A Guideline to Video Codecs Delay

Malek Maged Al-N’awashi
Department of Information Technology, Al-Balqa Applied University / Al-huson College,
PO box 50, Al-Huson, Irbid – Jordan

Obaida Mohammad Al-Hazaimeh
Department of Information Technology, Al-Balqa Applied University / Al-huson College,
PO box 50, Al-Huson, Irbid – Jordan
dr_obaidam@yahoo.com

Abstract
Due to the explosive growth of the internet and increasing demand for multimedia information on the web, streaming video over the Internet has received tremendous attention from academia and industry. Continuous media has a lot of issues and challenges. The most important among of them are delay. Delay is an expression of how much time it takes for a packet of data to get from one point to another. Some of delay source are fixed while the other is not. This paper describes the delay sources and magnitude of the most common video codecs and thus provides a guideline for the choice of the most suitable codec for a given application.

Keywords: Continuous media, Processing delay, Encoder, Decoder

1. Introduction
Telecommunication worldwide has experienced a significant revolution over recent years. The rapid convergence of data, voice, and continuous media (CM), such as streaming video and videoconferencing (video) using IP-based network is delivering advanced services at lower cost across the spectrum including residential users, business customers of varying size, and service providers [1-3]. However, continuous media has a lot of issues and challenges. The most important among of them is delay which effect on the quality of services (QoS) [1]. In term of delay, CM suffers from delay introduced by many sources such as digital conversion (coder), compression, packetization, transmission, decompression, digital conversion vice versa (decoder), encryption, and decryption [3 - 4].

2. Packet flow for continues media
As mentioned in the previous section, the most of the delay source in real time application is: handling delay (coder delay, decoder delay, and other Digital Signal Processing (DSP) such as encryption and decryption processes, serialization delay, queuing delay, propagation delay, and network delay [4-5]. Figure1 shows the delay sources.

![Figure 1. Delay Source of the Video Packet](image-url)
One of the keys problems is keeping end-to-end delay blow human perception. Therefore, The International Telecommunications Union (ITU) has recommended the end-to-end delay of the video should be blow 100 ms [4] as shown in equation 1:

\[
\text{End-to-end Delay} = \text{PD} + \text{ND} + \text{NRD} + \text{PRD} \leq 100 \text{ ms}
\]  

(1)

2.1 Processing delay

Processing delay consider one of the most delay sources of real time application such as video conferencing [1]. Processing delay consist of coder delay, decoder delay, and other digital signal processing.

2.1.1 Coder delay

A video codec is a device or software that enables compression or decompression of digital video [6]. In details, video coders operate on collections of video samples known as frames. Each block of input video samples is processed into a compressed frame. The coded video frame is not generated until all video samples in the input block have been collected by the encoder (encoder buffer). Thus, there is a delay of one frame before processing can begin. In addition many coders also look into the succeeding frame to improve compression efficiency [7-8].

When the video frames are generated, packetization delay problem rises up. Packetization is the time taken to fill a packet of encoded or compressed video frames. It can be also called accumulation delay, as the video samples accumulate in a buffer (encoder buffer) before they are released. Since each voice sample experiences look-ahead and packetization delays, then these processes are overlapping [9].

The video stream requires high capacity and since network resources are limited, the video pictures should be compressed, in this paper, we have assumed a set of codecs, since they are the most popular techniques, and we have calculated the delay time for each one of them in term of minimum codecs delay in IP environments in two phases: phase one, single frame per packet, and phase two, multi frames per packet where size of the packet around 25 ms.

2.1.1.1 Single frame per packet

The time required to process an input frame is assumed to be the same as the frame length since efficient use of processor resources will be accomplished when an encoder and decoder pair fully uses the available processing power [4]. Thus, the delay through an encoder is normally assumed in the equation 2.

\[
\text{Delay} = 2 \times \text{Frame Size} + \text{Time to look at the succeeding frame}
\]  

(2)

2.1.1.2 Multi frames per packet

If multiple video frames are grouped together into a single IP packet, the delay through an encoder is assumed in the equation 3 [4].

\[
\text{Delay} = \# \text{ frames} \times \text{Frame Size} + \text{Time to look at the succeeding frame}
\]  

(3)

Based on above, we have calculated the delay introduced by each video codec's. As shown in the Table 1.

<table>
<thead>
<tr>
<th>Video codec</th>
<th>Application</th>
<th>Resolution</th>
<th>Bit-rate (kb/s)</th>
<th>Frame size (s)</th>
<th>Frame size (ms)</th>
<th>Delay (ms)</th>
<th>No. of frames / Packet</th>
<th>Delay (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. H.264</td>
<td>Live Messenger</td>
<td>G640x480</td>
<td>600</td>
<td>24</td>
<td>0.024</td>
<td>0.048</td>
<td>1042</td>
<td>25.024</td>
</tr>
<tr>
<td>2. H.264</td>
<td>G-talk</td>
<td>G512x380</td>
<td>1000</td>
<td>30</td>
<td>0.03</td>
<td>0.06</td>
<td>834</td>
<td>25.03</td>
</tr>
<tr>
<td>3. H.264</td>
<td>Skype</td>
<td>640x480</td>
<td>400</td>
<td>30</td>
<td>0.3</td>
<td>0.6</td>
<td>834</td>
<td>25.03</td>
</tr>
<tr>
<td>4. VP4</td>
<td>Skype</td>
<td>640x480</td>
<td>560</td>
<td>20</td>
<td>0.02</td>
<td>0.04</td>
<td>1250</td>
<td>25.02</td>
</tr>
<tr>
<td>5. MPEG-4</td>
<td>X-lite</td>
<td>240x180</td>
<td>100</td>
<td>10</td>
<td>0.01</td>
<td>0.02</td>
<td>2500</td>
<td>25.01</td>
</tr>
<tr>
<td>6. MPEG-4</td>
<td>Yahoo</td>
<td>320x240</td>
<td>256</td>
<td>30</td>
<td>0.03</td>
<td>0.06</td>
<td>834</td>
<td>25.03</td>
</tr>
<tr>
<td>7. WMA</td>
<td>Yahoo</td>
<td>320x240</td>
<td>350</td>
<td>20</td>
<td>0.03</td>
<td>0.06</td>
<td>834</td>
<td>25.03</td>
</tr>
<tr>
<td>8. WMA</td>
<td>Telebox</td>
<td>320x240</td>
<td>1356</td>
<td>30</td>
<td>0.03</td>
<td>0.06</td>
<td>834</td>
<td>25.03</td>
</tr>
<tr>
<td>9. WMA</td>
<td>Aim</td>
<td>176x144</td>
<td>22</td>
<td>5</td>
<td>0.06</td>
<td>0.01</td>
<td>5000</td>
<td>25.005</td>
</tr>
<tr>
<td>10. WMA</td>
<td>Aim</td>
<td>240x180</td>
<td>34</td>
<td>6</td>
<td>0.006</td>
<td>0.012</td>
<td>4167</td>
<td>25.006</td>
</tr>
</tbody>
</table>

Note: To make it clear, the following column charts showing the delay time for each of selected video codecs in Figure 2 and 3 respectively.
2.1.2 Decoder delay

At the receiver side, decompression and decoding processes take place [7]. Based in our experimental test, decoder delays can be assumed to be almost half of the encode delays. To make it clear, the following column charts showing the delay time for each of selected video codecs for encoder and decoder processes.
Figure 4: Encoder and Decoder Delay for Variety of Video Codec's

It is clear from figure 4, decoder delay in the most of video codec's can be assumed half of the encode delay.

3. Conclusion
It can be seen that there is a wide variety of video coders available, for a wide variety of applications. This paper described the basic delay sources of compressed video transmission system to provide a guideline for the choice of the most suitable codec. In other words, we have assumed a set of video codec, since they are the most popular techniques, and we have calculated the delay time for each one of them in two phases: phase one, single frame per packet, and phase two, multi frames per packet. And we have found coders where made for high quality, high delay, and high compression ratio as well as decode delays can be assumed to be almost half of the encode delays.

4. Acknowledgement
The authors would like to thank all the people who have supported this work, as well as special thanks to our colleagues from Al-Balqa Applied University.

References