MPEG's Dynamic Adaptive Streaming over HTTP -
An Enabling Standard for Internet TV

Thomas Stockhammer
Qualcomm Incorporated

ABSTRACT

Internet video is experiencing a dramatic growth in both fixed and mobile environments. All available data points indicate that Internet video is likely to be the long-awaited killer application for mobile data networks. It will also create significant new business opportunities for the fixed Internet access providers. Internet video is currently dominated by a handful proprietary technologies that exploit and reuse existing Web infrastructure to deliver live and On-Demand video services. To address these market demands, MPEG in coordination with 3GPP has taken the lead to define an enabler standard for streaming video and multimedia delivery over the Internet. MPEG Dynamic Adaptive Streaming over HTTP (DASH) specifies a delivery format that enables interoperability between different servers and client from various vendors. DASH clients can select and dynamically switch Representations in order to provide a seamless user experience. The standard evolves the proprietary HTTP streaming solutions to provide a universal format for Internet video delivery by reusing, building on and interfacing with a significant amount of existing state-of-art technologies.

1. INTRODUCTION

Wide usage of Internet video is imminent - it has become the long-awaited killer application in fixed and mobile access networks. In the mobile world with the growing popularity of smart phones, smartbooks, tablets, connected netbooks and laptops the Mobile Internet use is dramatically expanding. According to just recently updated studies Cisco White Paper: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010-2015, http://bit.ly/bwGY7L, between 2010 and 2015 the mobile data traffic will grow by a factor of 26, i.e., more than doubling every year. The study also shows that the video traffic will by then account for 66% of the total amount of the mobile data, exceeding regular web traffic by a factor of 3 and VoIP by a factor of more than 40. The popularity of Internet video also leads to dramatic and astonishing numbers in the fixed Internet: According to Sandvine Incorporated (2011, June). Global Internet Phenomena Report Spring 2011, http://www.sandvine.com/news/global_broadband_trends.asp, in North America real-time entertainment traffic (excluding p2p video) today contributes more than 50% of the downstream traffic at peak periods, with notably 30% from Netflix and 11% from Youtube. Averaged over 24 hours, Netflix with 22.2% has overtaken BitTorrent as the largest component of Internet traffic on North America's fixed access networks. Beyond on-demand services, also the opportunity of distributing live events over the open Internet drives new business models. Similar trends are observed in other regions of the globe.

A couple of key aspects drive the growth of video delivery over the Internet, while requiring continuous improvements. One success factor is user experience: Video consumers demand high-quality video experience in terms of video quality, start-up time, reactivity to user interaction, trick mode support, etc. Another key aspect is the confidence and trust of content providers and end users in the Internet delivery platform. Only in the collaboration of the entire ecosystem including content providers, network operators, service providers, device manufacturers and technology providers these high demands can be met. Affordable and mature technologies are required to create a successful ecosystem. One step into this direction is a common, efficient and flexible distribution platform that scales to the rising demands and is based on standardized components. MPEG's Dynamic Adaptive Streaming over HTTP (DASH) standard is built to serve this purpose and is considered to be an enabler to catalyse Internet video services over the next years. MPEG committee, the Working Group 11 of ISO/IEC/JTC1/SC29, has previously defined several widely used international multimedia standards such as MPEG-1, MPEG-2, MPEG-4 and etc. The committee recently developed the DASH standard that defines a delivery format allowing DASH-based clients interoperating with various standard based services and different service providers. MPEG DASH enables the creation of a multi-vendor ecosystem in which each client can receive high quality content from any service provider and therefore addressing both mentioned aspects. This paper introduces some key features of this emerging standard and discusses some aspects beyond standardization.

2. DASH STANDARDIZATION

Dynamic Adaptive Streaming over HTTP (DASH) combines the strength of and eliminates the weaknesses of RTP/RTSP-based streaming and progressive download. In the context of this document we only use the term DASH to refer to technologies also known as for example HTTP Streaming, Adaptive HTTP Streaming, HTTP Adaptive Streaming or HTTP Live Streaming. Many of the introduced principles also apply to the emerging proprietary technologies such as Apple HTTP Live Streaming R. Pantos, W. May. HTTP Live Streaming, http://tools.ietf.org/id/draft-pantos-http-live-streaming-06.txt, March 2011 or Microsoft Smooth Streaming IIS Smooth Streaming Transport Protocol: http://www.iis.net/community/files/media/smoothspecs/[MS-SMTH].pdf. MPEG's DASH standard should not be considered as competitive to the proprietary solutions, but an enabler to provide a natural evolution from successful proprietary technologies to address a broader set of requirements for a larger community using a well-defined global standard.

Figure shows the typical media distribution architecture for DASH-based video delivery. The media encoding process typically generates segments that each contains different encoded versions of one or several of the media components of the media content. Each segment contains streams required for decoding and displaying a time interval of the content. The segments are then hosted on one or several media origin servers along with a manifest, known as Media Presentation Description (MPD). The media origin server may be a plain HTTP server conforming to RFC2616 IETF RFC 2616: "Hypertext Transfer Protocol – HTTP/1.1", Fielding R. et al., June 1999. as any communication with the server is HTTP-based. The MPD information provides instructions on the location of segments as well as the timing and relation of the segments, i.e. how they form a media presentation. Based on this information in MPD, a client requests the segments using HTTP GET or partial GET methods. The client fully controls the streaming session, i.e., it manages the on-time request and smooth playout of the sequence of segments, potentially adjusting bitrates or other attributes, e.g. to react to changes of the device state or the user preferences. As long as the MPD provides RESTful HTTP-URIs for the Segment locations, the HTTP-based delivery infrastructure may be kept unaware of the actual data that is delivered. This feature permits the reuse of existing CDNs for massively scalable Internet video distribution.

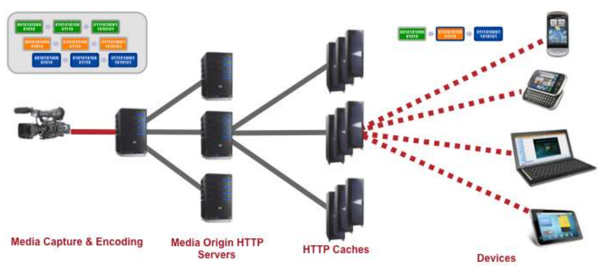

Figure DASH-based Media Distribution Architecture

Based on these considerations and a set of detailed requirements ISO/IEC JTC1/SC29/WG11 N11338, Call for Proposals on HTTP Streaming of MPEG Media, April 2010, Dresden, Germany MPEG has initiated a standardization process to provide specifications to enable scalable and flexible video distribution that addresses fixed and mobile networks. The work had been in close coordination with a parallel effort in 3GPP such that the two standards are aligned for broad industry support across different access networks. 3GPP's Release-9 specification on Adaptive HTTP Streaming (AHS) 3GPP TS 26.234: "Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs"., completed in 2010 served as a baseline for MPEG's DASH as well as for 3GPP's Release 10's DASH 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)". specification. Both specifications are technically completed and are expected to be released in the within the next few weeks.

The emphasis in this document is on MPEG's DASH specification as it serves as an generic format specification providing. The specification is not attempted to specify a full end-to-end service, but to provide a base building block to enable efficient and high-quality streaming services over the Internet. MPEG DASH defines the delivery format framework - complete system definitions that include for example definitions of detailed delivery protocols, codecs, content protection systems, etc. are left to other organizations that are interested in creating and end-to-end service in their respective eco-system. MPEG DASH specifically provides an excellent set of tools by what is essential for a future-proof online streaming technology, namely it enables

- the reuse of and combination with existing technologies such as container formats, codecs, content protection systems and therefore existing content;
- the deployment on top of HTTP-CDNs to exploit existing web infrastructures and support efficient caching at different locations in the networks;
- achieving higher user-experience by enabling seamless switching of different bitrate/quality versions, low start-up, no rebuffering, advanced trick modes support;
- selecting multimedia versions of the content based on network and device capability as well as user preferences including aspects such as accessibility, rating, multiview capabilities, support of audio channels and such;
- entirely move of the decision intelligence in terms of initial selection and dynamic adaptation to the client and enabling differentiation of client implementations to provide maximum user experience;

- deployment flexibility by supporting live, on-demand, and time-shift viewing services with the same infrastructure and formats, providing live experience with low start-up and end-to-end delay and providing DVD-like experience by flexible combination of media tracks (different languages, view-points, sub-titles, etc.);
- and finally, providing different interoperability points for different profiles to serve different environments and application domains.

To address the above, MPEG DASH primarily defines two formats as shown in Figure :

- The Media Presentation Description (MPD) describes a Media Presentation, i.e. a bounded or unbounded presentation of media content. In particular, it defines formats to announce resource identifiers for Segments as HTTP-URLs and to provide the context for these identified resources within a Media Presentation.
- The Segment format that specifies the format of the entity body of the request response when issuing a HTTP GET request or a partial HTTP GET with the indicated byte range through HTTP/1.1 as defined in RFC 2616 to a resource identified in the MPD. Segments typically contain efficiently coded media data and metadata according to or aligned with common media formats.

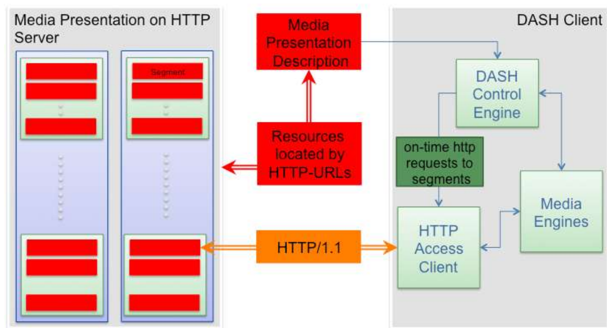

Figure Standardized aspects in DASH (marked in red)

Other aspects, such as client implementations of control and media engines are not part of the normative specification.

3. DASH - TECHNICAL HIGHLIGHTS

The design of the formats defined in DASH is based on the client model as shown in Figure . The figure illustrates the logical components of a conceptual DASH client model: The DASH access engine receives the MPD and possibly a Segment Index (information that describes time and byte maps of a media segment), constructs and issues requests and receives segments or parts of segments. The segment formats are designed in such a way that the output of the DASH access engine is aligned to MPEG container formats (ISO/IEC 14496-12 ISO Base Media File Format ISO/IEC 14496-12, Information technology – Coding of audio-visual objects – Part 12: ISO base media file format (technically identical to ISO/IEC 15444-12) or ISO/IEC 13818-1 MPEG-2 Transport Stream ITU-T Rec. H.222.0 | ISO/IEC 13818-1, Information technology – Generic coding of moving pictures and associated audio information: System), together with timing information that maps the internal timing of the media to the timeline of the Media Presentation. This design permits significant reuse of existing media engines that are already deployed for example for digital TV and mobile video services.

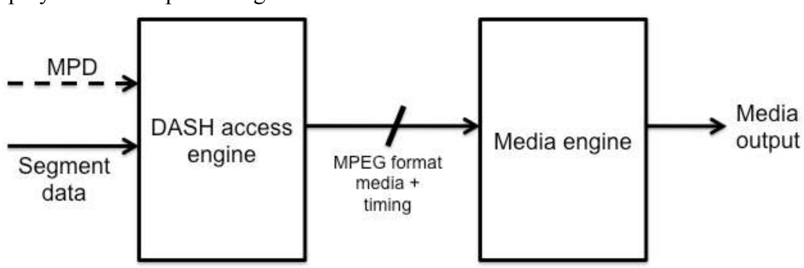

Figure Design principle of specification

In the context of DASH, encoded versions of media content are offered to DASH clients that are accessible through HTTP. DASH clients use the information in the MPD to provide a streaming service by selecting and downloading the appropriate encoded versions and make those available to the media decoding engines. To enable a streaming service, the media content and the components that compose the media content need to be prepared for delivery and sufficiently described for selection, switching and for other streaming related actions such as seeking, fast-forward, trick modes, etc.. Media content is composed of one or multiple media content components (e.g. a main video, a native audio in English, a subtitle text in French and a dubbed audio in German).

Media content component are encoded into media streams. A media stream inherits the properties of media content, the media content component from which it was encoded and in addition it gets assigned the properties of the encoding process. All these properties may be provided with each media stream in form of metadata for selection and switching. A media component may be encoded into multiple media streams, each of which represents the media component at different bitrate or resolution or sample rate, etc..

All this information is mapped into a hierarchical data model that is described by the MPD. An example is provided in Figure . One or more media streams are prepared in Segments. These Segments together with describing metadata form a Representation. A Representation may include one or more media streams. An Adaptation Set collects Representations that include media streams with containing the same media content components. Therefore, each Adaptation Sets contains dynamically switchable Representations. The collection of media content components provided in all Adaptation Sets within one Period comprises a media content. Periods may be spliced to compose a full Media Presentation.

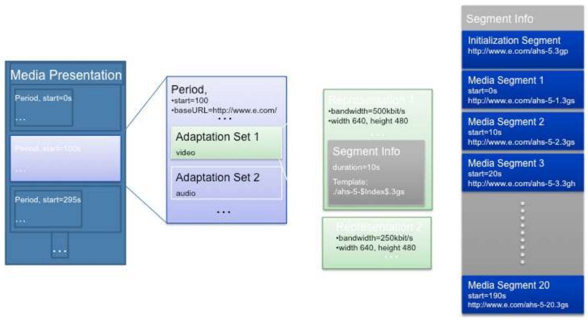

Figure Hierarchical data model

In summary, the MPD provides information of Media Streams for the purpose to initially select or reject Representations based on information such codecs, content protection, language, resolution, bandwidth, etc.. In addition, MPD provides access and timing information for Segments. For live and on demand services, Segments get assigned and availability start and end time in wall-clock time. Finally, dynamic switching and splicing relationships across Representations are provided in the MPD.

In addition to the MPD information, the DASH access client also has access to the Segment Index providing binary information in ISO box structure on each media segment. The Segment Index includes information on accessible units of data in a media segment, referred to as subsegments, in terms of byte ranges, accurate start times, accurate duration and random access information. This information may for example be used for seamless switching as well as for intelligent request scheduling at the subsegment level and is therefore independent of the segment size.

In comparison to HLS and IIS Smooth Streaming (SS) published features, DASH supports the features supported by those two technologies and also provides additional aspects such as extensibility to other formats, scalable coding (SVC) and multiview coding (MVC) support, support for efficient trick modes by using sub-segment indexing and temporal subsequences, simple splicing of advertisement using the Period concept, a set of well-defined metrics and a set of well-defined content descriptors as well as flexible content protection and a set of profiles.

4. DASH Profiles

Profiles of DASH are defined to enable interoperability for different deployment scenarios by limiting the number of features for each profile. Profiles in DASH are aligned to the profile concept in the ISO base media file format [10]: A profile has an identifier and refers to a set of specific restrictions. Those restrictions might be on features of the Media Presentation description (MPD) document, usage of the network, media format(s), codec(s) used, protection formats, or on quantitative measures such as bit-rates, segment sizes, and so on. Profiles defined in MPEG DASH define restrictions on features of MPEG's DASH, but not on other aspects. Externally defined profiles may additionally impose restrictions on other aspects of media delivery. A profile is a claim and a permission; it claims that the Media Presentation (MPD and Segment formats) conforms to the profile, and gives permission to a client that implements that profile to read the Media Presentation, interpret what it recognizes, and ignore the material it does not understand. A Media Presentation may conform to more than one profile.

The profiles in DASH are considered to address different deployment scenarios. Beyond the full profile, five additional profiles are currently considered, differentiated in the use of different encapsulation formats (MPEG-2 TS vs. ISO base media file format), in constraining the encoding to support legacy devices, in constraining the client devices in order to support the distribution of legacy content as well as to support delivery of live services and large libraries of On-Demand content.

5. Beyond DASH Standardization

Beyond the initial MPEG DASH specification, MPEG is also addressing the issue of conformance. The conformance specification will define the conformance requirements for each DASH profile. A detailed conformance specification is being developed with a delay of 6 months to the main specification and will be supported by appropriate test software and test vectors. Other organizations that build on DASH are also establishing conformance procedures in order to provide system tests.

Also outside the scope of the specification are aspects dealing with content provisioning guidelines, detailed client implementation and the definition of detailed end-to-end services integrated into delivery systems. While it is expected that MPEG will provide a certain set of general implementation guidelines, most of these efforts will be part of further standardization by other organizations. Relevant aspects in content provisioning are for example providing different sets of bitrates and video resolutions, providing appropriate segment sizes and durations, applying a suitable frequency of switch points, etc. System integration will mostly deal with the use of DASH formats with codecs, media component formats, protocols and content protection schemes. Other aspects of immediate interest are the combination of DASH with presentation formats such as HTML5 and the integration into web browsers.

Quite relevant, but also not in scope of the MPEG DASH specification, are DASH client implementations. The client is offered the Media Presentation with a full set of alternative and complementary representations of the media content components. Figure shows an example of the task of a client, which needs to download the segments for the adequate Representations to make them available for real-time consumption. Clients task is appropriate selection and dynamic switching, also reacting to user interactions. The detailed usage of the access networks and the provided HTTP and TCP/IP protocol options are part of the optimized client implementations.

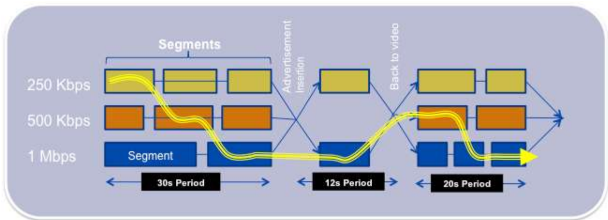

Figure DASH client process

Finally, commercial issues are discussed outside the MPEG standardization process. Individual companies have started to provide information on favourable licensing terms for the MPEG DASH standard with the ambition to spur universal adoption of MPEG DASH. More information is expected to be available at the work shop.

6. REFERENCES

Cisco White Paper: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010-2015, http://bit.ly/bwGY7L

Sandvine Incorporated (2011, June). Global Internet Phenomena Report Spring 2011, http://www.sandvine.com/news/global_broadband_trends.asp

R. Pantos, W. May. HTTP Live Streaming, http://tools.ietf.org/id/draft-pantos-http-live-streaming-06.txt, March 2011

IIS Smooth Streaming Transport Protocol: http://www.iis.net/community/files/media/smoothspecs/[MS-SMTH].pdf

IETF RFC 2616: "Hypertext Transfer Protocol – HTTP/1.1", Fielding R. et al., June 1999.

ISO/IEC JTC1/SC29/WG11 N11338, Call for Proposals on HTTP Streaming of MPEG Media, April 2010, Dresden, Germany

3GPP TS 26.234: "Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs".

3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

ISO/IEC 14496-12, Information technology – Coding of audio-visual objects – Part 12: ISO base media file format (technically identical to ISO/IEC 15444-12)

ITU-T Rec. H.222.0 | ISO/IEC 13818-1, Information technology – Generic coding of moving pictures and associated audio information: System