SMPTE Technology Series Webcast

SMPTE Enabling Global Education

Live Low-Latency Streaming

Russell Trafford-Jones

SMPTE Technology Webcast Series Sponsored by:

AJA VIDEO SYSTEMS

Your Host

SMPTE Enabling Global Education

Joel E. Welch

Director of Education
SMPTE

SMPTE Technology Webcast Sponsors

SMPTE Enabling Global Education

- *Thank you to our sponsor for their generous support:*

AJA® VIDEO SYSTEMS

SMPTE Technology Webcasts

SMPTE Enabling Global Education

- Series of monthly 60- to 90-minute online, interactive webcasts covering a variety of technical topics
- Free professional development benefit for SMPTE members
- Sessions are recorded for member viewing convenience.

Housekeeping

SMPTE Enabling Global Education

- Please indicate you want to ask verbal question by indicating such in the chat box
  - If you do not have a microphone, please submit your questions via text
- SMPTE provides a PDF of select slides used during webcasts in exchange for your feedback
  - Once your feedback is submitted, you will automatically be redirected to the PDF for downloading
- Please feel free to post or blog about today's webcast on your social media platform of choice

@smpteconnect

#SMPTEWebcast

Views and opinions expressed during this SMPTE Webcast are those of the presenter(s) and do not necessarily reflect those of SMPTE or SMPTE Members.

This webcast is presented for informational purposes only. Any reference to specific companies, products or services does not represent promotion, recommendation, or endorsement by SMPTE

Today's Guest Speaker

SMPTE Enabling Global Education

Russell Trafford-Jones
Manager, Services & Support, Techex
Editor, TheBroadcastKnowledge.com

Agenda

SMPTE

- What is live streaming?
- Standardisation within Streaming
- How is it achieved?
- What is *Low* Latency?
- What are the sources of Latency?
- How can we improve on HLS?
- WebRTC
- CMAF
- Conclusion

What is Live Streaming

- Streaming using the public internet to the end-user
- One to many
- Example use cases:
  - Streaming a TV channel to journalists in the field
  - Delivering live TV to viewers at home

Contribution vs Distribution Streaming

- Contribution
  - Business to business or movement within a company
  - Need to keep the quality high
  - Ultra-low latency may be critical
  - Often Point-to-point - low number of receivers
- Distribution
  - Goal is to deliver to viewers
  - Bandwidth & other factors require higher compression
  - Scalability very important

What is Live Streaming

- Each frame needs to be ready in time - presents a challenge
  - Restricts time to encode if source is live
  - Dependant on method of transmission
- More dependant on the network
  - Bitrate of video needs always to fit into the bandwidth - limits VBR
  - Even short network interruptions are an issue

Standardisation Bodies

- IETF and W3C for WebRTC
- MPEG for MPEG DASH
- HLS several versions as an IETF RFC (e.g. 8216)
- RTMP ‘Specification Document’
- MPEG for CMAF
- Defacto standards

Chunked Streaming

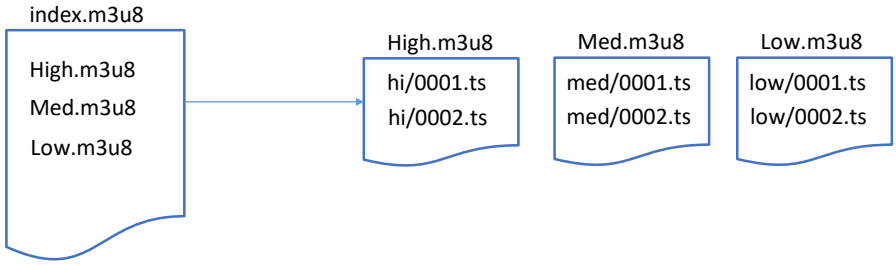
- Why not just download the file you want to watch?
  - Size of file can be over 500Mb
  - Startup time in the minutes
  - Seeking forward is problematic
- Cutting the file up
  - Small files
  - Startup time in seconds
  - Random access

Navigating Nasty Networks

- Delivery is tough due to varying bandwidth
- Adaptive Bitrate (ABR) makes available several versions
  - Player can choose a bitrate of file to fit its current bandwidth
- Imagine pressing play on 4 VT machines at once
  - Need to synchronise position to avoid jumps

Keeping Track of Chunks

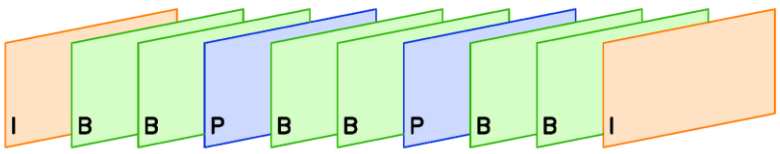
- Playlists are text files
- HLS is based on .m3u8 files
- MPEG-DASH is based on .mpd files
- Contains path and filename of available chunks
- Describes bitrate, resolution etc.



MPEG GOP Primer

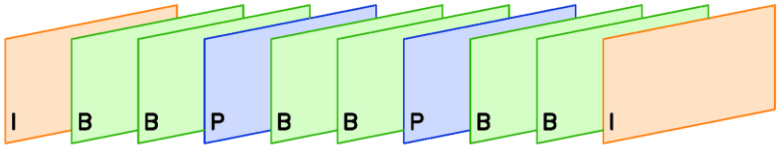
- Intra-frame codecs look at a single frame at a time
- Inter-frame codecs look for similarities across many frames

- Group of Pictures (GOP)



MPEG GOP Primer

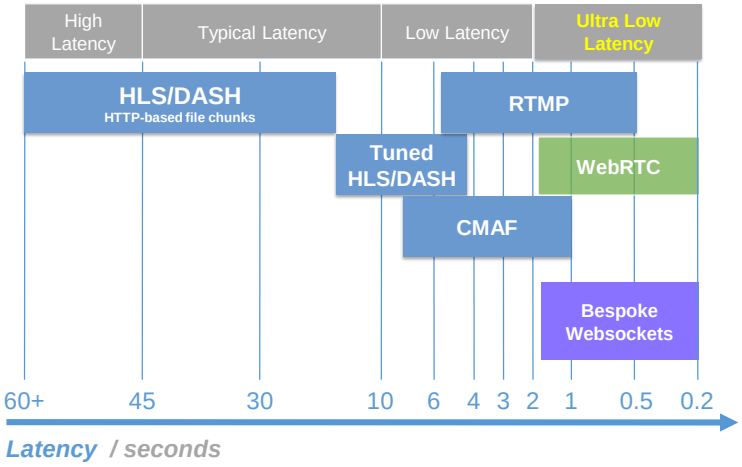
- I frames describe a complete frame of video
- P frames describe new information compared to previous I
- B frames show how parts of the image have moved referring to I and P frames.



- The only place a decoder decoding is an IDR frame

Latency - What should we expect?

Common Latencies



Business Case Check

- How much does latency matter?
  - Not everyone is Netflix
  - Not everyone is streaming the Super Bowl
  - Live Sports
  - Betting
  - Auctions
  - E-Gaming
  - Return Vision as part of a production

Business Case Check

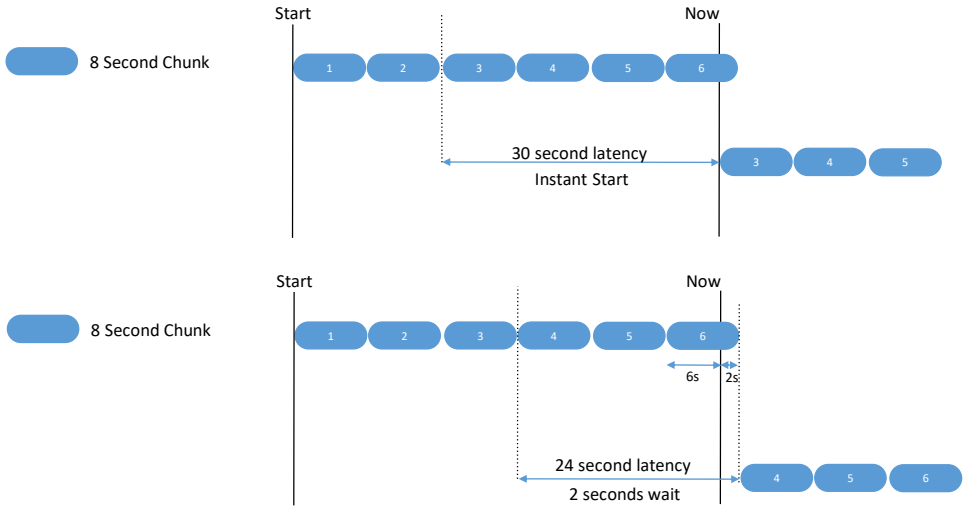
- Catch up with Broadcasters
- Beat broadcasters
- Live is a USP
- Innovate in different types of programming
  - Live overlays
  - Interactivity

Where Does Latency Come From?

The Latency Chain

- Encoder
- Contribution Propagation
- Packaging
- Transcoding
- CDN Propagation
- Delivery Propagation
- Playback Buffer
- Playback Start algorithm

Playback latency



Challenges Reducing Latency

Challenges Reducing Latency

- Chunks have to be aligned with MPEG's GOP
- Smaller chunks mean shorter GOPs
- Compression efficiency decreases with larger GOP size

Approaches to Achieve Low Latency

1. Send the video as a continuous Stream
2. Emulate a stream

Benefits of Sending a Stream

- A stream sends frames as soon as they are ready
- This is how HLS started; as an MPEG 2 Transport Stream
- Eliminated file requests
- Sent over UDP instead of TCP
- Great for ingest into the cloud

Downsides to Unicast Streaming

- Streaming is a push technology
- Firewalls don’t generally allow unsolicited incoming data
- Pull technology like HLS doesn’t have this limitation
- Harder to switch bitrates

Multicast vs Unicast

- Consider a MPEG 2 Transport Stream
  - It can be sent direct to one endpoint - unicast.
  - Use a networking technology called Multicast,
    - IGMP - Internet Group Management Protocol
    - Used extensively for ST 2110, 2022-6
  - The internet, in general has no support for IGMP

WebRTC

- Created for video conferencing
- Widely used by Facebook Messenger, Google Hangouts etc.
- Peer-peer networking with 2-way data over websockets
- Continues to evolve and is increasingly relevant to broadcast
- Great replacement for RTMP distribution
- Adobe support ending

WebRTC – Does it Scale?

- There are large-scale CDNs
- Can use cascading within WebRTC
- Can convert to WebRTC at the edge

WebRTC – Pros

- Sub-Second Latency
- Data channels provide good analytics
- ABR delivered by Simulcast
- Interoperability is high
- New features coming with WebRTC NV
- Low-latency allows for innovation in productions
- Low-latency helps avoid live-data beating the video

WebRTC – Cons

- DRM is restricted –though often not necessary
- NAT Traversal Uses ICE (Interactive Connectivity Establishment)
- ABR delivered by Simulcast
- Young, evolving technology
- CDNs need to be adapted
- Lost packets are an issue
- Not intrinsically suited to VOD, Cloud DVR etc.

Approaches to Achieve Low Latency

1. Send the video as a continuous Stream
2. Emulate a stream

LHLS -Low-delay HLS

- Used by Periscope amongst others
- Can work down to 4 seconds
- Based on:
  - Smaller Chunk Sizes
  - Advertises chunks ahead of time

Chunked Transfer Encoding

- Arrived in HTTP 1.1
- Doesn’t know how long the file it is it is sending.
- It just starts
- Finishes, like with variables, when a null is sent. In this case, a zero-length chunk
- Since data is sent as soon as available, bitrate on the wire is very consistent

LHLS – Smoother Bitrate

Breaking the chunk size trade-off

ISO Base Media File Format

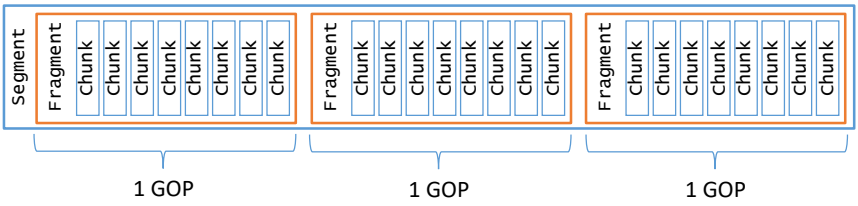
- An ISO/MPEG standard for containing media
- Based on Apple's QuickTime
- Part of the MP4 family
- Object-orientated using the concept of 'boxes' to contain media
- Timing is based on what went before, not recorded timestamps
- Extensible for new metadata, codecs etc.

CMAF: Common Media Application Format

- A container for holding video and other essences
- Based on ISO BMFF
- As a container - it, in itself, does not provide 'low latency'
- Defines a way to use ISO BMFF for streaming
- Overhead of extra file structure. Fixed, so high bitrate <<2% of video. Audio, not so lucky.
- Compatible with HLS and DASH
- Capable of containing HEVC

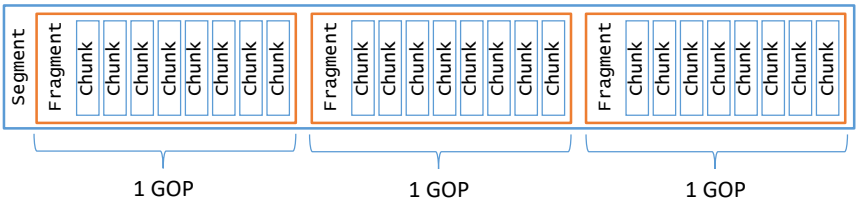
CMAF: Common Media Application Format

- CMAF files are broken down into:
  - Segments which can hold 1 or more Fragments
  - Fragments which are usually a number of Chunks
  - Chunks are the smallest amount of payload, be that 500ms of video, 200ms or just one frame.
- Chunks are transferred as they become available.

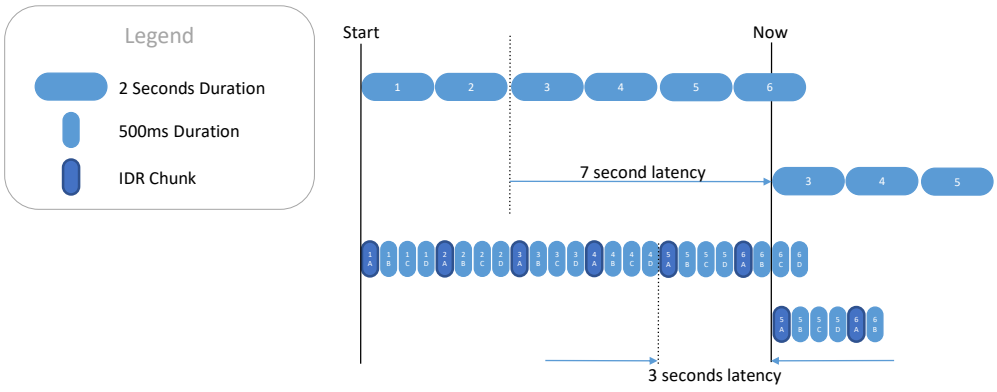


Low Latency with CMAF

- CMAF provides the file structure needed to deliver parts of the MPEG GOP separately
- No need to wait for the GOP to finish
- No need to wait for a fragment to finish (Chunked Transfer)



CMAF latency



Benefits of CMAF

- Latency between 0.5 and 1 second in proof of concepts
- Real world networks implementations are around 2 seconds
- Based on mature technologies
- Support for Common Encryption (CENC) standard for DRM
- More efficient than MPEG-2 Transport Streams
- One, widely supported format
- Reduces duplicate CDN storage

Transportation

- QUIC is a new protocol for transferring data to replace HTTP
- It has optimisations to be quicker
- Built on UDP
- Handles its own retransmissions
- Multiplexes data to avoid dependencies

- Jury is still out on how much it improves streaming performance

WebRTC All the Way?

- WebRTC is the only standards based technology for sub-second live streaming over public internet.
- CMAF performs strongly in 1-4 seconds range
- Look at your use case
  - Quality, DRM can point towards CMAF
  - Live-is king (look at FEC and NACK to ensure resilience)

Conclusion

We have seen:

- Why HLS has been so successful
- Why ultra-low latency isn't possible with HLS
- The pros and cons of using WebRTC
- The benefits of decoupling chunk size from MPEG GOP length
- That ultra-low latency streaming is possible

Questions

Russell Trafford-Jones
Manager, Services & Support, Techex
Editor, TheBroadcastKnowledge.com

SMPTE Technology Webcast Sponsors

- *Thank you to our sponsor for their generous support:*

SMPTE Technology Series Webcast
SMPTE – Enabling Global Education

Live Low-Latency Streaming

Russell Trafford-Jones

SMPTE Technology Webcast Series Sponsored by:

AJA VIDEO SYSTEMS