Adaptive Streaming of Traditional and Omnidirectional Media

ACM SIGCOMM Tutorial – Los Angeles, CA Aug. 2017

Ali C. Begen, Comcast/OzU

Christian Timmerer, AAU/Bitmovin

Upon Attending This Tutorial, You Will Know About

- Principles of HTTP adaptive streaming for the Web/HTML5
- Principles of omnidirectional (360°) media delivery
- Content generation/distribution/consumption workflows for traditional and omnidirectional media
- Standards and emerging technologies in the adaptive streaming space
- Current and future research on traditional and omnidirectional media delivery

This tutorial is public, however, some material might be copyrighted
Use proper citation when using content from this tutorial
(Thanks to T. Stockhammer, J. Simmons, K. Hughes, C. Concolato, S. Pham, W. Law
and many others for helping with the material)

Presenters Today

Ali C. Begen

- Electrical engineering degree from Bilkent University (2001)
- Ph.D. degree from Georgia Tech (2006)
  - Video delivery and multimedia communications
- Research, development and standards at Cisco (2007-2015)
  - IPTV, content delivery, software clients
  - Transport and distribution over IP networks
  - Enterprise video
- Consulting at Networked Media since 2016
- Assistant professor at OzU and IEEE Distinguished Lecturer since 2016

Christian Timmerer

- Associate Professor at the Institute of Information Technology (ITEC), Multimedia Communication Group (MMC), Alpen-Adria-Universität Klagenfurt, Austria
- Co-founder and CIO of Bitmovin
- Research interests
  - Immersive multimedia communication
  - Streaming, adaptation, and
  - Quality of experience (QoE)
- Blog: http://blog.timmerer.com; @timse7

ACM MMSys 2018 CfP is Out!

- City: Amsterdam
- Dates: June 12-15, 2018
- Co-located with
  - NOSSDAV
  - MoVid
  - MMVE
- http://www.mmsys2018.org/

Agenda

- Part I: Over-the-top (OTT) video and HTTP adaptive streaming
  - HTML5 standard and media extensions
  - HTTP adaptive streaming building blocks
  - Multi-bitrate encoding, encapsulation and encryption workflows
  - De jure and de facto standards (DASH, HLS, CMAF)
  - Experiences from the existing deployments
- Part II: Omnidirectional (360°) media
  - Acquisition, projection, coding and packaging
  - Delivery, decoding and rendering
  - Emerging standards and industry practices
  - Current research trends, open issues, efforts that are underway in the streaming industry

First Things First

IPTV vs. IP (Over-the-Top) Video

IPTV	IP Video
Managed delivery	Best-effort delivery
Emphasis on quality	Quality not guaranteed
Mostly linear TV	Mostly on demand
Always a paid service	Paid or ad-based service

Part I: OTT Video and HTTP Adaptive Streaming

Internet (IP aka OTT) Video Essentials

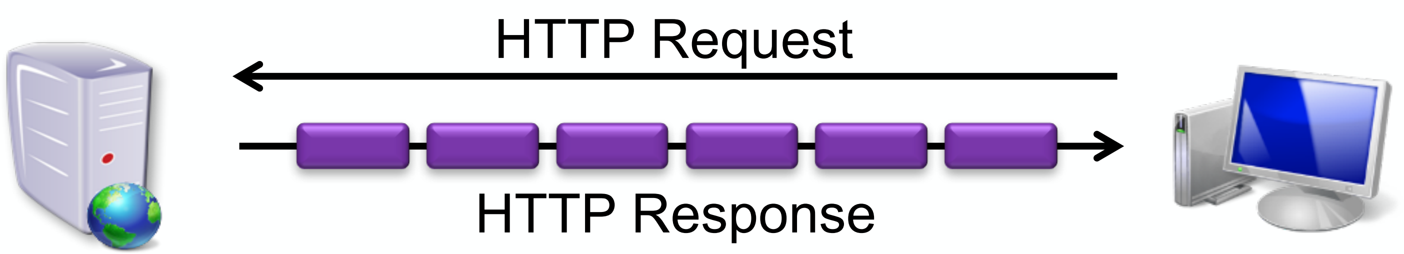
Reach	• Reach all connected devices
Scale	• Enable live and on-demand delivery to the mass market
Quality of Experience	• Provide TV-like or even better and richer viewer experience
Business	• Enable revenue generation thru paid content, subscriptions or targeted advertising
Regulatory	• Satisfy regulations for captioning, ratings, parental control and advertising

Creating Revenue – Attracting Eye Balls

- High-end content
  - Hollywood movies, TV shows
  - Sports
- Excellent quality
  - HD/3D/UHD audiovisual presentation w/o artifacts such as pixelization and rebuffering
  - Fast startup, fast zapping and low glass-to-glass delay
- Usability
  - Navigation, content discovery, battery consumption, trick modes, social network integration
- Service flexibility
  - Linear TV
  - Time-shifted and on-demand services
- Reach
  - Any device, any time

Progressive Download

One Request, One Response



What is Streaming?

Streaming is transmission of a continuous content from a server to a client and its simultaneous consumption by the client

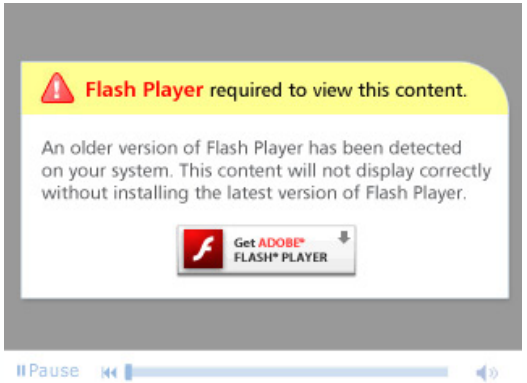
Two Main Characteristics

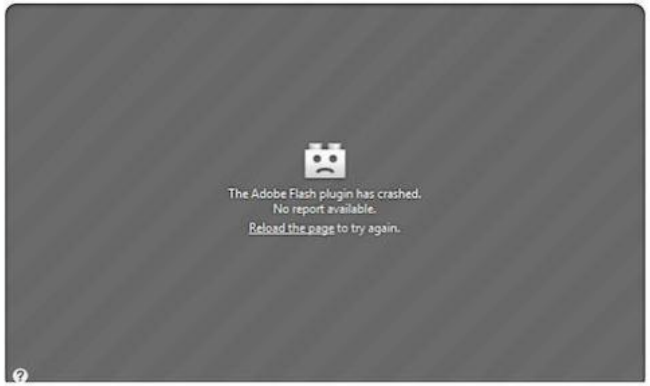
1. Client consumption rate may be limited by real-time constraints as opposed to just bandwidth availability
2. Server transmission rate (loosely or tightly) matches to client consumption rate

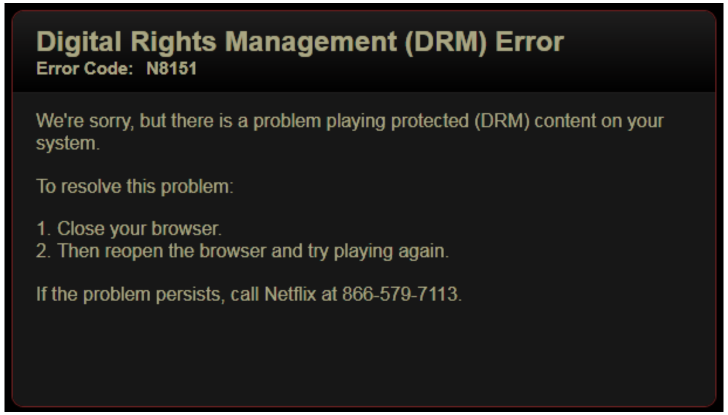
Common Annoyances in Streaming

Stalls, Slow Start-Up, Plug-In and DRM Issues

- Unsupported/wrong
  - protocol
  - plug-in
  - codec
  - format
  - DRM
- Slow start-up
- Poor quality, quality variation
- Frequent freezes/glitches
- Lack of seeking







Video Delivery over HTTP

Progressive Download

- Enables playback while still downloading
- Server sends the file as fast as possible

Pseudo Streaming

- Enables seeking via media indexing
- Server paces transmission based on encoding rate

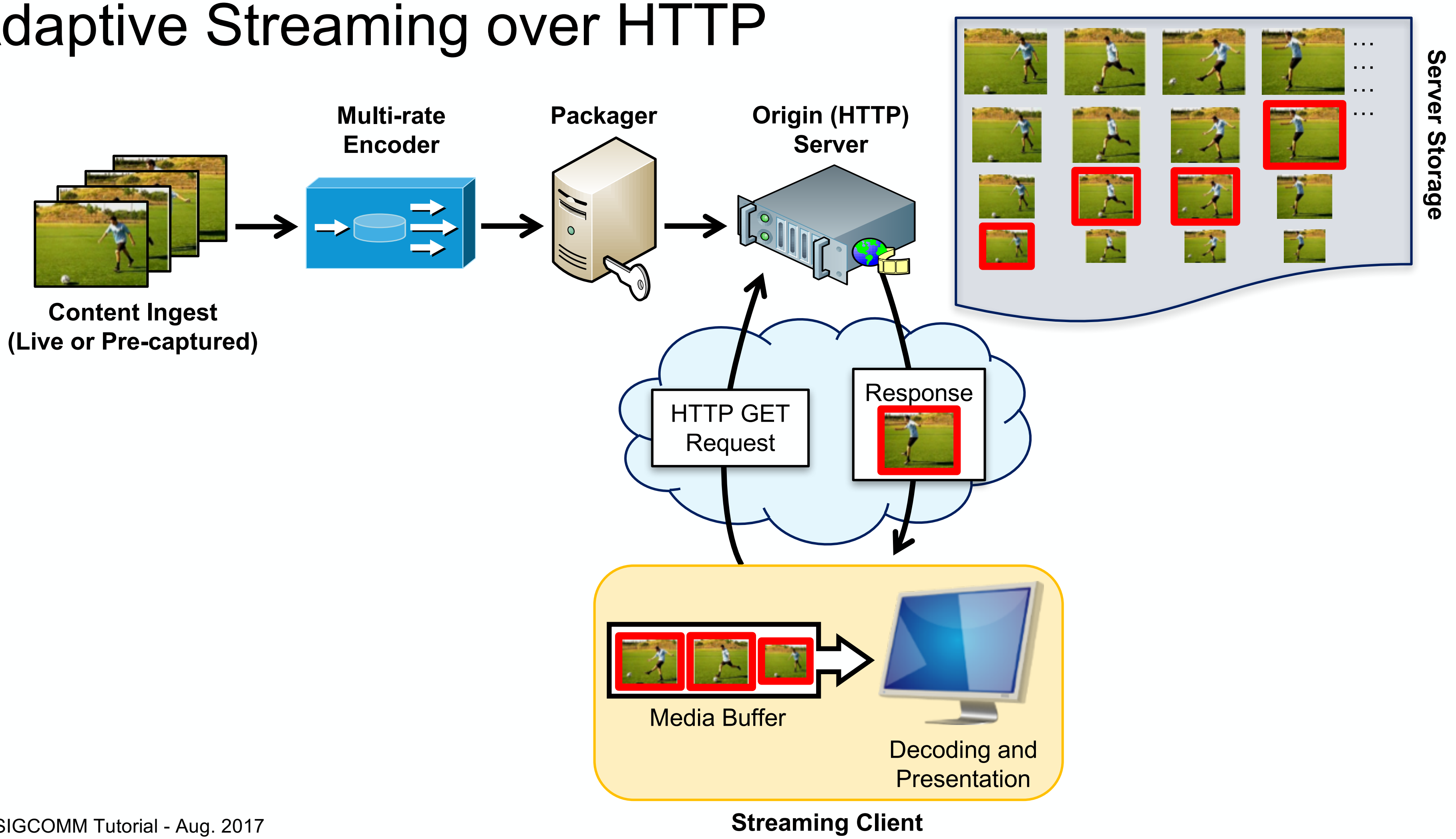
Chunked Streaming

- Content is divided into short-duration chunks
- Enables live streaming and ad insertion

Adaptive Streaming

- Multiple versions of the content are created
- Enables to adapt to network and device conditions

Adaptive Streaming over HTTP



Adaptive Streaming over HTTP

Adapt Video to Web Rather than Changing the Web

- Imitation of streaming via short downloads
  - Downloads small chunks to minimize bandwidth waste
  - Enables to monitor consumption and track the streaming clients
- Adaptation to dynamic conditions and device capabilities
  - Adapts to dynamic conditions in the Internet and home network
  - Adapts to display resolution, CPU and memory resources of the streaming client
  - → Facilitates "any device, anywhere, anytime" paradigm
- Improved quality of experience (<u>not always improved average quality</u>)
  - Enables faster start-up and seeking, and quicker buffer fills
  - Reduces skips, freezes and stutters
- Use of HTTP
  - Well-understood naming/addressing approach, and authentication/authorization infrastructure
  - Provides easy traversal for all kinds of middleboxes (e.g., NATs, firewalls)
  - Enables cloud access, leverages the existing and cheap HTTP caching infrastructure

Dead, Surviving, Maturing and Newborn Technologies

- Move Adaptive Stream (Surviving in Sling TV)
  - http://www.movenetworks.com
- Microsoft Smooth Streaming (Legacy)
  - http://www.iis.net/expand/SmoothStreaming
- Adobe Flash (Almost dead)
  - http://www.adobe.com/products/flashplayer.html
- Adobe HTTP Dynamic Streaming (Legacy)
  - http://www.adobe.com/products/httpdynamicstreaming
- Apple HTTP Live Streaming (The elephant in the room)
  - http://tools.ietf.org/html/draft-pantos-http-live-streaming (Soon to be an RFC)
- MPEG DASH and CMAF (The standards)
  - http://mpeg.chiariglione.org/standards/mpeg-dash
  - http://mpeg.chiariglione.org/standards/mpeg-a/common-media-application-format

Example Representations

Vancouver 2010

	Encoding Bitrate	Resolution
Rep. #1	3.45 Mbps	1280 x 720
Rep. #2	1.95 Mbps	848 x 480
Rep. #3	1.25 Mbps	640 x 360
Rep. #4	900 Kbps	512 x 288
Rep. #5	600 Kbps	400 x 224
Rep. #6	400 Kbps	312 x 176

Sochi 2014

	Encoding Bitrate	Resolution
Rep. #1	3.45 Mbps	1280 x 720
Rep. #2	2.2 Mbps	960 x 540
Rep. #3	1.4 Mbps	960 x 540
Rep. #4	900 Kbps	512 x 288
Rep. #5	600 Kbps	512 x 288
Rep. #6	400 Kbps	340 x 192
Rep. #7	200 Kbps	340 x 192

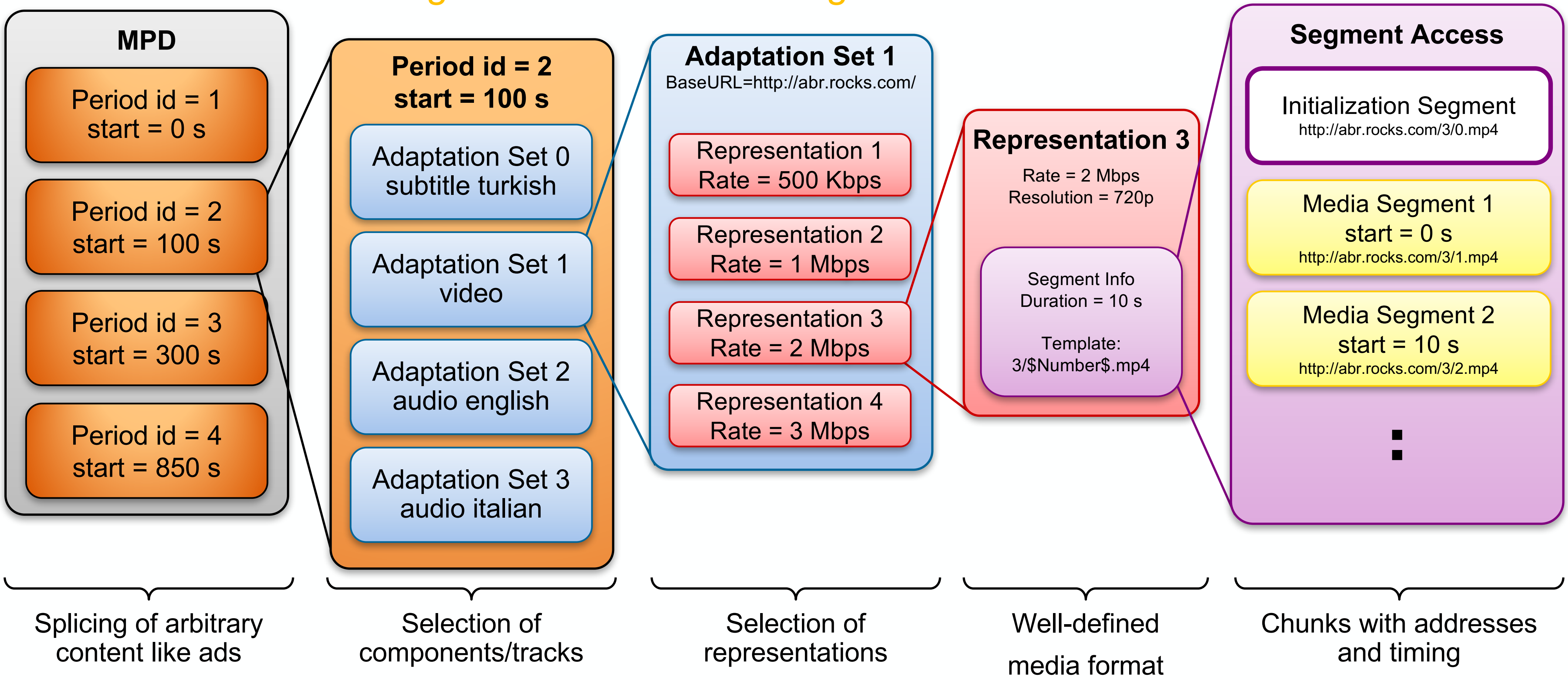
FIFA 2014

	Encoding Bitrate	Resolution
Rep. #1	3.45 Mbps	1280 x 720
Rep. #2	2.2 Mbps	1024 x 576
Rep. #3	1.4 Mbps	768 x 432
Rep. #4	950 Kbps	640 x 360
Rep. #5	600 Kbps	512 x 288
Rep. #6	400 Kbps	384 x 216
Rep. #7	250 Kbps	384 x 216
Rep. #8	150 Kbps	256 x 144

Source: Vertigo MIX10, Alex Zambelli's Streaming Media Blog, Akamai

An Example Manifest Format

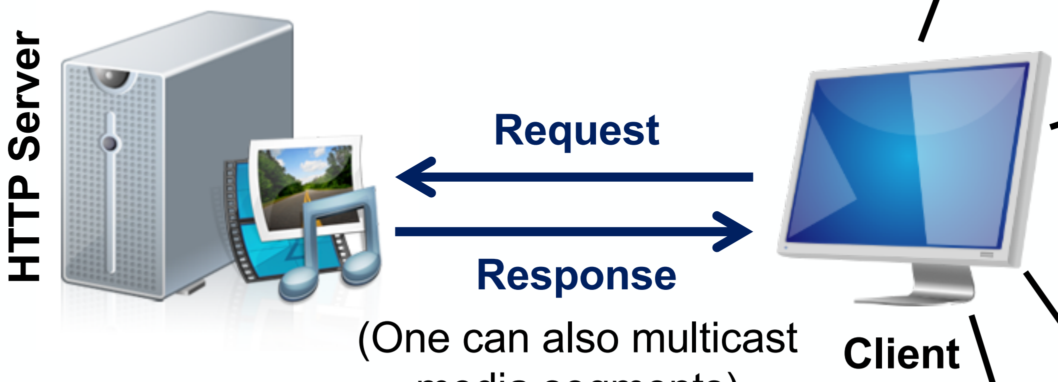
List of Accessible Segments and Their Timings



Smart Clients

- Client fetches and parses the manifest
- Client uses the OS-provided HTTP stack (HTTP may run over TCP or QUIC)
- Client uses the required decryption tools for the protected content

Client monitors and measures
- Size of the playout buffer (both in bytes and seconds)
- Chunk download times and throughput
- Local resources (CPU, memory, window size, etc.)
- Dropped frames

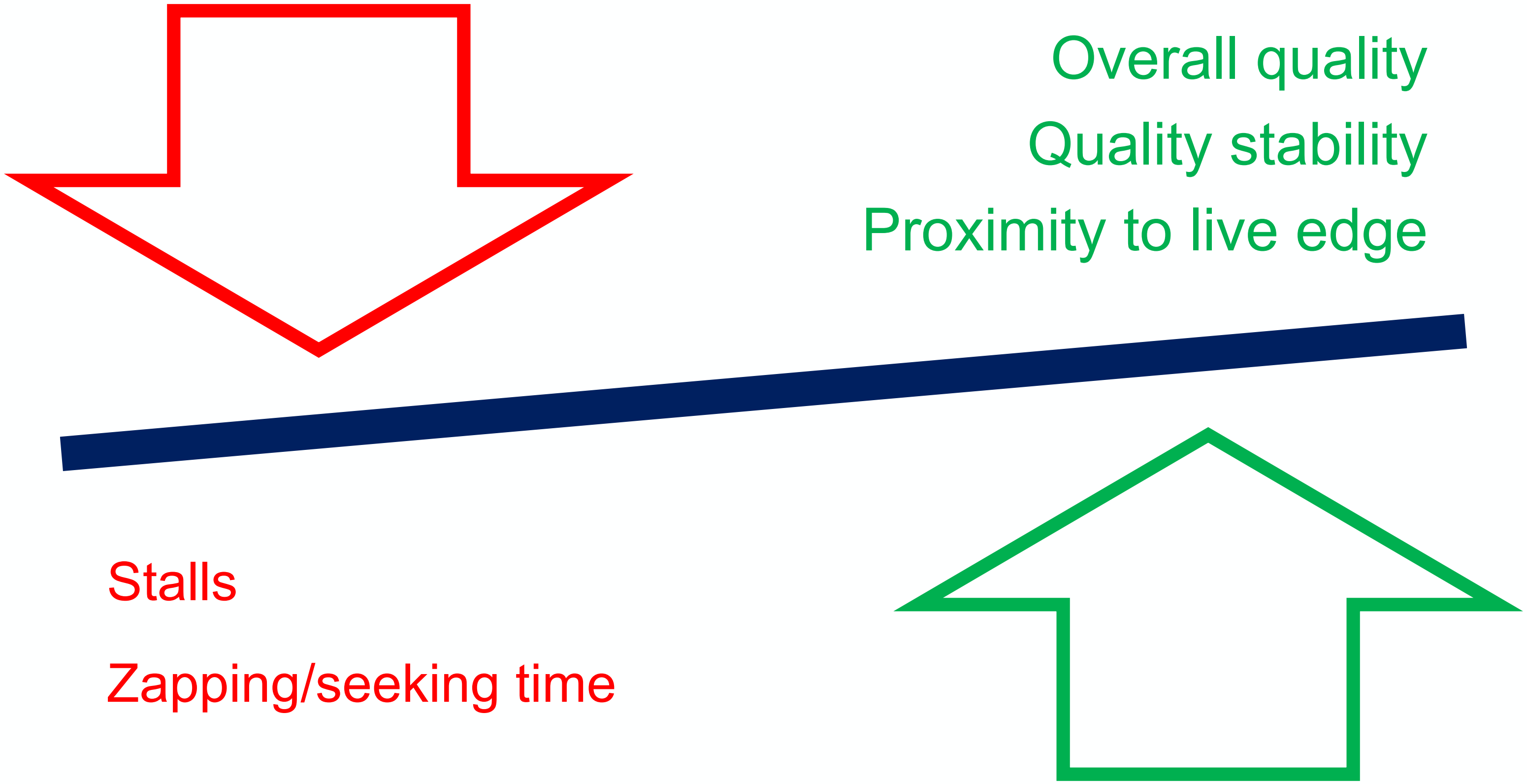


(One can also multicast media segments)

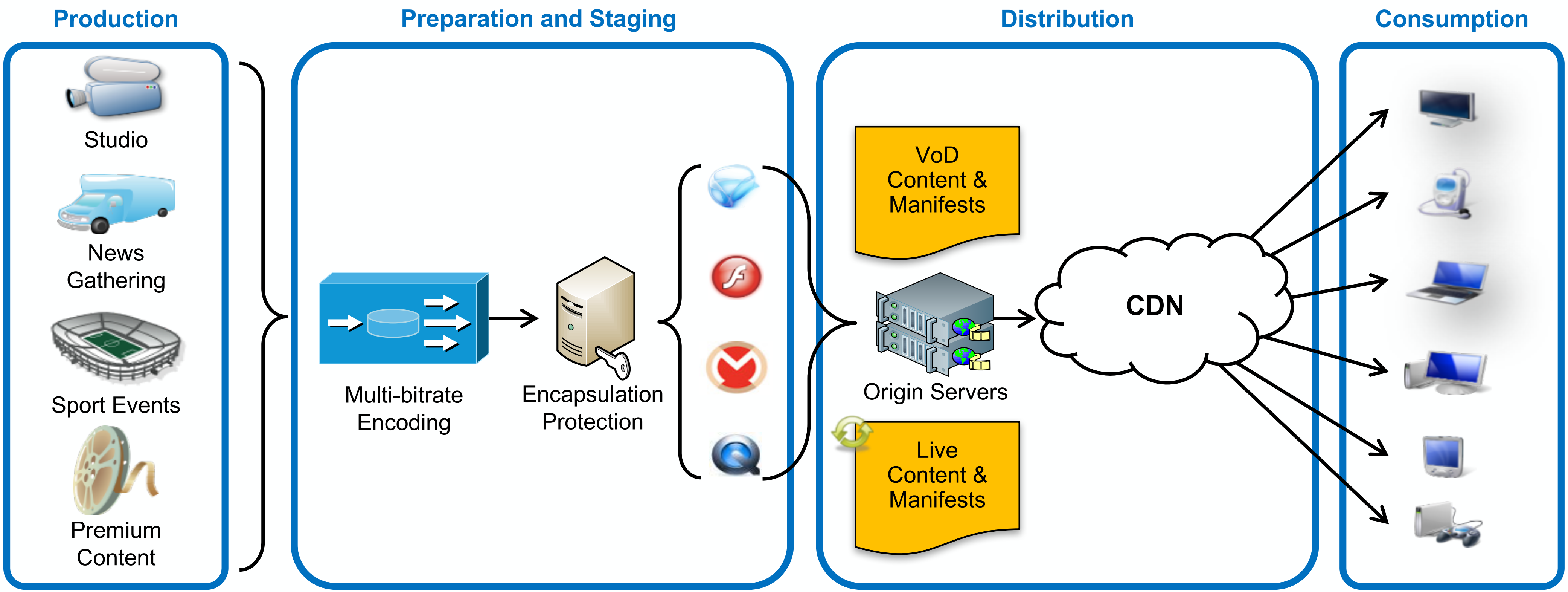
Client performs adaptation

Client measures and reports metrics for analytics

Tradeoffs in Adaptive Streaming



End-to-End Workflow for OTT




1
Source
The input files can start from a variety of sources such as AWS S3, Google Cloud Storage of FT/SFTP
amazon web services
AWS S3
Google
Google Cloud Storage
FTP & SFTP
FTP/SFTP
2
Encoding
Source files encoded into adaptive streaming formats e.g. MPEG-DASH and HLS
3
Storage
Files are stored on the Bitmovin storage which could be used as an origin for your CDN
4
CDN
Global delivery of your video files to provide the best possible quality
5
HTML5 Playback
HMTL5 playback of HLS and MPEG-DASH in any browser and any device
6
Analytics
Monitoring and reporting helps to quickly identify problems and improve user experience
VR - 360°
BITMOVIN

HTML5

A Common Platform across Devices

- HTML5 is a set of technologies that allows more powerful Web sites and applications
  - Better semantics
  - Better connectivity
  - Offline and storage
  - Multimedia
  - 2D/3D graphics and effects
  - Performance and integration
  - Device access
  - Styling
- Most interesting new elements
  - Semantic elements: <header>, <footer>, <article> and <section>
  - Graphic elements: <svg> and <canvas>
  - Multimedia elements: <audio> and <video>

Source: Mozilla MDN, w3schools.com

Hosted and Progressive Web Apps

- These apps can
  - leverage native APIs and offline capability
  - provide cross-device compatibility (iOS, Windows, Android, etc.)
  - be packaged and published to app stores
- For streaming:
  - To a user, it looks like a regular native app they have downloaded
  - The app is actually a container and loads loads dynamically from the web server
  - Server-side changes propagate automatically to installed apps

Source: Google PWA, Microsoft HWA

Types of Browser-Based Playback

- Type 1: Minimum control architectural model for HTML5 support of adaptive streaming where manifest and heuristics are managed by the user agent
- Type 2: Adaptation control architectural model for HTML5 support of adaptive streaming providing script manageable features
- Type 3: Full media control architectural model for HTML5 support of adaptive streaming allowing script to explicitly send the media segments (MSE + EME)

Source: CTA WAVE

MSE Support in Web Browsers

Media Source Extensions - CR

Global 74.53% + 3.32% = 77.85%

unprefixed: 74.43% + 3.32% = 77.75%

API allowing media data to be accessed from HTML **`video`** and **`audio`** elements.

Current aligned | Usage relative | Date relative | Show all

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android Browser *	Chrome for Android
		52	49			9.3		4.4	
	14	53	58		45	10.2		4.4.4	
[2] 11	15	54	59	10.1	46	10.3	all	56	59
	16	55	60	11	47	11			
		56	61	TP	48				
		57	62						

Notes | Known issues (0) | Resources (7) | Feedback

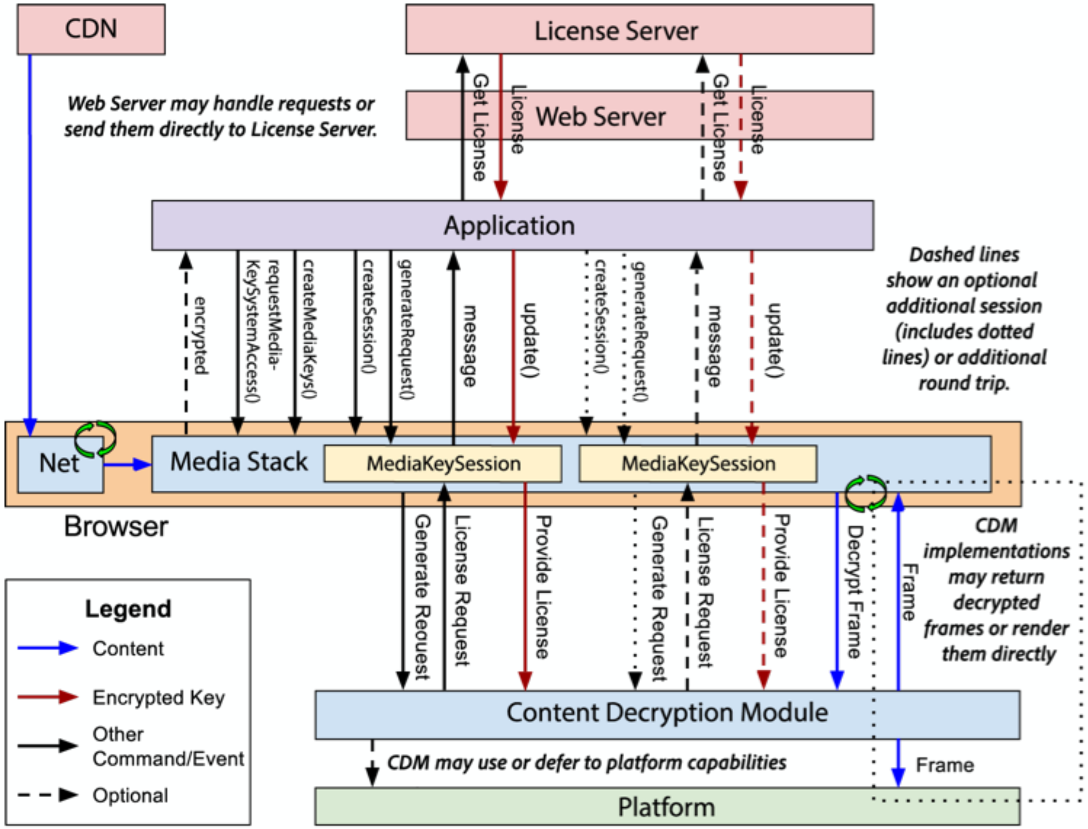
[2] Partial support in IE11 refers to only working in Windows 8+

Source: http://caniuse.com/#search=mse

Content Protection – Famous Quotes

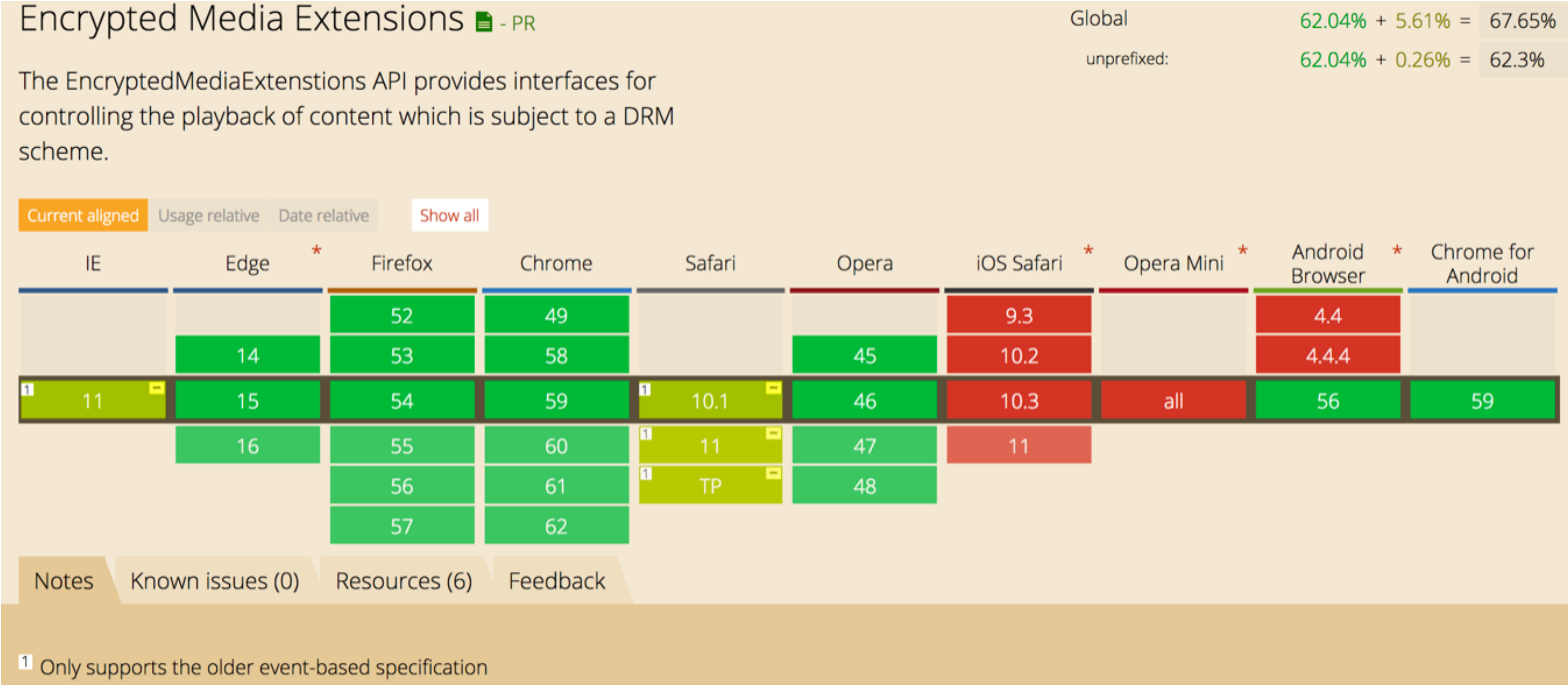
- “Digital files cannot be made uncopiable, any more than water can be made not wet”
  - *Bruce Schneier (cryptographer), May 2001*
- “We have Ph.D.'s here that know the stuff cold, and we don't believe it's possible to protect digital content”
  - *Steve Jobs, December 2003*
- “If we’re still talking about DRM in five years, please take me out and shoot me”
  - *eMusic CEO David Pakman, February 2007*
- “This is unethical”
  - *Ian Hickson, HTML5 editor, upon learning of the Netflix-Google-Microsoft EME proposal - February 2012*

W3C Encrypted Media Extensions

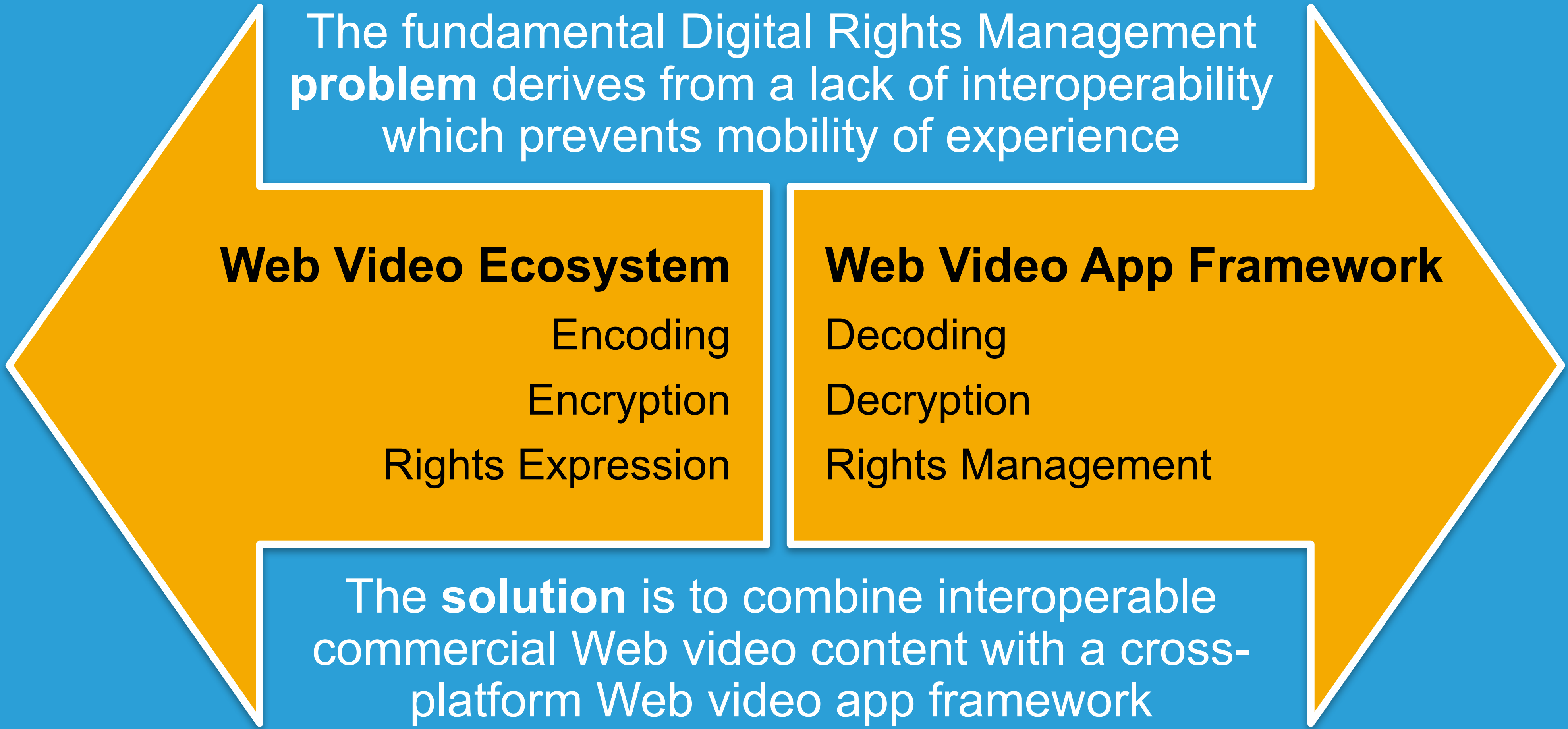


Source: https://www.w3.org/TR/encrypted-media/

EME Support in Web Browsers



Source: http://caniuse.com/#search=eme



Source: John Simmons

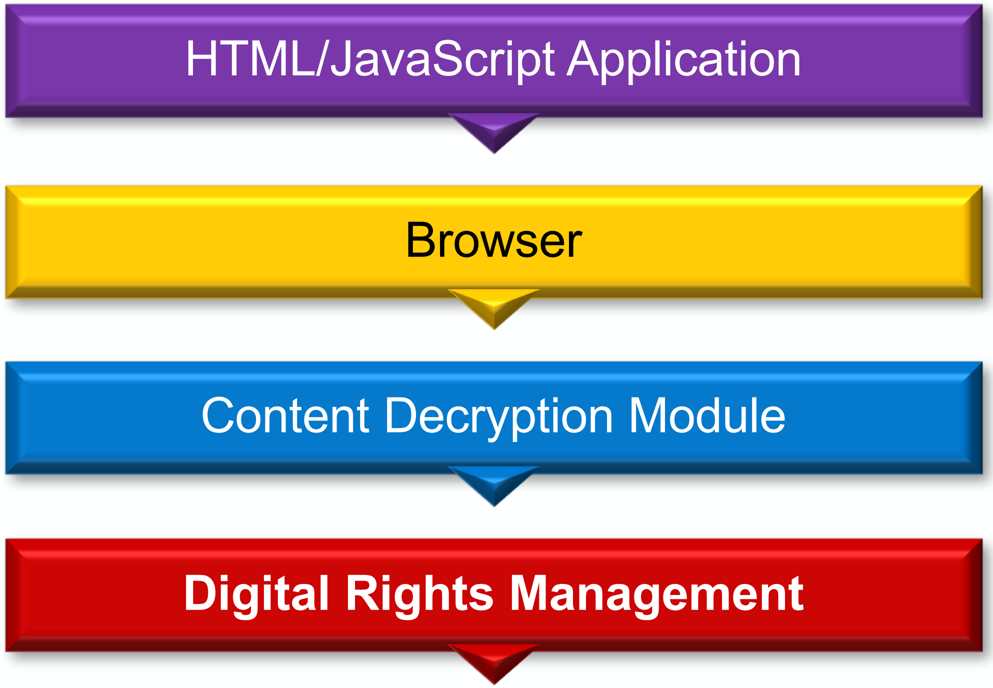
The Web, EME and Enhanced Content Protection (ECP)

The HTML/JS app selects the DRM and controls key exchange between DRM client and server

Browser extends HTML5 media element to allow JavaScript handled key acquisition

A CDM exposes a key system to JavaScript; it is transparent whether the CDM is in the browser

Contrary to a common misconception, with EME DRM functionality is not in the HTML/JS app
There is no DRM in HTML5 with EME, and ECP requires that this be the case



Source: John Simmons

DRM Support on Desktop Browsers

Browser	OS	EME/CDM	Flash Player	NPAPI/ Silverlight 5
Chrome	Win	Widevine	(Yes)	No
	OS X		(Yes)	No
	Linux		(Yes)	No
Firefox	Win	Widevine	Yes	No
	OS X		Yes	No
	Linux		Yes	No
Safari	> OS X Yosemite	(Fairplay)	Yes	Yes
	< OS X Yosemite	No	Yes	Yes
IE/Edge	< Win 7	No	Yes	Yes
	Win 10	PlayReady	Yes	No

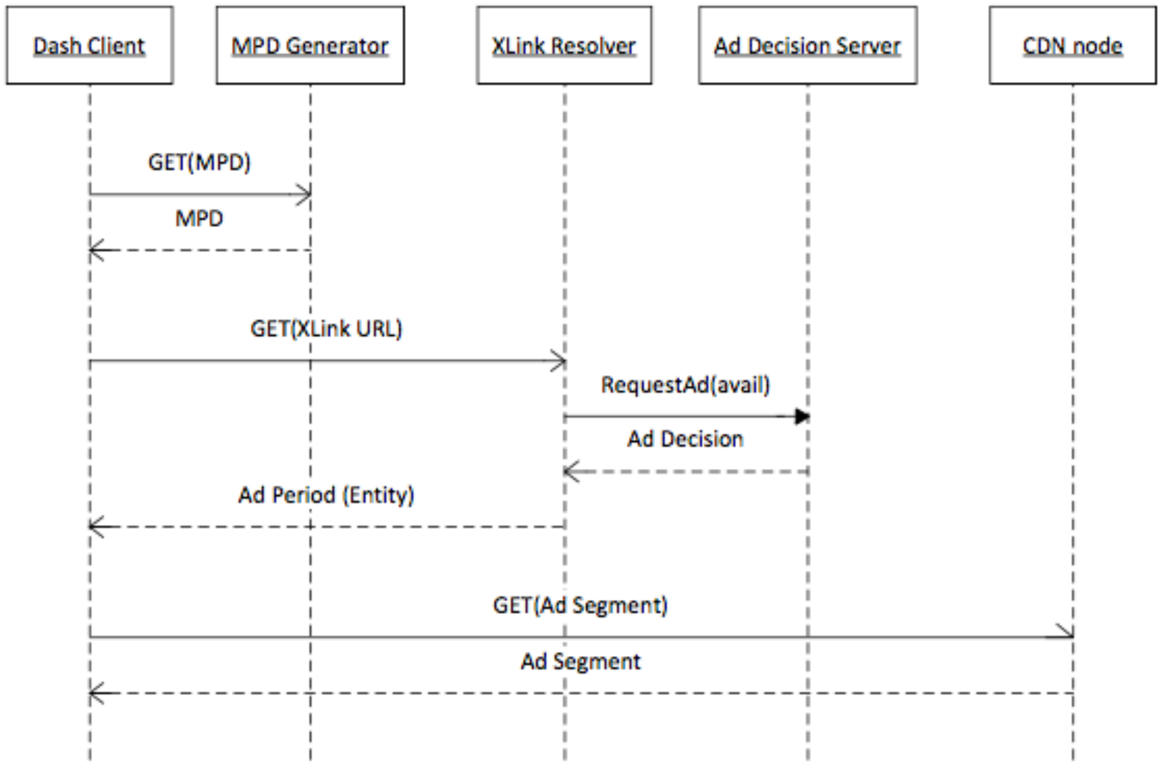
DRM Support on Mobile Platforms

Platform	Browser	Native App / DRM
iOS	No MSE/EME yet HLS (AES-128 CBC) via <video>	Native SDK and WebView: FairPlay
Android	MSE/EME	Native + Android SDK (MediaDRM APIs) - (Widevine + OMA v2) ExoPlayer Native + WebView with Widevine
Windows 10	MSE/EME	WebViews/Hosted Web Apps with PlayReady

Ad Insertion Methods

- Server-based
  - MPD is constructed or updated based on the ads
  - For multi-period MPDs, we can use
    - XLink
    - MPD chaining
  - For single-period MPDs
    - Ads are spliced into a continuous stream
    - Splicing can be offline or just-in-time
- App-based
  - Requires additional support in the application
  - Uses DASH-specified user-defined events

Ad Insertion – XLink Resolution



Streaming over HTTP – The Promise

- Leverage tried-and-true Web infrastructure for scaling
  - Video is just ordinary Web content!
- Leverage tried-and-true TCP
  - Congestion avoidance
  - Reliability
  - No special QoS for video

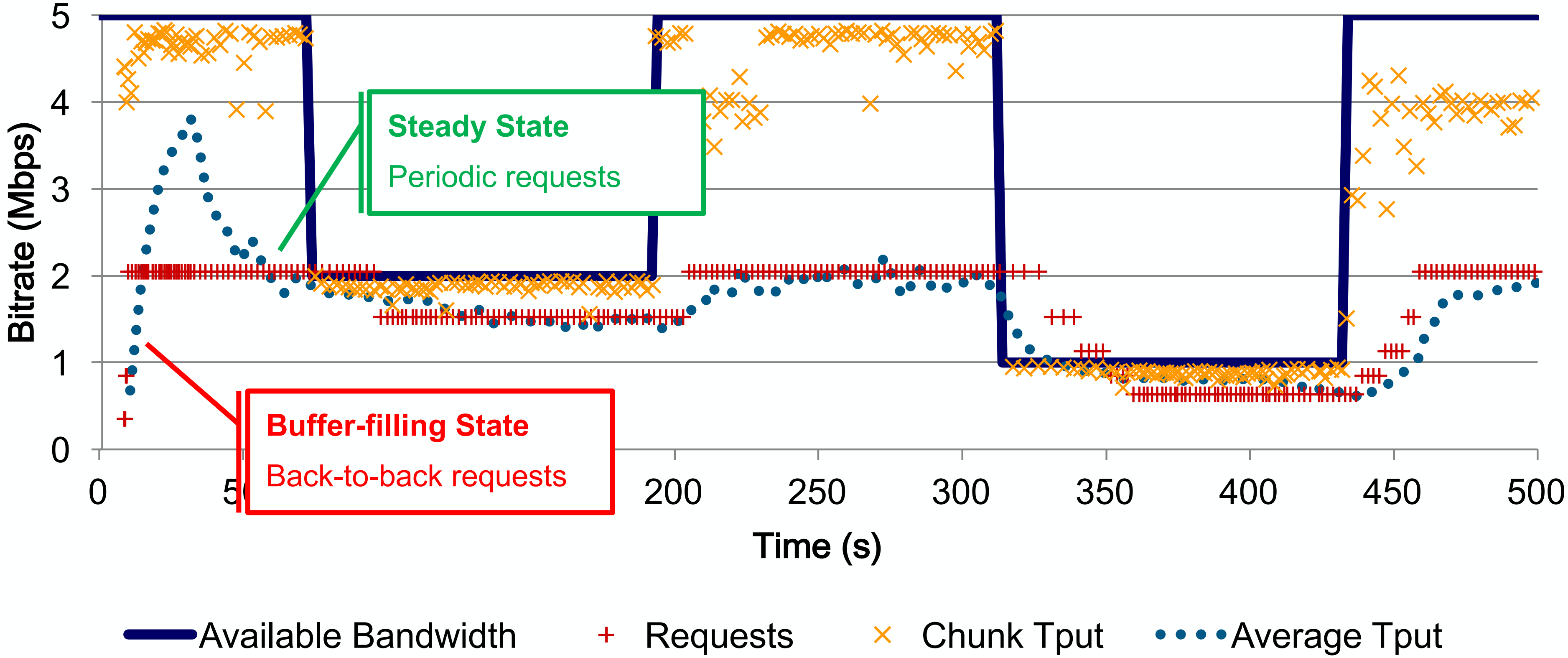
It should all "just work" ☺

Does It Just Work?

- When streaming clients compete with other traffic, mostly yes
- When streaming clients compete with each other, we begin to see problems:
  - The clients' adaptation behaviors interact with each other
  - The competing clients form an "accidental" distributed control-feedback system
- Unexpected behaviors will result in places like
  - Multiple screens within a household
  - ISP access and aggregation links
  - Small cells in stadiums and malls

Demystifying a Streaming Client

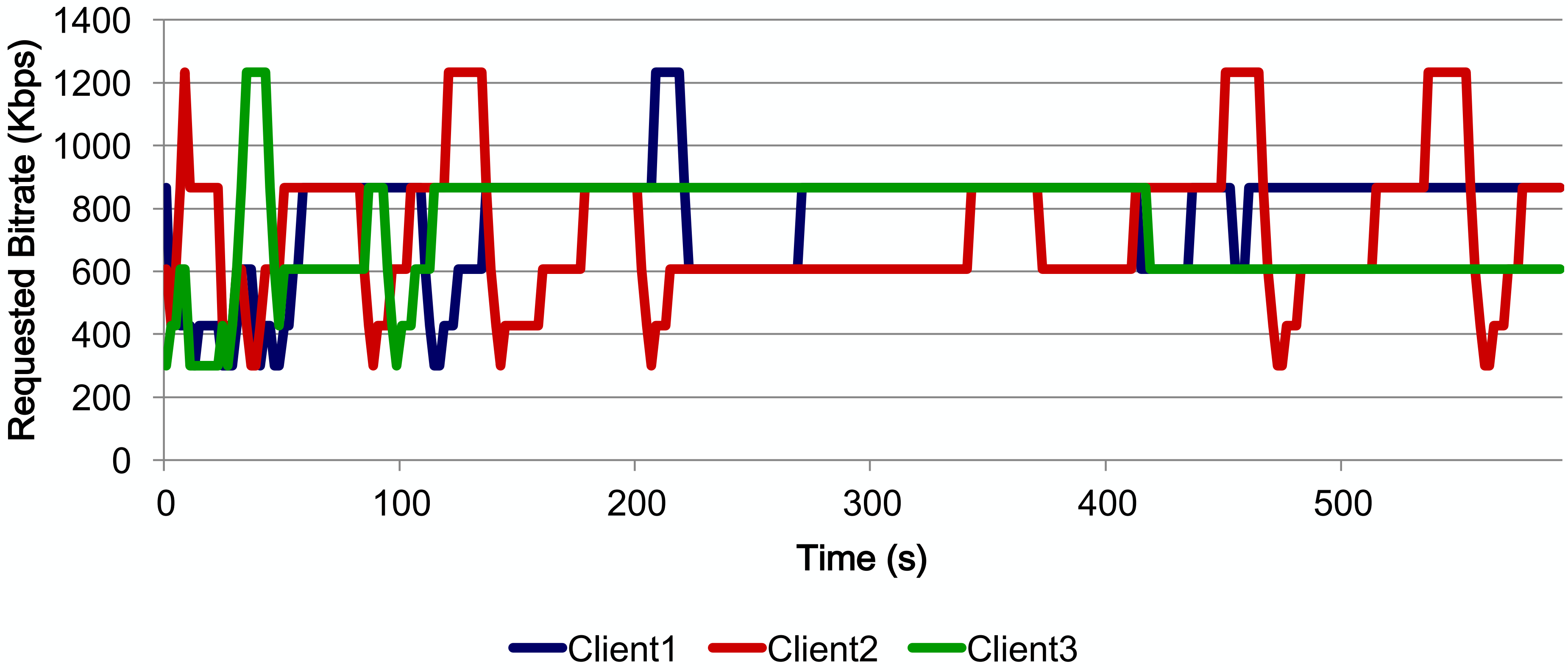
A Single Microsoft Smooth Streaming Client under a Controlled Environment



Reading: "An experimental evaluation of rate-adaptation algorithms in adaptive streaming over HTTP," ACM MMSys 2011

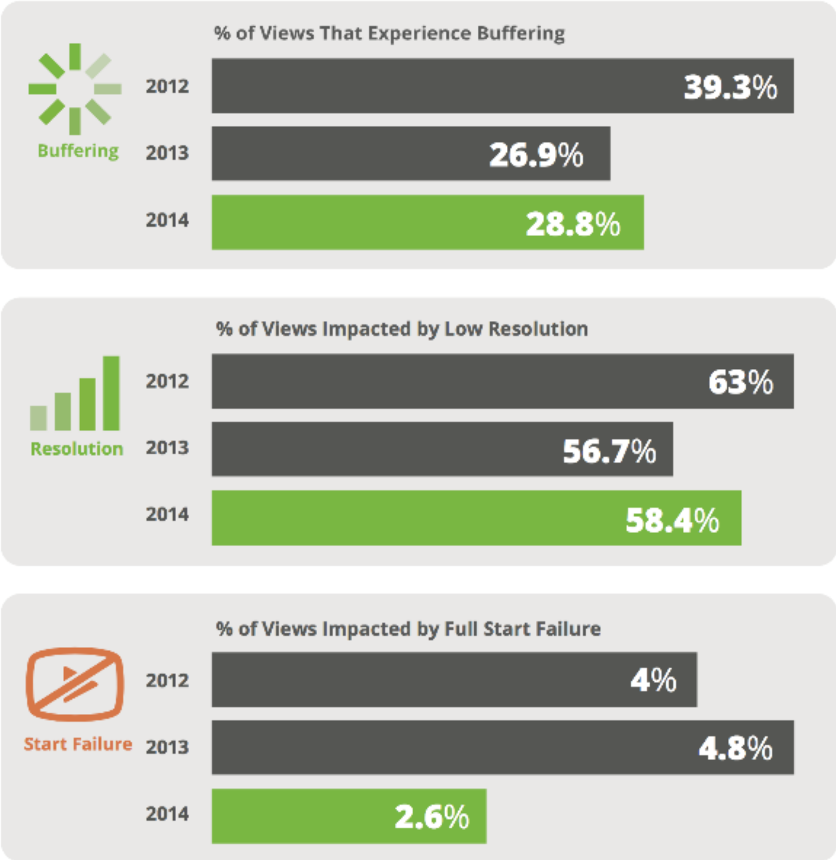
A Simple Test Scenario

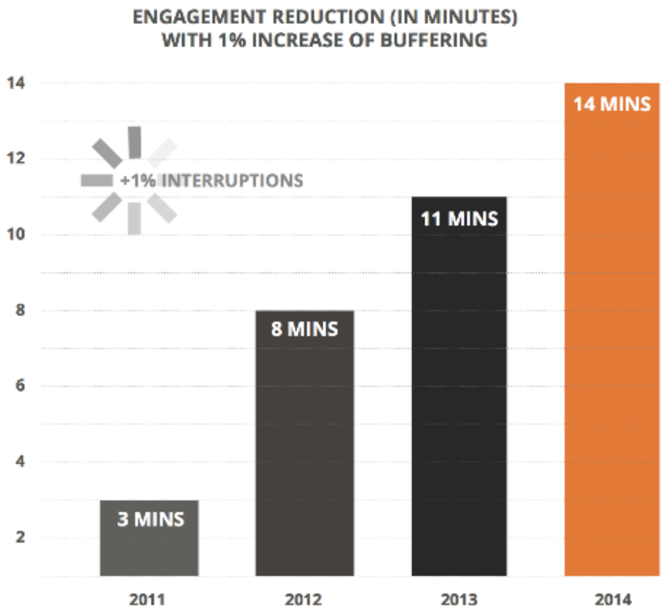
10 (Commercial) Streaming Clients Sharing a 10 Mbps Link

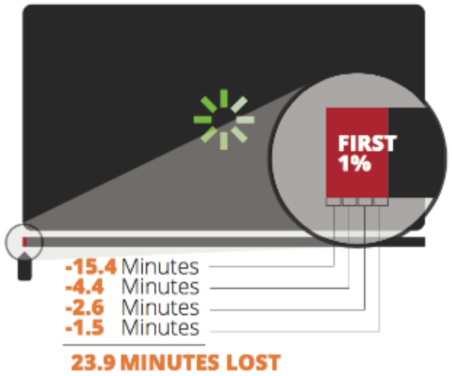


Selfishness Hurts Everyone

Viewer Experience Statistics

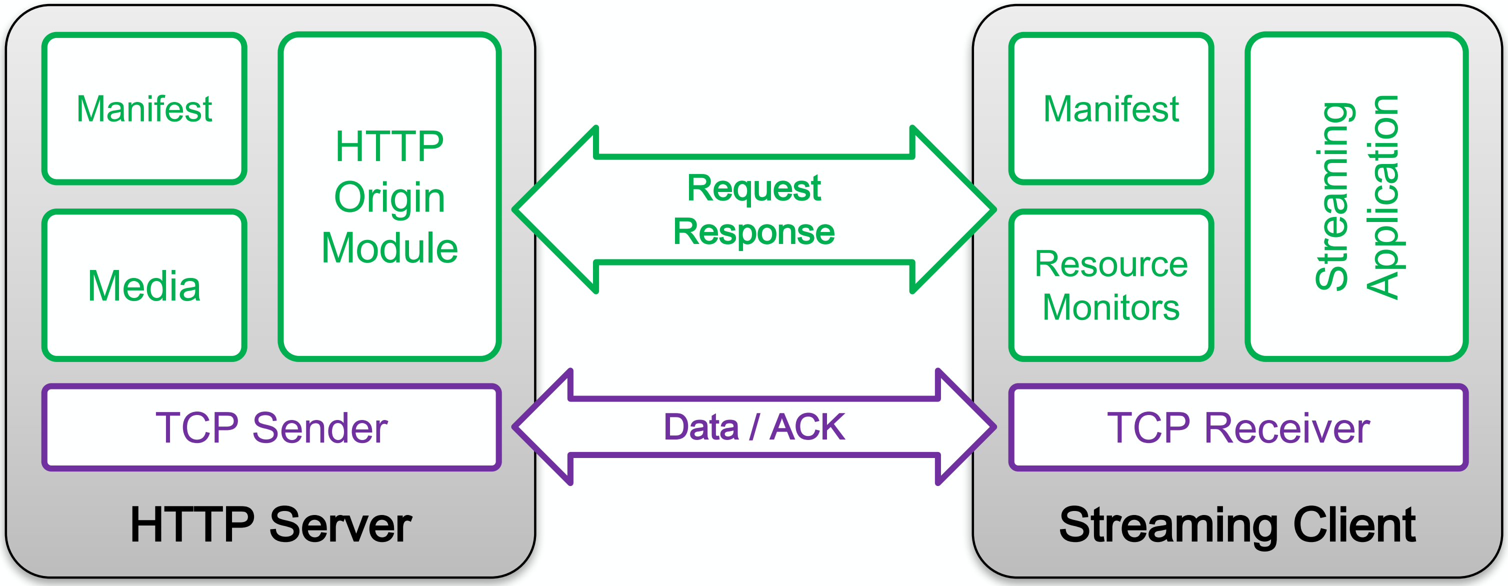






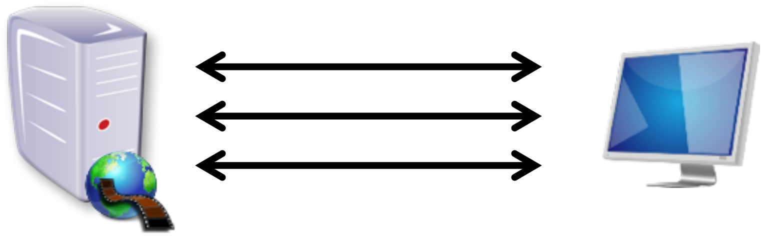
Source: Conviva Viewer Experience Report, 2015

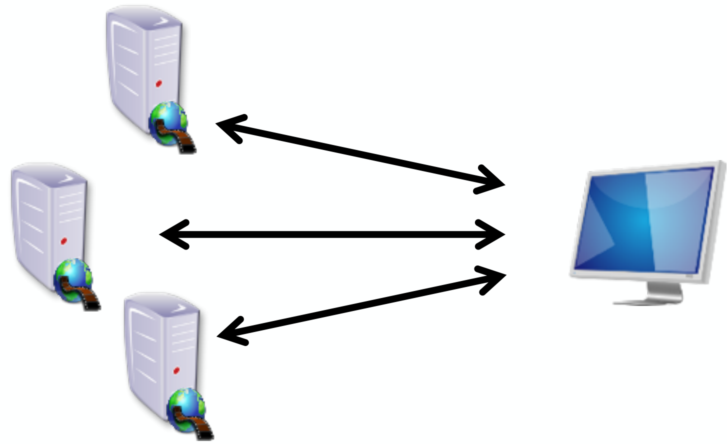
Inner and Outer Control Loops



There could be multiple TCPs destined to the same or different servers

Streaming with Multiple TCP Connections





- Using multiple concurrent TCPs
  - Should not be used to greedily get a larger share of the bandwidth
  - Can help mitigate head-of-line blocking
  - Allows fetching multiple (sub)segments in parallel
  - Allows to quickly abandon a non-working connection without having to slow-start a new one

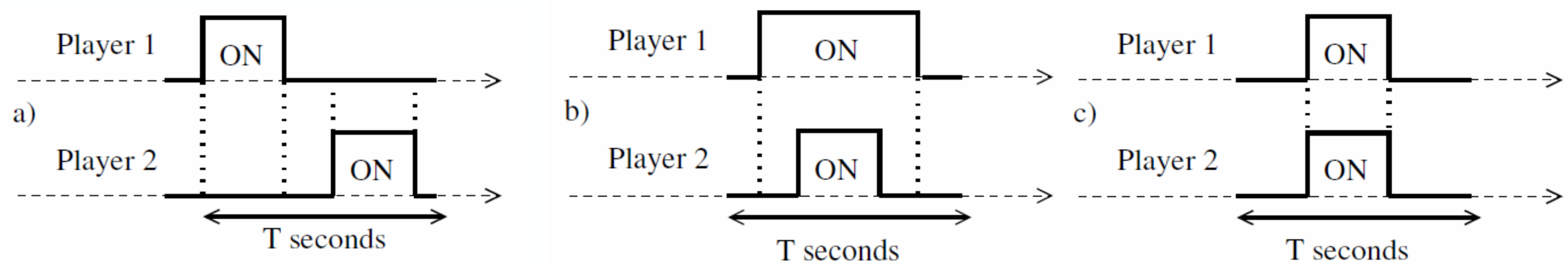
System performance deteriorates very quickly if many clients adopt this approach without limiting the aggregated bandwidth consumption

Understanding the Root Cause

Two Competing Clients

- Depending on the timing of the ON periods:
  - Unfairness, underutilization and/or instability may occur
  - Clients may grossly overestimate their fair share of the available bandwidth

Clients cannot figure out how much bandwidth to use until they use too much



Reading: "What happens when HTTP adaptive streaming players compete for bandwidth?," ACM NOSSDAV 2012

How to Solve the Issues?

Fix the clients

- Use a better algorithm like PANDA or BOLA

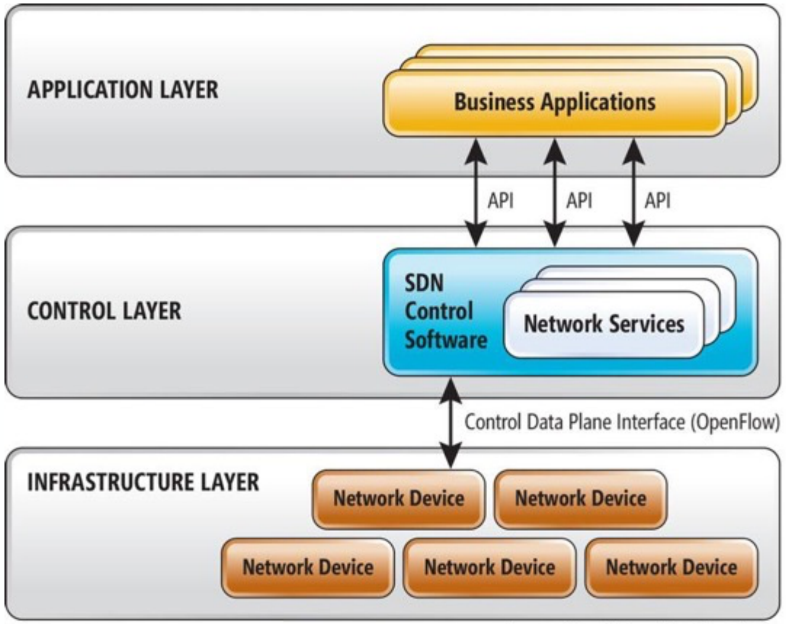
Enable a control plane

- E.g., 23009-5 SAND

Get support from the network

- QoS in the core/edge, SDN, etc.

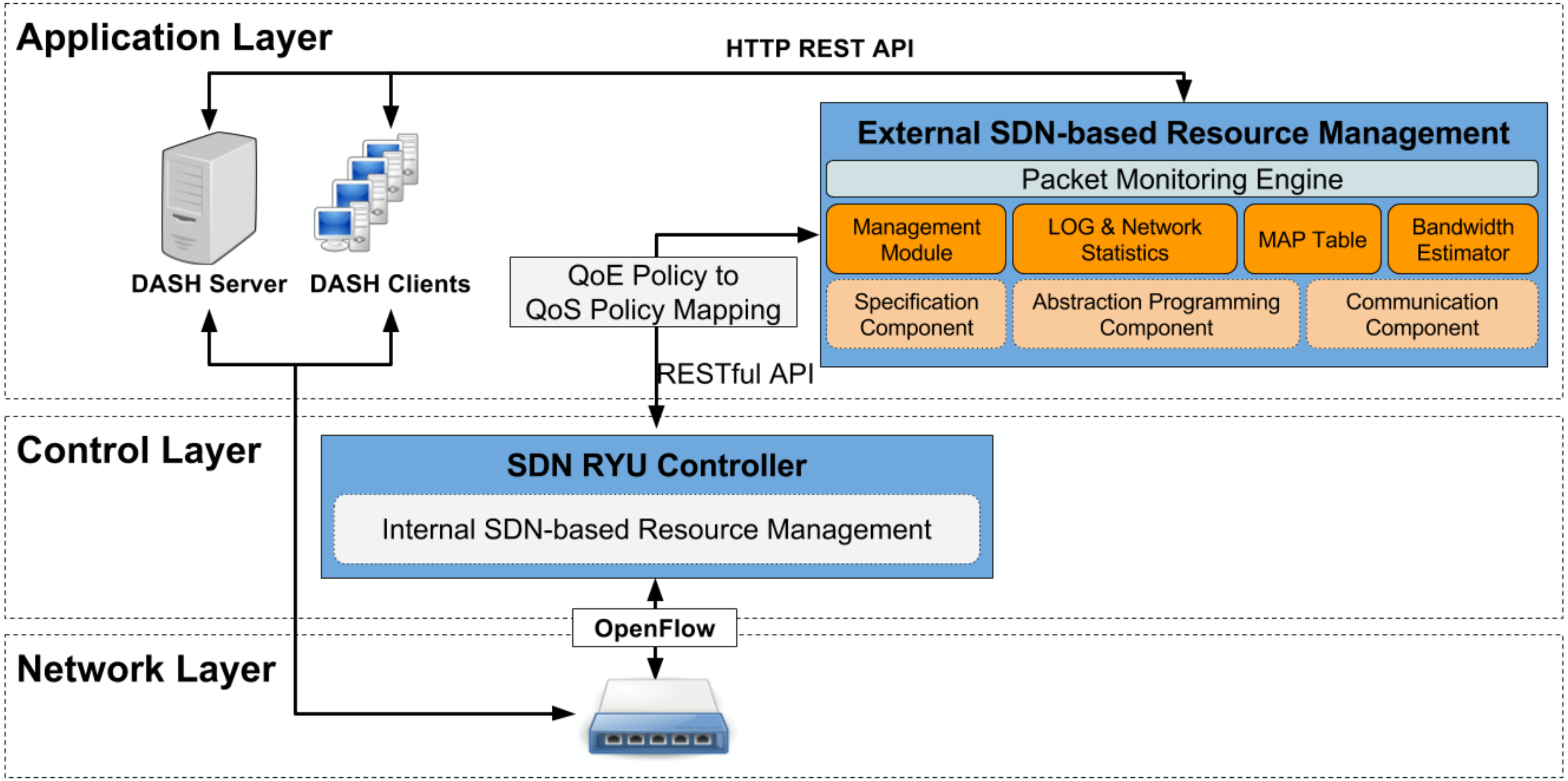
One Slide on Software Defined Networking (SDN)



Source: Open Networking Foundation

Control and data planes are decoupled, network intelligence and state are logically centralized, and the underlying network infrastructure is abstracted from the apps

SDN-Based Bitrate Adaptation



Reading: "SDNDASH: improving QoE of HTTP adaptive streaming using software defined networking," ACM MM 2016

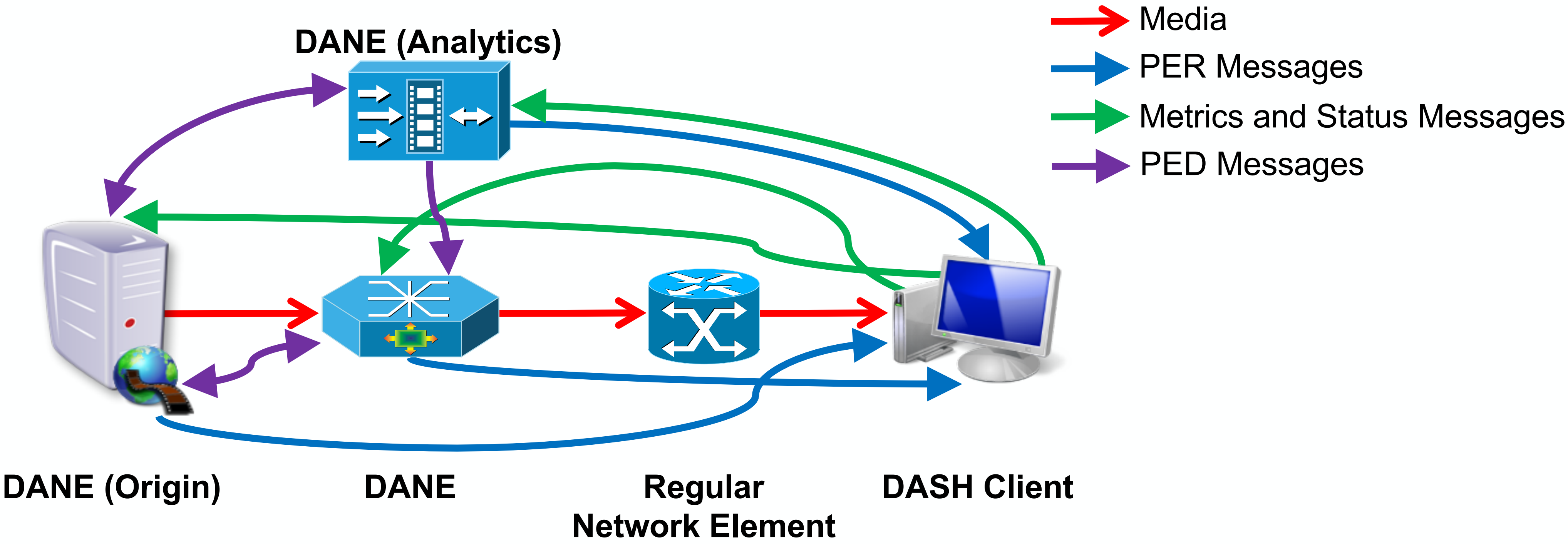
Control Points in the Ecosystem

Content Provider | Advertiser | CDN Provider | ISP | Device/ App | Consumer

- I want to make sure that my content is protected and looks awesome on any device
- I want to make sure that my ads are viewed, trackable and measurable
- I want to make sure that my servers are properly used and latency is low
- I want to control the QoE of all my customers, differentiate my own services, make $$$ from OTT services
- I want to make sure that my device/app provides the best possible video quality
- I want the best quality for minimal $

A Control Plane Approach

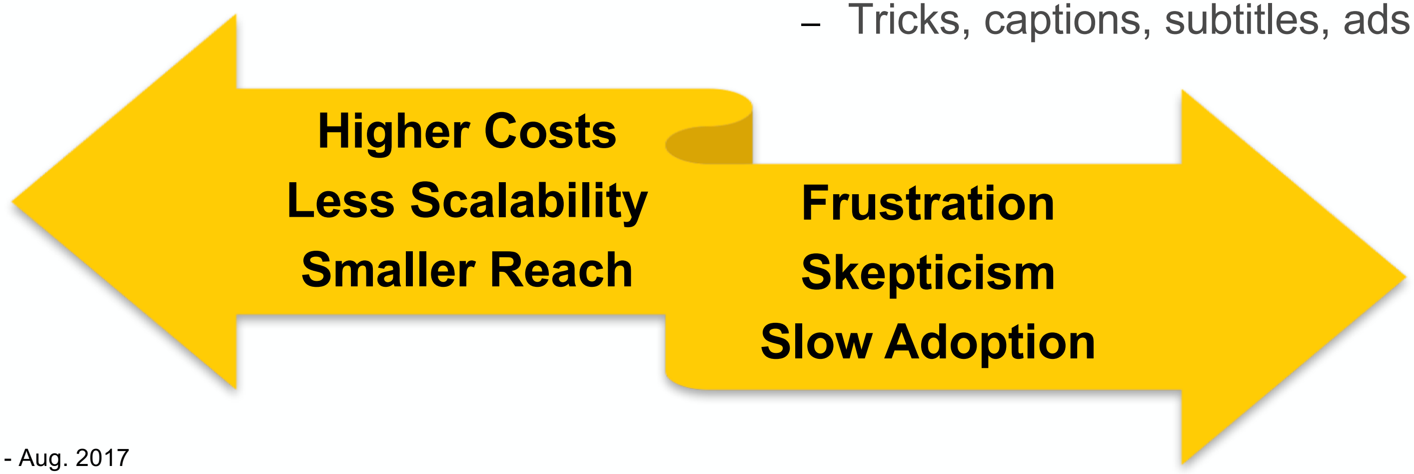
DASH Part 5: Server and Network Assisted DASH (SAND)



DANE: DASH-assisting network element
PER: Parameters for enhancing reception
PED: Parameters for enhancing delivery

What does Fragmentation Mean?

- Fragmented architectures
  - Advertising, DRM, metadata, blackouts, etc.
- Investing in more hardware and software
  - Increased CapEx and OpEx
- Lack of consistent analytics
- Preparing and delivering each asset in several incompatible formats
  - Higher storage and transport costs
- Confusion due to the lack of skills to troubleshoot problems
- Lack of common experience across devices for the same service
  - Tricks, captions, subtitles, ads, etc.



DASH intends to be to the Internet world …

what MPEG2-TS has been to the broadcast world

DASH intend~~s~~ed to be to the Internet world …
what MPEG2-TS has been to the broadcast world

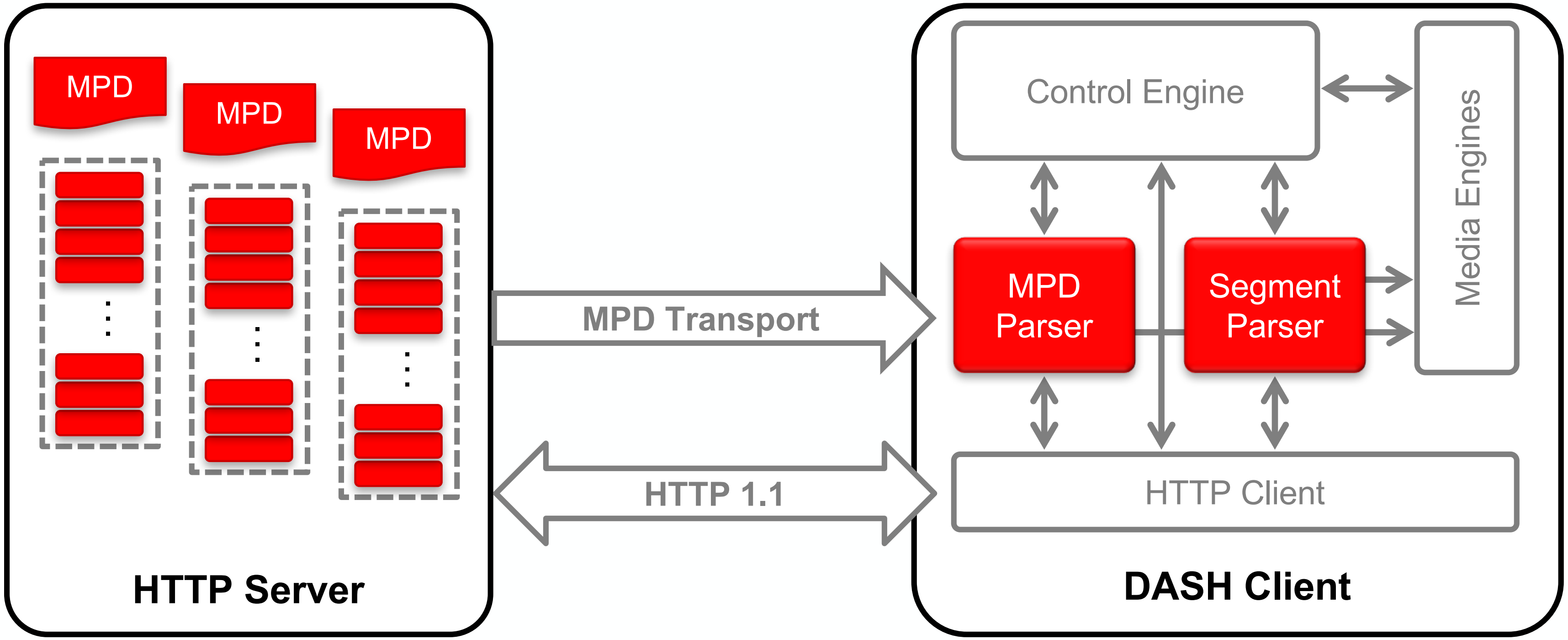

WHAT IS HAPPENING IN MPEG ?
1. DASH
2. CMAF

DASH Design Principles

- DASH is not
  - system, protocol, presentation, codec, interactivity, DRM, client specification
- DASH is an enabler
  - It provides formats to enable efficient and high-quality delivery of streaming services over the Internet
  - System definition is left to other organizations
- Design choices
  - Enable reuse of existing technologies (containers, codecs, DRM, etc.)
  - Enable deployment on top of CDNs
  - Enable live and on-demand experiences
  - Move intelligence from network to client, enable client differentiation
  - Provide simple interoperability points (profiles)

Scope of MPEG DASH

Shown in Red



DASH (ISO/IEC 23009) Parts

- 23009-1: Media Presentation Description and Segment Formats
  - Amd. 1: NTP sync, extended profiles
  - Amd. 2: SRD, URL parameter insertion, role extensions
  - Amd. 3: External MPD link, period continuity, generalized HTTP header extensions/queries
  - Amd. 4: TV profile, MPD chaining/resetting, data URLs in MPD, switching across adaptation sets
  - → 3rd edition (FDIS), to be published soon
- 23009-2: Conformance and Reference Software
  - 2nd edition (FDIS), to be published in 2017
  - Amd. 1: SAND conformance rules
- 23009-3: Implementation Guidelines (Informative)
  - 2nd edition Amd. 1: Multi-track content guidelines and rules, fix on dependent representations
  - 3rd edition is planned
- 23009-4: Segment Encryption and Authentication
  - 2nd edition in progress (DIS)

DASH (ISO/IEC 23009) Parts

- 23009-5: Server and Network Assisted DASH (SAND)
  - 1st edition, published in Feb. 2017
- 23009-6: DASH over Full Duplex HTTP-based Protocols (FDH)
  - 1st edition (FDIS), to be published in 2017
- 23009-7: Delivery of CMAF Contents with DASH (Informative)
  - WD available, PDTR to be published after the Oct. meeting

Part 1 – 3rd Edition

- TV profile
  - Broadcast TV content can be distributed over broadcast and/or broadband networks using DASH
  - Segments do not need to be a switching access point
- MPD chaining
  - Multiple MPDs can be chained to each other for easier ad insertion
- Data URLs in MPD
  - Initialization segments can be included in the MPD (only in the TV profile)
- Switching across adaptation sets
  - A capable client can switch between AVC and HEVC tracks

Part 6 – DASH over Full Duplex HTTP-Based Protocols

- Part 6 describes how a client can signal various types of “push” behaviors to a server in HTTP/2 and for WebSockets
- This is primarily targeted for low-latency live streaming
- Similar activities
  - DASH over LTE broadcast (3GPP SA4)
  - DVB ABR multicast
  - ATSC 3.0 hybrid delivery

Ongoing Work as of MPEG 119 (July 2017)

- Core Experiments
  - High quality VR delivery with DASH (DASH-VR)
- Technologies under Consideration
  - Support for Controlled Playback in DASH
  - Playback control
  - Owner defined content identifiers within MPDs
  - Usage of HEVC tile tracks in DASH
  - Relation between presentation time and EPT
  - Service-level service protection using segment encryption
  - Explicit segment loss indication
  - Using segment templates for forensic watermarking

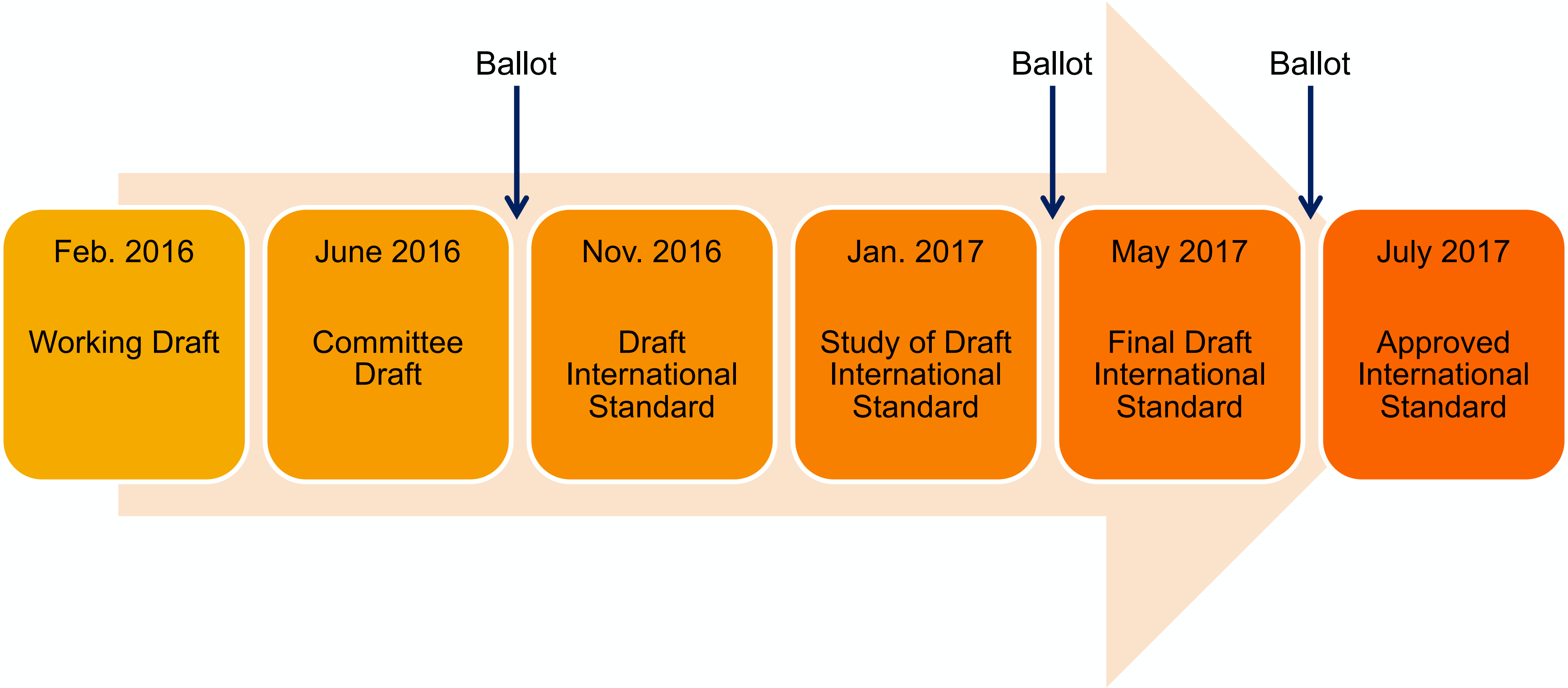
Common Media Application Format (CMAF)

- Media delivery has three main components:
  - Media format
  - Manifest
  - Delivery
- CMAF defines the media format only (Fragments, headers, segments, chunks, tracks)
  - CMAF does not specify a manifest format
  - CMAF does not specify a delivery method, either
- CMAF uses ISO-BMFF and common encryption (CENC)
  - CENC means the media fragments can be decrypted/decoded by devices using different DRMs (aka multi-DRM support)
  - CMAF does not mandate CTR or CBC mode (which makes CMAF useful only for unencrypted content for now)
- MPEG technologies (DASH and MMT) may be used for delivering CMAF content

Common Media Application Format (CMAF)

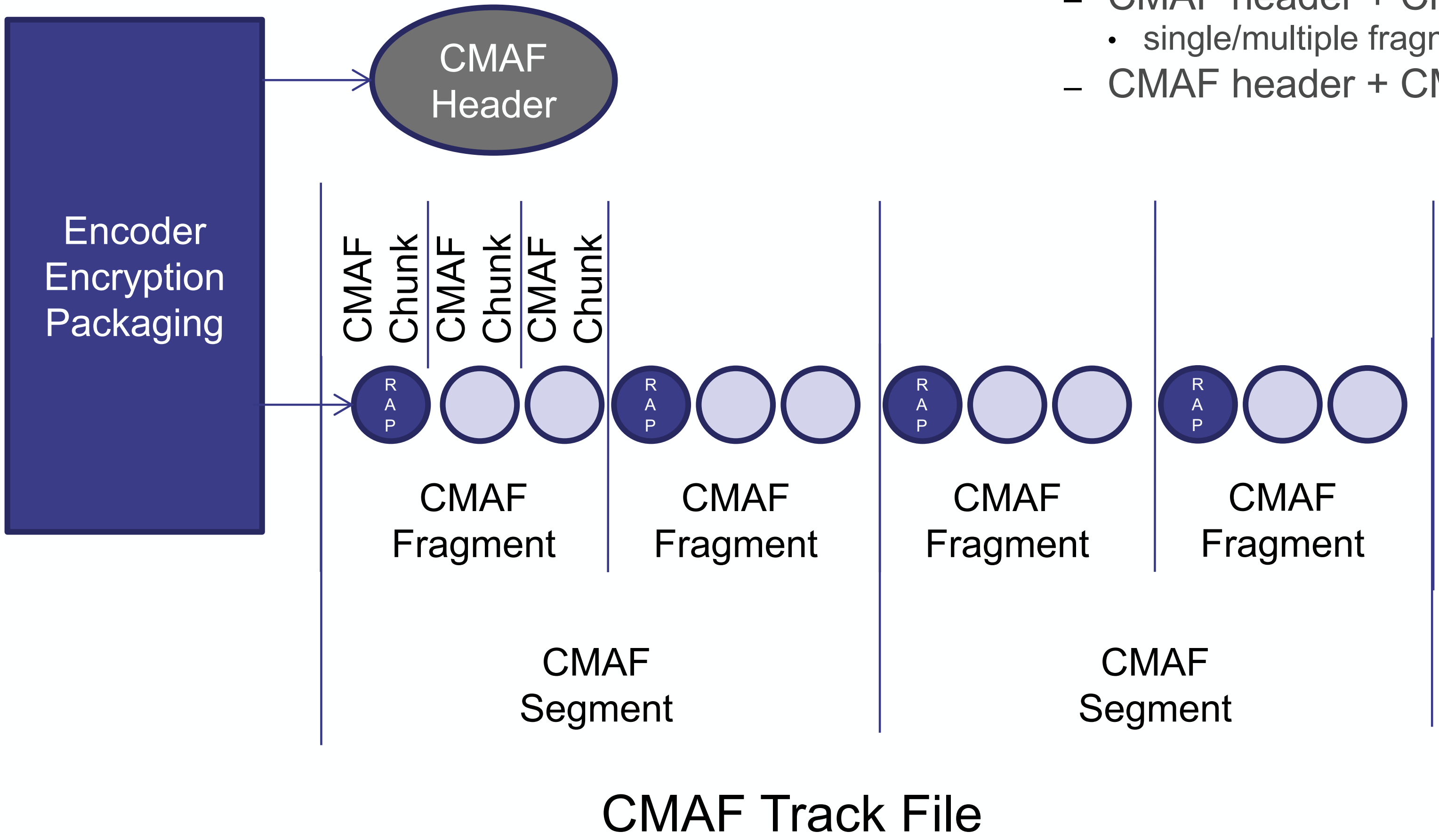
- CMAF defines media profiles for
  - Video
  - Audio
  - Subtitle
- CMAF defines presentation profiles by selection a media profile from each category
- Current status
  - iOS 10 supports CMAF with HLS manifests
  - PDAM1 (w16821)
    - SHVC support
    - AAC 7.1 support
  - Conformance (w16999)
  - Work in progress along with DASH-IF test vectors
  - PDAM2 considered for USAC and other media profiles

CMAF Development Timeline

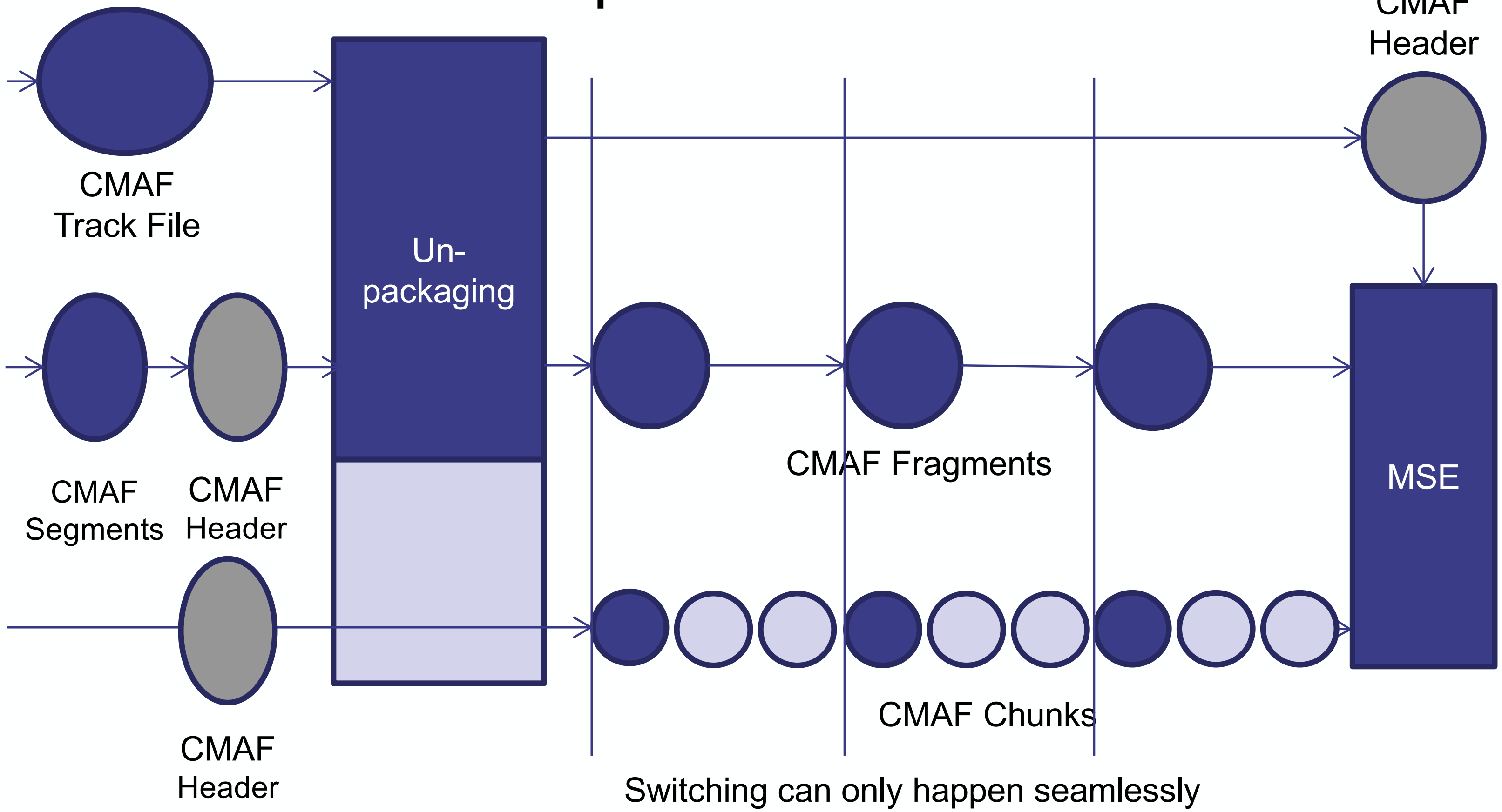


CMAF ISO-BMFF Media Objects

- Manifests typically provide URLs to
  - CMAF track files
  - CMAF header + CMAF segments
    - single/multiple fragment(s)
  - CMAF header + CMAF chunk



CMAF Content Consumption

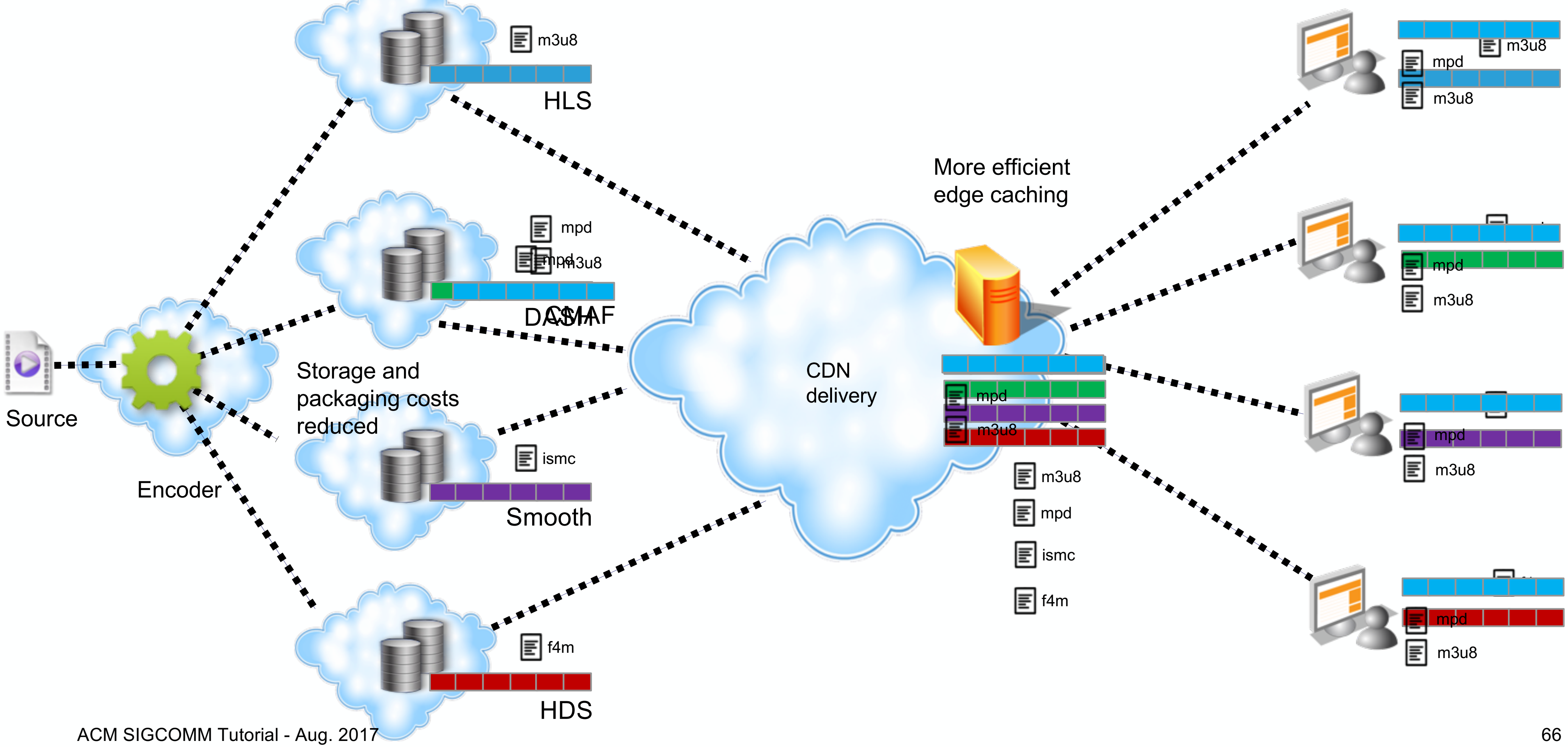


Switching can only happen seamlessly at CMAF Fragment boundaries

Commercial OTT Issues

- Content format issues
  - Each asset is copied to multiple media formats
    - different video codecs
    - different audio codecs
    - different (regional) frame rates
  - Cost to content creators and distributors
  - Inefficiencies in CDNs and higher storage costs
- Platform issues
  - Lack of consistent app behavior across platforms
  - Varying video features, APIs and semantics across platforms
- Playback issues
  - Codec incompatibility
  - Partial profile support
  - Switching bitrate glitches
  - Audio discontinuities
  - Ad splicing problems
  - Long-term playback instability
  - Request protocol deficiencies
  - Memory problems, CPU weaknesses
  - Scaling (display) issues
  - Variable HDR support
  - Unknown capabilities

Multi-Platform OTT Workflow with CMAF



CTA Web Application Video Ecosystem (WAVE)

Content Specification	Device Playback Requirements	HTML5 Reference Platform
Content specification based upcoming CMAF, compatible with DASH and HLS	Testable requirements covering the most common device playback interoperability issues	Reference application framework based on HTML5 providing functional guidelines for playback interop

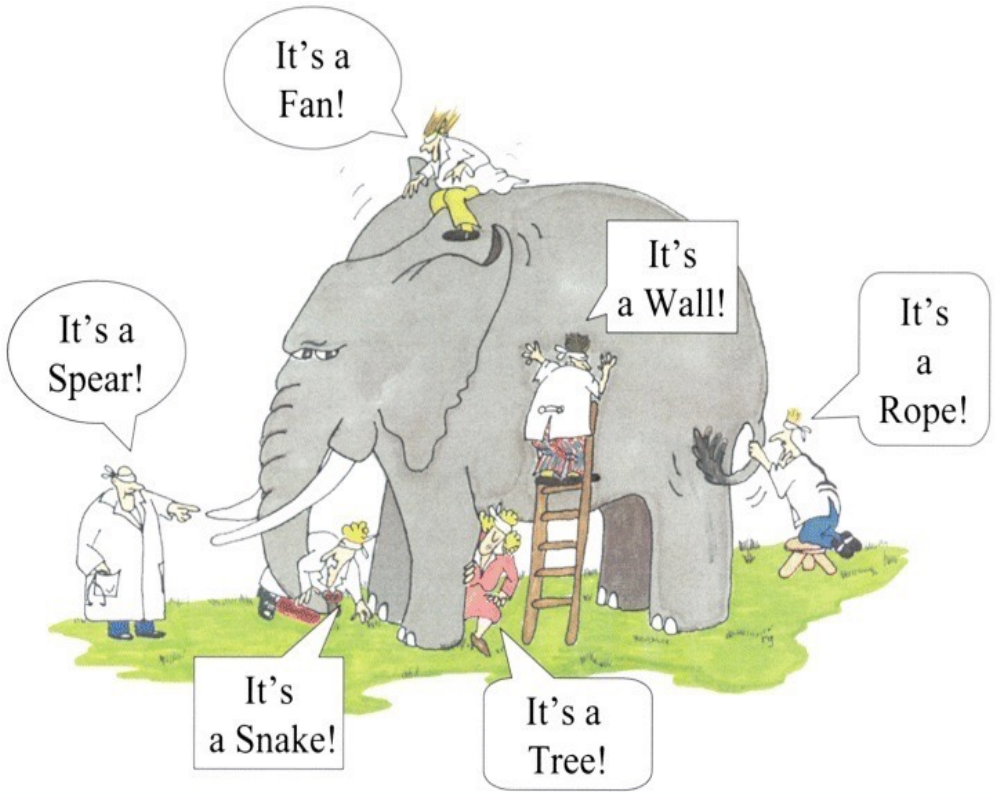
Test Suite

Web Media API Community Group

- Developers want to deploy their content on a range of devices and platforms, e.g., TVs, set-tops and mobile devices
- To ensure a smooth user experience across devices, these user agents need to support a minimum set of Web technologies
- This Community Group plans to specify such a set of Web technologies and additionally plans to provide guidance for developers and implementers
- Web Media APIs 2017
  - Details the Web APIs that should be included in device implementations in 2017
  - https://www.w3.org/community/webmediaapi/
- Web Media Application Developer Guidelines
  - Companion guide to outline best practices for implementing Web media apps
  - https://w3c.github.io/webmediaguidelines/

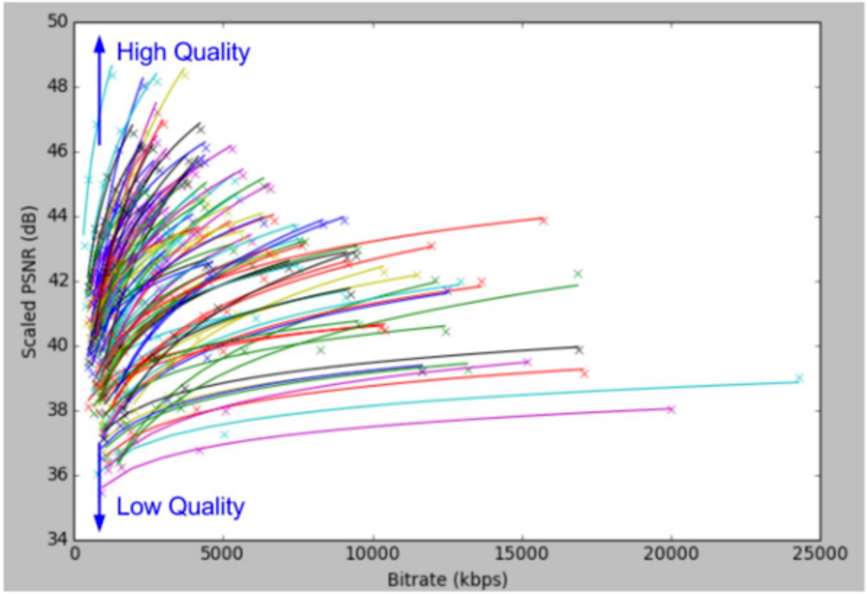
What is Quality?

Many Definitions Do Exist – It is Like an Elephant



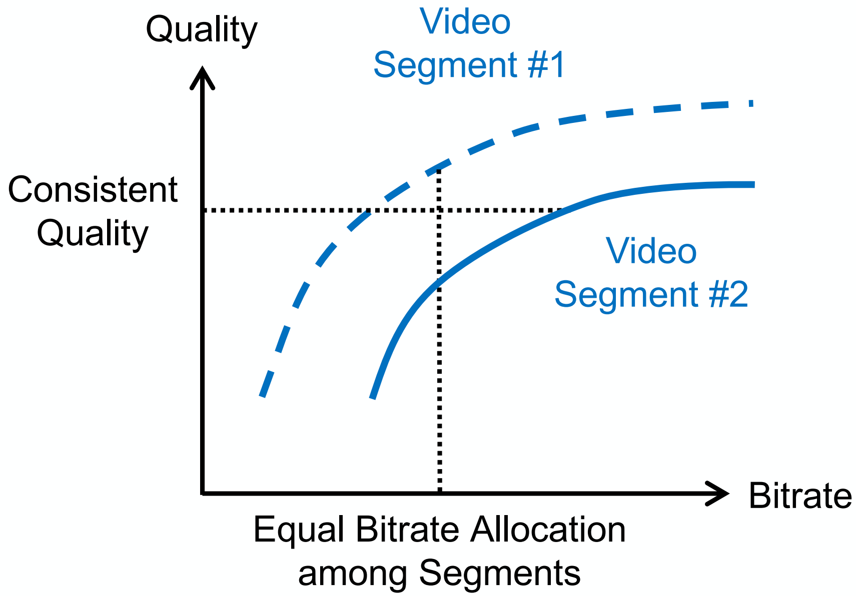
Increased bandwidth brings new opportunities

Drives overall maturity of IP/ABR delivery technologies, introduces new opportunities – Per Title Encode Optimization



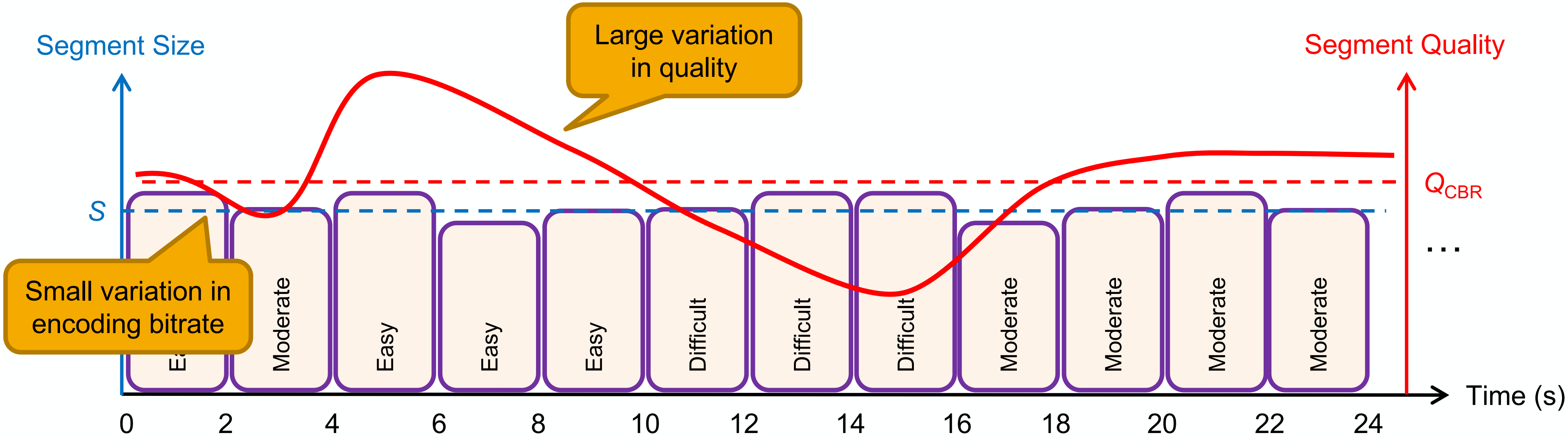
Source: Netflix Tech Blog, December 2015

Segments Have Different Complexities



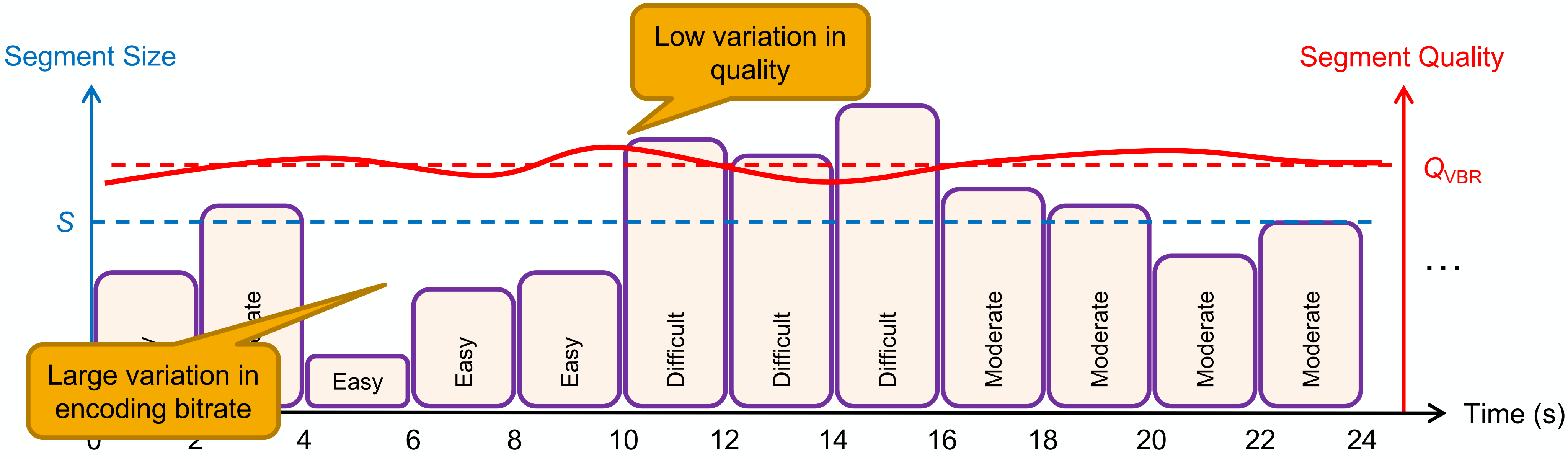
Adaptation Feature Does Not Deliver Consistent Quality

Guidelines Limited Bitrate Variability to (Mostly) 10% So Far



If there is something worse than having to watch a video at a lousy quality, it is to watch that video with varying quality

What if We Encode in a More Subtle Fashion?



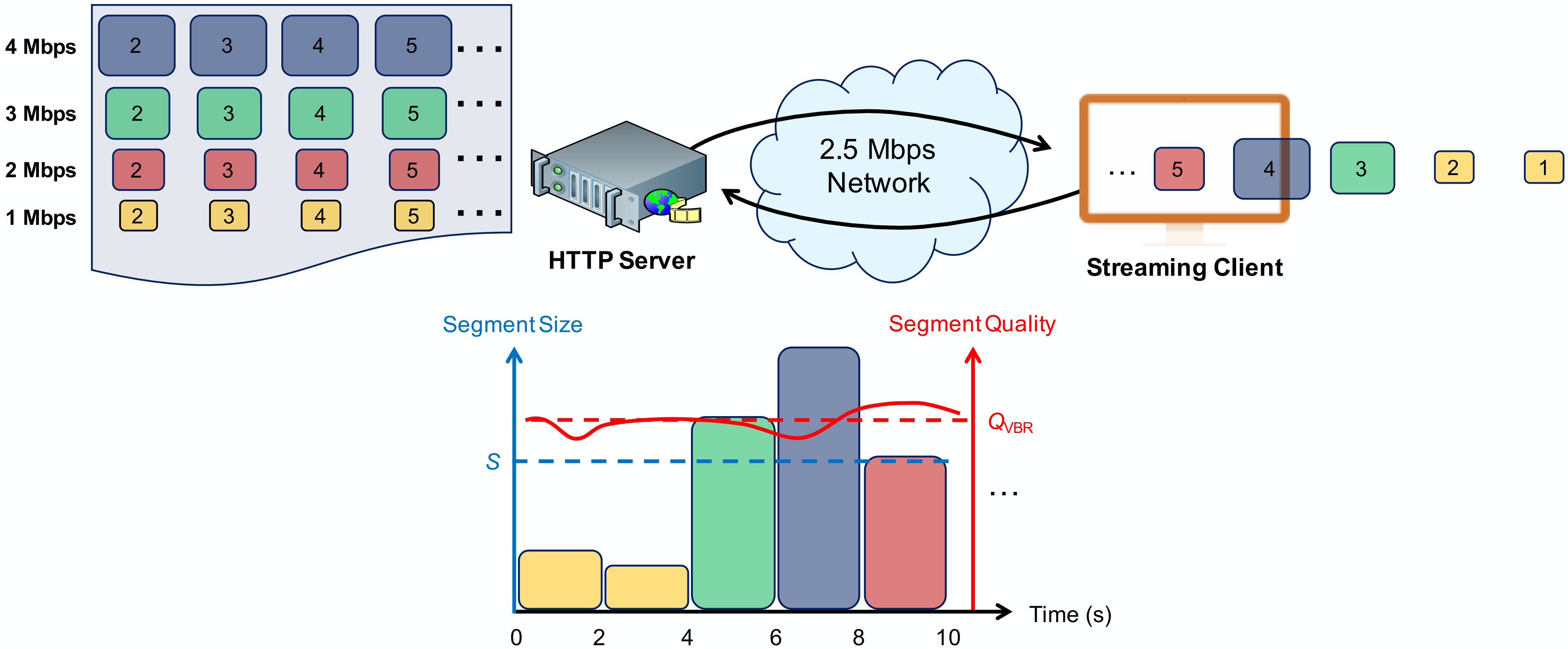
While we spend the same total amount of bits, we not only increase average quality but also reduce quality variation

HLS authoring spec for ATV allows 2x capping rate for VoD. For linear content, variability is limited to 10-25% range.

Generating VBR-encoded segments is easy, but streaming them is not!

Content Aware Encoding	≠	Content Aware Streaming

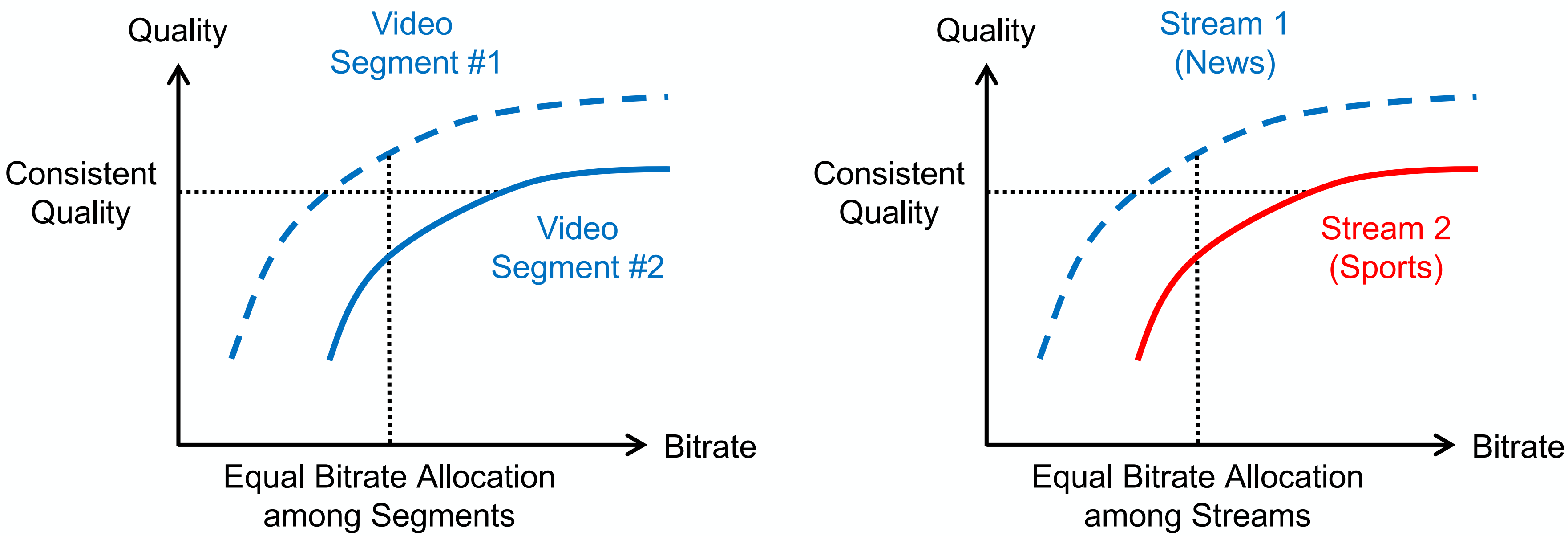
Multiple Representations Naturally Enable "Cherry-Picking"



Reading: "Streaming video over HTTP with consistent quality," ACM MMSys 2014

Dimensions – In-Stream vs. Across-Streams

- Same principle applies to both:
  - In-stream: Temporal bit shifting between segments
  - Across-streams: Bit shifting between streams sharing a bottleneck link



Reading: "Spending quality time with the Web video," IEEE Internet Comput., 2016

Deployment Challenges

- Challenge 1: Development of quality metrics and temporal pooling models
  - A common metric that is suitable for a variety of content types
  - A temporal pooling model that will reliably work for different viewer profiles, devices and networks
    - Different viewers have different sensitivity levels to glitches for different content types, and they are also forgiving in different time scales
      - A young viewer (likely to have longer-term memory) watching sports on a big screen vs. an elder viewer (likely to have short-term memory) watching news on a smaller screen
- Challenge 2: Integration into popular streaming client implementations
  - Many ecosystems are closed or proprietary, and one may not have access to the client algorithm to make the necessary changes (e.g., Apple HLS in iOS)
  - Standards bodies and industry consortiums may lead the way to develop certain guidelines

Reading: "Quality-aware HTTP adaptive streaming," IBC 2015

Deployment Challenges

- Challenge 3: Development of metadata standards
  - Computing the quality metric for each segment in each representation for each content is a tedious task, which is the easiest to deal with at the encoder or packager
  - Packing the metric values and conveying this information to all the clients in a timely and scalable manner is an equally important task
  - The timed metadata spec in MPEG (ISO/IEC 23001-10) is a good candidate for this task

Deployment Challenges

- Challenge 4: Expansion to multi-client scenarios
  - We need controlled unfairness (which is fairness in quality not bitrate) among clients adaptively streaming the same or a different content over a network sharing resources (e.g., access network)
    - Easy scenario: One adult watching sports on a big screen vs. one adult watching a food show on a tablet
    - More complex scenario: One adult watching sports on a phone vs. three adults watching news on a big screen
  - The optimization across a number of streaming clients has to be done based on the utilities of the streamed videos, which depend on factors such as:
    - Spatial pooling model
    - Content types
    - Content features
    - Rendering devices
    - Audience profiles and sizes
  - SAND can help deploy controlled unfairness that we need in quality-aware streaming in multi-client scenarios

This is Why We Need Analytics

OS Vendor:
“Your Internet connection must be bad”

Service Provider:
“Your video or CDN provider must be slow”

Video/CDN Provider:
“Your home network must be slow”

Consumer:
“The device or the app is slow”

Device/App Vendor:
“It must be the OS”

Four Major Areas of Further Exploration

Content Preparation	• How to choose bitrate/resolution pairs to make up/downshifts least visible • How to pick the segment durations • How to achieve content aware streaming and not just content aware encoding
Distribution and Delivery	• What information could the network provide to streaming clients • How to achieve controlled unfairness • MPTCP or QUIC better than TCP? What about multicast and HTTP/2?
QoE Modeling and Client Design	• How to model streaming dynamics for different genres • How to model the impact of faster zapping and trick modes on the QoE • Understanding the impact of QoE on viewer engagement
Analytics, Fault Isolation, Diagnostics	• Understanding the interaction of adaptive streaming with caching in CDNs • Extracting actions based on real-time analytics • Fixing issues faster and remotely

Part II: Omnidirectional (360°) Media

What is Virtual Reality (VR)

MPEG's Definition

VR is a rendered environment (visual and acoustic, pre-dominantly real-world) providing an immersive experience to a user who can interact with it in a seemingly real or physical way using special electronic equipment

Virtual Reality (VR) Puts Us in a Virtual World

Source: Phil Chou

Augmented Reality (AR) Puts Virtual Objects in Our World

Source: Phil Chou

The VR Challenge

Delivering High-Quality VR in Economic Scale is Extremely Challenging

Bit depth Frame rate Stereoscopic

$$\underbrace{30K \times 24K}_{\text{Resolution}} \times \overbrace{36}^{\text{Bit depth}} \times \overbrace{60}^{\text{Frame rate}} \times \overbrace{2}^{\text{Stereoscopic}} / \underbrace{600}_{\text{Compression gain}}$$

~<u>5.2 Gbps</u>

Resolution Compression gain

Ultimate Level of Immersion

Visuals
So vibrant that they are eventually indistinguishable from the real world

Interactions
So intuitive that they become second nature

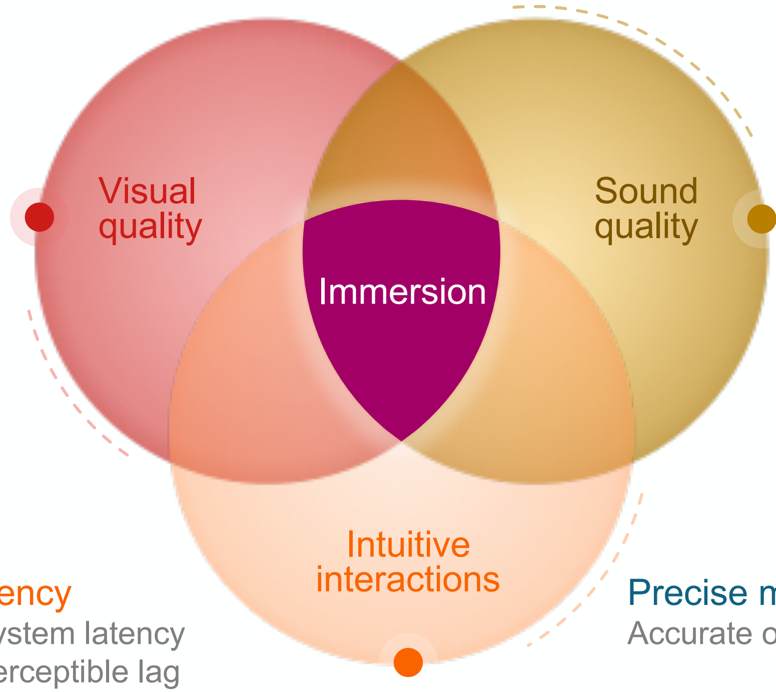
Sounds
So accurate that they are true to life

Immersive VR Has Extreme Requirements

Extreme pixel
quantity and quality
Screen is very close to the eyes

Spherical view
Look anywhere with a
full 360° spherical view

Stereoscopic display
Humans see in 3D



High resolution audio
Up to human hearing capabilities

3D audio
Realistic 3D, positional, surround
audio that is accurate to the real world

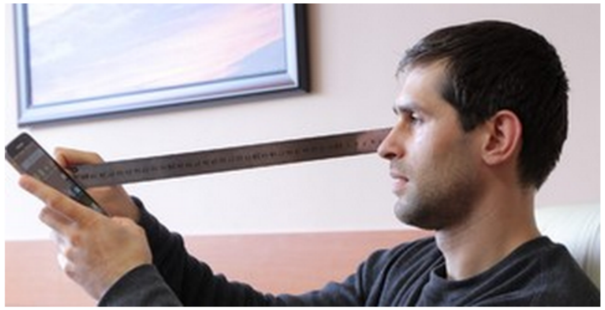
Minimal latency
Minimized system latency
to remove perceptible lag

Precise motion tracking
Accurate on-device motion tracking

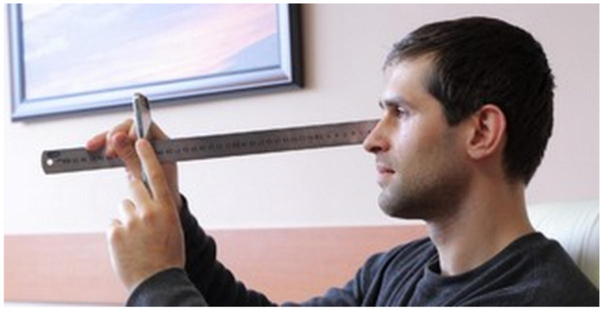
Natural user interfaces
Seamlessly interact with VR using
natural movements, free from wires

Source: Thomas Stockhammer

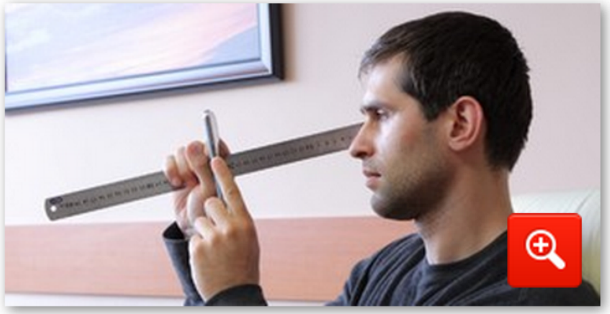
Need for Higher Resolutions on Mobile Phones?



480p, average person starts noticing pixelization at around 14.7 inches (here - 37.4cm) in a 4-inch 480p phone



720p, average person starts noticing pixelization at around 11 inches in a 4.7-inch 720p phone



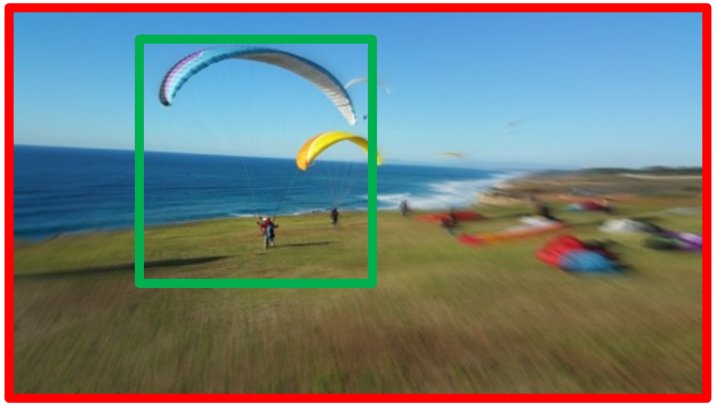
1080p, average person starts noticing pixelization at around 7.8 inches in a 5-inch 1080p phone

Quad HD, average person starts noticing pixelization at around 6.44 inches in a 5.5-inch 1440p phone

Source: Thomas Stockhammer

Foveated Rendering Reduces Pixel Processing

The Human Eye can Only See High Resolution Where the Fovea is Focused



- High resolution
- Low resolution

The GPU renders a small rectangle at a high resolution and the rest of the FoV at a lower resolution

Generating and Consuming 360° Spherical Video

Generate Video

- Simultaneously capture video with multiple cameras from different views to generate 360° spherical video
- Use twice the cameras for stereoscopic video
- Undistort, stitch together and map the discrete images to a equirectangular or cubemap format
- Encode video

Playback Video

- Decode video
- Apply an equirectangular or cubemap UV projection
- Determine the pose and show appropriate view of 360° spherical video

Discrete unstitched camera images for 360° spherical view

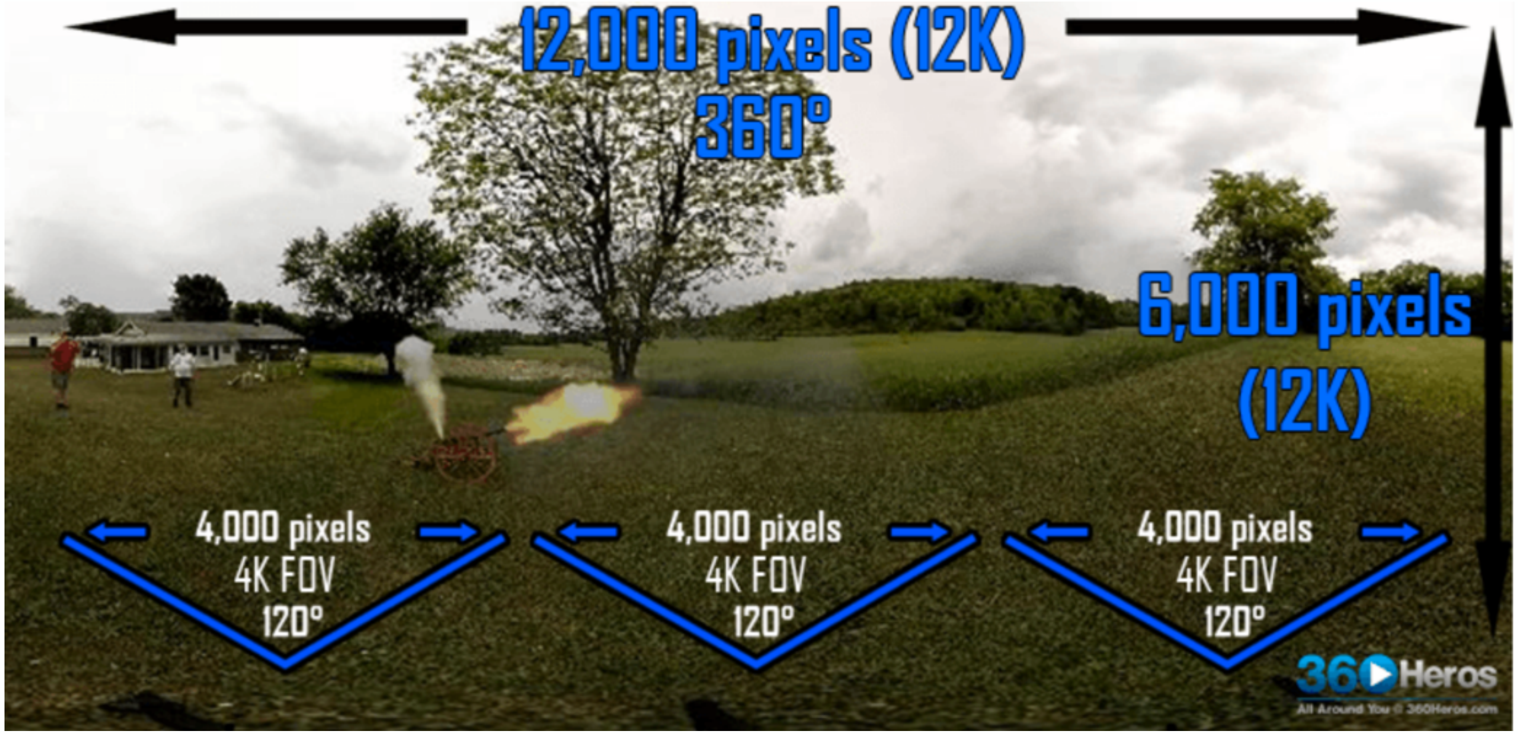
Equirectangular image

Cubemap image

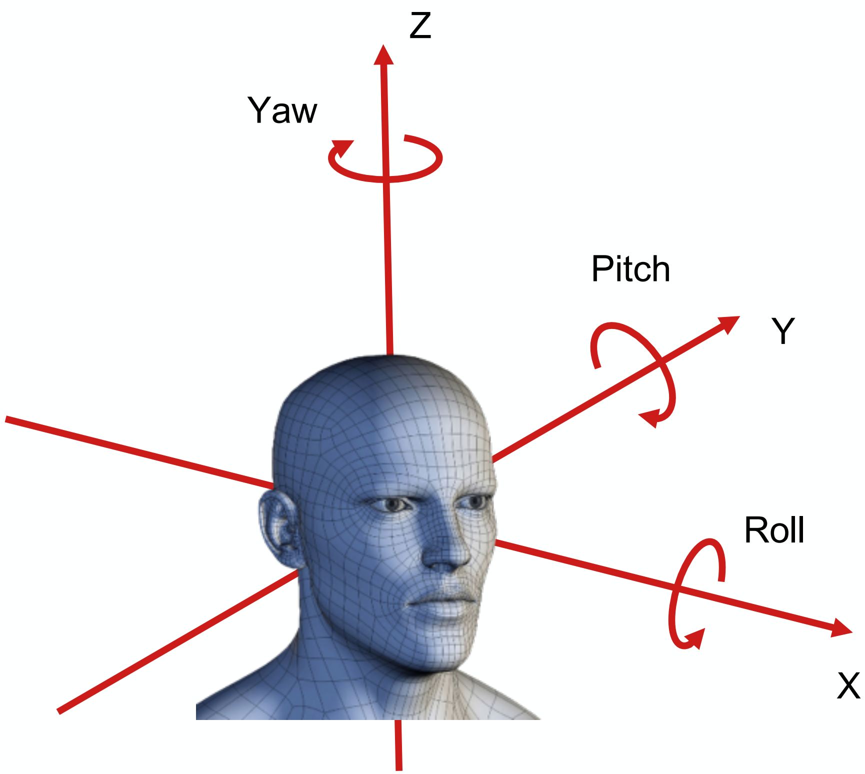
Left eye VR headset view

Adaptive Streaming of VR Content

- Required resolution of a panoramic video for achieving 4K resolution for viewport with 120° field of view (FoV)



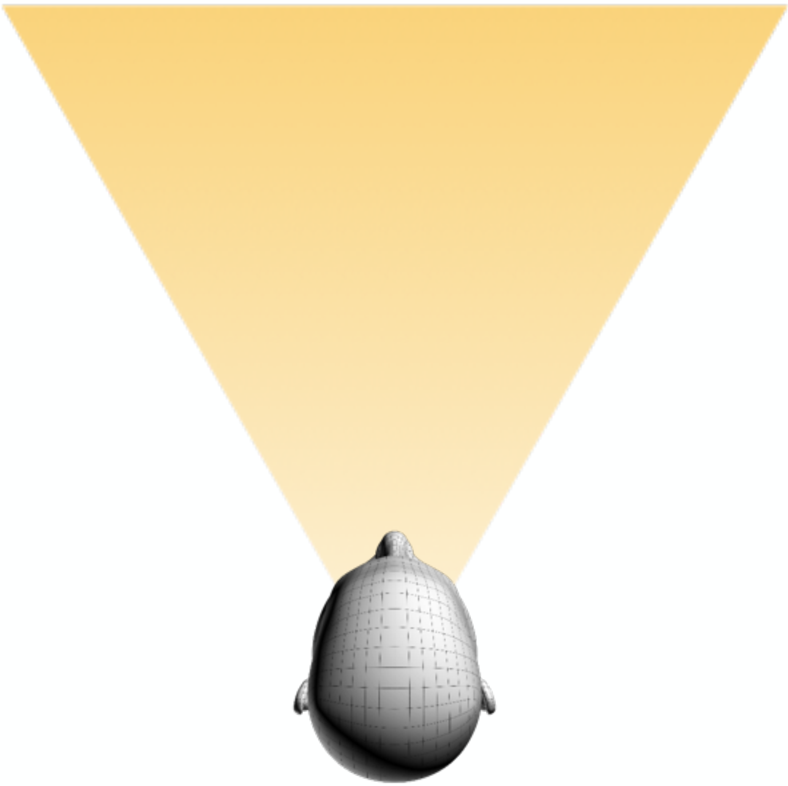
Precise Motion Tracking of Head Movements and Gaze



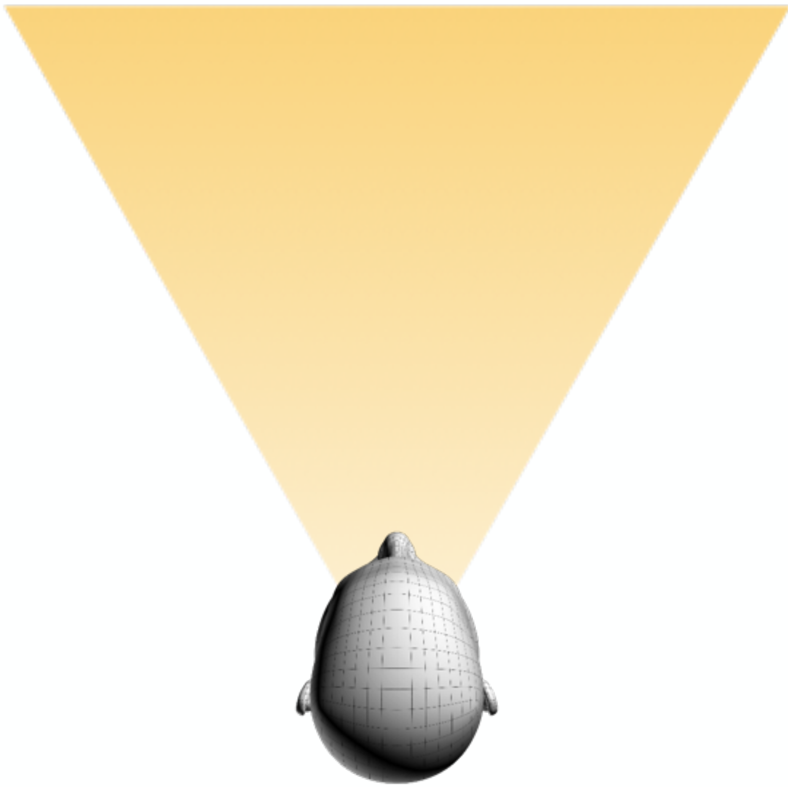
- 3-DoF
  - "In which direction I look," look around the virtual world from a fixed point
  - Detect rotational movement
- 6-DoF
  - "Where I am and in which direction I look," move freely in the virtual world and look around corners
  - Detect rotational movement and translational movement

Minimizing Motion-to Photon Latency is Important

Lag Prevents Immersion and Causes Discomfort



Low latency



Noticeable latency

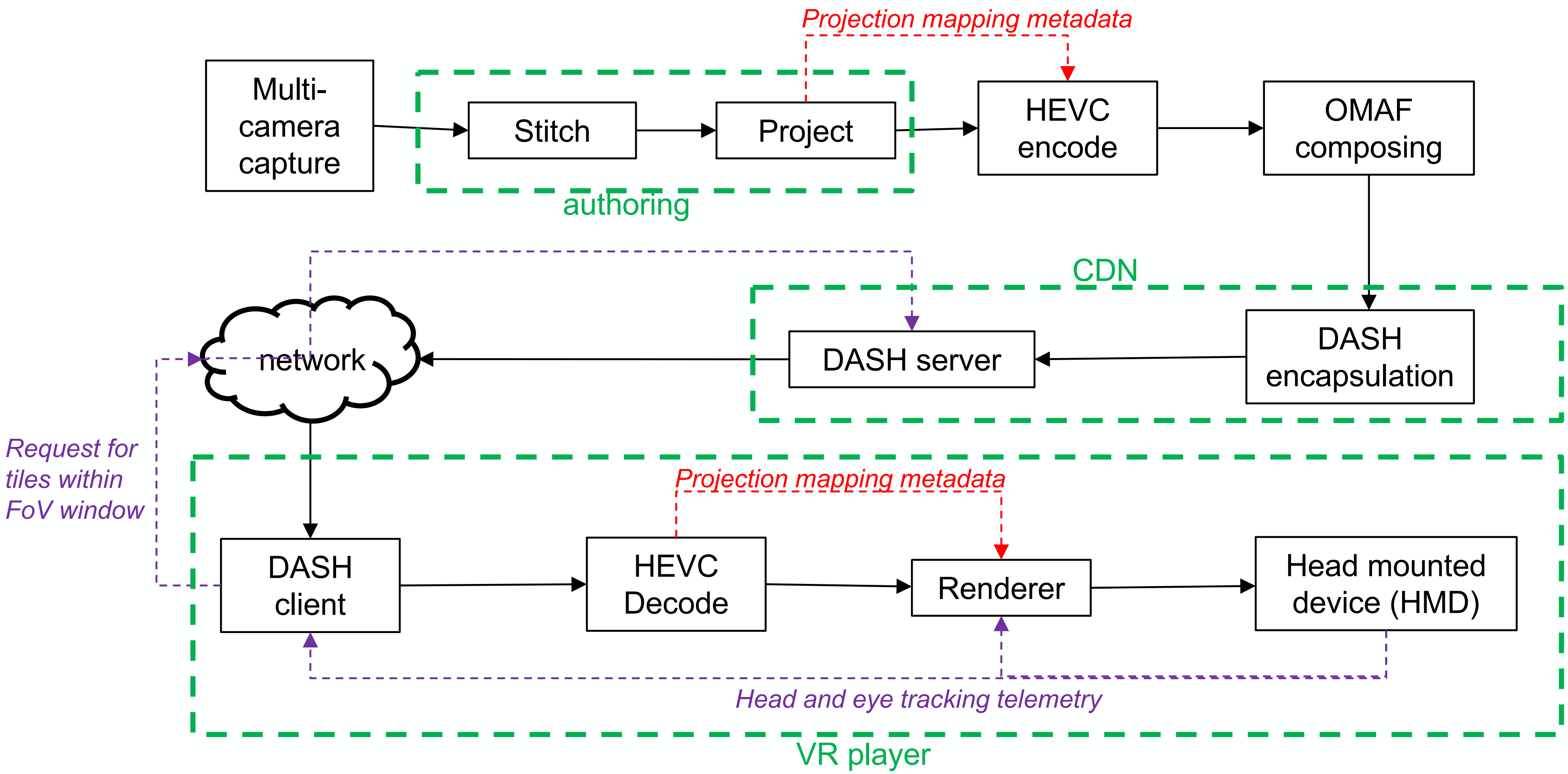
MPEG-I

- New MPEG project
  - ISO/IEC 23090: Coded Representation of Immersive Media
- Five parts are currently in progress:
  - Technical report: use cases, requirements and architectures
  - Omnidirectional media format (OMAF)
  - New and immersive video coding
  - New and immersive audio coding
  - Point cloud coding
- Two more parts are considered
  - VR metadata
  - VR metrics

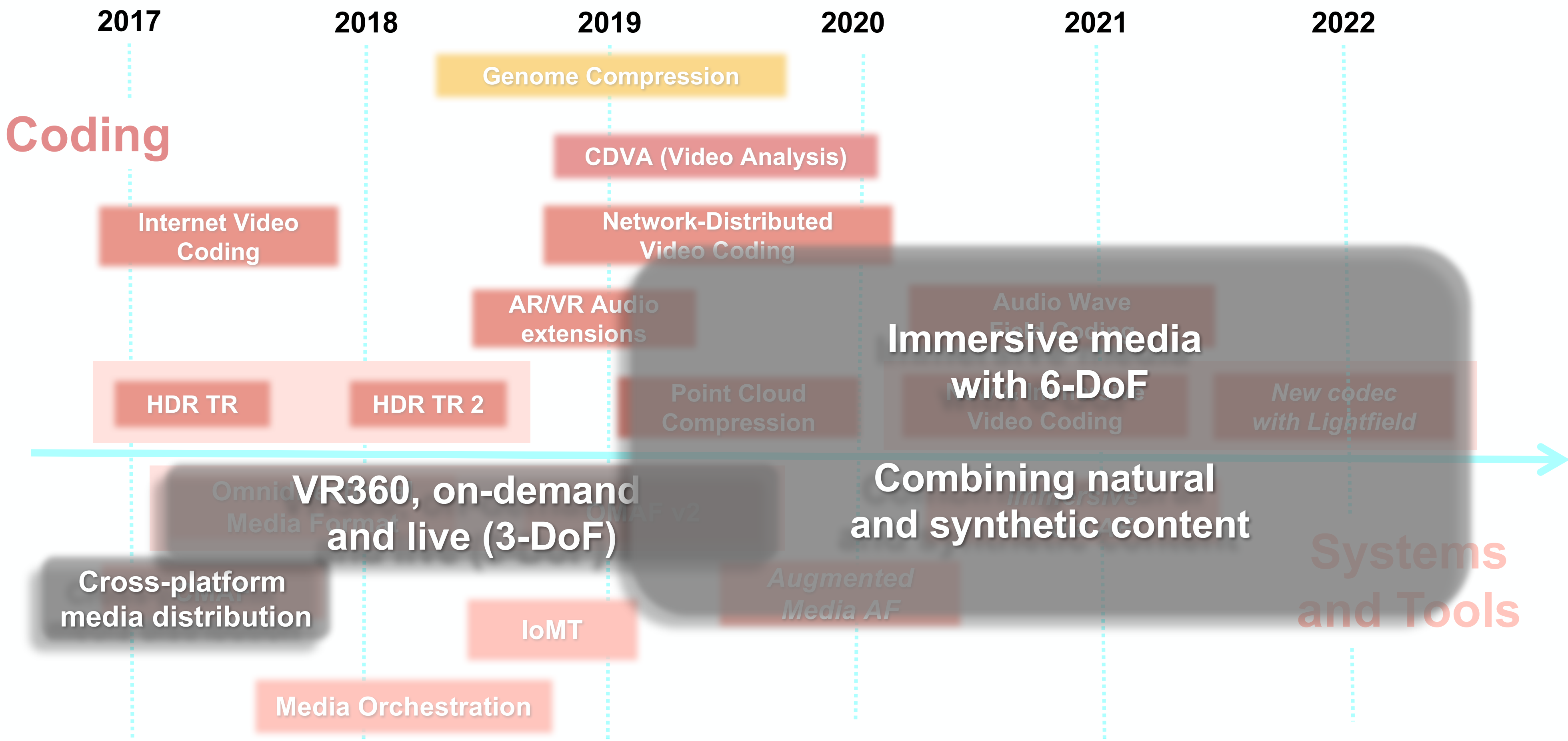
OMAF

- Scope: 360° video and associated audio, 3-DoF only
- Focus on the media format that enables VR video production, delivery and consumption
  - Projection and region-wise packing that are used to generate 2D video from the sphere signal
  - File format encapsulation and metadata signaling
  - DASH encapsulation and metadata signaling
  - Video/audio codecs and profiles/brands
  - Informative examples of viewport-dependent approaches

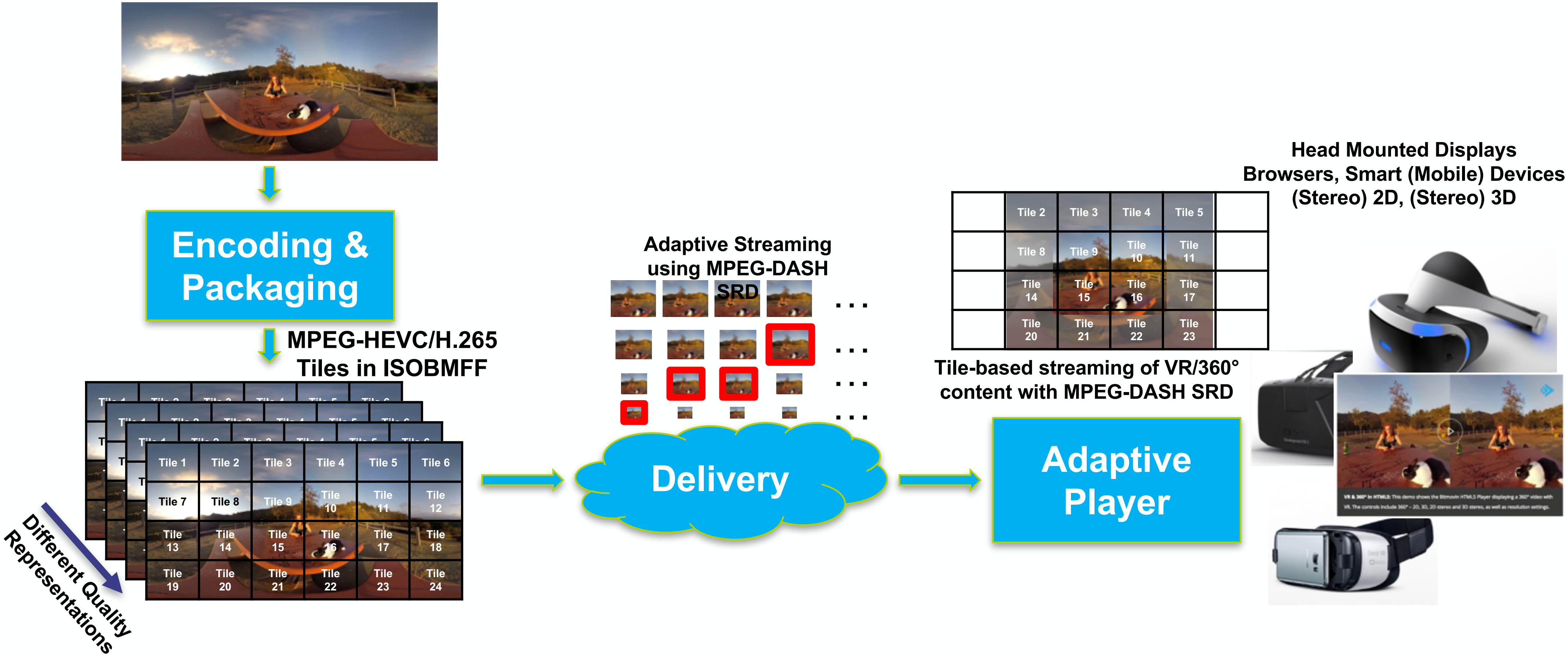
DASH/OMAF System Diagram



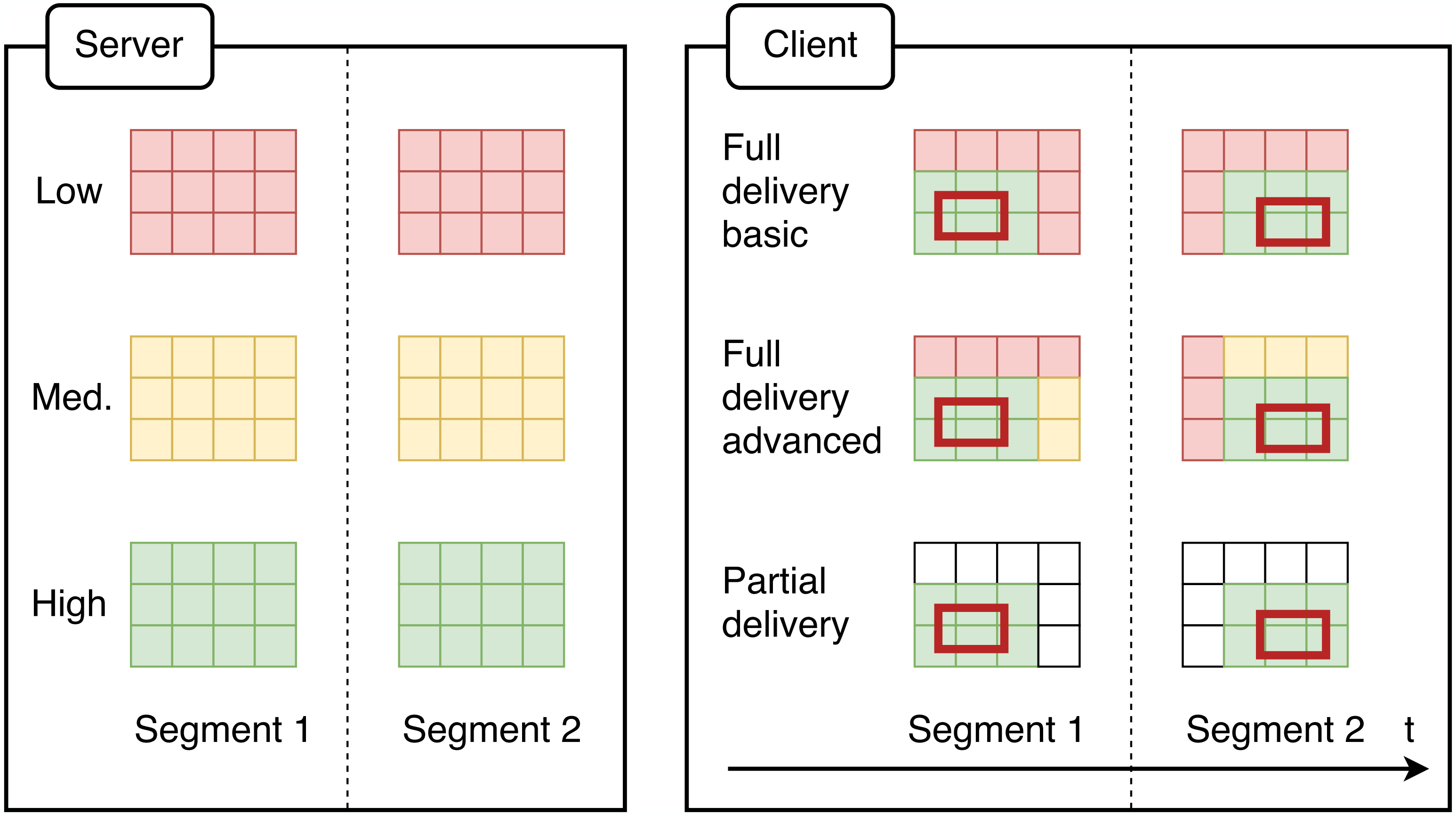
MPEG's Five-Year Roadmap



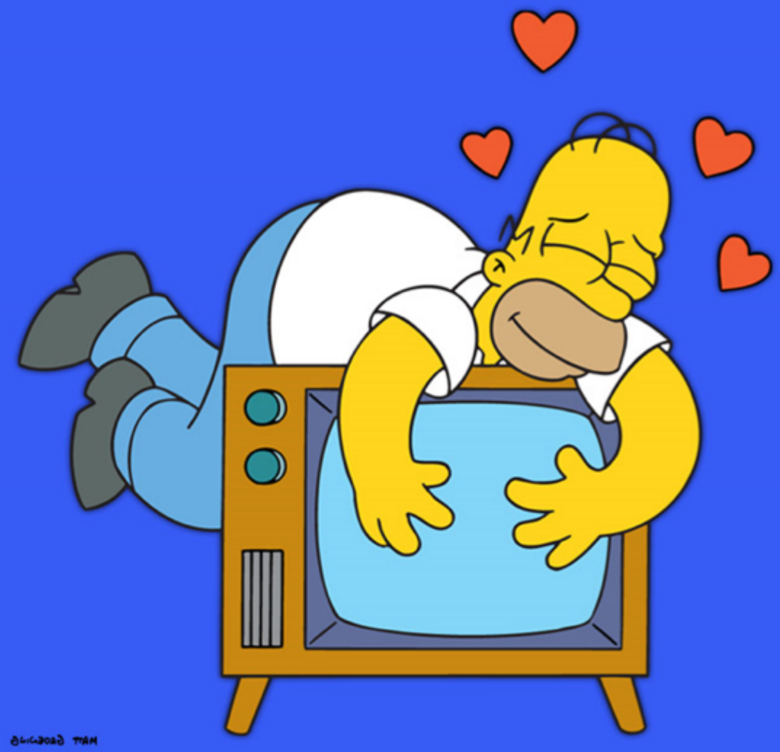
Tile-Based Streaming



Tile-Based Streaming Options



Thank You



“Television! Teacher, mother, secret lover”

– Homer Simpson