On the impact of adaptive RED in IP networks transporting H.264/MPEG-4 AVC video streams


Article info

Abstract

This paper investigates the impact on the QoS offered by IP networks that transport real-time H.264/MPEG-4 AVC video streams when applying adaptive random early detection (ARED) as an active queue management technique instead of traditional drop-tail. Simulation results indicate that the ARED technique reduces the loss of video packets but degrades a higher number of video frames if compared to the drop-tail technique, while showing a small qualitative gain regarding delay and jitter. From these results it may be reasonably concluded that H.264/MPEG-4 AVC video streams do not benefit from employing ARED technique as VoIP streams do.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Multimedia applications are getting more and more common, consuming an ever growing share of the resources of the world wide web. Two main reasons explain this growth. The first one is the dissemination of broad-band access. Real-time video streams require large bandwidth, so access networks with larger transmission capabilities naturally stimulate the transmission of large amounts of information like digital video streams. The second reason is the improvement level that the technology of voice, audio and video compression reached in the last few years, particularly in the case of the H.264/MPEG-4 AVC video compression standard.

Considering the growing importance that H.264/MPEG-4 AVC streams have in packet switched wireline networks that use the internet protocol (IP) for data transport, as well as the high variability of the traffic generated by this codec, this work assesses the gain in terms of quality of service (QoS) of H.264/MPEG-4 AVC streams when the active queue management (AQM) technique called adaptive random early detection (ARED) [1,2] is applied. The results are then compared to the ones obtained through the application of the traditional drop-tail technique. The proposed topology was simulated in the network simulator 2 (NS2) software and the parameters of QoS measured were loss of packets, delay and jitter. This work is an extension of a previous paper by the present authors [3].

The rest of this paper is organized as follows: Section 2 contains an overview of video transmissions over IP networks and some of the background concepts that led to the development of this work; Section 3 describes the experiment setup in terms of network topology and details of the video sequences and tools used for coding and generating the trace files used in the simulations; Section 4 shows the results obtained and discusses them comparatively between the ARED and drop-tail technique; and Section 5 shows the conclusions of this work.
2. Overview of H.264 video transmission over IP networks

Since the end of the 1980s, TCP has been used to transport the majority of internet content, like e-mail and web traffic, obeying the so-called TCP traffic paradigm [4]. Nevertheless, when the best-effort characteristic of IP and the reliability mechanisms of TCP are combined, the usage of TCP/IP for the transmission of real-time streams might not be the best choice. A large part of the multimedia applications running over the internet, such as VoIP, real-time audio/video streams and video-conference, have very strong conversational characteristics and restrictions in terms of transmission times and bandwidth. UDP is usually more suitable to transport these data. But UDP was not originally developed for the transport of real-time audio, voice and video through the internet, it is just the best choice after TCP. For UDP to be used in the transmission of real-time multimedia contents, another protocol is necessary. With that objective in mind, real-time transport protocol (RTP) was developed, offering end-to-end delivery services for data with real-time characteristics. These services include the identification of the type of data being transported, sequence number, time markers and delivery monitoring. Applications usually run RTP on top of UDP in order to use its multiplexing services and error detection capabilities. There is an extensive study [5] that points out native RTP, together with UDP/IP, as the best choice to transport real-time H.264/MPEG-4 AVC streams.

2.1. H.264/MPEG-4 AVC standard overview

Reaching compression rates 50% better than MPEG-2 standard and 30% better than MPEG-4 Part 2 for the same level of perceived quality, the H.264/MPEG-4 AVC standard adds up the most modern techniques of video coding and decoding [6]. The dramatic increase in the compression rate has made H.264/MPEG-4 AVC the standard elected for the majority of the new products and services for video transmission and storage. Examples are video services over DSL lines, video broadcasting to wireless mobile devices, internet video players and BluRay/HD-DVD optical disks standards [7].

The H.264/MPEG-4 AVC standard defines two layers: the video coding layer (VCL) and the network abstraction layer (NAL). The VCL consists of the central compression mechanism, where the syntactic levels commonly known as block, macro-block and slice are found. It was developed to be as independent of the network as possible. On the other hand, the NAL abstracts and adapts the data generated by the VCL so as to adequately them to the transport through a great variety of communication channels, as well as storage media. Looking to facilitate the integration with the communication channel, the NAL specifies both byte stream and packet stream formats. The packet stream format is used in applications based on the combined use of the protocols RTP/UDP/IP [8] as is the case of this work.

2.2. Constant bit rate versus variable bit rate

There are two main ways of coding video signals: with fixed quantization scales, which result in almost constant quality at the expense of variable bit rate; and with rate control, which adapts the quantization scales so as to keep the transmission rate almost constant, at the expense of variable quality.

This work focuses on video coding using a fixed quantization scale in order to examine the fundamental traffic characteristics of the H.264/MPEG-4 AVC standard [6,9] and its implications when the drop-tail and ARED techniques are applied.

2.3. Traffic variability

Auwera et al. [6] showed that the traffic variability generated by the H.264/MPEG-4 AVC standard is significantly higher than the traffic variability generated by the MPEG-2 and MPEG-4 Part 2 standards. While the frame size variation coefficient (defined as the standard deviation divided by the average) reaches levels above 2.4 in the case of H.264/MPEG-4 AVC, the same coefficient is not higher than 1.5 when considering MPEG-4 Part 2. Such a coefficient level above 1.5 is unprecedented.

The same authors demonstrate in [10] that the higher traffic variability generated by the H.264/MPEG-4 AVC coder significantly increases the frame losses if compared to the MPEG-4 Part 2 standard in the transmission of video streams over a bottleneck link.

Such conclusion led us to consider the application of ARED technique in order to evaluate a possible minimization of packet losses and frame damages, if compared to the traditional drop-tail technique.

2.4. Congestion control techniques

The objective of AQM is to reduce the size of the queues and their oscillations, maintaining a high level of link utilization through a fair allocation of resources. TCP traffic sources react to dropped packets reducing its transmission window and consequently its transmission rate. AQM techniques that are more efficient and fair than drop-tail in the congestion control of IP networks have been proposed by several authors.

Reguera et al. [11] present an extensive study on the impact of AQM in the QoS of VoIP applications. One of the more representative AQM algorithms is analyzed through computational simulation, assessing its effect on the perceived voice quality. The degradation of the transmission is linked to the users’ perception through well known algorithmic models,
expressing their satisfaction in the mean opinion score (MOS) scale [12]. The main results show that AQM mechanisms such as ARED significantly improve the perceived voice quality. This study is enlarged in [13], incorporating other congestion control mechanisms such as random exponential marking (REM) [14] and adaptive virtual queue (AVQ) [15].

The present work focuses on the application of the ARED technique in order to assess the QoS of H.264/MPEG-4 AVC because this technique showed an expressive qualitative gain in the transmission of VoIP streams, which, like real-time video streams, have serious restrictions in terms of timing and loss of information packets [11,13].

3. Experiment setup

The methodology chosen for this work was the computational simulation of an IP network that transmits real-time H.264/MPEG-4 AVC video streams, disputing network resources with TCP traffic which is the most commonly type of traffic found in the internet.

The video samples were coded off-line using the JM reference software version 13.0 [16]. Simulations were implemented in the network simulator version 2 (NS2) [17], using the additional code for the transmission of video streams from trace-files developed by Chih-Heng Ke et al. [18]. Data analysis was performed off-line from the simulation’s trace files.

3.1. Simulated network topology and configuration

Over a dumbbell topology, where routers R0 and R1 form the main link, secondary links were created between the routers and the traffic sources as shown in Fig. 1. The secondary links for the TCP sources were configured with 10 Mbps data rate and different fixed propagation delays selected between 0 and 50 ms, while the link used by the video source was configured with a data rate of 10 Mbps and a fixed propagation delay of 2 ms. The main link was configured with 2 Mbps data rate and propagation delay of 5 ms, characterizing the transmission bottleneck. This topology was selected because it has been traditionally used to evaluate the performance of AQM techniques [19,20].

Each round of simulation lasts 50 s. The TCP streams were started in times that vary from 0.1 to 0.7 s. Each video sample is started at \( t = 5 \) s and stopped at \( t = 49 \) s, i.e., each 10 s video sequence is repeated 4.4 times.

Initially a set of five TCP streams was created in the same direction of the video stream, with packets of different sizes (120, 300, 500, 900 and 1200 bytes). Then this set of five streams was successively multiplied by 2, 3, 6 and 12, reaching a total of 60 TCP concurrent streams in the fifth simulation round. The sizes of the TCP packets were chosen based on empirical data obtained through the capture of packets during laboratory tests using routers and switches working together with Web and VoIP devices. We could observe that sizes around those values were the most common. This could also be considered a worst case scenario since the AQM algorithm suffered a higher level of stress because it should handle the queue with packets of varied sizes.

The two routers that form the main link were simulated in two different scenarios regarding queue management: first scenario considered traditional drop-tail queue while the second scenario considered ARED technique. In both scenarios the queues were set to 120 packets. The ARED implementation is in accordance with the proposal that appears in [20]. Since, ARED has the capacity of dynamically adjust its configuration in response to network condition the only parameter that needs to be specified is the average queue length or equivalently the mean queue delay. In the simulations, the average queue delay for the ARED algorithm was set to 40 ms. Besides that, in the ARED scenario, the two possible modes of operation were simulated: one mode where queue size is calculated in bytes and other mode where queue size is determined in

Fig. 1. Simulated network topology.
This means that the five simulation rounds (one for each set of TCP loads) were carried out in each of the three modes of queue management under consideration, giving a total of 15 simulation runs.

3.2. Simulated video streams

The H.264/MPEG-4 AVC video streams were generated from raw video sequences in the CIF format ($352 \times 288$ pixels, 4:2:0 YUV), originally sampled at 30 frames per second. The chosen sequences were news_cif.yuv and coastguard_cif.yuv, both available at [http://trace.eas.asu.edu/yuv/index.html](http://trace.eas.asu.edu/yuv/index.html).

These two video sequences were chosen because they present rather diverse characteristics, both in terms of compression ratio and traffic variability. While coastguard_cif presents a high rate of image variation between frames, the sequence news_cif presents images with low variation rate. The result is high compression ratio and high traffic variability in the case of the video sequence news_cif, and low compression ratio and low traffic variability in the case of the video sequence coastguard_cif. Table 1 lists the most relevant statistical data of the encoded video sequences.

Fig. 2 shows the frame size distribution for both video sequences. Video sequence news_cif has a greater number of small frame sizes with low variation in their sizes. The other video sequence, coastguard_cif has larger frame size and more variation in their sizes. The two clusters for each sequence correspond to P frames (smaller size) and I frames (larger size).

Encoding was done through the JM reference software version 13.0, using the ‘High’ profile. The GOP (group of pictures) used was one intra-coded frame (type I) for every 11 inter-coded frames (type P), with no bi-directional inter-coded frames (type B), resulting in a closed GOP of the form IPPPPPPPPPPP. The quantization parameter was set to 28, and each frame was segmented in packets of 200 and 1024 bytes, respectively, generating the video samples called ‘News 200’, ‘News 1024’, ‘Coastguard 200’ and ‘Coastguard 1024’.

The video encoding bit-rate control in the JM reference software was disabled for two reasons: (i) coding was done offline; (ii) the objective of the simulation is to assess the behavior of the other concurrent streams interacting with a high variability stream. So the output of the encoder resulted in a video stream of unconstrained variable bit rate (VBR), without any interaction between the encoder and network conditions.

Through the MP4TRACE tool, the encoded video sequence was used as the input for the generation of a text-type trace file, which contains all the necessary information for the NS2 to create a simulated stream that behaves like the original encoded stream in terms of frame sequence number, type of frame, quantity and size of frame segments (packets) sent, and transmission timing [21]. The tools that were used are described in [22].

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Statistical data of the encoded video sequences.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>News_cif</td>
</tr>
<tr>
<td>Number of frames</td>
<td>300</td>
</tr>
<tr>
<td>Compression ratio</td>
<td>11.54</td>
</tr>
<tr>
<td>Average frame size (bytes)</td>
<td>1392.54</td>
</tr>
<tr>
<td>Minimum frame size (bytes)</td>
<td>70</td>
</tr>
<tr>
<td>Maximum frame size (bytes)</td>
<td>8603</td>
</tr>
<tr>
<td>Peak-to-mean ratio</td>
<td>6.18</td>
</tr>
<tr>
<td>Total length of the sequence (bytes)</td>
<td>417761</td>
</tr>
<tr>
<td>Average transmission (bits/s)</td>
<td>334208.8</td>
</tr>
</tbody>
</table>

Fig. 2. Frame size distribution of the video sequences.
4. Results and discussion

Video information losses were calculated as percentages of the total number of video packets and video frames transmitted on one simulation. All frames that suffered the loss of one or more packets were considered damaged.

4.1. Average delay and standard deviation of delays

Figs. 3–6 show the average delay and standard deviation of delays from the video samples ‘News 200’, ‘News 1024’, ‘Coastguard 200’ and ‘Coastguard 1024’, respectively, comparing drop-tail and ARED in both byte and packet modes.

ARED operating in both modes shows a significant advantage over drop-tail under the conditions simulated, as expected from its characteristic of diminishing the oscillations at the end of the queue.

The results indicate that the ARED technique is able to keep the average delay of the video packets relatively close to the configured queuing delay (40 ms) even under severe congestion conditions. Besides that, measured jitter, in most of the traces was comparable or lower than the achieved by the drop-tail technique. Such results lead to the following perceptions: (i) the losses caused by excessively late arrivals can be minimized; (ii) the size of the jitter buffer needed to absorb the delay variations can also be minimized. However this metrics do not reflect the damages caused by packet losses due to the early discarded mechanism of ARED. The impact of packet losses on the video stream is presented in the next subsections.
**Fig. 5.** Video sample news 1024 – average delay and standard deviation.

**Fig. 6.** Video sample coastguard 1024 – average delay and standard deviation.

**Fig. 7.** Video sample news 200 – percentage of packets lost.
4.2. Packet losses and video information losses

In a direct comparison between drop-tail and ARED techniques (in both byte and packet modes), the results obtained for the video samples segmented in packets 200 bytes long, which show a smaller percentage of packets lost when using the ARED technique, are in accordance with the behavior expected from this technique (Figs. 7 and 8).

On the other hand, the results obtained for the video samples segmented in packets 1024 bytes long show a slight advantage when using the drop-tail technique in comparison to the ARED technique, in both byte or packet modes (Figs. 9 and 10).

These results clearly show how the ARED technique favors the small packets (200 bytes), and how indifferent it is for relatively large packets (1024 bytes) within the simulated conditions.

---

**Fig. 8.** Video sample coastguard 200 – percentage of packets lost.

**Fig. 9.** Video sample news 1024 – percentage of packets lost.

**Fig. 10.** Video sample coastguard 1024 – percentage of packets lost.
Figs. 11–14 show a direct comparison between drop-tail and ARED techniques (in both byte and packet mode) regarding the percentage of damaged frames affected by the loss of one or more packets.

Except for one scenario, the measurements in all the simulated scenarios showed that ARED caused damages in a larger number of frames than drop-tail did. The only scenario in which ARED damaged less frames than drop-tail was when using the video sample 'News 200' operating in byte mode.

These are interesting results because they show a negative characteristic of the ARED technique when transporting video streams, one that is masked by the smaller number of lost packets that it imposes to the video sequence if compared to drop-tail: the number of damaged video frames is significantly higher due to the distribution of the discards in regular time intervals.

**Fig. 11.** Video sample news 200 – percentage of damaged video frames.

**Fig. 12.** Video sample news 1024 – percentage of damaged video frames.

**Fig. 13.** Video sample coastguard 200 – percentage of damaged video frames.
Figs. 15–18 show the percentage of damaged ‘I’ frames, according to the congestion control mechanism under consideration. Again, the negative characteristic of distributing the discards over time of the ARED technique appears, affecting a larger number of ‘I’ frames in all scenarios if compared to drop-tail.

4.3. Peak signal-to-noise ratio (PSNR)

Among all the simulated scenarios the one that best represents, in terms of an objective metric like PSNR, what the above results indicate is when the video sample ‘News 200’ is used with ten concurrent TCP sources. As previously shown in Figs. 7 and 11, in this scenario the percentage of lost packets and damaged frames is lower when applying ARED in byte mode if compared to the drop-tail technique. However, the measurement of PSNR over 300 video frames of ‘News 200’ shows a worse performance of ARED in byte mode, which is coherent with the percentage of damaged ‘I’ frames (Fig. 15), which
shows a slight advantage of drop-tail. This last comparison is represented in the figures below. Fig. 19 shows the PSNR measure-
ment when applying the drop-tail technique and Fig. 20 shows the same measurement under ARED in byte mode. The
former averages to 13.55 dB and the latter to 12.65 dB, so the reconstructed video frame is less distorted when drop-tail is
configured in the routers.

The observed performance of ARED in terms of PSNR suggests that a differentiating drop queuing mechanism could im-
prove the quality of the video streams. In [24] weighted random early detection (WRED) was proposed. It is an extension of
RED where different queue thresholds are associated to eight different classes of traffic. This cause that lower priority pack-
ets are dropped early, hence protecting the higher priority packets as congestion increase. Using WRED and marking video
packets as priority packets, certainly could produce an improve in the PSNR. Due to the classification procedure this mech-
anism increases system complexity and processing load as compared to ARED. However, the current state of the art of rout-
ing technology permits its deployment in practical scenarios.
5. Conclusion

The present work assessed QoS issues related to H.264/MPEG-4 AVC video transmission over best-effort IP networks. It was motivated by the expressive performance gain obtained by VoIP applications when the ARED mechanism is used instead of the traditional drop-tail as demonstrated by Reguera et al. using simulated bottleneck links [11,23]. A similar network topology was employed and the results obtained in the simulations show that ARED queue management yields more packet loss than drop-tail queue on many scenarios. In one scenario (200 bytes packets) ARED exhibited less packets loss, but in all simulated scenarios it damaged a higher percentage of video frames yielding smaller video PSNR than drop-tail queue. Also relevant is the fact that the percentage of damaged ‘I’ frames (the most important type of frame in compressed video streams) is significantly higher in all scenarios when ARED is applied, which has a great impact in the PSNR values of the video frames. Regarding delay and jitter, there is a slight advantage of ARED over drop-tail, but this advantage is of little significance under the conditions considered for this work.

Considering all the above, we may reasonably conclude that H.264/MPEG-4 AVC video streams do not benefit from ARED technique as much as VoIP streams do. Actually ARED causes a poorer performance in terms of objective quality (PSNR) when compared to drop-tail, at least under the simulated conditions.

Acknowledgement

Daniel Pioli Torres would like to thank the Brazilian agency CAPES for granting the Masters Degree scholarship that allowed for the development of this research.

References

Daniel Pioli Torres received the B.Eng. and M.Sc. degrees in Electrical Engineering from Federal University of Parana, Brazil, in 1994 and 2009, respectively. He worked at Telepar from 1996 to 1999, before the privatization of the Brazilian telecommunications sector. After the privatization, he worked at Telepar mirror-company GVT from start-up, in 2000 to 2005. In 2006 he entered the M.Sc. course at UFPR and worked as professor in a private institution until 2008, when he joined Huawei in Campinas, working as Access Network Engineer. In 2010 he left Huawei to join the family business back in Curitiba.

Evelio M. G. Fernandez received the B.Eng. degree in Electrical Engineering from the Central University of Las Villas (UCLV), Cuba, in 1985. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical engineering from the State University of Campinas, Brazil in 1997 and 2001, respectively. He is currently an associate professor at the department of Electrical Engineering at the Federal University of Parana. His current research interests include channel coding techniques, digital communications and wireless networks.

Eduardo Parente Ribeiro is a Professor at Electrical Engineering Department of Federal University of Parana, Brazil, since 1997. He received Bachelor degree on Electrical Engineering (1990), M.Sc. (1992) and Ph.D. (1996) from Pontifical Catholic University of Rio de Janeiro. He did research stage at Vanderbilt University in 1995 and post-doctoral stage at The University o British Columbia in 2006. His interests include data communication, multimedia modeling and signal processing.

Vitalio A. Reguera received the B.Eng. degree in Telecommunications and Electronics Engineering from the Central University of Las Villas (UCLV), Cuba. He received the M.Sc. degree in Telecommunications Engineering and the Ph.D. degree in Electrical Engineering from the same university, in 2000 and 2007, respectively. He is currently Professor at the Department of Telecommunications and Electronics, UCLV. His current research interests include cognitive radio, communication protocols, quality of service and wireless networks.

Cezar A. G. Oliveira is currently a B.Eng. degree senior student at the Department of Electrical Engineering, Federal University of Parana, Brazil. He was part of the “Information Processing and Transmission Research Group” from 2008 to 2010. His current interests include network infrastructures and project management.