SERVER-SIDE PLAYOUT DELAY MANAGEMENT FOR VIDEO STREAMING

Peter van Beek and Louis J. Kerofsky

SHARP Laboratories of America, Camas, WA 98607, USA, Email: pvanbeek@sharplabs.com

Summary

We examine two schemes for server-side management of playout (startup / preroll) delays due to receiver buffering:

1. Stream Acceleration with Rate Adaptation (SA-RA): combines stream acceleration with bit rate adaptation
2. Adaptive Media Playout by Server Media Processing (AMP-SMP): achieves adaptive media playout (AMP) by server-side frame interpolation combined with stream acceleration and bit rate adaptation

Both schemes reduce playout delay without decreasing playout robustness.

NS-2 simulations with streaming CIF, 4CIF and HD video over wireless IEEE 802.11 channels show:

- End-to-end delay can easily be grown dynamically using such schemes, e.g. from 100 to 500 ms;
- PSNR is close to scheme with constant 500 ms end-to-end delay, but with 100 ms playout delay;
- PSNR is up to several dB better than scheme with constant end-to-end delay of 100 ms.

1. Background

Streaming video systems utilize a network or dejitter buffer at the receiver to absorb variations in packet network transmission time. Buffering reduces the probability of packets arriving late for decoding (or interruptions of decoding and playout). Also, a pre-decoder buffer is needed to absorb variations in the size of coded video frames, and may be combined with the dejitter buffer.

Initial buffering of video data introduces undesirable playout delay experienced by the viewer. In conventional systems, the end-to-end delay is constant, and reducing the probability of buffer underflow (by increasing the amount of data buffered) results in increased playout delay.

Delay and buffering constants for conventional video transmission over CBR and VBR channels is discussed in [1] and [2].

Adaptive Media Playout (AMP) [3][4] is a client-side technique, where the video playout rate is adapted to the fullness of the client playout buffer. AMP can be utilized to reduce playout delay without decreasing playout reliability.

Playout rate at startup may be reduced temporarily, such that the rate at which video frames are retrieved from the client buffer for decoding and playout is lower than the rate at which video frames arrive from the channel. Once the amount of video data in the buffer has grown to a sufficient level, playout at the nominal rate starts.

Potential disadvantage of AMP: additional processing required at the receiver, e.g. for audio time scaling. Also, a server may be required to operate with legacy clients incapable of AMP.

Stream Acceleration (SA) is a server-side technique included in well-known commercial streaming media products. The basic SA technique is to adapt the rate at which video frames are transmitted by the server. At startup, the number of frames transmitted per second may be temporarily increased beyond the nominal frame rate. This results in a rapid increase of the client buffer fullness. SA can only be applied to streaming stored video. Potential disadvantage of SA: requires additional channel bandwidth, which may not be available. If available, it remains unused once the number of frames transmitted per second returns to the nominal frame rate.
2. Stream Acceleration with Rate Adaptation (SA-RA)

SA-RA combines basic stream acceleration with bit rate control. During stream acceleration (e.g. when starting a stream), the server:
- increases the number of video frames transmitted per second, and:
- reduces the bit rate of the video stream if necessary.

Example is illustrated in figure on the right:
- Initial end-to-end delay \( \Delta T_{S} \) can be small: \( \Delta T_{S} = t_{2} - t_{0} \).
- End-to-end delay is grown dynamically (to increase playout reliability) to its final value \( \Delta T_{E} = t_{2} - t_{0} \).

Advantages:
- Compared to a constant end-to-end delay schedule starting at \( t_{0} \), SA-RA achieves reduced play-out delay, while minimizing the impact on the bit rate during the startup stage.
- Compared to a constant end-to-end delay schedule starting at \( t_{1} \), SA-RA achieves increased playout reliability and possibly higher video bit rate.

Growing the receiver buffer size over time and the impact of this on rate control is also discussed in [5].

3. Adaptive Media Playout by Server Media Processing (AMP-SMP)

AMP-SMP is a novel technique to achieve adaptive media playout.
- Video frame repetition/interpolation and audio time scaling are performed at the server instead of the client. This increases the duration of a portion of the original stream. The server may also modify the presentation time stamps accordingly.
- Server transmits the converted stream at increased number of frames per second. This results in rapid growth of the client buffer.
- Also, rate control is applied to adapt the video bit rate depending on the increase in number of frames and channel conditions.
- The client plays the video frames at nominal frame rate.

Example is illustrated in figure on the right:
- Initial end-to-end delay (and playout delay) \( \Delta T_{S} \) is equal to \( t_{1} - t_{0} \).
- During startup stage, original and interpolated frames generated at server are adapted and transmitted within the duration used normally to code and transmit only the original frames.
- After startup stage, the end-to-end delay \( \Delta T_{E} \) is equal to \( t_{2} - t_{0} \).

Advantages:
- AMP-SMP can be applied to a live audio/video streaming scenario (like AMP itself).
- A server can operate with legacy clients without AMP capability.
- Computational cost for AMP removed from client. Computational resources for AMP-SMP at server only needed at startup of the client. May scale reasonably with multiple clients because clients may not tune in simultaneously.
- Frame interpolation and rate adaptation processes at server can be jointly optimized for best visual quality and least computational complexity.

4. Video Streaming over Wireless LAN using SA-RA or AMP-SMP

Our main application: robust transmission of high-quality video streams over wireless home network, based on IEEE 802.11 a/b/g [6][7]. Video is in single-layer (non-scalable) MPEG-2 format.

Video transcoder implements real-time bit rate scaling. Transcoder must operate at increased frame rate. In experiments, we use low-complexity open-loop requantizing transrater to reduce bit rate.

Bit rate is adapted to current channel conditions, buffer backlog at server and instantaneous delay constraints [7]. Available bandwidth is estimated at the client and fed back to server every time a frame is received.
5. Simulation Results

Server and client implemented in network simulator NS-2. NS-2 simulates UDP, IP and MAC layers. Channel traces captured from real-world measurements of wireless links to simulate PHY and wireless medium:

- **802.11B (2.4 GHz)** – average bandwidth = 4.5 Mbps (100 sec. trace)
- **802.11A (5 GHz)** – average bandwidth = 14 Mbps (10 sec. trace)

Test video streams – MPEG-2 IBP encoded with TM-5 MSSG reference software:

- **Mobile**: 352x288 (CIF) 30p @ 4 Mbps 802.11B
- **Crew**: 704x576 (4CIF) 30p @ 6 Mbps 802.11B
- **Harbour**: 1280x720 (HD) 60p @ 16.9 Mbps 802.11A
- **Raven**: 1280x720 (HD) 60p @ 16.9 Mbps 802.11A

Lost and late packets incur losses in video bit-stream at slice level (with concealment from a previous frame).

Schemes compared:

- **Reference**: Basic delay-constrained bit rate adaptation to channel, with constant end-to-end delay of 100 ms.
- **Reference**: Basic delay-constrained bit rate adaptation to channel, with constant end-to-end delay of 500 ms.
- **SA-RA** with delay-constrained bit rate adaptation to channel, with end-to-end delay growing from 100 ms to 500 ms. Stream acceleration is applied only for 4 seconds (this corresponds to a 10% frame rate increase).
- **AMP-SMP** with delay-constrained bit rate adaptation to channel, with end-to-end delay growing from 100 ms to 500 ms. Frame repetition was achieved by duplicating 2 B-frames in every GOP for the first 12 GOPs. This increases the duration of the first 3 seconds of video by 400 ms (corresponding to a 13% playout rate reduction).

<table>
<thead>
<tr>
<th>Average PSNR-Y (dB)</th>
<th>Mobile CIF / 802.11b</th>
<th>Crew 4CIF / 802.11b</th>
<th>Harbour HD / 802.11a</th>
<th>Raven HD / 802.11a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic RA: ( \Delta T_s = \Delta T_e = 100 \text{ ms} )</td>
<td>28.4</td>
<td>33.4</td>
<td>29.9</td>
<td>38.4</td>
</tr>
<tr>
<td>Basic RA: ( \Delta T_s = \Delta T_e = 500 \text{ ms} )</td>
<td>31.7</td>
<td>34.2</td>
<td>32.1</td>
<td>39.7</td>
</tr>
<tr>
<td>SA-RA: ( \Delta T_s = 100 \text{ ms}; \Delta T_e = 500 \text{ ms} )</td>
<td>31.8</td>
<td>34.1</td>
<td>31.8</td>
<td>39.3</td>
</tr>
<tr>
<td>AMP-SMP: ( \Delta T_s = 100 \text{ ms}; \Delta T_e = 500 \text{ ms} )</td>
<td>-</td>
<td>-</td>
<td>31.6</td>
<td>39.2</td>
</tr>
</tbody>
</table>

6. Conclusions

- **SA-RA** and **AMP-SMP** are two server-side schemes for reduction of playout delays without reducing playout robustness.
- These schemes extend existing techniques by utilizing video bit rate adaptation (and frame interpolation) at the server, considering a varying end-to-end delay, and considering available channel bandwidth.
- Compared to reference scheme with 100 ms end-to-end delay, a PSNR gain of several dB can be obtained.
- Compared to reference scheme with 500 ms end-to-end delay, some of the video quality during a short period of time is traded off for a large reduction in playout delay.

7. References