MPEG-DASH: The Standard for Multimedia Streaming Over Internet[1]

Iraj Sodagar
Microsoft Corporation, 1 Microsoft Way, Redmond, WA, USA 98052

Watching the Olympics live over the Internet? Streaming last week's episode of your favorite TV show to your game console? Watching a 24-hours news TV channel on your mobile phone? These use cases might already seem possible as parts of our daily lives. In fact, during the 2008 Olympics, NBC reported delivering of 3.4 petabytes of video content over the Internet [1]. But multimedia streaming over the Internet is still its infancy compared to its potential market. One reason is that today every commercial platform is a closed system: it has its own manifest and content formats and streaming protocols. In other words, there is no interoperability between devices and servers of various vendors. A recent study indicated that in a few years video content could make up the vast majority of Internet traffic [2]. One of the main enablers would be an adopted standard that provides interoperability between various servers and devices. Achieving such interoperability will be instrumental for the growth of the market, since a common ecosystem of content and services will be able to provision a broad range of devices such as PCs, TVs, laptops, set-top boxes, game consoles, tablets and mobiles phones. MPEG-DASH was developed to do just that.

Why HTTP Streaming?

Delivery of video content over the Internet had started in the 1990s with timely delivery and consumption of large amounts of data being the main challenge. The IETF's Real-time Transport Protocol (RTP) was designed to define packet formats for audio and video content along with stream session management which allowed delivery of those packets with very low overhead. RTP works well in managed IP networks. However, in today's Internet, managed networks have been replaced by Content Delivery Networks (CDN), many of which do not support RTP streaming. In addition, RTP packets are often not allowed through firewalls. Finally, RTP streaming requires the server to manage a separate streaming session for each client, making large scale deployments resource intensive.

With the increase of Internet bandwidth and the tremendous growth of the World Wide Web, the value of delivering audio or video data in very small packets has diminished. Multimedia content can now be delivered very efficiently in larger segments (in contrast to small packets) using HTTP. HTTP streaming has several benefits. First, Internet infrastructure has evolved to efficiently support HTTP. For instance, CDNs provide localized "edge" caches which reduce long-haul traffic. Also, HTTP is "firewall friendly" since almost all firewalls are configured to support its outgoing connections. HTTP server technology is a commodity and therefore supporting HTTP streaming for millions of users is very cost effective. Second, with HTTP streaming the client manages the streaming without the need of maintaining a session state on the server. Therefore, provisioning a large number of streaming clients does not impose any additional cost on server resources beyond standard web usage of HTTP and can be managed by a CDN using standard HTTP optimization techniques.

Because of the above reasons, HTTP streaming has become a popular approach in commercial deployments. For instance, streaming platforms such as Apple's HTTP Live Streaming (HLS) [3], Microsoft's Smooth Streaming [4], and Adobe's HTTP Dynamic Streaming [5] all use HTTP streaming as their underlying method of delivery. However, each implementation uses different manifest and segment formats and therefore, to receive the content from each server, a device must support its corresponding proprietary client protocol. A standard for HTTP streaming of multimedia content would allow a standard-based client to stream content from any standard-based server, thereby enabling interoperability between servers and clients of different vendors.

Observing the market prospects and requests from the industry, MPEG issued a Call for Proposal for an HTTP streaming standard in April 2009. Fifteen full proposals were received by July 2009, when MPEG started the evaluation of the submitted technologies. In the two years that followed, MPEG developed the specification with participation of many

[1] Preprint of the article published in IEEE Multimedia, Oct-Nov 2011.

experts and collaboration with other standard groups such as 3GPP[6]. The resulting standard, known as MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH), is currently at the Draft International Standard stage [7][2].

A Simple Case of Adaptive Streaming

Figure 1 illustrates a simple example of on-demand dynamic adaptive streaming. In this figure, the multimedia content consists of video and audio components. The video source is encoded at three different alternative bitrates: 5M, 2M and 500K bits/sec. Additionally, an I-frame only bitstream with low frame-rate is provided for streaming during the trick mode play. The accompanying audio content is available in two languages: Audio 1 is a dubbed English version of the audio track and is encoded in surround, AAC 128K and AAC 48K bits/sec alternatives, while Audio 2 is the original French version, encoded at AAC 128K and AAC 48K bits/sec alternatives only.

Assume that a device starts streaming the content by requesting segments of the video bitstream at the highest available quality (5M) and the English audio at 128K AAC since for instance the device does not support surround Audio (1). After streaming the 1st segments of video and audio, and monitoring the effective network bandwidth, the device realizes that the actual available bandwidth is lower than 5M bits/sec. So, at the next available switching point, it switches the video down to 2M bits/sec by streaming the next segments from the mid-quality track while it continues streaming of the 128K AAC English audio (2). The device continues to monitor the actual network bandwidth and realizes that the network bandwidth has further decreased to a value lower than 2M bits/sec. Therefore, in order to maintain continuous playback, the device further switches the streams down to 500K bits/sec video and 48K bits/sec audio (3). It continues playing the content at these rates until the network bandwidth increases and then it switches the video up to 2M (4). After a while, the user decides to pause and rewind. At this point, the device starts streaming the video from the trick mode track to play the video in the reverse order, while audio is muted (5). At the desired point, the user clicks to play the content with the original French audio. At this point, the device resumes streaming the video from the highest quality (5M) and audio from 128K French audio (6).

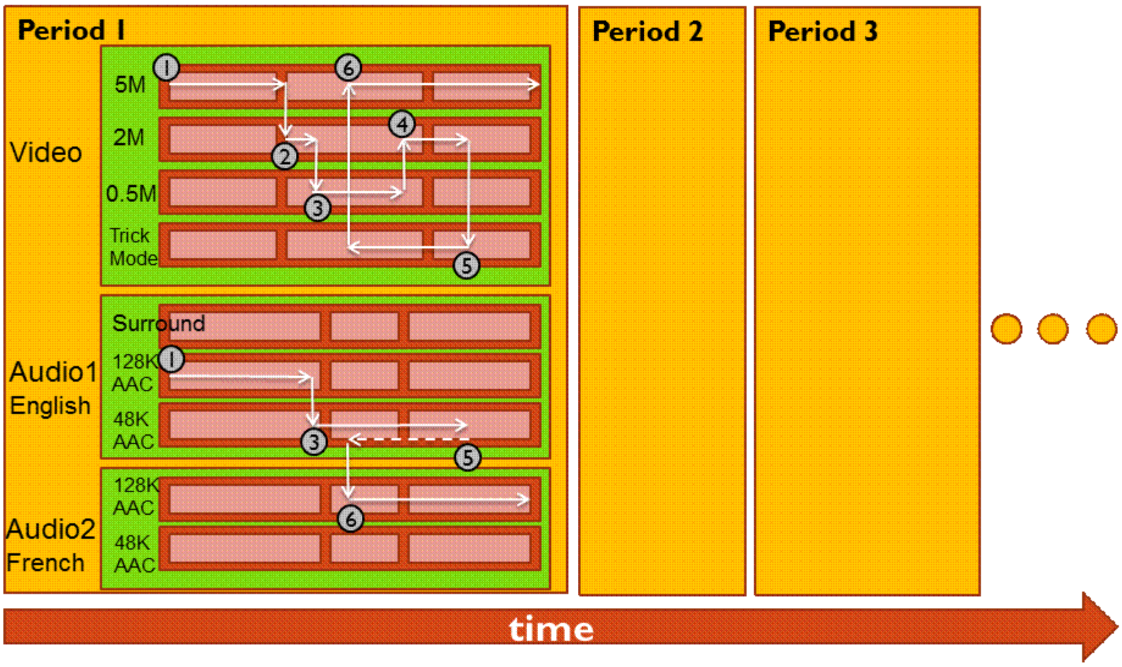

Figure 1. A simple example of dynamic adaptive streaming. Numbered circles demonstrate the action points taken by the device.

[2] At the time of publishing this article, only the referenced draft is publically available. Note that the specification was further revised in August 2011 and is expected to be published as ISO/IEC 23009-1.

The above example perhaps is one of the most simple use cases of dynamic streaming of multimedia content. More advance uses cases may include switching between multiple camera views, 3-D multimedia content streaming, video streams with subtitles and captions, dynamic ad-insertion, low latency live streaming, playback of mixed streaming and pre-stored content, and others.

Scope of MPEG-DASH

Figure 2 illustrates a simple streaming scenario between an HTTP server and a DASH client. In this figure, the multimedia content is captured and stored on an HTTP server and is delivered using HTTP. The content exists on the server in two parts: 1) Media Presentation Description (MPD) which describes a manifest of the available content, its various alternatives, their URL addresses and other characteristics, and 2) Segments which contain the actual multimedia bitstreams in form of chunks, in single or multiple files.

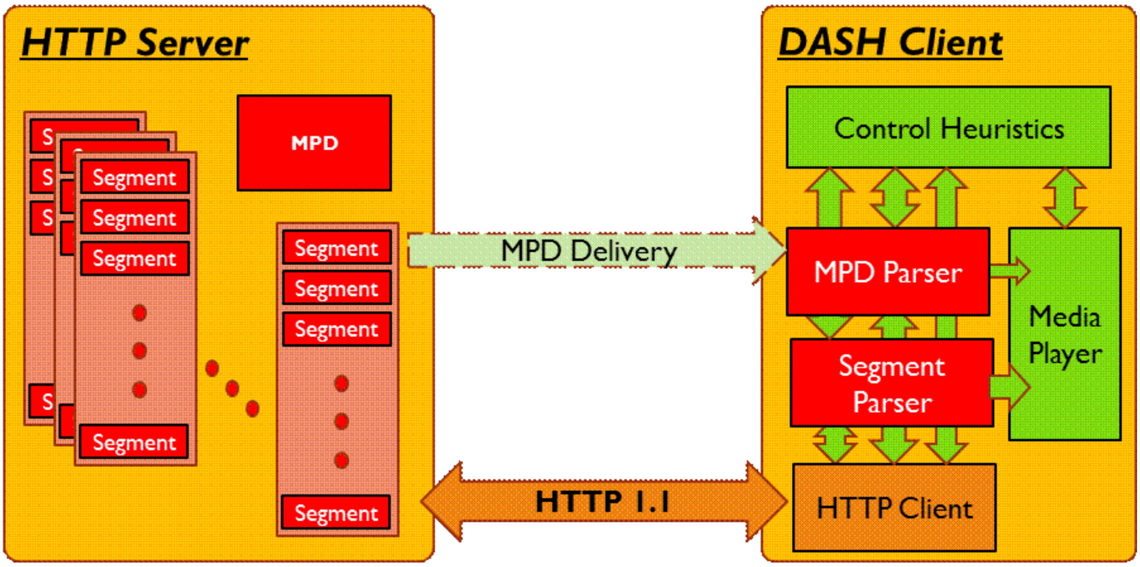

Figure 2. Scope of the MPEG-DASH Standard. The formats and the functionalities of the red blocks are defined by the specification. The clients control heuristics and media players are not within the scope of the standard.

In order to play the content, the DASH client first obtains the MPD. The MPD can be delivered using HTTP, email, thumb drive, broadcast or other transports. By parsing the MPD, the DASH client learns about the timing of the program, the availability of media content, the media types, resolutions, minimum and maximum bandwidths and the existence of various encoded alternatives of multimedia components, the accessibility features and the required digital right management (DRM), the location of each media component on the network and other characteristic of the content. Using this information, the DASH client selects the appropriate encoded alternative and starts streaming of the content by fetching the segments using HTTP GET requests.

After appropriate buffering to allow for network throughput variations, the client continues fetching the subsequent segments and also monitors the bandwidth fluctuations of the network. Depending on its measurements, the client decides how to adapt to the available bandwidth by fetching segments of different alternatives (with lower or higher bitrate) to maintain an adequate buffer.

The MPEG-DASH specification only defines the MPD and the segment formats. The delivery of the MPD and the media encoding formats containing the segments as well as the client behavior for fetching, adaptation heuristics and playing the content are outside of MPEG-DASH's scope.

Multimedia Presentation Description

Dynamic HTTP streaming requires various bitrate alternatives of the multimedia content to be available at the server. In addition, the multimedia content may consist of several media components (e.g. audio, video, text), each of which may have different characteristics. In MPEG-DASH, these characteristics are described by MPD which is an XML document.

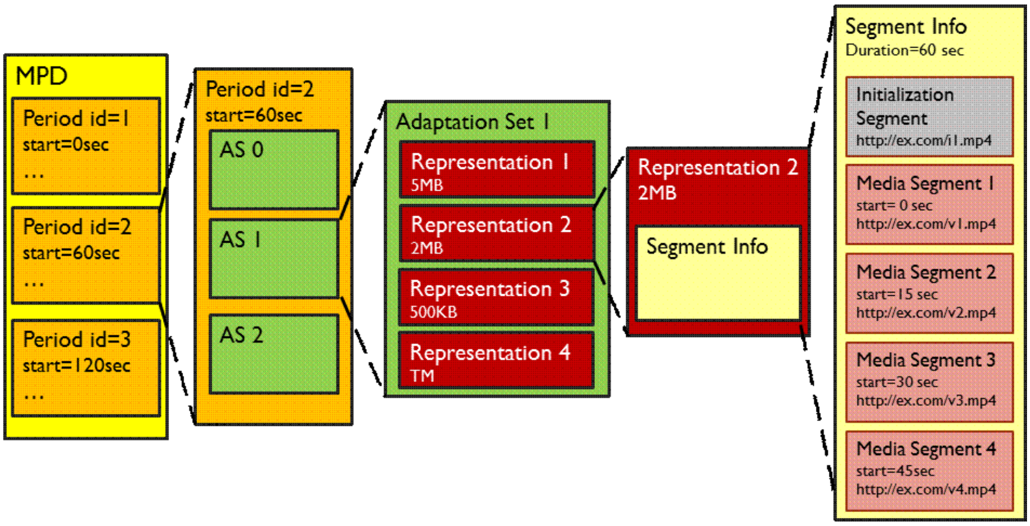

Figure 3. The MPD hierarchical data model.

Figure 3 demonstrates the MPD hierarchical data model. The MPD consists of one or multiple periods, where a period is an interval of the program along the temporal axis. Each period has a starting time and duration and consists of one or multiple adaptation sets. An adaptation set provides the information about one or multiple media components and its various encoded alternatives. For instance, an adaptation set may contain the different bitrates of the video component of the same multimedia content. Another adaptation set may contain the different bitrates of the audio component (e.g. lower quality stereo and higher quality surround sound) of the same multimedia content. Each adaptation set usually includes multiple representations.

A representation is an encoded alternative of the same media component, varying from other representations by bitrate, resolution, number of channels or other characteristics. Each representation consists of one or multiple segments. Segments are the media stream chunks in temporal sequence. Each segment has a URI, i.e. an addressable location on a server which can be downloaded using HTTP GET or HTTP GET with byte ranges.

To use this data model, the DASH client first parses the MPD XML document. The client then selects the set of representations it will use based on descriptive elements in the MPD, the capabilities of the client and choices of its user. The client then builds a time-line and starts playing the multimedia content by requesting appropriate media segments. Each representation's description includes information about its segments, which enables requests for each segment to be formulated in terms of HTTP URL and byte range. For live presentations, the MPD also provides segment availability start time and availability end time, approximate media start time, and fixed or variable duration of segments.

Segment Format

The multimedia content can be accessed as a collection of segments. A segment is defined as the entity body of the response to the DASH client's HTTP GET or a partial HTTP GET. A media component is encoded and divided in multiple segments. The first segment might be an initialization segment which contains the required information for initialization of the DASH client's media decoder. It does not include any actual media data.

The media stream then is divided to one or multiple consecutive media segments. Each media segment is assigned a unique URL (possibly with byte range), an index, and explicit or implicit start time and duration. Each media segment contains at least one Stream Access Point (SAP), which is a random access or "switch-to" point in the media stream where decoding can start using only data from that point forward.

To enable downloading segments in multiple parts, the specification defines a method of signaling sub-segments using a segment index box [8]. This box describes sub-segments and SAPs in the segment by signaling their durations and byte offsets. The DASH client can use the indexing information to request sub-segments using partial HTTP GETS. The indexing information of a segment can be put in the single box at the beginning of that segment, or spread among many indexing boxes in the segment. Different methods of spreading are possible, such as hierarchical, daisy-chain and hybrid.

This technique avoids adding a large box at the beginning of the segment and therefore prevents a possible initial download delay.

MPEG-DASH defines segment container formats for both ISO Base Media File Format [9] and MPEG-2 Transport Streams [10]. MPEG-DASH is media codec agnostic and supports both multiplexed and unmultiplexed encoded content.

Multiple DRM and Common Encryption

In MPEG-DASH, each adaptive set can use one content protection descriptor to describe the supported DRM scheme. An adaptive set can also use multiple content protection schemes and as long as the client recognizes at least one, it can stream and decode the content.

In conjunction with the MPEG-DASH standardization, MPEG is also developing a common encryption standard, ISO/IEC 23001-7, which defines signaling of a common encryption scheme of media content. Using this standard, the content can be encrypted once and streamed to clients, which support different DRM license management systems. Each client gets the decryption keys and other required information using its particular supported DRM system, which is signaled in the MPD, and then streams the commonly encrypted content from the same server.

Additional Features

The MPEG-DASH specification is a feature-rich standard. Some of the additional features are:

- Switching and selectable streams: the MPD provides adequate information to the client for selecting and switching between streams, e.g. selecting one audio stream from different languages, selecting video between different camera angles, selecting the subtitles from provided languages, and dynamically switching between different bitrates of the same video camera.
- Ad-insertion: advertisements can be inserted as a period between periods or segment between segments in both on-demand and live cases.
- Compact manifest: the segments' address URLs can be signaled using a template scheme resulting in a compact MPD.
- Fragmented manifest: the MPD can be divided into multiple parts or some of its elements can be externally referenced, enabling downloading MPD in multiple steps.
- Segments with variable durations: the duration of segments can be varied. With live streaming, the duration of the next segment can also be signaled with the delivery of current segment.
- Multiple base URLs: the same content can be available at multiple URLs, i.e. at different servers or CDNs, and the client can stream from any of them to maximize the available network bandwidth.
- Clock drift control for live sessions: the UTC time can be included with each segment to enable the client to control its clock drift.
- Scalable Video Coding (SVC) and Multiview Video Coding (MVC) support: the MPD provides adequate information regarding the decoding dependencies between representations which can be used for streaming any multi-layer coded streams such as SVC and MVC.
- A flexible set of descriptors: for describing content rating, components' roles, accessibility features, camera views, frame-packing and audio channels configuration.
- Subseting of adaptation sets into groups according to the content author's guidance.
- Quality metrics for reporting the session experience: a set of well-defined quality metrics for the client to measure and report back to a reporting server.

What's Next?

The specification defines five specific profiles, each addressing a different class of applications. Each profile defines a set of constraints, limiting the MPD and segment formats to a subset of the entire specification. Therefore, a DASH client conforming to a specific profile is only required to support those required features and not the entire specification.

Some profiles are specifically designed to use the legacy content and therefore provide a migration path for the existing non-standard solutions to a standard one.

Several other standard organizations and consortia are collaborating with MPEG to reference MPEG-DASH in their own specifications. At the same time, it seems that industry is moving quickly to provide solutions based on MPEG-DASH. Some open source implementations are also on the way. It is believed that the next two years will be a crucial time for the industry, i.e. content and service providers, platform providers, software vendors, CDN providers and device manufactures, to adopt this standard and build an interoperable ecosystem for multimedia streaming over Internet.

REFERENCES

[1] Miller, P., SVP/GM Digital Media at NBC Sports, keynote, http://blogs.msdn.com/b/tims/archive/2009/03/18/mix09-day-1-keynote-pt-2-scott-guthrie-on-advancing-user-experiences.aspx, March 2009.

[2] Cisco Networks, “Cisco’s Visual Networking Index Global IP Traffic Forecast 2010-2015,” http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html#~forecast, June 2011.

[3] Pantos, R. and May, E. W., “HTTP Live Streaming”, Informational Internet Draft, http://tools.ietf.org/html/draft-pantos-http-live-streaming-06, March 2011.

[4] Microsoft, “IIS Smooth Streaming Transport Protocol,” http://www.iis.net/community/files/media/smoothspecs/[MS-SMTH].pdf, September 2009.

[5] Adobe, “Adobe HTTP Dynamic Streaming,” http://www.adobe.com/products/httpdynamicstreaming/.

[6] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP".

[7] “MPEG systems technologies — Part 6: Dynamic adaptive streaming over HTTP (DASH), ISO/IEC FCD 23001-6,” http://mpeg.chiariglione.org/working_documents/mpeg-b/dash/dash-dis.zip, January 2011.

[8] ISO/IEC 14496-12:2008/DAM 3, “Information technology – Coding of audio-visual objects – Part 12: ISO base media file format – Amendment 3: DASH support and RTP reception hint track processing”, January 2011.

[9] ISO/IEC 14496-12, “Information technology – Coding of audio-visual objects – Part 12: ISO base media file format”, 2008.

[10] ITU-T Rec. H.222.0 | ISO/IEC 13818-1, “Information technology – Generic coding of moving pictures and associated audio information: Systems”.