Countering video packet loss due to buffer overflow by means of retransmissions

Frederik Vanhaverbeke[1] and Marc Moeneclaey[1], Koen Laevens[2], Natalie Degrande[2], and Danny De Vleeschauwer[2]
[1]TELIN/DIGCOM, Ghent University
Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium
[2]NTS, Alcatel-Lucent Bell, Copernicuslaan 50,
B-2018 Antwerp, Belgium

Abstract- **We investigate the impact of retransmissions on the HDTV video packet loss ratio over a network in which video packets are at the risk of getting lost due to buffer overflow. The use of retransmissions in these circumstances is a mixed blessing. On the one hand, retransmissions add to the packet sending rate associated with each video flow, which increases the original load on the network, resulting in a higher buffer overflow probability. On the other hand, retransmissions allow for the recovery of lost packets, which alleviates the bad video quality due to buffer overflow. We find in this paper that the beneficial effect of basic retransmission strategies vastly outweighs the detrimental effect caused by the load increase.**

I. INTRODUCTION

Conceptually a packet-based network transporting video often has a tree structure as illustrated in figure 1 [1]. In its simplest form it consists of 1) a common link between an ingress node (in which the video flows are injected) and a fan-out node, and 2) individual links from the fan-out node towards each individual user. For video flows transported over a packet-based network it is of paramount importance that the Packet Loss Ratio (PLR) over the network be kept as low as possible as each packet that does not reach a user translates in an unappreciated visible distortion on the user's display.

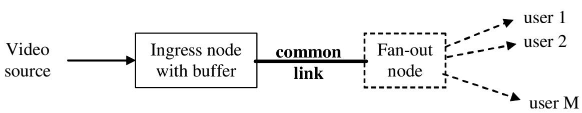

Fig. 1: Tree structure of a typical packet-based network.

In previous work [2] we concentrated on controlling the impact of packet loss that occurred beyond the fan-out node (i.e., the last mile link and the home network), where the packet loss was due to noise phenomena during transport. In this paper we focus on the packet loss on the common link between the ingress node and the fan-out point. Generally speaking the latter link is free of transport errors. However, since this link needs to transport the packets of all individual video flows destined for different users there may be some contention in the ingress node. Indeed, the arrival of the packets of the aggregate of all video flows in the ingress node is governed by a random process, and it may happen that a packet arrives at the same time an earlier packet is being transported. To deal with this kind of contention, the ingress node must be equipped with a "First In First Out" (FIFO) buffer that temporarily stores packets awaiting transport over the common link [3]. In practice, such buffer always has a limited storage capacity, which implies that an arriving packet can encounter a full buffer. When this happens, the packet will simply be discarded at the buffer (i.e., it gets lost), which will (if this is not corrected) be experienced further downstream by some user(s) as a visible distortion.

In this paper we examine whether or not retransmissions [4,5] can alleviate the packet loss problem in a video stream caused by buffer overflow in the ingress node. For that purpose the video packets received by a user are temporarily stored for a certain period before they are played out. This gives the receiver ample time to detect that a packet is lost (because it did not arrive and the following packet did) and to ask the video source for a retransmission via a Negative ACKnowledgment (NACK). In this way, each packet that is lost due to buffer overflow is retransmitted and a second (or even third) attempt is made to get it (over the ingress node) to the receiver. The drawback of this strategy, however, is that retransmissions introduce an overhead bit rate, which increases the overall load in the ingress buffer – and the attendant packet loss due to buffer overflow – vis-á-vis the classical system without retransmissions. Moreover, an additional delay is introduced (due to the fact that packets need to be temporarily stored at the receiver side), but in an IPTV system this hardly matters.

So, in summary, although retransmissions provide us with a means to fight packet losses due to buffer overflow, it backfires as it results in an increase in buffer load and a concomitant increase of the PLR owing to overflow. The pivotal question is now how these two influences work out together. Put another way, we would like to know whether or not the overall impact of retransmissions is beneficial or detrimental. The aim of this paper is to sort this question out. In Section 2, we give the mathematical framework used to evaluate the retransmission-loading trade-off that is central in this paper. Section 3 presents the results of the evaluation along with a discussion. And finally in Section 4, we draw our conclusions.

II. Retransmissions as a means to counter buffer overload

As explained in the introduction, the impact of retransmissions on a buffer system is twofold. On the one hand, it increases the load, which leads to a higher probability of buffer overflow. But on the other hand, retransmissions make it possible to recover lost (erased) packets. We quantify the impact of both effects in the following sections, after we have justified the simplifying assumptions we made.

A. System under study

In the system at hand, users receive a video flow over a network with a tree topology as illustrated in figure 1. All packetized video flows are fed into an ingress node, travel over a common link towards a fan-out node and are further transported over an individual link towards the end-user. At the receiver side the video packets are retained for some period before they are played out in order for the receiver to ask for one or more retransmissions. We focus on High Definition Television (HDTV) quality requiring a bit rate of about 7.6Mb/s [6]. With the usual payload size of (7x188=) 1316 bytes, this results in a packet rate of about 720 packets/s [7]. The Round Trip Time (RTT) between receiver and HDTV encoder (the sum of the delay from receiver to HDTV encoder and back) is typically below a few 10s of ms. If the RTT is larger a dedicated retransmission buffer can be implemented close to the ingress node.

As we do not concentrate on the individual user links in this paper, we assume that they are error-free. To determine the packet loss ratio (PLR) in the ingress node buffer, we rely on a M/D/1/L queueing model [3]. This so-called Kendal notation for queues, summarizes the following underlying assumptions:

- M: a memoryless, i.e., Poisson packet arrival process
- D: fixed service times (i.e., at Q bps, a packet of size L bits requires about L/Q sec to be transmitted)
- 1: a single outgoing link
- L: a buffer that can hold up to L packets (including the one being transmitted)

Except for the first one, these assumptions are obvious. The assumption of a Poisson packet arrival process is a simplifying approximation that can be justified to some extent by noting that the packets are generated by a rather large number of independent variable-bit rate (VBR) video sources, that each contribute only a small fraction to the total traffic.

There are several ways to determine (or approximate) the packet loss ratio in such a queue. We opted for a straightforward numerical approach that consists of two steps. First, one determines the equilibrium probability mass function (pmf) of the system occupancy, i.e., the number of packets in the buffer, as observed at the beginning of a packet transmission time (i.e., a service time), once transient effects have died out. This can be done in a recursive way, which turned out to be numerically stable for the parameter values we will consider further on. Second, one derives the PLR by calculating the average number of packets that get lost during a packet transmission time, taking into account the number of packets already in the buffer at the start thereof, and weighing with the probabilities derived in step one.

One can easily show that the PLR is determined by two parameters only: the size of the system L, and ρ, the so-called load. The latter, a dimensionless parameter, is given by $\rho = \lambda \cdot D$, where λ is the arrival rate of the Poisson process, expressing the average number of packets arriving at the system per unit of time, and D is, as before, the packet transmission time (measured in the same time units as λ). The load measures how much of the capacity of the system is requested by the arriving packets. A load of 1.0 indicates that, on average, as much packets arrive per time unit, as the system can transmit.

In figure 2 we illustrate the behavior of the PLR due to buffer overflow $PLR_{M/D/1/L}^{overflow}$ as a function of the load ρ and for various buffer sizes L. From this figure we can see that the buffer overflow probability curve sharply increases if the load is increased over the range $\rho \in [0.5;1]$, in particular for buffer sizes L in excess of about 40.

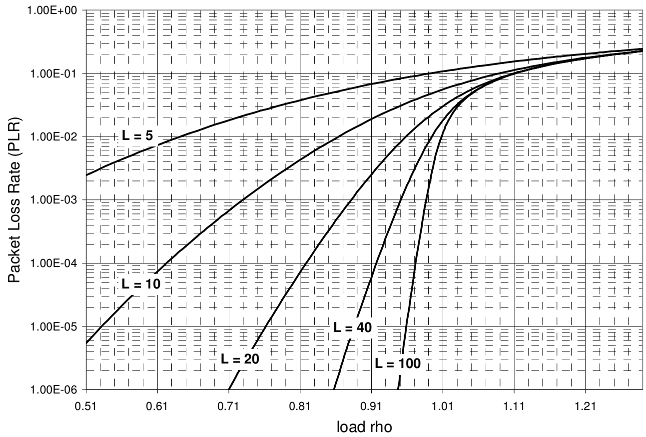

Fig. 2: PLR due to buffer overflow in a M/D/1/L buffer system.

B. PLR increase due to overhead

A load increase is, unfortunately, an inevitable by-product of the introduction of retransmissions. By observation of figure 2 one can expect a substantial increase in PLR due to buffer overflow when the load is increased in an M/D/1//L system, especially for common buffer sizes L around 40. In fact, when the load in the *absence* of retransmissions is given by ρ, the load in the *presence* of retransmissions will be given by

$$\rho_{retrans} = \rho . E[\#] \quad (1)$$

where E[#] is the expected value of the number of transmitted packets per original video packet.

If we apply a retransmission procedure where one copy is retransmitted for each NACK and where no more than N retransmissions are allowed for each original video packet, the value of E[#] will be given as:

$$E_N[\#] = \sum_{n=1}^{N+1} n.\Pr_{tr+re}[n] \qquad (2)$$

where $\Pr_{tr+re}[n]$ is the probability that a video packet is transmitted (transmission + retransmissions) n times by the video source. With p the probability that a packet gets lost due to buffer overflow, we can express $\Pr_{tr+re}[n]$ as:

$$\Pr_{tr+re}[n] = \begin{cases} (1-p)\,p^{n-1} & 1 \le n \le N \\ p^N & n = N+1 \end{cases} \qquad (3)$$

which, upon substitution in (2), yields:

$$E_N[\#] = (1-p)\sum_{n=1}^{N} n.p^{n-1} + (N+1)p^N = \sum_{n=0}^{N} p^n = \frac{p^{N+1}-1}{p-1} \qquad (4)$$

Hence, the load (1) with this retransmission procedure will be the solution of the following equation:

$$\rho_{retrans}(\rho) = \rho \; \frac{p^{N+1}-1}{p-1} \qquad (5)$$

with $p = PLR^{overflow}_{M/D/1/L}[\rho_{retrans}(\rho)]$.

C. PLR reduction with retransmissions

Of course, the overhead introduced by the retransmissions serves a purpose, which is that it allows for the restoration of erased (lost) packets. If we assume that the round-trip time between the video source and the ingress node is sufficiently high such that there is no correlation in the overflow process between successively retransmitted video packets[1], the packet loss rate after the application of the retransmission protocol explained in the previous section will be:

$$PLR^{N}_{retrans}(\rho) = \left\{ PLR^{overflow}_{M/D/1/L}[\rho_{retrans}(\rho)] \right\}^{N+1} \qquad (6)$$

III. RESULTS

Based on the results derived in section II, we can examine the overall impact of retransmissions on the PLR caused by buffer overflow in the ingress node. Although the focus of this paper is on the application of retransmission strategies for HDTV, we will first present some general results in section A, while section B will interpret these results in light of the application for HDTV.

A. General results

Based on the expressions (5), and (6), we can evaluate the relationship between the load ρ of the original video packets and the PLR after retransmissions. In figure 3, this relationship is shown for a common buffer with size L = 40. We immediately observe the following striking results:

- The retransmission strategies result in a tremendous gain of several orders of magnitude in terms of PLR for values of $\rho < 1$. However, when ρ exceeds one, $PLR_{retrans}^{N}$ quickly converges to $PLR^{overflow}_{M/D/1/L}$ for all values of N.
- Although the maximum number of allowed retransmissions N has a very substantial impact on the PLR for $\rho < 1$, the performance of all retransmission schemes (N = 1, 2, 3, and 4) converges for $\rho > 1$.

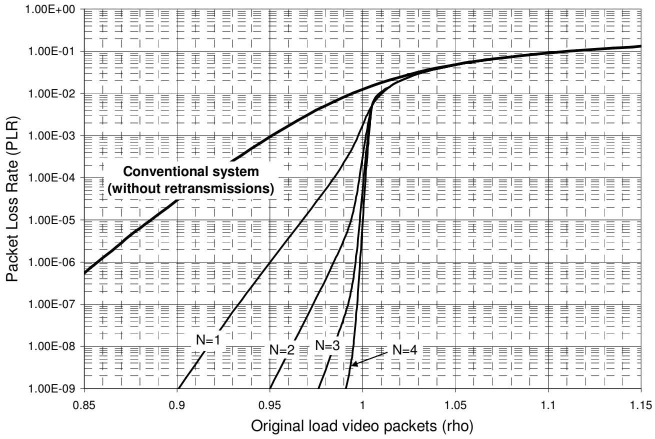

Fig. 3: Comparison of PLR with and without retransmissions for L = 40.

The previous two conclusions imply that the increased performance thanks to retransmissions overcompensates for the increased PLR due to the increased load, especially for $\rho < 1$. Once the original load ρ increases beyond one, there is only a minor gain that tends to dissipate with increasing load.

Although this observation was made for L = 40, similar results can be found for other values L ≤ 150. In figure 4, we show for L ≤ 150 the value of the maximal load $\rho_{max}(L)$ where the system with a single retransmission (i.e., N = 1) outperforms the system without retransmission by a margin of 10%, 1% and 0.1%, where this margin Δ is defined as:

$$\Delta = \frac{PLR^{overflow}_{M/D/1/L}(\rho) - PLR^{1}_{retrans}(\rho)}{PLR^{overflow}_{M/D/1/L}(\rho)} \qquad (7)$$

This figure shows that $\rho_{max}(L)$ decreases as L increases. Nevertheless, the value remains higher than 1 for all considered values of L. This indicates that the retransmission strategies are beneficial for all buffer sizes (L ≤ 150) up to loads in excess of 1.

B. Results for a HDTV system

In the HDTV system described in section II.A, about 720 video packets are transmitted per second per user. As a result, when the occurrence of packet losses is assumed to be

[1] This is a fair assumption since we only consider the case where the total load ρ is constituted of many sources that each contribute only a small fraction to the total load, such that many packets of other flows are fed into the buffer between consecutive packets of one particular video stream. As such, one stream of retransmitted packets "observes" the queue at time instants far enough apart to consider the packet loss of consecutively retransmitted packets as statistically independent.

random[2], the mean time between visible distortions T_{MTBVD} [h] observed by a user for a given load ρ will be given as

$$T_{MTBVD}(\rho) \approx \frac{1}{PLR(\rho).720.3600} h \qquad (8)$$

where PLR(ρ) is $PLR^{overflow}_{M/D/1/L}(\rho)$ or $PLR^{N}_{retrans}(\rho)$ in the absence, respectively presence of retransmissions.

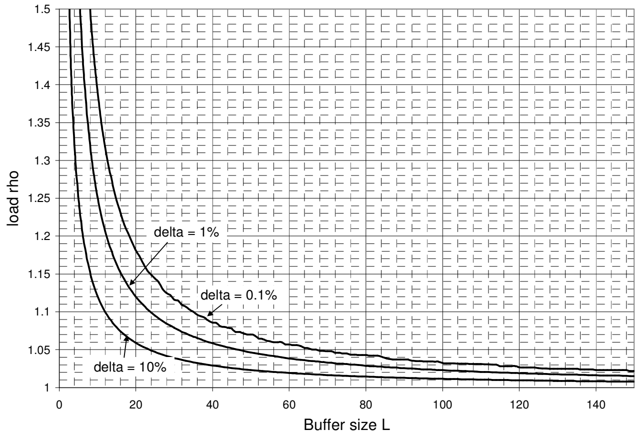

Fig. 4: The value of $\rho_{max}(L)$ for Δ = 10%, 1%, and 0.1%.

The Quality of Experience (QoE) constraint in HDTV is typically that a user should experience no more than one visible distortion per day (1 day = 12 hours) [6]. In figure 5 we compare this QoE constraint with the value of $T_{MTBVD}(\rho)$ for the conventional unprotected system and the systems with retransmissions (N = 1, 2, 3, and 4) for ρ > 0.85. We notice immediately that the conventional system is unable to meet the QoE requirement of HDTV for loads this high. The systems with retransmissions, on the other hand, do meet the QoE requirement for loads up to more than 0.92 (N = 1) or even more than 0.99 (N = 3, 4).

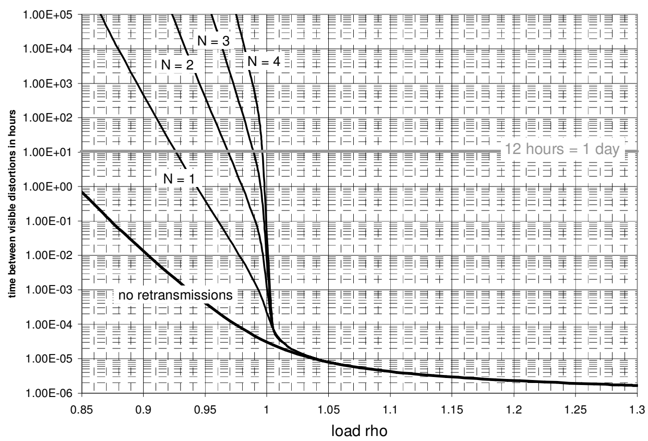

Figure 5: T_{MTBVD} [h] for the systems with and without retransmissions for L = 40.

[2] Note that in case the packet losses are correlated, these losses will tend to occur in clusters where each cluster will be observed as one single visible distortion. In that case, (8) will correspond with a worst-case scenario.

Nevertheless, a real-time video stream cannot tolerate too high a latency. Although the constant time the video packets are kept in the receiver buffer (in order to allow for retransmissions) is not that problematic in steady state, this buffer has to be build up when the user tunes in to the channel. In that way this latency at the receiver side contributes to the zapping delay, which should be kept low enough. This implies that we cannot apply just any retransmission schedule. Depending on the Round Trip Time (RTT) between the video source and the ingress node, we have to restrict the maximum number of retransmissions. In fact, under the assumption that one copy of the packet is retransmitted per retransmission event, the maximum allowed number of retransmissions is the (additional) zapping delay a user is willing to tolerate divided by the RTT.

IV. Conclusions

The object of this paper was to evaluate the impact of retransmissions on the packet loss ratio due to buffer overflow in a network transporting video flows. From the evaluation it was obvious that retransmissions have real merit in fighting the problem of buffer overflow for a M/D/1/L buffer system. In fact, they allow for a reduction of the PLR by several orders of magnitude for video loads in excess of one. This superior performance allows for the achievement of the QoE constraint of HDTV up to video loads close to one, substantially outperforming the unprotected conventional system. This superior performance is even achieved when no more than one single retransmission is permitted. Although the performance can be stepped up further when more retransmissions are allowed, this higher number of retransmissions cannot prevent the rather poor performance (in terms of QoE) at video loads in excess of one.

Acknowledgement

The second author is a Postdoctoral Fellow with the Fund for Scientific Research, Flanders (FWO-Vlaanderen), Belgium.

References

[1] "Beyond HSI: Reinventing the B-RAS", Heavy Reading Report, December 2005.

[2] F. Vanhaverbeke, M. Moeneclaey, D. De Vleeschauwer, "Retransmission strategies for an acceptable quality of HDTV in the wireless and wired home scenario", International Journal on Communication Systems, No. 20, pp. 297-311, 2006.

[3] A. Willig, "A Short Introduction to Queueing Theory", http://www.tkn.tu-berlin.de/curricula/ws0203/ue-kn/qt.pdf, 1999.

[4] D. R. Doll, *Data Communications, Facilities, Networks, and System Design,* New York: Wiley, 1978.

[5] H. O. Burton and D. D. Sullivan, "Errors and error control," *Proc. IEEE*, vol. 60, pp. 1293-1301, November 1972.

[6] "Parameter values for the HDTV+ standards for production and international program exchange," ITU-R recommendation BT. 709-5, 2002.

[7] D. Hoffman, V. Goyal, M. Civanlar, and G. Fernando, "RTP Payload Format for MPEG1/MPEG2 Video," *IETF RFC2250*, January 1998.