The Content-Aware Video Adaptation Service for Mobile Devices

Stefan Wilk, Wolfgang Effelsberg
TU Darmstadt, Germany
{stefan.wilk, effelsberg}@cs.tu-darmstadt.de

ABSTRACT

In most adaptive video streaming systems adaptation decisions rely solely on the available network resources. As the content of a video has a large influence on the perception of quality our belief is that this is not sufficient. Thus, we have proposed a support service for content-aware video adaptation on mobile devices: Video Adaptation Service (VAS). Based on the content of a streamed video, the adaptation process is improved by setting a target quality level for a session based on an objective video quality metric. In this work, we demonstrate VAS and its advantages of a reduced data traffic by only streaming the lowest video representation which is necessary to reach a desired quality. By leveraging the content properties of a video stream, the system is able to keep a stable video quality and at the same time reduce the network load.

CCS Concepts

•**Computer systems organization** → *Client-server architectures;*

Keywords

DASH, Adaptation, Content-Awareness, Video Quality.

1. INTRODUCTION

With the arrival of adaptive video streaming systems, streaming sessions can be adjusted to available network resources and device properties.

In contrast to non-adaptive video streaming systems, adaptive approaches are able to switch the bitrate level during the playback of the video and thus avoid video freezing events (stalling). Stalling occurs when video data is not received by the client in time. Yet, those switches from one representation to another should be planned carefully to ensure a high quality streaming experience. Solely relying on the throughput for adaptation decisions can be inefficient, as the precise estimation of network conditions on the application layer is difficult [5]. The streamed video content has a significant impact on the perceived quality of a video stream. Thus, we argue that adaptation schemes should consider the characteristics of the streamed video. Previously, we have introduced a system, which addresses content-aware adaptation for mobile streaming clients [15]. It leverages the video streaming protocol MPEG-DASH [12], which standardizes the network communication between the streaming client and the server as well as the description of the different video qualities in a manifest. Different video representations are encoded with different target bitrates. The video bitrates are affected by the video's encoding dimensions: the spatial dimension (resolution), quality dimension (signal-to-noise ratio) and the temporal dimension (frame rate). Different DASH representations of a video are stored independent of each other and then split into segments of equal duration, which usually last between two and thirty seconds. An arbitrary web server can store the video segments, which are retrieved by the clients using HTTP. Depending on the available network resources the client can decide at runtime which quality to select for the next segment.

The proposed technical demo shows the contributions of the Video Adaptation Service (VAS) for mobile devices. VAS performs content-aware adaptation on mobile devices by calculating the video quality on a per-segment or per-shot basis for a DASH client. This content-aware streaming achieves significant data reductions since depending on the content of a video stream, higher bitrate representations may not increase the perceived quality above the current level. VAS categorizes new video segments using structural, temporal and color information and thus enables our system to react to changing content properties.

Our technical demo uses different mobile devices as well as a laptop. The laptop represents the video streaming and the VAS server. The VAS is used to support the mobile devices in a content-aware manner. The audience attending the technical demo have access to a video website including different videos of different genres and can select which video to watch. One mobile device is using VAS and another device is streaming video without VAS. The VAS-enabled device adapts the stream according to the content properties of the video. The VAS-disabled device neglects content-aware streaming; it is based on the available bitrate only. Users can set the desired subjective quality of the stream on the VAS-enabled device and will immediately see the differences between the played back video between the two devices: when selecting the highest subjective quality no or only minimal quality degradations will be shown in

MMSys'16 May 10-13, 2016, Klagenfurt, Austria

ACM ISBN 978-1-4503-4297-1/16/05.
DOI: http://dx.doi.org/10.1145/2910017.2910639

comparison to the VAS-disabled device, but the data traffic also decreases.

2. RELATED WORK

In comparison to prior work, we propose a content-aware adaptation scheme for MPEG DASH-like systems. Sophisticated video adaptation schemes have been proposed with the AMES [13] and QDASH [9] systems. Both approaches leverage cloud agents to transcode video streams when being requested by mobile devices. Our proposed server-based VAS offers recommending adaptations to a mobile client and thus does not execute adaptations itself. Thus, the MPEG DASH standard, stating that adaptations should be executed by the client, is not broken. Prior works in the area of content-adaptive video stream include Fiandrotti et al.'s [3] content-adaptive traffic scheduling schemes. A scalable video stream is used for which different layers aren assigned with differing priorities depending on the detected motion in a video stream. In contrast to Fiandrotti's work, we take three dimensions of a video into account and inspect the structures and the motion of a video sequence. A machine-learning approach for adaption heuristics for scalable video codecs [11] has been proposed by Akyol et al. [2]. Content-awareness in adaptive video streaming protocols similar to DASH is not new. Adzic et al. [1] show that by content-aware segmentation at scene boundaries and an associated placement of reference video frames one can increase the perceived quality of a streamed video. In contrast to this approach we design a system analyzing content to give recommendations on how to adapt in three content dimensions: resolution, frame rate and signal-to-noise ratio.

3. VAS - VIDEO ADAPTATION SERVICE

To better understand the technical demo, we first introduce the concepts and architecture of VAS. A detailed description of VAS and an evaluation can be found in our prior work [15]. VAS is a server-based system service designed for assisting mobile streaming clients in adapting video streams (see Figure 1). VAS has the goal to evaluate the video quality of a stream on behalf of the mobile device. We have placed VAS on a server in order to reduce the client's computational burden. The service evaluates different video representations and the quality differences between them. As the perceived quality may change over the course of a video, VAS classifies segments of a video regarding content properties such as motion, color and structure. By video content inspection it provides instructions for the adaptation of a mobile client to reduce the network traffic while keeping a consistent quality. The video adaptations are then executed on the mobile device. A mobile device can request the adaptation plan for the remaining part of the stream from VAS periodically. Thus, existing MPEG DASH clients have to undergo minimal changes.

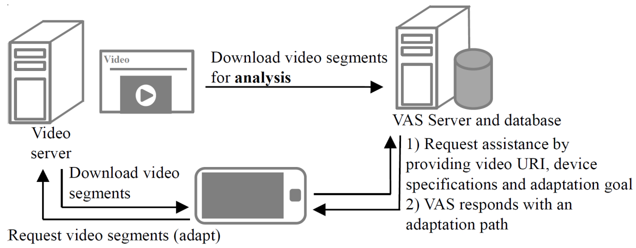
Video
Download video segments for analysis
VAS Server and database
Video server
Download video segments
1) Request assistance by providing video URI, device specifications and adaptation goal
2) VAS responds with an adaptation path
Request video segments (adapt)

Figure 1: A brief overview on the setup of VAS and its links to the video streaming client and the video hosting server. Taken from [15].

As the basis for the technical demo the DASH Industry Forum player[1] is used. Upon the start of a streaming session, the client redirects the URL of the video to the VAS server. This triggers the VAS server which starts the evaluation of the video stream in order to support adaptation in real-time. To retrieve the information how to adapt the streaming client regularly contacts the VAS server. This allows the client to quickly request new adaptation recommendations if the network conditions change.

In the remaining subsection, we will discuss the tools used for estimating the perceived video quality using an objective video quality metric, an approach to classify DASH video segments and the implemented content-aware adaptation strategy.

3.1 Video Quality Measurement

VAS needs to understand the perceived quality of a video representation over the course of time. Our assumption is that this quality can not always be approximated by the bitrate alone. Thus, we leverage the objective video quality metric VQM (General Model) by Pinson et al. [10] to estimate the subjective quality perceived by the user. Furthermore, we map the VQM model to the Mean Opinion Score (MOS) model based on Zinner et al. [16].

The standard VQM metric is not designed for real-time quality assessment of different video representations. Furthermore, it cannot conduct an assessment if the reference and test sequences have different video resolutions and frame rates. Thus, we leverage an extension of the VQM model which can be used on GPUs, based on the one proposed by Wichtlhuber et al. [14].

3.2 Classification of DASH Segments

The perceived quality of a video stream may quickly change depending on the video content. Thus, we determine the video quality for individual video shots. Yet, the video quality calculation using VQM may still be challenging to be completed in real-time when the number of requests is high. Thus, VAS classifies video shots according to metrics representing the content of a video shot, which can be calculated with less computational effort. As a video shot represents an uninterrupted sequence of frames recorded from a single camera, we assume that the perceived quality of DASH representations in two video shots is similar, if the shots are very similar in terms of motion, structures and color. This has been shown, e.g., by Adzic et al. [1] and Akyol et al. [2]. Regarding the adaptation plans produced by VAS, this means that video shots which are similar in these three respects benefit from similar video adaptations. Metrics once calculated for a video are stored in a lookup database for easy retrieval and comparison.

The metrics we use are based on the recommendations of the ITU to use the spatial perceptual information (SI) and the temporal perceptual information (TI) [6].

$$SI = max_{time}\{std_{space}[Sobel(F_n)]\} \tag{1}$$

[1] DASH IF Reference: https://github.com/Dash-Industry-Forum/dash.js

SI uses a Sobel filter on each video frame F_n. The luminance plane of a frame is used to retrieve the standard deviation over all pixels (std_{space}). The maximum value in a frame series (max_{time}) in a video shot is chosen as the representation of the structural information.

The TI is used to represent the motion in a video shot:

$$TI = max_{time}\{std_{space}[M_n(i,j)]\} \tag{2}$$

The motion is calculated between two adjacent frames as the difference $M_n(i,j) = F_n(i,j) - F_{n-1}(i,j)$ where $F(i,j)$ is a pixel in a frame of the video and n is the current number of the frame. i and j represent the row and the column index of the frame pixel. TI is computed as the maximum over a video shot (max_{time}) of the standard deviation over space (std_{space}) of $M_n(i,j)$ over all i and j.

Additionally, the colorfulness of videos (Co) is evaluated. It was proposed by Hassler [4]:

$$Co = \sqrt{\omega_{rg}^2 + \omega_{by}^2} + 0.3\sqrt{\mu_{rg}^2 + \mu_{by}^2} \tag{3}$$

in the RGB color space where, rg is represented by $rg = R - G$ and $yb = 0.5(R+G) - B$. Given that structures as well as motion may change significantly over the duration of a video, a single value of SI, TI or Co is inadequate to classify the entire content of a video.

SI, TI and Co profiles are generated per video shot and used to classify it. As mentioned above, we assume in this work that the same adaptations in different videos with similar characteristics result in the same perceived quality. The calculated adaptation paths, available video representations and SI,TI and Co combinations of a video are stored in a database of VAS.

The selection is then conducted by using the Euclidian distance for SI, TI and Co:

$$ED = \sqrt{(SI_{\mathrm{D}} - SI_{\mathrm{S}_i})^2 + (TI_{\mathrm{D}} - TI_{\mathrm{S}_i})^2 + (Co_{\mathrm{D}} - Co_{\mathrm{S}_i})^2} \tag{4}$$

SI_{D} represents the SI value for an arbitrary video shot stored in the VAS database, whereas SI_{S} represents the respective value of the requested video shot S_i. The aim is to minimize ED and choose the reference video shot with the smallest Euclidian distance. For the video shot selected (VS_D) the adaptation information is retrieved to produce an adaptation plan for S_i. By using SI/TI/Co we classify the video independent of the temporal, spatial or quality dimension (e.g., by using quantization parameters). These dimensions are then used for executing the adaptation, so that, e.g., a video shot which consists of fast motion would be adapted in order to ensure a high frame rate representation, whereas in a structure-intensive sequence an adaptation to high resolution and quality (SNR) dimension representations would be favored.

3.3 Minimum MOS Adaptation

The proposed minimum MOS adaptation scheme ensures that a client streams the minimum bitrate representation offering a specified perceived quality level (MOS). Thus, a client which wants to stream a high quality video streaming session picks a MOS of 5 (highest possible MOS value). This does not mean that only the highest bitrate representation is streamed, but the representation which just achieves a MOS of 5, which has the minimum average bitrate. The proposed scheme requires that the modified DASH clients states his desired minimum quality upon which VAS recommends video adaptations for the upcoming DASH segments in order to keep the specified perceived quality.

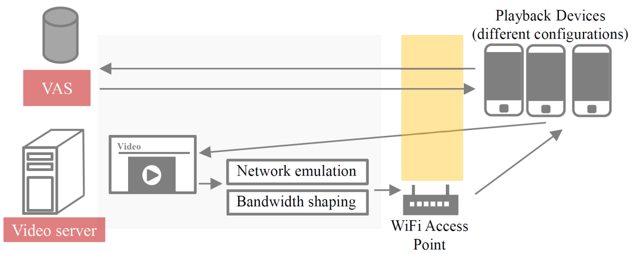

Figure 2: Physical setup of the VAS technical demo.

The minimum MOS strategy calculates the three metrics SI, TI and Co for the video shots as well as the video qualities of each representation and maps them to DASH segments. When a video adaptation request is received by the client, the VAS checks if the quality calculation is completed for the respective video shot. A timely completion allows the VAS to calculate the adaptation paths based on the current video properties. On this basis, our system can recommend the client to select specific representations for the next segments.

This scheme implies that adaptations are necessary either due to the fact that the minimum MOS recommends to switch the representation, or the available bandwidth requires an adaptation. In those cases we aim for a seamless adaptation from the current to the target representation. As VAS knows the the perceived quality values of all video representations for the *upcoming* DASH segments it can compute an adaptation path which minimizes the quality differences with each adaptation.

4. SETUP OF THE TECHNICAL DEMO

The technical demo consists of one laptop, which runs a web server hosting DASH videos and the VAS. In addition, four mobile devices (smartphones and tablets) are used to run the DASH video streaming client. The communication between the mobile devices and the laptop is using a dedicated WiFi Access Point. Available resources for the streaming are virtually limited using a traffic shaping module. The setup is depicted in Figure 2.

The web server shows a video sharing site which we call 'streamlet' (see Figure 3), which consists of five video sequences from different genres and contents, including The Swiss Account, Big Bucks Bunny, Valkaama, Of Forest and Men and the Elephant's Dream. The videos are taken from the DASH datasets provided by Lederer et al. [7, 8].

The mobile devices run an application accessing the mobile version of 'streamlet'. Once a user of the mobile device starts streaming a specific video on 'streamlet', the video details page shows the played back video as well as statistics. These statistics include the currently played back video representation and its average bitrate, the absolute generated data traffic of this streaming session, as well as the currently played back perceived video quality measured by VQM [10]. In addition, metrics including the startup delay and the number and duration of stalls are shown.

Mobile devices are either set up to use the VAS service or plain DASH. Devices using VAS include additional statistics like the relative data savings when using VAS as well as the

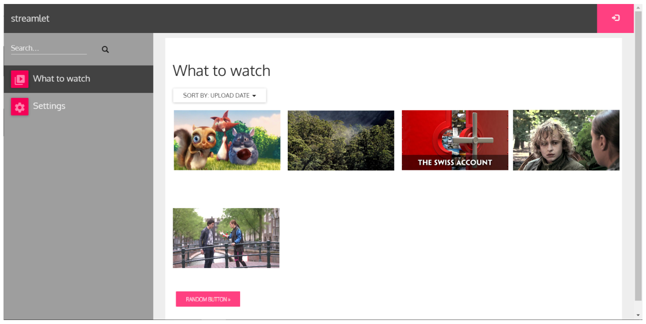

Figure 3: Streamlet website showing the set of different videos that can be used in the VAS demo.

representation and bitrate which would be played back when VAS was not used. These additional metrics illustrate the data savings achieved by VAS.

Users of the VAS technical demo use two mobile devices – one using VAS and one without VAS – to select one of the five videos available. On the VAS-enabled device the user can set the desired perceived quality of the streaming session by choosing a MOS of 3.5 (acceptable video quality), 4 (good video quality) or 5 (best video quality). This desired video quality influences which representations the minimum MOS adaptation scheme is selected for adaptation. The client will adapt accordingly, and statistics given in the mobile device show the achieved data savings. The VAS-disabled device is being used so that users can observe the impact of VAS on the perceived video quality. Whereas the VAS-disabled device optimizes the average streamed bitrate, i.e., playing back the highest, possible bitrate representation, VAS usually recommends to play back lower bitrate representations. Thus, we demonstrate that with a MOS of 5 the quality differences between the VAS-disabled stream and the VAS-enabled stream are nearly unnoticeable. After each streaming session users are able to re-run the technical demo with other videos showing that depending on the content of a video different data savings can be achieved.

5. SUMMARY

Our technical demonstration of VAS illustrates that significant data savings can be achieved when adapting video in a content-aware manner. A user can participate in this technical demo by using a mobile streaming device and streaming video either in a content-aware or -unaware manner. During the streaming session, statistics show that VAS achieves significant data savings while maintaining a constant perceived quality compared to a non-VAS device. Our technical demo thus illustrates the advantages of respecting video content when making adaptation decisions in MPEG DASH-like streaming systems.

6. ACKNOWLEDGEMENTS

This work has been funded by the DFG as part of the CRC 1053 MAKI in the projects C03. We thank Matthias Wichtlhuber and Gregor Wicklein for providing the source code of RT-VQM which is used for calculating VQM scores on CUDA-enabled graphic cards.

7. REFERENCES

[1] V. Adzic, H. Kalva, and B. Furht. Optimizing video encoding for adaptive streaming over HTTP. *IEEE Transactions on Consumer Electronics*, 2012.

[2] E. Akyol, A. M. Tekalp, and M. R. Civanlar. Content-Aware Scalability-Type Selection for Rate Adaptation of Scalable Video. *EURASIP Journal on Advances in Signal Processing*, 2007.

[3] A. Fiandrotti, D. Gallucci, E. Masala, and J. De Martin. Content-adaptive traffic prioritization of spatio-temporal scalable video for robust communications over QoS-provisioned 802.11e networks. *Signal Processing: Image Communication*, 2010.

[4] D. Hasler and S. E. Suesstrunk. Measuring Colourfulness in Natural Images. In B. E. Rogowitz and T. N. Pappas, editors, *IS&T/SPIE Electronic Imaging 2003: Human Vision and Electronic Imaging VIII*, volume 5007, 2003.

[5] T.-Y. Huang, N. Handigol, B. Heller, N. McKeown, and R. Johari. Confused, timid, and unstable. In *ACM Internet Measurement Conference (IMC)*, pages 225–238, 2012.

[6] ITU. ITU-R Recommendation P.910, 2008.

[7] S. Lederer, C. Mueller, C. Timmerer, C. Concolato, J. Le Feuvre, and K. Fliegel. Distributed dash dataset. In *ACM Multimedia Systems Conference*, 2013.

[8] S. Lederer, C. Müller, and C. Timmerer. Dynamic adaptive streaming over http dataset. In *ACM Multimedia Systems Conference*, 2012.

[9] R. K. P. Mok, X. Luo, E. W. W. Chan, and R. K. C. Chang. QDASH: A QoE-aware DASH system. In *ACM Multimedia Systems Conference*, 2012.

[10] M. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3), 2004.

[11] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007.

[12] T. Stockhammer. Dynamic Adaptive Streaming over HTTP: Standards and Design Principles. In *ACM Multimedia Systems Conference*, 2011.

[13] X. Wang, M. Chen, T. T. Kwon, L. Yang, and V. C. M. Leung. AMES-Cloud: A Framework of Adaptive Mobile Video Streaming and Efficient Social Video Sharing in the Clouds. *IEEE Transactions on Multimedia*, 2013.

[14] M. Wichtlhuber, G. Wicklein, S. Wilk, and W. Effelsberg. Rt-vqm: Real-time video quality assessment for adaptive video streaming using gpus. In *ACM Multimedia Systems Conference (MMSys)*, 2016.

[15] S. Wilk, D. Stohr, and W. Effelsberg. Vas: A video adaptation service to support mobile video. In *ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*, 2015.

[16] T. Zinner, O. Hohlfeld, O. Abboud, and T. Hossfeld. Impact of frame rate and resolution on objective QoE metrics. In *International Workshop on Quality of Multimedia Experience*, 2010.