper tackles the issue of optimizing the transport of video over packet networks with respect to both resource utilization and user perceived quality. Previous work showed that the quality of service requirements of multimedia applications can be satisfied by Pipeline Forwarding (PF) of packets. However, the current Internet is not based on such technology and its incremental introduction raises questions on how to handle video packets generated by PF unaware sources at the interface between a subnetwork deploying conventional packet scheduling techniques and one implementing PF. This work proposes to use the perceptual importance of the carried video samples to determine which packets shall be transferred with PF — thus receiving deterministic service — and which with a traditional, e.g., best effort or differentiated, service. Simulation results with the first implemented variants of this solution are presented. Distortion-aware video communication with PF demanding applications, such as 3D video on demand, high quality videoconferencing, high definition TV, distributed gaming, (3D) virtual reality, and remote surveillance, might become the dominant traffic sources in the future Internet.

Previous research showed that the requirements of multimedia applications can be satisfied by PF of packets [1][2], which bases its properties on network nodes sharing a common time reference (CTR). As demonstrated in [2], optimal performance for interactive real-time applications can be achieved on a PF network where end systems and applications share the CTR. However, the operation of today's Internet is inherently asynchronous and for the immediate future it is not realistic to assume that applications make use of a CTR. It is not straightforward for asynchronous end-systems and applications to take advantage of the features of PF, although potentially extremely beneficial. This is certainly the case with the transmission of video encoded with constant quality, which results in a packet flow at an irregularly variable bit rate.

This paper addresses the problem of optimizing the transport of packetized video generated by a PF unaware source over a PF network with respect to the user perceived quality. The perceptual importance of the carried video samples is used at the interface to the PF network to determine which packets shall be transferred with PF — thus receiving deterministic service — and which with a traditional, e.g., best effort or differentiated, service. An extensive evaluation work is required for the many implementation variants of the proposed solution. This paper reports on simulation results with the first implemented variants.

Section 2 focuses on PF by presenting its operating principles and properties. The approach deployed to model perceptual importance associated to and distortion resulting from the loss of video samples is presented in Section 3. Section 4 first presents two existing low complexity heuristic algorithms for generic sender-driven scheduling of packet transmission aiming at minimizing expected distortion, then discusses their application at the interface to a PF network. Section 5 presents the network model. Initial simulation results obtained by applying the proposed solution are provided in Section 6. Finally, the proposed approach and research directions are discussed in Section 7.

2. PIPELINE FORWARDING

Pipeline forwarding is a known optimal method that is widely used in computing and manufacturing. The necessary requirement for PF is having common time reference (CTR). An extensive and detailed description of UTC-based PF is outside the scope of this paper and can be found in [1].

In order to perform PF of packets all switches are synchronized, while utilizing a basic time period called time frame (TF). The TF duration (T_f) may be derived, for example, as a fraction of the UTC second received from a time-distribution system such as the global positioning system (GPS) and, in the near future, Galileo. TFs are grouped into time cycles (TCs) and TCs are further grouped into super cycles, each super cycle lasts for one UTC second. TFs are partially or totally reserved for each flow during a resource reservation phase. The TC provides the periodicity of the reserved flow. This result in a periodic schedule for IP packets to be switched and forwarded, which is repeated every TC. A synchronous virtual pipe (SVP) is a predefined schedule for forwarding a pre-allocated amount of bytes during one or more TFs along a path of subsequent UTC-based routers.

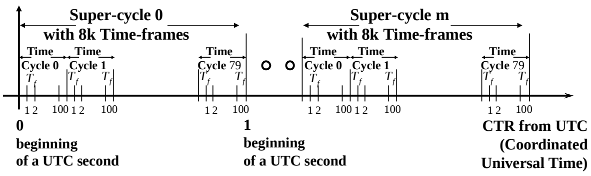

Fig. 1 Common time reference structure

Pipeline forwarding guarantees that reserved real-time traffic experiences: (I) bounded end-to-end delay, (II) delay jitter lower than one TFs, and (III) no congestion and resulting losses.

Non-pipelined (i.e., non-scheduled) IP packets, i.e., packets that are not part of a SVP (e.g., IP best-effort packets), can be transmitted during any unused portion of a TF, whether it is not reserved or it is reserved (partially or completely) but currently unused. Consequently, links can be fully utilized even if flows with reserved resources generate fewer packets than expected. Any service discipline, such as for example the ones deployed in the context of the Differentiated Services framework, can be applied to packets being transmitted in unused TF portions. In summary, PF is a best-of-breed technology combining the advantages of circuit switching (i.e., predictable service and guaranteed quality of service) and packet switching (statistical multiplexing with full link utilization) that enables a true integrated services network providing optimal support to both multimedia and elastic applications.

3. ANALYSIS-BY-SYNTHESIS DISTORTION ESTIMATION

Multimedia data, and video in particular, exhibit non uniform perceptual importance. When video is transmitted over a noisy channel, each loss event causes a decrease of the video quality that depends on the perceptual importance of the lost data. Such importance can be defined a priori based, for instance, on the average importance of the elements of the compressed bit stream, as with the data partitioning approach. At a finer level of granularity, the importance of a video coding element, such as a macro block or a packet, could be considered proportional to the distortion that would be introduced at the decoder by the loss of that specific element.

A practical method to compute the distortion (D_i) caused by the loss of each packet i is the analysis-by-synthesis one presented in [3], which is reported in the following: (I) Decoding (including concealment) of the bit stream simulating the loss of the packet being analyzed, (II) Computation of the distortion D_i between reconstructed and original sequence, (III) Storage of the obtained value D_i as an indication of the perceptual importance of the analyzed video packet.

The analysis-by-synthesis distortion estimation method is independent of the video coding standard. Since it includes the synthesis stage in its body, it can accurately evaluate the effect of both the error propagation and the error concealment. Note that this method assumes isolated packet losses; nevertheless, this leads to a useful approximation as demonstrated by some applications of the analysis-by-synthesis approach to MPEG coded video [3].

The complexity and delay of the analysis-by-synthesis method depends on the frame types the sequence is composed of. If only I-type frames are present, the technique is quite simple since each frame is coded independently of the others. If the sequence contains also predicted frames, such as in the case of H.264, the method has higher complexity because error propagation must be taken into account; a model-based approach, however, can be used to drastically reduce complexity [4]. Moreover, note that in the case of stored video (e.g. non-live streaming scenarios), the distortion values can be precomputed and stored.

4. DISTORTION OPTIMIZED SCHEDULING ALGORITHMS

Two scheduling algorithms have been designed to determine which are the best packets to transmit from the distortion viewpoint at each transmission opportunity, that is during each TF in which a reservation for the given video flow exists.

The first algorithm keeps the packets waiting to be transferred through the PF network ordered in decreasing order of distortion (D_i), then it selects the ones with the highest D_i value until no more packets fit in the allocation for the video stream during the current TF. This will be referred to as the *Distortion-based Algorithm* (*DA*).

The second algorithm aims at determining the best rate-distortion solution to the packet selection problem formulated as follows.

Given a certain rate constraint (the allocation within a TF), the distortion D_i and size P_i of the packets available for the transmission, send with PF those packets which minimize the expected distortion at the decoder.

The resulting scheduling problem is hence converted into a classic rate-distortion optimized transmission problem, which can be easily solved by recasting it into a Lagrangian minimization problem: the Lagrange multiplier is determined using the bisection algorithm and the solution produced by Lagrangian optimization is further refined, as already suggested in [5], adding, if possible, not yet selected packets in decreasing order of distortion. This algorithm will be referred to as *LA*.

Note that packet reordering issues stemming from the application of the above algorithms can be easily handled at the destination node by means of the RTP sequence number.

5. NETWORK MODEL

Fully benefiting from PF requires providing network nodes and end-systems with a CTR and implementing CTR-aware applications to maximize the quality of the received service [2]. Since this is not realistic for the near future, as discussed in Section 1, this work assumes traditional video sources. The PF unaware packet stream must be time-shaped at the ingress of the packet forwarding network, i.e., packets are forwarded during the TFs in which resources have been allocated to their SVP. The network element implementing such functionality is called *SVP interface*.

Each frame is assumed to be encoded, packetized, and immediately sent by the source. For the sake of simplicity, video packets are assumed to reach the SVP interface without loss and after a negligible delay. This model is reasonable in the currently realistic scenario of a lightly loaded (asynchronous) broadband access network. Consequently, all packets belonging to a frame are fully available at the SVP interface every $1/F_r$ seconds, where F_r is the frame rate of the video sequence.

Assuming that non-pipelined traffic receives a best effort service[1], the SVP and, based on the above assumption, the whole PF channel from source to destination, is modeled as an independent time-invariant channel with constant delay and loss/late probability equal to zero. In fact, all pipelined packets arrive, without loss, on time at the destination. The forward trip time FTT can be calculated as follows[2]:

$$FTT = \sum_{i=1}^{N}\left(\left\lceil \frac{Cd_i}{T_f} \right\rceil \cdot T_f + T_f\right) + Cd_{N+1} + J$$

where N is the number of PF nodes on the path, Cd_i is the propagation delay between the $(i-1)_{th}$ node and the i_{th} node (the ingress SVP interface being node 0 and the egress SVP interface being node N+1), T_f is the duration of the TF and J is the jitter with maximum value equal to one T_f. Given that in normal operating conditions T_f is at most on the order of hundreds of microseconds, for all purposes of video transmission the end-to-end delay can be considered constant and is deterministically known in advance given the path the video flow takes through the network as:

$$FTT = \left(\sum_{i=1}^{N}\left\lceil \frac{Cd_i}{T_f} \right\rceil + N + 1\right) \cdot T_f + Cd_{N+1}$$

To find a global optimum minimizing the expected distortion, the scheduling algorithms presented in Section 4 should run on the entire video sequence, which is obviously not possible in a live video scenario. Normally, in order to avoid packets to get to the destination beyond their play-out deadline because of the delay introduced by both the network and the distortion optimized scheduling algorithm (DOSA) waiting a large number of frame periods $1/F_r$, a trade-off is found by running the algorithms on a small part of a video sequence, which results is a locally optimal schedule. The length of the video sequence on which the algorithm is run is determined based on the maximum delay packets experience in the network, or a percentile thereof.

As mentioned before, due to PF the delay introduced by the network is known in advance and it is typically smaller than the maximum delay introduced by networks deploying other packet queuing techniques [2]. Hence the sender can determine in advance the maximum delay the DOSAs can introduce on a packet while making sure that it arrives at the destination by its play-out deadline, taking into account the fixed network latency. This enables running the DOSAs on longer sequences of video frames compared to when traditional network solutions are being deployed, which results in a potentially more optimized solution.

In practice, the SVP interface assigns to each packet a *transmission deadline*, which is the latest time at which the packet can be scheduled for transmission. The DOSA schedules for transmission as pipelined traffic the most perceptually important packets that, given the reservation to their flow, can be transmitted by their transmission deadline. The remaining packets (i.e., the least perceptually important packets) are handled by the SVP interface as not pipelined traffic. The advanced knowledge of the transmission deadlines make it possible to perform scheduling optimization as soon as packets arrive at the SVP interface so that packets that do not have a chance to be handled as pipelined traffic by their transmission deadline can be immediately forwarded as non-pipelined traffic. This improves the probability of their timely reception and minimizes the resulting distortion.

The forwarding policy (e.g., burstiness, prioritization, etc.) applied by the SVP interface to non pipelined traffic during the unused portion of TFs might affect the performance of the presented solution. However, it is not addressed in this paper and will be the subject of future work.

6. SIMULATION MODEL AND RESULTS

The proposed network model and DOSAs have been implemented and simulated in NS2. For the sake of simplicity this first study considers a single video flow and all non pipelined packets are dropped at the sender. Consequently, the obtained results represent a sort of minimum performance bound. The TC consists

---

[1] As mentioned in Section 1, this is not necessarily the case. However, this paper reports preliminary work and studies only this scenario. The evaluation of more sophisticated approaches will be the subject of further research (see Section 7).

[2] Without loss of generality, the presented FTT calculation assumes the time required by network nodes to move packets from input to output to be null. See [2] for details on the pipeline forwarding FTT calculation.

of 100 TFs, each having a duration of 125 μs. The capacity of the SVP interface output link is 100 MB/s, hence at most 1562 bytes can be transmitted during a TF. The overall buffering space at the SVP interface, which limits the number of packets possibly waiting for service, is large enough to avoid packets drops due to buffer overflow.

The presented results are not dependent on a specific network topology because the deterministic service offered by PF only depends on the number of traversed nodes and the propagation delay through the links.

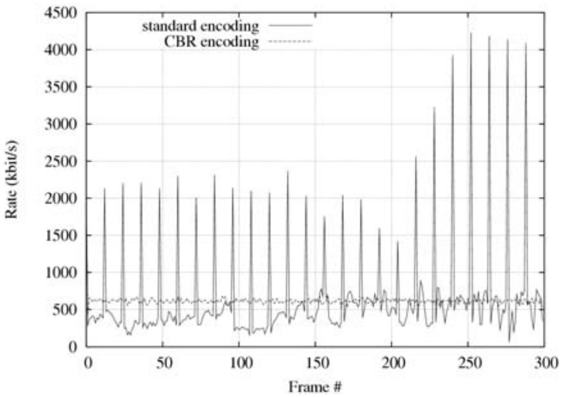

Fig. 1 Bit rate of the compressed video sequences.

The well-known *foreman* video sequence have been encoded at 30 fps by means of the H.264 test model software v. JM10 using two different schemes. The first encoding scheme, called *standard encoding*, employs 12-frame GOPs, composed by the first I-type frame followed by P-type frames only, the quantization parameter has been set to a fixed value, which results in high variability of the bit rate sequence, as shown by Fig. 1. Another encoding scheme, called *CBR encoding*, has been optimized to achieve the smoothest bit rate. Hence, rate control has been activated using a target bit rate equal to the average bit rate achieved by the standard encoding scheme. Moreover, only P-type frames have been used, with an intra refresh mechanism which achieves a full refresh every 12 frames. These settings lead to a much more constant bit rate, as shown in Fig. 1.

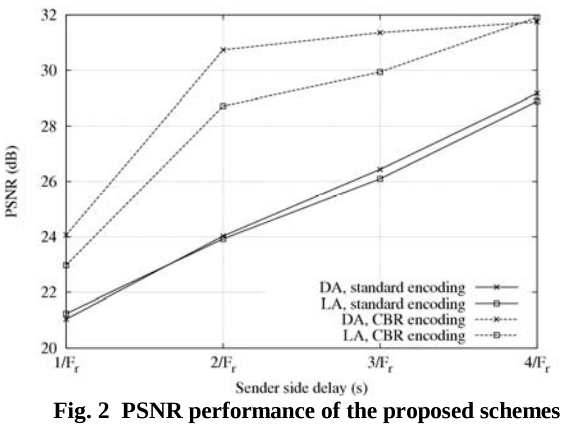

Fig. 2 PSNR performance of the proposed schemes

The proposed DOSAs have then been applied to both the standard and CBR sequences transmitted over an SVP with a resource reservation of 1500 bytes in a single TF during each TC to the video flow, which results in an average allocated bandwidth of 960 kb/s. This value has been selected as a tradeoff between the guaranteed bandwidth available to the multimedia flow and the bandwidth waste when the instantaneous rate of the encoded video is lower than the allocated bandwidth. The resulting Peak Signal-to-Noise Ratio (PSNR) is shown in Fig. 2 and was obtained with a confidence level of 99% and a confidence interval of 3%.

Clearly the PSNR value increases as the sender side delay increases. Note also that the performance of the proposed algorithms strongly depends on the characteristics of the video sequence. The CBR sequence, in fact, achieves a PSNR value which is up to 7 dB higher than the standard encoded sequence.

While analyzing these results, it is worth highlighting that the number of packets that can be transmitted in the allocated TF, thus benefiting from PF, is limited. For instance, for the standard sequence case, such number is typically one, hence the two DOSAs do not have many options for selecting packets and end up performing similar scheduling decisions. In the CBR sequence case, two or more packets often fit in the allocated TF, which leads to a larger number of admissible solutions, and the two algorithms have the possibility of differentiating. The Lagrangian based approach, in fact, cannot always find a good solution because it limits its search to those on the convex hull (as explained in [5]). On the contrary, the DA algorithm, despite it is not guaranteed to achieve the optimal solution, often achieves a very good if not optimal solution, which explains the performance gain compared to the LA.

7. CONCLUSIONS AND FUTURE WORK

In this work the first results of video transmission over pipeline forwarding have been presented, showing that performance strongly depends on the characteristics of multimedia flows.

Future work will be devoted to study different forwarding policies for the non-PF traffic both by simulation and by modeling the service it receives in order to achieve topology-independent results. Moreover, we plan to improve the scheduling algorithms so they can deal with different types of multimedia flows and to optimize the source encoding process to make it pipeline forwarding aware, so that the video flow can fully benefit from it.

8. REFERENCES

[1] C-S. Li, Y. Ofek, A. Segall, K. Sohraby, “Pseudo-Isochronous Cell Forwarding,” *Computer Networks and ISDN Systems*, Vol.30, No.24, Dec. 1998, pp. 2359-2372.

[2] M. Baldi, Y. Ofek, “End-to-end Delay of Video-conferencing over Packet Switched Networks,” *IEEE/ACM Trans. on Networking*, Vol.8, No.4, Aug. 2000, pp. 479-492.

[3] E. Masala, J.C. De Martin, “Analysis-By-Synthesis Distortion Computation For Rate-Distortion Optimized Multimedia Streaming”, *IEEE Int. Conf. on Multimedia & Expo (ICME)*, Baltimore, MD, Jul 2003, vol. 3, pp. 345-348.

[4] F. De Vito, D. Quaglia, J.C. De Martin, “Model-based Distortion Estimation for Perceptual Classification of Video Packets”, *Proc. of IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, Sep. 2004,Vol.1, pp.79-82.

[5] A. Gersho, Y. Shoham, “Efficient Bit Allocation for an Arbitrary Set of Quantizers”, *IEEE Transaction of Acoustics, Speech and Signal Processing*, Vol.36, No.9, Sep. 1988.