MPEG Encoding and IPTV SLA Requirements

John Evans, Distinguished Consulting Engineer
john@cisco.com
MPEG Encoding and IPTV SLA Requirements

- The primary SLA requirements for IP-based video streaming services are defined in terms of:
  - Throughput
  - Delay/Jitter
  - Loss
  - Service Availability

- An understanding of the principles of MPEG compression is required in order to understand the SLA requirements for a transported MPEG video stream
Video Compression

- Typical uncompressed video bit rates are 270Mbps for SD and 1.485Gbps for HD sources
- Compression is achieved by removing redundant information or information to which the eye is less sensitive
  - Reduce color information (sub-sampling)
  - Remove spatial redundancy (intra-coding)
  - Remove temporal redundancy (inter-coding)
**MPEG Video Structure**

- **Sequence**
  
  The highest structure of the coded video bitstream is the video sequence.

  This is a sequence of frames containing a GOP or GOP’s with common format and content attributes.

- **GOP**

  A Group of Pictures (GOP) includes an I-frame and all frames up to the next I-frame.

  GOP sizes of 15 (30 fps) and 12 (25 fps) are typical.
MPEG Video Structure

- Frames

There are three main picture/frame types in MPEG-2:

**I-Frame**: An Intra-coded (I) frame is coded using information only from itself, i.e. without reference to other frames

**P-Frame**: a Predictive-coded (P) picture is a picture which is coded using motion compensated prediction from a previous reference frame

**B-Frame**: a Bidirectionally predictive-coded (B) picture is a picture that is coded using motion compensated prediction from a past or future reference frame
A **slice** is a series of an arbitrary number of consecutive macroblocks.

A **macroblock** contains a section of the luminance component and the spatially corresponding chrominance components.

A **block** is an 8x8 matrix of pixels or DCT values representing a small chunk of luma or chroma.
Frame Referencing

- **I-frames** provide reference to B and P-frames
  - Intra-coded only - reference frame for the GOP as well as last B-frames in the previous GOP (open GOP)
  - Limits propagation of errors, supports random access

- **P-frames** provide reference to B-frames and subsequent P-frames
  - Forward prediction from either previous I or P-frames

- **B-frames** do not provide reference to other frames (MPEG-2)
  - Bi-directional interpolated prediction from two sources:
    - Either previous or future reference I or P-frames
  - H.264/MPEG-4 does allow for B-frames to provide reference
GOP Picture Ordering

Two Picture Orders:

- **Display Order** – input to the encoder, output from the decoder

- **Coded (transmission) Order** – output of the encoder, input to the decoder

In an open GOP the B-frames \(B_{14}, B_{15}\) from the previous GOP may reference the next I-frame, and hence will need to be transmitted/received after the reference I-frame (Blue).
MPEG Encapsulation within IP

- A typical IP packet for the transport of MPEG video contains seven 188-byte MPEG Transport Stream (TS) packets.
- MPEG frames may traverse multiple IP packets and each IP packet may contain service information as well as audio and video MPEG packets.
MPEG-2 Stream Analysis

Avg. bytes per frame type

- High motion clips require larger P-frames and B-frames as the differences between the frames are significant and the temporal redundancy will be low and hence the I-frame size is reduced to fit the imposed encoding rate.

- A low motion clip has much greater difference between the I-frame size, and the P and B-frame sizes.

- Analysis for HD 1080i and 720p clips displayed the same characteristics.
MPEG-2 Stream Analysis

Total bytes per frame type

- In the low motion clip the I-frame constitutes almost half of the total number of bytes in the frame.

- In the high-motion clip, the B-frames constitute 59% of the total and the I-frame only accounts for 11%.
Video SLA Requirements

- The key SLA requirements for video streaming services are defined in terms of:
  - Throughput
  - Delay/Jitter
  - Loss
  - Service Availability
Video SLA Requirements

Throughput

- Throughput requirements depend upon encoding scheme used and quality
  - MPEG-2:
    - SD: 3-4Mbps
    - HD: 16-20Mbps
  - MPEG-4:
    - SD: 2-3Mbps
    - HD: 8-10Mbps

- Throughput requirements are addressed through capacity planning and QOS (i.e. Diffserv)
Video SLA Requirements

Delay / Jitter

- For video streaming services, network delay can impact end-user “interactivity”, or the “finger-to-eye” delay.

- For Broadcast Television services (such as IPTV or Cable TV), the impact that network delay has is on the time it takes for the end user to change from one TV channel to another a.k.a the “channel change time” or “channel zapping time”

- For VOD services the network delay impacts the finger-to-eye delay, i.e. the response time it takes for user requests to be translated into actions visible to the end user
Broadcast video channel change time

Legend
- Red = STB processing delays
- Blue = Network transmission delays (fixed | variable)
- Orange = Multicast processing

Remote control and STB processing
Network transmission delays (fixed and variable)
Multicast processing
Network transmission delays (fixed and variable)
Playout
Decryption
I-frame delay
Decode and render

0 200 400 800 600 1000 time (ms)
VOD response time

Legend
- **Red** = STB processing delays
- **Blue** = Network transmission delays (fixed | variable)
- **Orange** = Middleware processing

Remote control and STB processing

Network transmission delays (fixed and variable)

Middleware processing

Network transmission delays (fixed and variable)

Playout

Decryption

Decode and render

1

2

3

4

5

6

7

Legend

0 200 400 800 600 time (ms)
Video SLA Requirements

Delay / Jitter

- Delay variation absorbed by jitter buffer at set top box (STB)
- Network delays represent a small proportion over the overall channel change time or VOD response time compared to other factors
- Nonetheless, desire to minimise jitter buffer to optimise user responsivity
- One-way network delays from the first hop router to STB or from the VoD server to the STB and vice versa of <<100ms are typically targeted in order to try to achieve overall channel change times of 1-2 seconds.
- Diffserv IP QOS mechanisms are used to control queuing delays and ensure that video control messages and MPEG packets are not dropped
- Advanced capabilities such as Cisco Visual Quality Experience (VQE) technology, can enable rapid channel-change for improved viewer QOE
Impact of Packet Loss

- Unlike voice, any single unrecovered video packet loss may result in an impairment.

- Depending upon specific encoding and compression scheme etc., losses of different packet types are realised as different types of visual impairment, or artefact.
  
  Slice errors, blocking / pixelisation, ghosting or freeze-frame.

- Video SLA requirements for real-time and streaming video may be expressed in terms of loss events as impairments/time.
  
  Can also be expressed as mean time between the artefacts (MTBA).

- A commonly used industry benchmark for entertainment-grade IPTV service quality is to deliver MTBA of 2hrs or greater, i.e. no more than one perceivable error during a two-hour movie.

- You may decide that the most important metric is in terms of number of support calls received.
Slice Errors

- A slice is seen when at least one IP packet is dropped within an I,B,P-frame.
- The IP packet dropped could impact part of a single slice or multiple slices, dependent on the size of the frame impacted.
- Receipt of a new I-frame will clear the error.
Blocking / Pixelisation

- Blocking/pixelization is seen when there is a loss of a reference frame (either I-frame or P-frame) or when there is loss within a reference frame.
- Receipt of a new I-frame will clear the error
Ghosting

- Ghosting is seen when loss occurs on a scene change.
- Ghosting results from the loss of the I-frame around the scene change.
- Receipt of a new I-frame will clear the error.
Freeze Frame

- The loss of a continuous string of MPEG frames will result in a freeze frame artifact.

- A small number of frames such as 2 or 3 frames lost will result in more of a jerkiness to the displayed clip rather than a full on freeze frame.

- Pixelisation and blocking will usually follow the freeze frame as the newer frames received would not have suitable reference to work from.
**I-frame Impairments**

- Loss of the I-frame or information contained in an I-frame will cause errors that will be propagated through the remainder of the GOP.

- One IP packet lost which removes the I-frame header information has the same effect as removing the whole I-frame.

- The impact on the viewer’s QOE will be dependant on
  - GOP length for the length of impairment
  - Motion in the clip for the degree of blocking/pixelisation
P-frame Impairments

- Loss of P-frame or information contained in a P-frame will cause errors that will be propagated through the remainder of the GOP.
- One IP packet lost which removes the P-frame header information has the same effect as removing the whole P-frame.
- The impact of an impairment to the first P-frame in a GOP can be almost as significant as an I-frame impairment.
- Impairments to subsequent P-frames will produce similar artifacts, but which last for less time.
B-Frame Impairments

- B-frames do not provide reference to any other frames
- Removing the B-frame produces just a single frame of impairment
  
e.g. 1/30s of impairment in a 30 fps clip, usually perceptible as a slight jerkiness or blip to the viewer.
- One IP packet lost which removes the B-frame header information has the same effect as removing the whole B-frame
Study on the impact of packet loss

- What are the common affects of a network outage on an IP-encapsulated MPEG (MPEGoIP) stream?
- What happens if an I-frame, P-frame or B-frame is lost?
- What happens if one or more MPEGoIP packets is lost?
- What happens if contiguous IP packets are lost for 50ms, 200ms, 500ms etc.?
Study on the impact of packet loss

- Jason Greengrass, John Evans, Ali C. Begen, “Not All Packets Are Equal: The Impact of Network Packet Loss on Video Transport” – to be published in IEEE Internet Computing, Nov 08
  
  http://www.employees.org/~jevans/videopaper/videopaper.html
Study on the impact of packet loss

Conclusions

- Even a single IP packet loss can result in a significant visual impairment.
- The impairment resulting from 50ms of packet loss can be as significant as for 500ms of packet loss.
- Receipt of a non-impaired I-frame will restore the video quality after any impairment due to packet loss.
- No significant difference between the impairments seen in SD compared to HD, for an equivalent duration of packet loss.
- Impact of equivalent loss with H.264/MPEG-4 will likely be greater in practise
  - Longer GOPs - over a second
  - Hierarchal GOP structure Possible B-frames used as references
Video SLA Requirements

Packet Loss

- Therefore, in practise, networks supporting video streaming services should typically be designed for very close to zero percent video packet loss

- Four primary causes for packet loss
  
  Excess Delay
  
  Congestion
  
  PHY-Layer Errors
  
  Network Reconvergence Events
Four Primary Causes for Packet Loss

- **Excess Delay**
  - Packets delayed beyond an acceptable bound are effectively lost
  - Must be prevented with appropriate QoS (i.e., Diffserv)

- **Congestion**
  - Considered a catastrophic case, i.e., a fundamental failure of service
  - Must be prevented with appropriate QoS (i.e., Diffserv) and admission control (i.e., RSVP)

- **PHY-Layer Errors**
  - Can occur in both core and access – although generally more prevalent in the access
  - FEC or retransmission can be used to recover from loss experienced – as supported by Cisco Visual Quality Experience (VQE) technology

- **Network Reconvergence Events**
  - Network level approaches such as fast convergence or fast reroute can be used to reduce any loss experienced
  - Application or transport level approaches can be used to recover from loss experienced: FEC, Temporal Redundancy, Spatial Redundancy
Video Quality of Experience (QOE)

- QOE metrics define the perception of application performance, experienced from the perspective of the end-users.

- For IP based video applications the QOE is a compound metric dependent upon the quality of the encoder used, the quality of the service delivered by the IP network, and the quality of the decoder used.

- QOE targets do not directly define the delay, jitter, loss etc. characteristics that a network should provide, but rather for a specified application, using a defined encoder / decoder, the network characteristics may be implied given a particular QOE target.

- QOE metrics can be measured subjectively or objectively.
  
  Subjective measures rely upon end user feedback of their perception of the quality of the service.
  
  Objective measures use measurements of characteristics of the received stream, and possibly also of the transmitted stream, in order to infer the subjective quality that would be experienced by the end user.
Video SLA Requirements – Summary

- One-way network delays from the first hop router to STB or from the VoD server to the STB and vice versa of <=100ms are typically targeted in order to achieve responsivity targets.
  - Diffserv IP QoS mechanisms and admission control (RSVP) are used to control queuing delays.

- Even a single IP packet loss can result in a significant visual impairment – hence networks supporting video streaming services should typically be designed for very close to zero percent video packet loss.
  - Diffserv IP QoS mechanisms are used to control queuing delays and ensure that video control messages and MPEG packets are not dropped.

- With appropriate network engineering and capacity planning, the only network events that should result in a visual impairment to the video service are packet losses due to lower-layer errors or network element failures.
  - Packet loss due to network failure events can be minimised with technologies such as IP routing protocol fast convergence, MPLS TE FRR, IPFRR and MoFRR, but packet loss may still be experienced, which will impact the perceived quality of experience.
  - Where the underlying network transport infrastructure cannot meet the loss SLA targets required for the video service, end-to-end loss concealment techniques may be required, such as FEC, spatial redundancy or temporal redundancy.

- Advanced capabilities such as Cisco Visual Quality Experience (VQE) technology, can enable rapid channel-change and support retransmission and FEC to recover from packet loss and improve viewer QOE.