VR-IF Distribution Guidelines
CES Master Class

Ozgur Oyman, Intel
Sr. Standardization Manager
VR-IF Board Member and Distribution TF Co-Chair

Scope

- The Distribution Task Force (TF) Guidelines Charter:
  - Compression: Media codecs for VR, i.e., encoding of different production formats and related media profiles for video, audio and other media types such as text, graphics, etc.
  - Storage: Media formats for VR content (e.g., file/segment encapsulation) for different distribution means, including but not limited to storage, download, adaptive bitrate streaming and broadcasting
  - Delivery: Interfaces and protocols for Live, Linear and VOD delivery over streaming (unicast), and broadcast applications

Media Profiles

- Media profiles are aimed to provide interoperability points for:
  - media codecs and associated metadata for omnidirectional media content
  - media coding and encapsulation configurations that may be used for rendering, compression, streaming, and playback of the omnidirectional media content
- VR-IF guidelines contain 3 video/audio profiles based on MPEG's Omnidirectional MediA Format (OMAF) spec
  - (Video) HEVC-based viewport-independent OMAF video profile
  - (Video) HEVC-based viewport-dependent OMAF video profile
  - (Audio) OMAF 3D Audio Baseline Media Profile

HEVC Viewport-Independent OMAF Profile

INDUSTRY FORUM

- Relies on a video decoder based on HEVC Main 10 Profile, Main Tier, Level 5.1., e.g., 4K@60fps
- Both monoscopic and stereoscopic spherical videos up to 360°are supported.
- The whole sphere is encoded in a single Level 5.1 bitstream, thus the viewport (what user sees) is of significantly lower resolution compared to the whole sphere
- Advantages include simplicity, viewport-agnostic delivery and decoding, and minimal file format and DASH-level extensions
- Disadvantage: Lower quality at Viewport, only ~10% of Level 5.1 decoding is displayed
  - Example: with frame rate 60fps, for a display with an FOV of 90°×90, and given content coverage
    - Encode full 360 video (360x180) 8Mpix -> Viewport resolution: 1Kx1K (1 Mpix)
    - Better viewport quality can be achieved by reducing content coverage, i.e., a portion of the sphere

HEVC Viewport-Dependent OMAF Profile

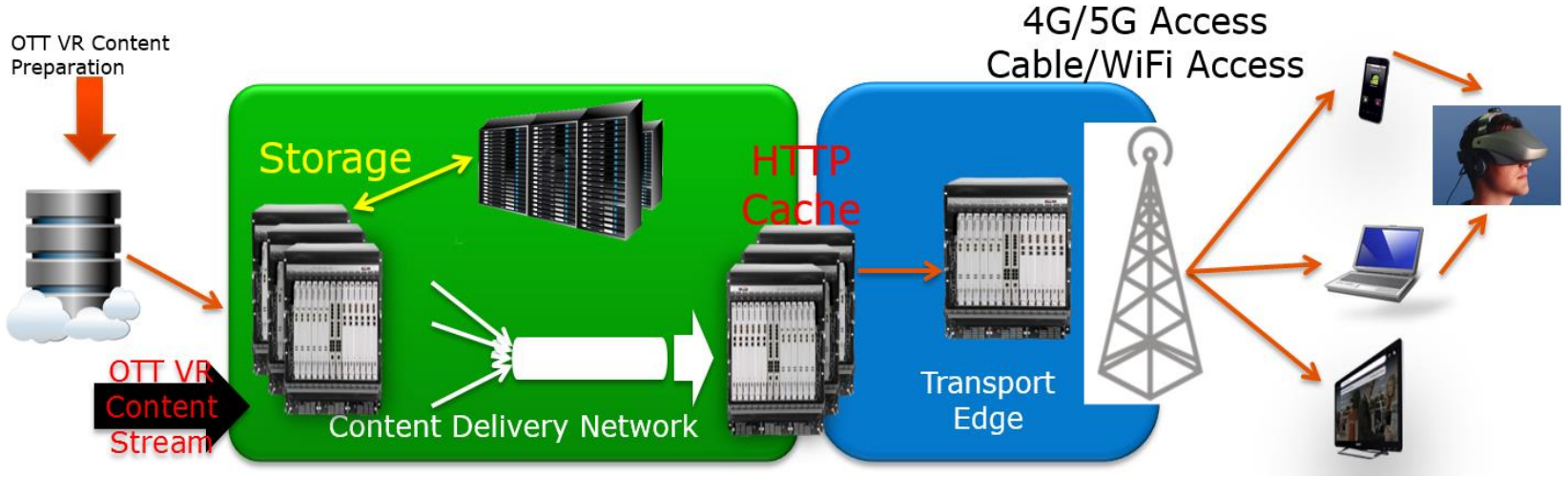
- Relies on a video decoder based on HEVC Main 10 Profile, Main Tier, Level 5.1., e.g., 4K@60fps
- Both, monoscopic and stereoscopic spherical video up to 360° is supported.
- Uses HEVC tiling concept to deliver content in tiles allowing streams to have a different quality or resolution for different areas/regions of the omnidirectional video
- Advantage: Allows for higher resolution (e.g., 6K) in the viewport and/or reduced bitrate, compared to Viewport-Independent Media profile
- Requires viewport-dependent delivery and support for ISOBMFF extractors to be able to adaptively select tiles with changing viewports
- Motion-to-high resolution latency is a key metric for the viewport-dependent profile, as this form of delivery comes with additional sources of latency such as
  - network request delay, origin-to-edge delay (in case of cache miss), transmission delay (accounting for access network delay) and delays incurred in the client device due to buffering, decoding and rendering
  - CDN’s VR content caching ability corresponding to different viewports is critical to help towards reducing download delays and also potentially improving the network bandwidth utilization efficiency.

OMAF 3D Audio Baseline Media Profile

INDUSTRY FORUM

- The profile is based on MPEG-H 3D Audio, which specifies coding of immersive audio material and the storage of the coded representation in ISOBMFF.
- Channels, objects and Higher-Order Ambisonics (HOA) are supported, as well as combinations of those.
- Content authors can synchronize audio and video portions of a media presentation, e.g. ensuring lip-synch.
- When orientation sensor inputs (i.e. pitch, yaw, roll) of an MPEG-H 3D Audio decoder change, there will be some algorithmic and implementation latency (perhaps tens of milliseconds) between user head movement and the desired sound field orientation.
- This latency will not impact audio/visual synchronization (i.e. lip synch), but only represents the lag of the rendered sound field with respect to the user head orientation.
- MPEG-H 3D Audio specifies methods for binauralizing the presentation of immersive content for playback via headphones, as is needed for 360° VR presentations.
- MPEG-H 3D Audio specifies a normative interface for the user's orientation, as Pitch, Yaw, Roll, and 3D Audio technology permits low-complexity, low-latency rendering of the audio scene to any user orientation.

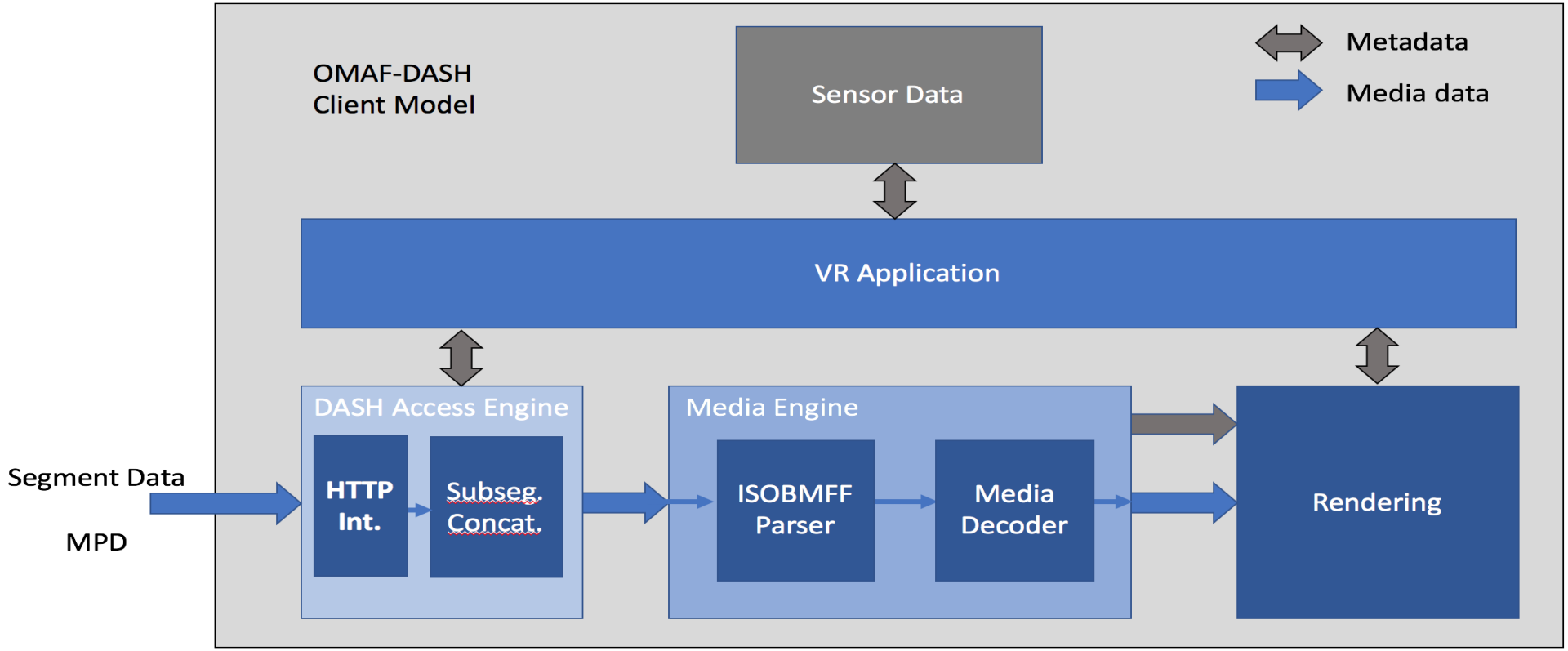
OTT VR Content Distribution over HTTP

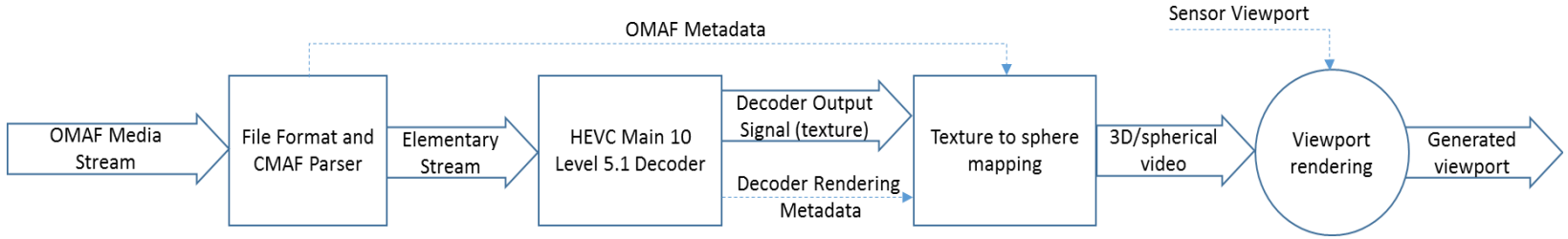


Both HTTP download and DASH-based streaming can be used to deliver OTT VR content.

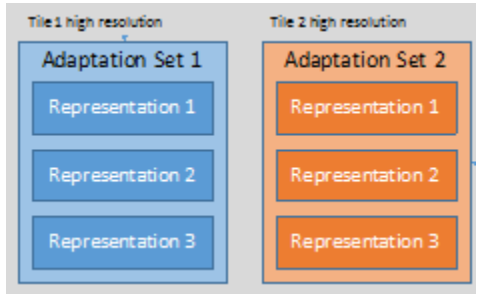
Video Audio	DASH Media Presentation Description
OMAF File Format	
HTTP	
TCP	
IP	

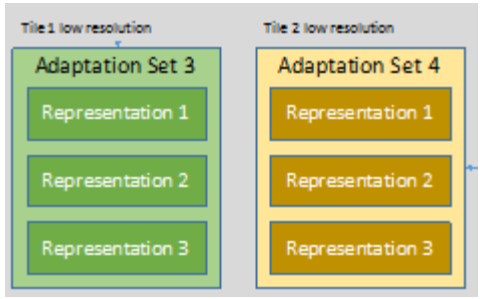
Example OMAF-DASH Client Model

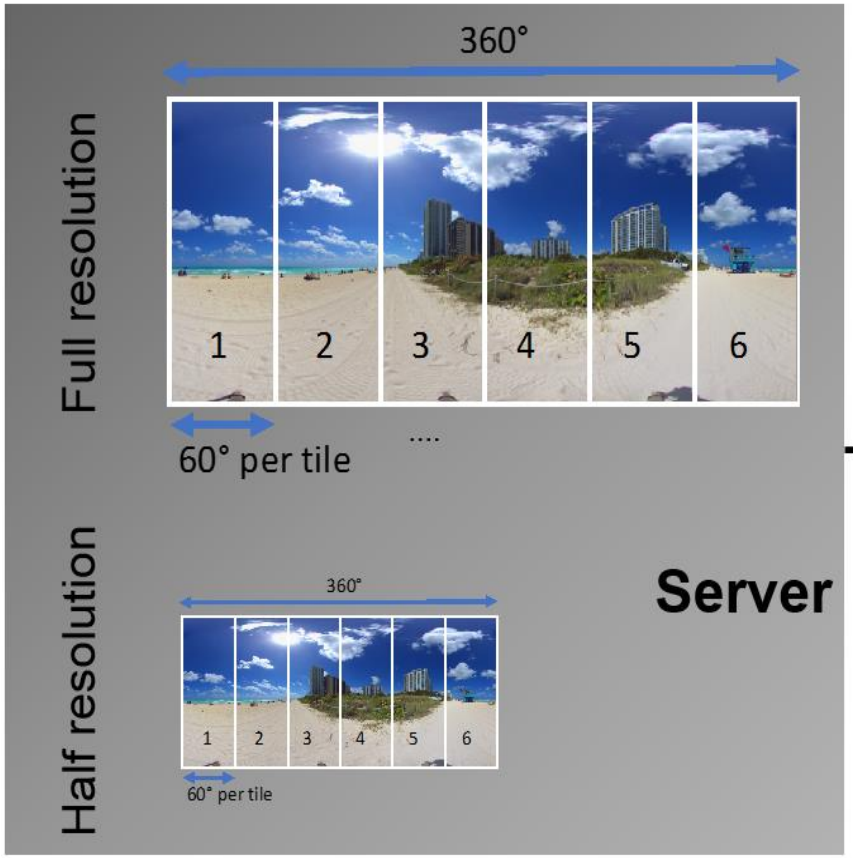


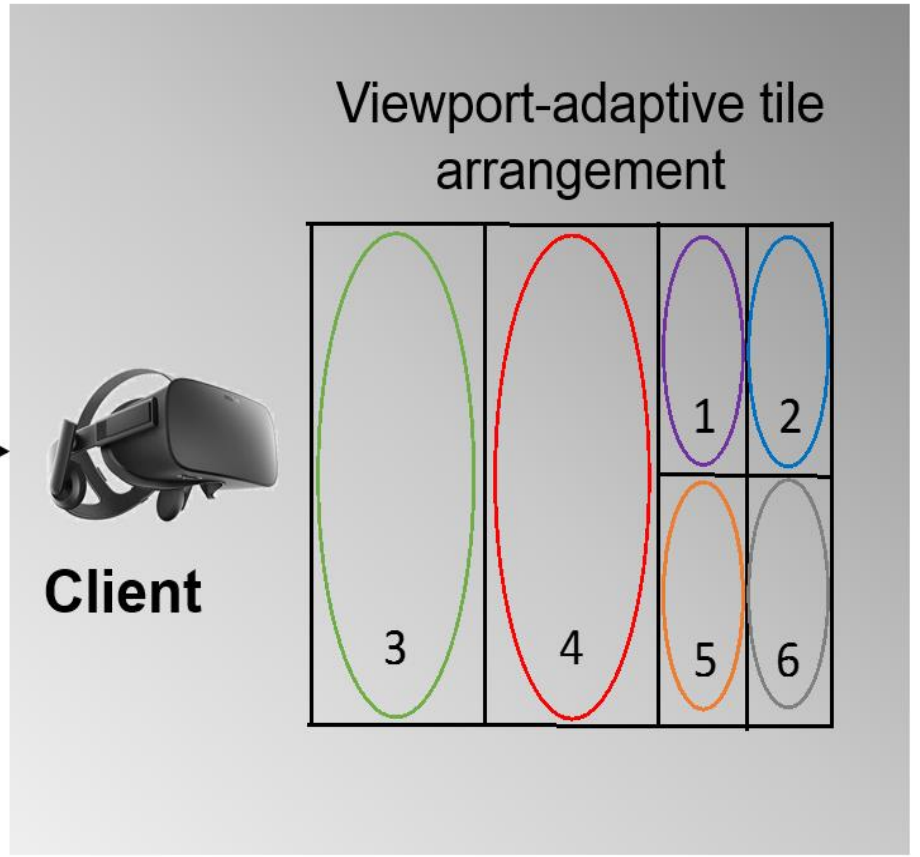


Example Viewport-Dependent DASH Streaming with Tiles









Additional Distribution Aspects in Guidelines

- Suitable production formats for various media profiles
- Sphere-to-texture mapping and SEI message generation
- Encoding and content preparation guidelines for various media profiles
- Download and streaming guidelines for the various media profiles
- Rendering guidelines using OMAF metadata or SEI messages