MPEG-DASH Enabling Adaptive Streaming with Personalized Commercial Breaks and Second Screen Scenarios

Stefan Kaiser
Fraunhofer Institute FOKUS
Kaiserin-Augusta-Allee 31
Berlin, Germany

Stefan Pham
Fraunhofer Institute FOKUS
Kaiserin-Augusta-Allee 31
Berlin, Germany

Stefan Arbanowski
Fraunhofer Institute FOKUS
Kaiserin-Augusta-Allee 31
Berlin, Germany

first.lastname@fokus.fraunhofer.de

ABSTRACT

The growing demand for video streaming over the Web has increased the importance of the recently published MPEG standard Dynamic Adaptive Streaming over HTTP (MPEG-DASH). With MPEG-DASH, video delivery will be harmonized across the Internet by enabling consumption and transport of adaptive bitrate content.

Future opportunities that come with this technology are discussed in this paper as well as the challenges that have to be mastered. The discussion is done on the basis of proof of concept implementations. The topics deal with various aspects of MPEG-DASH deployment ranging from the fundamental content creation over applied adaptation to smart use cases like personalized commercial breaks or session mobility-enabled adaptive video streaming.

Keywords

MPEG-DASH, IPTV, HTTP Adaptive Streaming, Second Screen, Personalized Commercial Breaks

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video—*algorithms, design, experimentation, standardization*

1. INTRODUCTION

Audio and video streaming over Hypertext Transfer Protocol (HTTP) has received much attention in recent months due to its ability to be accessed by any consumer device that is connected to the Web because of increasing bandwidths and affordable provider contracts. In addition, advances in streaming like dynamic adaptation on changed connection quality or serving media resolution depending on the current demands of the user's device are applied to HTTP streaming as well. The recently published ISO standard Dynamic Adaptive Streaming over HTTP (MPEG-DASH) [6] defines media presentation and content formats and is growing in popularity as well as becoming generally accepted by most vendors of pre-existing technologies. It has various possible use cases in the field audio and video content distribution, including live, on-demand and premium content. This area is as popular as never before with increasing customer numbers for services such as Spotify [3] or Netflix [2] (represents 29.7% of North American Peak Downstream Traffic [4]). In combination with ubiquitous computing, from computer via tablet and smartphone through to TV, over any network, the need for MPEG-DASH support in browsers for these devices is growing as well. Streaming media has become a commodity. Moreover open standards such as HbbTV have adopted MPEG-DASH in recent specification releases (HbbTV Version 1.5 [7]).

MPEG-DASH consists of a XML-based manifest file named Media Presentation Description (MPD) that refers to different representations of one media. Multiple representations of a specific media component are bundled to an adaptation set. The representations are provided in a segmented format and can be switched within an adaptation set adaptively. This format evolved from early efforts done by 3GPP with their proposal of Adaptive HTTP Streaming. At this time proprietary solutions like Microsoft Smooth Streaming, Apple HTTP Live Streaming and Adobe HTTP Dynamic Streaming existed, but only worked in the respective company's own component ecosystem. This led to the existence of multiple different formats that were not interoperable. To overcome this issue the MPEG-DASH specification was developed on the basis of the 3GPP approach with little innovation and mostly constraints on other existing industry solutions. Further, the standard specifies common encryption, which means the same short segments of video can be decrypted and decoded by devices using different DRMs. This feature is of great importance for future paid video business models.

2. STATE OF THE ART

The majority of currently available DASH clients consist of native implementations, such as the Osmo player [1], or media player plugins, e.g. VLC [5]. Regarding DASH support in Web browsers contributions of the Web Hypertext Application Technology Working Group (WHATWG) and the W3C HTML Working Group have led to the specification of the Media Source Extension [9] (MSE) and Encrypted Media Extensions [10] (EME) to enable seamless, DRM-interoperable HTTP adaptive streaming using the HTML5

EuroITV'13, June 24–26, 2013, Como, Italy.

video object. It describes a JavaScript API to dynamically push media segments into an HTML5 video or audio object or respectively, to exchange license keys. Using MSE and EME content providers can build their own MPEG-DASH players for each scenario that they want to support.

The Fraunhofer FOKUS solutions Famium and MPEG-DASH transcoder build a platform for the consumption and delivery of adaptive bitrate content. Famium is a Webkit-based Web browser with integrated MPEG-DASH support. The Famium extensions are based on W3C proposals of MSE and EME. For the usage in delivery networks, the MPEG-DASH transcoder supports MPEG-DASH-compliant content creation, including common encryption for ISO Base Media File Format (ISO BMFF). With an end-to-end solution for browser-based DRM for MPEG-DASH, Fraunhofer FOKUS addresses the need for a common DRM-interoperable encryption and HTTP adaptive streaming, which is a key factor in Internet-delivered video.

These components and the related proof of concept implementations are the basis for the discussion of the following opportunities and challenges.

3. OPPORTUNITIES

I. IP-based Broadcast TV

MPEG-DASH enables the delivery of video content over IP. These streams, no matter if on-demand or live video, can be received by multiple devices supporting MPEG-DASH, e.g. laptops, mobile devices and connected TV sets. The latter present a huge opportunity for broadcasters and content providers to transmit their already existing contents to the end customers by using the Internet Protocol.

During summer 2012 European broadcasters already experimented with a unicast HTTP-based live video transmission, but without using MPEG-DASH, as this approach was not supported by HbbTV (1.0) in the corresponding time period. For instance six different Olympic Games' live streams were shown on the red button presence of the German public broadcaster ARD, instead of using dedicated broadcast channels as was done for previous Olympic Games. The user could switch between different sports and see the corresponding live contents without leaving the current channel.

It is obvious that this approach is focusing on minimal broadcast and maximal broadband bandwidth. Broadcast channel slots and their corresponding bandwidth are sold by the DVB-T, DVB-C and DVB-S service providers, and are relatively expensive. IP delivery has the potential to be more cost-effective than broadcast, but investment in 'adequate' content delivery infrastructures (e.g. CDNs or multicast), which can scale to meet the demands of replacing broadcast, have to be considered.

The non-adaptive unicast streaming approach leads to several issues and consequently a bad user experience. For example, the video quality had to be decreased significantly to allow a fluent video playback. Moreover freezing and stuttering due to buffering during streaming have not been an exception. These issues may have been caused by an unexpectedly high number of viewers, but are also due to the nature of non-adaptive unicast streaming. Adaptive streaming offers a solution to this problem. It allows devices to seamlessly adapt to media representations (e.g. in terms of resolution, codec or bitrate) that fit the user's device and network connection best. Besides delivering the content in the highest quality available for a given network, faster start up times for media playback can be expected.

As part of our work and with the help of the MPEG-DASH transcodre we provide live DASH content to various organizations to ensure interoperability. Once devices have widely implemented new TV standards (e.g. HbbTV 1.5/2.0) adaptive live streaming using MPEG-DASH can become reality.

The usage of the new HbbTV 1.5 Standard [7] for IP-based Broadcast TV will improve the user experience by benefiting from MPEG-DASH features, such as adaptive streaming and content protection on connected TV sets. For instance the video quality will not decrease on cloudy days, when there is no DVB-T or DVB-S signal, and depends only on the Internet connection. The broadcasted streams can be protected with DRM-interoperable common encryption, thus validating the authorization of the current viewer. Moreover, broadcasters can ensure the delivery of specific content for specific regions, devices or user groups.

II. Personalized commercials using MPEG-DASH

IP-based Broadcast TV can leverage the bidirectional connection to personalize advertising, e.g. based on users' TV watching behavior or ratings. Consequently consumers are only shown commercials that are of interest to them. Thus, a child will not see the regular horror movie trailer but a video showing a toy commercial delivered over IP.

The exchange of ads in an exclusively IP-based environment can use all benefits of MPEG-DASH: a seamless start without video buffering, adaptive streaming and content protection. Therefore advertisement insertion is one of the main topics currently discussed in various standardization organizations dealing with MPEG-DASH. One approach that we have realized for using personalized commercials is to replace the relevant parts in the description file. A MPD-File, which contains the metadata of adaptive video streams to be played back, can be linked to other MPD-Files presenting an ad stream for a pre-, mid- and post-roll [6]. This second MPD-File is generated on demand on a recommendation server and contains the corresponding video information for a specific spot. The decision what content and therefore which description file shall be retrieved server side is based on collected information for the current user.

Taking the law compliance as given, broadcasters must collect usage data to predict the users' behavior and their likings to find the best fitting commercial. One method to get the needed data is the collection of automatically generated feedback such as the watching behavior. Another way is the collection of explicit feedback, as end customers can directly rate actual ads. Collaborative filtering algorithms defined in order to predict ratings and to recommend the potentially best fitting commercials can improve the commercial break experience.

III. Session mobility and Second screen

Second or more screens can leverage the experience of media consumption. While the main content of a live sport event may be presented on the main TV in the living room, the tablets of the people in the room can be used to show additional information synchronously. This can be other camera angles like on board cameras of single car drivers, or always having an overview of the whole racing track of a Formula 1 race. Additional statistics may also be presented, as well as targeted advertisement. Further, session mobility can be easily realized using MPEG-DASH. The media representations in MPEG-DASH use a common timeline, which enables simple timestamp synchronization between devices. For example, this allows a user to watch a movie on the TV, and when leaving the home, the TV session can be transferred to a mobile device to continue watching the movie. Further, session mobility and second screen features are in discussion for HbbTV 2.0. Our implementation of these features serves as proof of concept.

Another opportunity for using MPEG-DASH is the possibility to support additional content metadata. The Media Presentation Description (MPD) therefore supports the declaration of different roles for an Adaptation Set that contains a media component. A role of an adaptation set has an attribute schemeIdUri and a value. For the schemeIdUri "urn:mpeg:dash:role:2011" the following roles are defined: caption, subtitle, main, alternate, supplementary, commentary and dub. Each role shall be used for media content to describe it further, so a MPEG-DASH client can decide which media content to use. Other schemeIdUri and therefore other Role values can be used as well.

Supporting multiple screens with different content is possible using the role description and a single MPD file. This can be for example a car racing video. The MPD file contains one adaptation set with a role value "main". The media files in this adaptation set show the race from an on top view. Additionally other Adaptation Sets in the same MPD file are marked with the role value "alternate" or "supplementary". Media content within these adaptation sets may show videos from the cockpit perspective of different cars. A MPEG-DASH enabled Website and player can now use this extra information to play different content on different devices. Synchronized playback of these videos is one challenge that is addressed in this paper (see section III of the challenges).

There are different possibilities for MPEG-DASH players to choose the role and therefore the media to play. The players can show users the available roles to choose so every user can decide on his own what to see. A second approach is that there are players who know that they are supposed to only play media that is marked with certain Role values. Another way is that the players themselves do not get the whole MPD file.

Second or multiple screen scenarios can take advantage of MPEG-DASH features such as describing cross-platform media formats in one presentation description file (MPD). For example, we have implemented a node.js-based synchronization Web server with WebSocket support. Every device in a household connects to this server and is therefore joining a session. The server now knows the existence of different devices. After selecting a MPD file on one device, this screen shows the media that is described by the main role. All connected devices or devices that join the same server hosting the application after that get custom versions of the original MPD file. This means that some media marked as "alternate" in the original MPD file may now be marked as "main" and the original main Adaptation Set is removed from the MPD file for the other devices. This way the players do not need to decide anything but playing "their" main media.

A second screen scenario can also be realized using special MPEG-DASH players. The decision which part of the MPD should be played can be determined by the client. One approach is that the user manually selects the camera angle of the car race. Alternatively, each player on each device knows what to play. This can depend on other players discovered in the local network using, for example, UPnP or Bonjour. The decision can also depend on device capabilities and properties like the screen resolution or the ability to listen to touch events.

4. CHALLENGES

I. Algorithms for Adaptive Streaming

The first and most crucial challenge of using MPEG-DASH is the adaptivity algorithm of the client. The deployment of MPEG-DASH alone is not sufficient for getting all the benefits of this solution. Since this part is not defined in the standard, it is as distinctive feature of a MPEG-DASH client implementation. In general, the available bandwidth of a client defines the maximum possible bitrate for a representation to be used. An approach that we have implemented for reaching the best quality representation is to start with the lowest bitrate available and gradually increase the quality of the representation. This allows the player to converge to the maximum bitrate and considers the stability of the network condition.

Network condition can vary everywhere the Web is accessed. The most affected devices obviously are mobile phones and tablets connected via a cellular network. The coverage of a mobile network's base station is limited as well as its amount of clients that can be served. This quickly leads to a bottleneck and bandwidth losses for each user. The bandwidth changes in this kind of network can differ in type and intensity and vary from small spikes to long-term changes. Often used with mobile phones and tablets but also with laptops are wireless networks based on the IEEE 802.11 standard. Characteristic for this kind of network are small spikes [8]. In a local area network (LAN) the same Internet connection to the Internet service provider (ISP) is shared with other users. This can lead to unexpected bandwidth drops when other network members cause high traffic loads. This applies to a wireless LAN as well.

The length of the media segments, or more detailed, the distance of key frames, has an impact on the reaction time of the algorithm. Long key frame distances make the client react more slowly than short distances do.

Efficient switching is very complex and difficult to develop. Its behavior depends on the kind of change that is happening, whether it increases or decreases bandwidth, whether it is a small spike, or a long-term change. It also depends on the buffer settings that can be facilitated. And in the end it has to incorporate human aspects of usability. Big steps in quality changes are not expected by the user, therefore the quality should be smoothly shifted to the current possible one. Frequent changes are annoying and thus not recommended to happen. Bandwidth can cause these types of behavior, e.g. by a break-in of bandwidth from highest to the minimum quality level for several minutes on a mobile phone, or oscillating on the bandwidth threshold of a specific quality level. On short spikes that last only for a few seconds, the buffer can possibly compensate the occurred gap, depending on its settings. A proper switching algorithm needs to consider all of these properties to enable a smooth and pleasant viewing experience.

II. Content creation

The counterpart to consuming MPEG-DASH media is its creation. The MPEG-DASH specification is format and codec agnostic. Nevertheless, the standard currently defines rules for the two container formats: ISO Base Media File Format (ISO BMFF) and MPEG-2 Transport Stream (MPEG-2 TS). These rules are applied by different profiles for each container format and specify restrictions on the manifest file as well as the content.

As the profiles just limit the appearance of the container format, it is possible to convert current content to the MPEG-DASH format by just repacking it along the profile rules. However, adaptive streaming holds some challenges that need to be considered.

First of all, the content has to be encoded into different quality levels, i.e. bitrates. These bitrates should be chosen in an appropriate way to meet the desired target characteristics. The choice of appropriate bitrates is not the only issue that comes with content creation for adaptive streaming. During the creation process, the content will be split into segments and fragments. One segment can contain one or multiple fragments. As mentioned in the previous chapter about algorithms for adaptive streaming, the segment length plays an important role in the behavior of a MPEG-DASH client. Shorter segment lengths mean a more reactive adaptive behavior. At the beginning of a segment, there should be one key frame, i.e. one full picture. Using short segment lengths, the key frame rate might be higher than necessary. This leads to higher storage usage as well as higher bandwidth usage. Altogether, this raises the task of searching for the right tradeoff between effective resource usage and adaptivity reaction time.

MPEG-DASH content is based on existing adaptive streaming technologies like Microsoft's Smooth Streaming, Apple's HTTP Live Streaming (HLS) or Adobe's HTTP Dynamic Streaming. Conversion of content compliant to one of these technologies is less effort than a completely new transcoding/repacking process. The format used in Smooth Streaming is closely related to ISO BMFF using the live profile. Adobe uses the MP4 container format being a specialization of ISO BMFF and therefore also closely related to a MPEG-DASH format. Apple instead uses MPEG-2 TS in its HLS specification, which is related to the MPEG-2 TS profiles defined in the MPEG-DASH standard.

III. Synchronization of multiple devices

Using MPEG-DASH for multiple devices raises the challenge of synchronization, as this is important for the aforementioned second screen and session mobility scenarios. The shown content in the scenarios may need to have the same state. Furthermore some scenarios may need to play media in time synchronization, so actions in the media happen at the same playback time.

Communication and synchronization of different devices is not part of the MPEG-DASH standard. This means there exists no single recommended way of realizing this issue. The program part controlling this mechanism can be part of the MPEG-DASH player or application logic in a Website that communicates with a synchronization server, for example.

A synchronization scenario which cannot rely on extra information, can be implemented by only using current standard Web technologies. Using WebSockets a Website can send and receive data via the TCP protocol. The client devices open the Web application and will be connected to a server. The synchronization server is part of the service provider's delivery network. All information is sent to the server instead. The server then evaluates incoming data such as if a user wants to pause the video and sends this information to all connected clients, so they will pause their media as well.

A centralized controlling mechanism, as we have implemented, has the advantage of one "real" state that is only pushed to other devices. The synchronization server shares a common timeline for synchronization between the clients. The MPEG-DASH players do not need to have special implementations. Only standard functions like play, pause and seek are needed to synchronize the players on all devices.

5. CONCLUSION

Today MPEG-DASH client as well as server implementations are still in development. There are not many native MPEG-DASH clients available and implementation work for built-in browser support or corresponding APIs is still in progress. The paper gives an outlook on future scenarios and advantages that come with them by using MPEG-DASH technology. With our implementation of MPEG-DASH components and scenarios, valuable input to current standardization activities is given in the form of proof of concept implementations.

We described future, as well as already possible scenarios, which can benefit from Internet-delivered television. The ability to have session and context information as well as the possibility to deliver different content parts to different users or devices will greatly improve the user experience. With MPEG-DASH support growing and implementations improving, the opportunities previously mentioned in this paper can be realized. Further, the mentioned challenges need to be handled to improve user experience and to create new business models.

6. REFERENCES

[1] GPAC Osmo4 Player website. `http://gpac.wp.mines-telecom.fr/player`.

[2] Netflix Inc. website. `http://www.netflix.com`.

[3] Spotify Ltd. website. `http://www.spotify.com`.

[4] Global Internet Phenomena Report. *Sandvine*, 2011.

[5] *A VLC Media Player Plugin enabling Dynamic Adaptive Streaming over HTTP*, Scottsdale, Arizona, November 2011.

[6] Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media presentation description and segment formats. *ISO/IEC 23009-1:2012*, 2012.

[7] HbbTV Specification. `http://www.hbbtv.org/pages/about_hbbtv/HbbTV-specification-1-5.pdf`, March 2012. Version 1.5.

[8] S. Akhshabi, A. Begen, and C. Dovrolis. An Experimental Evaluation of Rate-Adaptation Algorithms in Adaptive Streaming over HTTP. *In Proceedings of the second annual ACM conference on Multimedia systems, MMSys*, 11:157–168, 2011.

[9] A. Colwell, A. Bateman, and M. Watson. Media Source Extension. `https://dvcs.w3.org/hg/html-media/raw-file/tip/media-source/meia-source.html`, 2012.

[10] D. Dorwin, A. Bateman, and M. Watson. Encrypted Media Extensions. *https://dvcs.w3.org/hg/html-media/raw-file/tip/encrypted-media/encrypted-media.html*, 2012.