MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 8, No. 2, March 2013

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from MMTC Chair

Dear MMTC colleagues:

Wish you all a pleasant holiday season and a fruitful new year in 2013. It is really a great honor for me to serve as the Asia vice-chair for this vital ComSoc Committee during the period 2012-2014! As part of my duties, I have contributed to the initial setting of the Interest Groups (IGs) and I am starting to work on the promotion of special issues with top journals and the webseminars.

Concerning our IGs, I really believe that these represent the core of our networking and scientific activities and I warmly invite all of you to select one or more IG(s) to get involved by contacting the chair(s) so as to take part as key member. The activities of the IGs include, among others, the organization of workshops, sessions and conferences with the involvement of the MMTC, the editing of special issues in major IEEE journals, the setting of invited talks trough conference calls that can be of interest for our community and the rest of the ComSoc members. While these are the major activities, some others can be carried out following the specific IG topics, such as the contribution to standardization activities.

The two iniatives will not success without strong support from our IG leaders and contributing members. For the special-issue effort, I will be working with Dr. Chonggang Wang to first identify a list of potential transactions, journals and magazines that are relavant to our TC. This list will be shared with our IG leaders who will in turn propose potential topics for a chosen venue We will first socialize the topics with EiC(s) and AEs and work with the IG(s) to develop the full proposal. It is our hope to achieve the largest efficiency via our collaborative efforts, ideally organizing up to 7-10 speciall issues. We also encourage multiple IGs to collaboratively propose topics that are relavent. For the webseminars, we would like to call for recommendations from our IG(s) and all the members with MMTC. Upon receiving your recommendations and/or volunteerings, we will work with IEEE ComSoc to set up the infrastructure and announcement the talks to potential audience. During the two-year term, we will aim to have a quarterly webseminar for our MMTC.

I would like to thank all the IG chairs and co-chairs for the work they have already done and will be doing for the success of MMTC and hope that any of you will find the proper IG of interest to get involved in our community!

Yonggang Wen
Asia Vice-Chair of Multimedia Communications TC of IEEE ComSoc

EMERGING TOPICS: SPECIAL ISSUE ON

QoE Aware Optimization in Mobile Networks

Guest Editors: Tasos Dagiuklas, TEI of Mesolonghi, Greece, Weisi Lin, Nanyang Technological University, Singapore, Adlen Ksentini, University of Rennes 1, France

Quality of Experience (QoE) is the overall performance of an end-to-end networking system from users' perspective. It is basically a subjective measure of end-to-end performance at the service level, from the point of view of users. As such, it is also an indicator of how well the network satisfies users' preferences. The QoE reflects the perceptive output of the network and its performance with respect to the expected quality by end users and it is the result of the perceived effects of all the Quality of Service (QoS) mechanisms across network and application layers. QoE optimization of multimedia applications across mobile networks faces the following challenges: physical impairments, bandwidth variability, session mobility and maintenance while the users hands-off across inter-technology networks.

This special issue of E-Letter focuses on the recent progresses of QoE Aware Optimization in Mobile Networks. It is the great honor of the editorial team to have five leading research groups, from both academia and industry laboratories, to report their solutions for meeting these challenges and share their latest research results.

In the first article entitled, "A QoE cross layer approach to model media experiences*",* Andrew Perkis, from Norwegian University of Science and Technology presents QoE optimization in Immersive Media Technology Experiences (IMTE) through cross layer dynamics. The IMTE represents a holistic view of a user's experience within the media sphere following content from acquisition, representation, interaction to delivery and usage and finally the business model. To optimize QoE, the following aspects for each of the five stages and their interconnections have been defined: the User, Content and Infrastructure.

Dialo from Orange Labs, Hassnaa Moustafa from Intel Labs, Affifi from Institut Mines Télécom and Marechal from Orange Labs author the second article, "*Context Aware Quality of Experience for Audio-Visual Service Groups*". The authors propose a new notion of user experience based on context-awareness with extended context information (network context, device context, user context, content context,…), aiming to improve the user satisfaction using the full media distribution chain (domestic, access and core networks along with content delivery). The authors argue that global QoE should be used for groups sharing a resource to optimize the overall distribution parameters. The paper analyzes the complete distribution chain of multimedia content (device, access network, content network) to identify which parameters should be first adjusted to have an immediate influence on the QoE improvement.

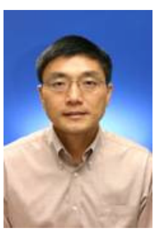
The third article is contributed by Ning Liao and Zhibo Chen from Technicolor Research & Innovation, and the title is *"No-reference IPTV Video Quality Assessment Based on End-to-End Visual Distortion Estimation"*. The authors present a model for parsing-mode P.NBAMS in IPTV and Mobile video streaming scenarios, which demonstrated the best performance in ITU-T P.NBAMS standard competition. In this model, the visual distortions of three types of above-mentioned artifacts are modeled separately, and then the mutual influence of perceptible compression artifacts and slicing/freezing artifacts is modeled by linearly combining the output of the two worst quality levels of the individual artifact models.

Martín Varela and Janne Seppänen presented a QoE-based traffic control system by VTT in the fourth article, *"Video Quality as a Driver for Traffic Management with Multiple Subscriber Classes"*. The authors propose a composite approach to manage the traffic in order to provide adequate QoE to the users. This is accomplished through a subscriber-based differentiation scheme (implemented, without loss of generality, with two classes of users, namely premium and normal ones), and a traffic management scheme based on both access control and application-based traffic differentiation.

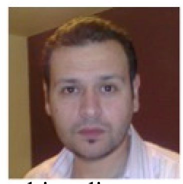
The last article is entitled *"Cross-layer design for quality-driven multi-user multimedia transmission in mobile networks"*, from Maria Martini at Kingston University. In this paper, the author presents the main aspects of QoE-driven cross-layer for multimedia transmission over mobile networks, highlighting design issues and open points.

While this special issue is far from a complete coverage on this exciting research area, we hope that the five invited articles give the audiences a taste of the main recent activities in this area, and provide them an opportunity to discuss, explore and collaborate in the related fields. Finally, we would like to thank all the authors for their great contribution and the E-Letter Board for making this special issue possible.

Tasos Dagiuklas received the received the Engineering Degree from the University of Patras-Greece in 1989, the M.Sc. from the University of Manchester-UK in 1991 and the Ph.D. from the University of Essex-UK in 1995, all in Electrical Engineering. He is Assistant Professor at the Department of Telecommunication Systems and Networks, TEI of Mesolonghi, Greece. He is the Leader of the CONES research group (http://www.tesyd.teimes.gr/cones). Dr Dagiuklas is a Vice-Chair for IEEE MMTC QoE WG and Key Member of IEEE MMTC MSIG and 3DRPC WGs. He is also an active member of IEEE P1907.1 Standardiazation WG. He is a reviewer for journals such as IEEE Transactions on Multimedia, IEEE Communication Letters and IEEE Journal on Selected Areas in Communications. His research interests include 2D/3D video transmission/adaptation/rate control across heterogeneous wireless networks, P2P video streaming and service provisioning across Future Internet architectures. He is a Senior Member of IEEE and Technical Chamber of Greece.



Weisi Lin received his Ph.D from King's College, London University, U.K. He was the Lab Head, Visual Processing, and the Acting Department Manager, Media Processing, Institute for Infocomm Research, Singapore. Currently, he is an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, perceptual quality evaluation, video compression, multimedia communication and computer vision. He has published 200+ refereed papers in international journals and conferences. He is on the editorial boards of IEEE Trans. on Multimedia, IEEE SIGNAL PROCESSING LETTERS and *Journal of Visual Communication and Image Representation*. He has been elected as a Distinguished Lecturer of APSIPA (2012/3). He is the Lead Technical Program Chair for PCM2012, and a Technical Program Chair for *IEEE* ICME2013. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow, Singapore Institute of Engineering Technologists.



Adlen Ksentini Adlen KSENTINI is an Associate Professor at the University of Rennes 1, France. He is a member of the INRIA Rennes team Dionysos. He received an M.Sc. in telecommunication and multimedia networking from the University of Versailles . He obtained his Ph.D. degree in computer science from the University of Cergy-Pontoise in 2005, with a dissertation on QoS provisioning in IEEE 802.11-based networks. His other interests include: future Internet networks, cellular networks, green networks, QoS, QoE and multimedia transmission. Dr. Ksentini is involved in several national and european projects on QoS and QoE support in Future Internet Networks. Dr. Ksentini is a co-author of over 50 technical journal and international conference papers. Dr. Ksentini has been in the technical program commitee of major IEEE ComSoc conferences, ICC/Globecom, ICME, WCNC, PIMRC. Dr. Ksentini is Vice-Chair for IEEE MMTC Interest Group on QoE.

A QoE cross layer approach to model media experiences

Andrew Perkis
NTNU, Norway
andrew@*iet.ntnu.no*

1. Introduction

A Media experience supports natural interactions between people and their environment. The media considered consists of audio and visual presentations and their interactions as well as user interactions including traditional interactivity as well as novel methods through Natural User Interfaces creating real world presence.

In order to find a measure for the user's perceived quality of the experience we need to shift from using simple Quality of Service (QoS) as a measure of the quality to the broader concept of Quality of Experience [8]. The current assumptions of QoE in the media representation and delivery community, with close links to other fields such as Psychology and social sciences, are shown in Figure 1.

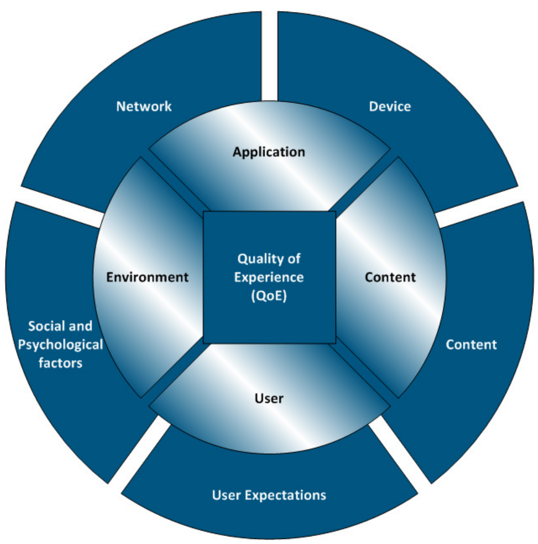


Figure 1 QoE influencing factors

A definition based on these assumptions is given in the Qualinet [4] White paper published in 2012 [1].

One of the earliest works on QoE can be found in [5] which define QoE as a measure of the impact of content on a specific user, in a specific context. This can be measured either through a subjective assessment, or estimated through a model based on the content, specific user and specific context parameters. Another early approach of modeling QoE is found in [2]. It follows a traditional methodology where the user's perception is measured by formal subjective evaluations. The results of these evaluations are considered as ground truth and used as a basis for developing highly correlated objective metrics for quality.

QoE optimisations in multimedia communications has lead into developing new digital media enabling more immersive experiences. The media considered still consists of audio and visual presentations, but now enriched by new functionality enabling interactivity. The ultimate goals are to digitally create real world presence and describe and define the work within a new field denoted Immersive Media Technology Experiences (IMTE).

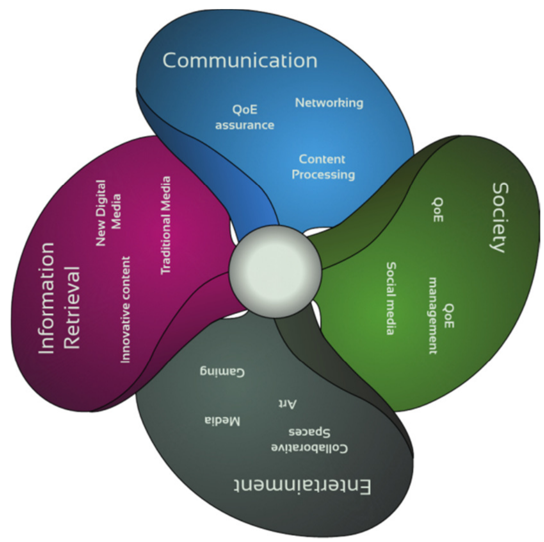


Figure 2 Cross disciplinary spiral research approach QoE

IMTE is highly cross disciplinary incorporating several disciplines including Media Technology, Information and Communication Technology, and Media Studies. In this way IMTE can encompass diverse core competencies covering fields such as communications, information retrieval, entertainment and social networks. Our approach is to use a spiral based research approach, as illustrated in Figure 2, where IMTE is the propeller advancing the field and maintaining the spin of new ideas and approaches.

In this paper, we present a new media experience model based on the spiral research approach and show how QoE is used as in the cross layer optimization between each layer in order to optimize the user's experience.

2. QoE optimization

To achieve QoE optimization through cross layer dynamics a holistic multidisciplinary cross layer effort is required. IMTE presents a holistic view of a user's experience within the media sphere following content from capture, representation, interaction to delivery and usage and finally into the business model as shown in Figure 3, where QoE is the driving force for quantifying the users experience at each stage.

To optimize QoE we need to identify the research domains for each of the five stages and their interconnections. Figure 4 shows three identified areas and their links; the User, Content and Infrastructure.

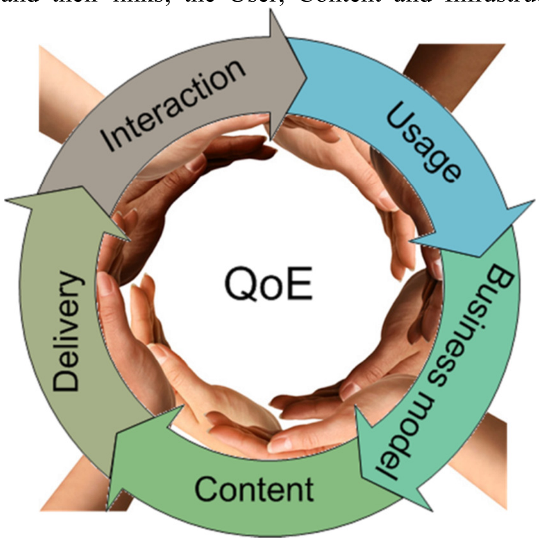


Figure 3 A Holistic view of the media experience

The glue in the linking is by the Users interacting with the Content through devices connected to the Infrastructure. Such Interactions and Devices demands New Digital Media and abilities for the Content to Adapt to User requests and the available Infrastructure (Networked Media Handling). This structure allows for piloting new advanced applications and services such as Digital Storytelling, Digital Art, Serious gaming, Presence and immersive experience (interactivity) etc. Some example devices are [3]:

- Mobile phones
- Pads (iPAD and Androids)
- Interactive tables
- Screens and larger displays
- Cinema

2. The media experience model

For the users media started with storytelling and wall drawing around the fire in the caves of early men. Media technology systems are evolved versions of this good old storytelling and wall drawing, which hopefully offer the same or more immersive and rich experience. Today multimedia is about sharing experiences (real or imaginary) with others. These experiences can be modeled as shown in the proposed media experience model in Figure 5.

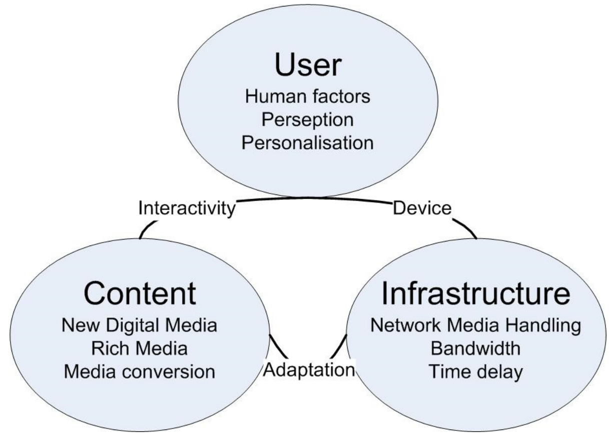


Figure 4 Research areas for IMTE

QoE is the overriding factor in the model and is seen as a tool for monitoring and managing the users experience at each of the interfaces between the model layers, providing cross layer optimization.

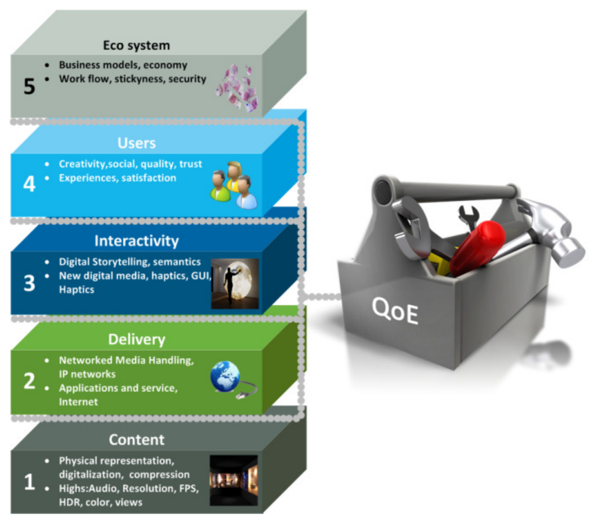


Figure 5. The media experience model.

The media experience model layers

The first layers in the model consider the physical representation and delivery of the content. Today's media content is evolving around optimal utilization of 2D media and has focused on HD (High Definition) issues of resolution, frame rates, dynamic range, color space and formats. There are numerous advances in these fields, amongst others Ultra High Definition TV

(UHDTV), High Dynamic Range (HDR) and 3D. The future looks at increasing the user's experience by moving to multi-scopic, multi-view, free viewpoint and omnidirectional. Together with the advances in audio technology all the way to auralization and 3D audio we see the possibility of offering Interactive holistic rendering of our real world to the User, with the ultimate goal being to digitally create real world presence where we can build business models and an economy based on the Eco system at the top layer. As an example of a concrete cross layer optimisation we see the interaction between the Content and Delivery layer by efforts within Networked Media Handling.

Media technology and art

In order to work on quality modeling and measurements between the layers in the model we have to work in an experimental setting and design our own novel content. The experimental dimension leads to our work bridging the technology and art where the content itself becomes an exhibition. An example of this is Chroma Space, where the experimental results are published as a scientific paper [6], while the content is exhibited as a piece of art [7].

4. Conclusion

In this paper we have proposed a new media experience model and motivated for an experimental approach for QoE modeling and assessment in order to optimize the users QoE.

References

[1] P. L. Callet, S. Möller and A. Perkis, "Qualinet White Paper on Definitions of Quality of Experience," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.1, June 3, 2012.

[2] Perkis, S. Munkeby and O. I. Hillestad, "A Model for Measuring Quality of Experience," 7th Nordic Signal Processing Symposium (NORSIG 2006), pp. 198-201, june 2006.

[3] Perkis, Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi and J. Chicaro, "Universal multimedia access from wired and wireless systems," vol. 20, no. 3, pp. 387-402, 2001.

[4] Qualinet, "COST action IC1003 – European network on Quality of Experience in multimedia systems and services," [Online]. Available: http://www.qualinet.eu/.

[5] T. Ebrahimi, "Quality of Experience - A new look into quality and its impact in future personal communications, Keynote at Workshop on Quality of Service in Mobile Multimedia Networks," YRP, Japan, February 2001., [Online]. Available: http://www.slideshare.net/touradj_ebrahimi. [Accessed 7 November 2012].

[6] Mansilla, Wendy Ann; Puig, Jordi; Perkis, Andrew; Ebrahimi, Touradj. Chroma space: affective colors in interactive 3d world. In Proceedings of the 18th ACM International Conference on Multimedia 2010 (MM 2010). ACM New York, NY, USA, Oct, 2010. Doi: 10.1145/1873951.1874239

[7] Mansilla, Wendy Ann; Puig, Jordi; Perkis, Andrew; Ebrahimi, Touradj. Chroma Space. [Exhibition catalogue]. Colorito: An Interactive Renaissance of Colours. ACM & Titivillus Mostre Editoria; Pisa, Italy, Oct, 2010.

[8] Perkis, "Quality of Experience," SPIE Newsroom, 23 January 2013. [Online]. Available: http://spie.org/x1004.xml.

Andrew Perkis received his Siv.Ing and Dr. Techn. Degrees in 1985 and 1994, respectively. In 2008 he received an executive Master of Technology Management in cooperation from NTNU, NHH and NUS (Singapore). Since 1993 he has held the position of Associate Professor at the Department of Telecommunications at NTNU and as full professor since 2003. Currently he is focusing on Multimedia Signal Processing, specifically within methods and functionality of content representation, quality assessment and its use within the media value chain in a variety of applications. He was one of the founding authors of concepts such as Universal Multimedia Access and Quality of Experience (QoE). He is also involved in setting up directions and visions for new research within media technology and entertainment as well as directions for innovations in Immersive Media Technology Experiences. He is member of The Norwegian Academy of Technological Sciences (NTVA), senior member of the IEEE, member of ACM and member of The Norwegian Society of Chartered Engineers (TEKNA). He is Vice Chair of the COST Action IC1003 (QUALINET) which gathers more than 150 researchers in a consortium related to QoE issues in multimedia systems and services.

Context Aware Quality of Experience for Audio-Visual Service Groups

M. Tourad Diallo[1], H. Moustafa[2], H.Afifi[3], N. Marechal[1]
(1) Orange Labs, France, (2) Intel Labs USA, (3) Institut Mines-Télécom France
{mamadoutourad.diallo, nicolas.marechal}@orange.com, hassnaa.moustafa@intel.com, hossam.afifi@mines-telecom.fr

1. Introduction

With the network heterogeneity and increasing demand for multimedia audio-visual services and applications, Quality of Experience (QoE) has become a crucial determinant of the success or failure of these applications and services. As there is burgeoning need to understand human hedonic and aesthetic quality requirements, QoE appears as a measure of the users' satisfaction from a service through providing an assessment of human expectations, feelings, perceptions, cognition and acceptance with respect to a particular service or application [1]. This helps network operators and service providers to know how users perceive the quality of video, audio, and image which is a prime criterion for the quality of multimedia and audio-visual applications and services. QoE is a multidimensional concept consisting of both objective (e.g., human physiological and cognitive factors) and subjective (e.g., human psychological factors) aspects. There are several methods to measure and predict QoE for multimedia in fixed and mobile networks [10,11]. There are also commercial tools as conviva, skytide and mediamelon. [4, 5 and 6].

Measuring client satisfaction is not something new within service providers business. What is new is the ability to deduce it automatically, dynamically and contextually. We define in this paper a new notion for contextual group QoE with more parameters in the user environment to accurately predict the QoE. To maximize the end-user satisfaction, and given those dynamic score evaluations it will be necessary to do some adaptation at both the application layer (e.g. choice of the compression parameters, change bitrate, choice of layers which will be send to client), the network layer (e.g. delivery means, unicast, multicast, choice of access network) and the delivery side (source server choice, CDN delivery from a cloud...) [7].

2. Context-Aware QoE

To improve the QoE and maximizing the user's satisfaction for a service, we extend the QoE notion through coupling it with different context concerning the user (preferences, content consumption style, level of interest, location, …..) and his environment including the used terminal (capacity, screen size, ..) and network (available bandwidth, delay, jitter, packet loss rate..).

Human moods, expectations, feelings and behavior could change with variation in his/her context [8]. The context-aware QoE notion presents an exact assessment of QoE with respect to contextual information of a user in a communication ecosystem. To measure user experience, context information needs to be gathered as a first step in a dynamic and real-time manner during services access [3]. This information includes: i): devices context (capacity, screen size, availability of network connectivities …), ii) network context (jitter, packet loss, available bandwidth …), iii) user context (Who is the user, his preferences, his consumption style, gender, age…), and iv) user localization. User localization considers both the physical location of the user (indoor or outdoor) and the user location within the network that we call the Geographical Position Within the Network (GPWN). For the physical localization, user's can be localized indoors (within their domestic sphere for examples) through the Wi-Fi or Bluetooth techniques [9]. For the outdoor localization of users, GPS (Global Positioning System), Radio Signal Strength Indicator (RSSI), and Cell-ID (based on the knowledge of base stations locations and coverage) are commonly used.

After context information gathering, a second step is to personalize and adapt the content and the content delivery means according to the gathered context information for improving user's satisfaction of services and for better resources consumption. Figure 1 illustrates our vision of content adaptation and Figure 2 describes the QoE evaluation process.

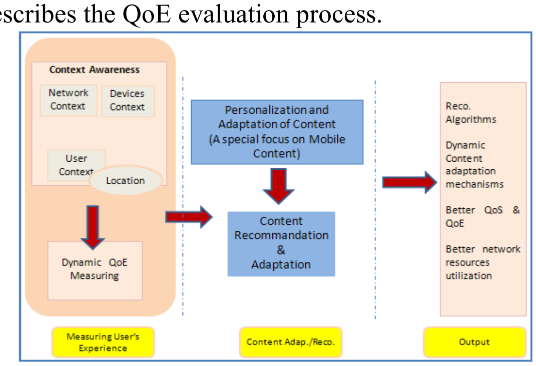


Figure 1: New vision of content adaptation

3. QoE Measuring Techniques

Internet Service Providers (ISPs) use Quality of Service (QoS) parameters such as bandwidth, delay or jitter to guarantee good service quality. QoS parameters [10] are not the only parameters affecting QoE. The challenging question is how to quantify the QoE measure. In general there are three main techniques for measuring the QoE as discussed in the following sub-sections.

- Objective QoE Measuring Techniques, based on network related parameters that need to be gathered to predict the users' satisfaction.
- Subjective QoE Measuring Techniques based on surveys, interviews and statistical sampling of users and customers to analyze their perceptions and needs with respect to the service and network quality.
- Hybrid QoE Measuring Techniques, with an objective measuring for identifying the parameters that have an impact on the perceived quality for a sample video database. Then the subjective measurement takes place through asking a panel of humans to subjectively evaluate the QoE while varying the objective parameters values. The method presented in [11] proposes a QoE measure with automatic QoE tool for SVC video coding mechanisms. The proposed module is based on PSQA (Pseudo Subjective Quality Assessment tool), which is a hybrid QoE measure technique (objective/subjective) assessment tool. PSQA uses RNN (Random Neural Network) to capture the non-linear relation between the video coding as well as the network parameters affecting the video quality, and QoE.

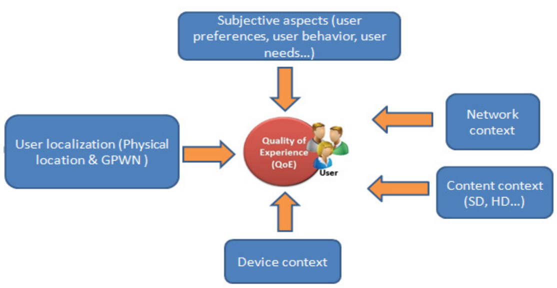


Figure 2: QoE Process & Model

4. Global QoE in the domestic sphere

If we take as example a family in the evening that is distributed in several rooms of a house with different access networks (optical and wireless). The family members watch different content on different kinds of devices ranging from high definition TVs to low resolution smartphones. If we evaluate the context of each user and calculate his best achievable contextual QoE, we will be able to know the adequate parameters for all the distribution chain. When we consider the QoE as a whole and try to optimize it, the context will be of great value to us. For example, if a user is using a low resolution device, we know that the MoS will not improve if we increase network parameters such as bandwidth (it is useless hence to give more capacity). This delta bandwidth can be beneficial to another family member for whom the MoS will increase if we transfer the bandwidth to his session. So the resulting architecture is depicted in the figure.

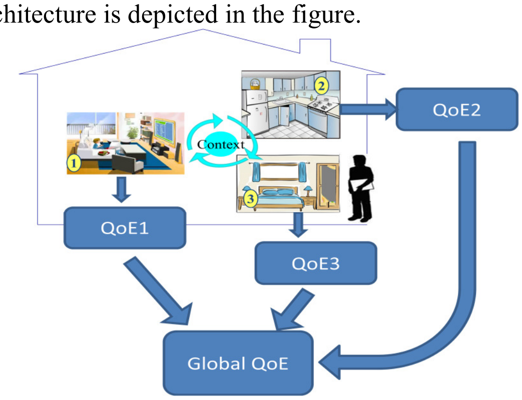


Figure 3: The Global contextual QoE

The figure does not show the back loop network control for clarity. The optimization process can be based on simple linear programming techniques. It could be also based on a simple game to find an optimal equilibrium point for bandwidth shares versus MoS. This is one of the axes that we try to deepen in our future studies. The following equation summarizes the evaluation process. We calculate the global Satisfaction (MoS) $S_{\mathrm{GlobalQoE}}$ based on single MoS satisfaction that is function of context information for user (i) Ci.

The global QoE in the domestic sphere is measured by this method:

$$S_{GlobalQoE} = \sum^{N} \frac{S_i(c_i)}{N}$$

N is the number of users and S_i is the user satisfaction. Ci is a complex function of different context parameters.

5. Conclusion and Perspectives

Quality of Experience (QoE) becomes crucial for service providers and network operators to continue gaining users' satisfaction. It needs to be analyzed in the specific context of the client and has to be compared with large numbers of customers sharing the same resource. This short paper presents a study on the QoE measuring means considering both the classical methods and the research contributions. We present a new approach for QoE with extended context information (network context, device context, user context, content context…). It aims at improving the user satisfaction. We also argue that global QoE should be used for groups sharing a resource to optimize the overall distribution parameters.

The perspectives of this work are to analyze in detail the complete distribution chain of multimedia content (device, access network, content network) to identify which parameters should be first adjusted to have an immediate influence on the QoE improvment.

References

[1] M. Diallo, H. Moustafa, H. Afifi. Quality of experience for audio-visual services. UP-TO-US '12 Workshop : User-Centric Personalized TV ubiquitOus and secUre Services, ACM, Berlin, Germany. Vol. 2. pp. 121-127. 04-06 july 2012.

[2] T.Tominaga, T.Hayashi, J.Okamoto, A.Takahashi, "Performance comparison of subjective quality assessment methods for mobile video", Quality of Multimedia Experience (QoMEX), Second International Workshop on Date of Conference: June 2010.

[3] Song Songbo, H. Moustafa, H. Afifi. Advanced IPTV Services Personalization Through Context-Aware Content Recommendation. IEEE Transactions on Multimedia. Volume: 14 , Issue: 6. December 2012.

[4] www.conviva.com

[5] www.skytide.com

[6] www.mediamelon.com

[7] M Tourad & Al. "Adaptation of Audio-Visual Contents and their Delivery Means" Accepted to appear in Communications of the ACM. 2013.

[8] V. George Mathew. (2001) Environmental Psychology. http://www.psychology4all.com/environmentalpsychology.htm.

[9] Deliverable D3 1 1 Context-Awareness- Version15 Dec 2011.pdf, Task 3.1 (T3.1): Context Definition, Gathering and Monitoring. CELTIC UP2US Project.

[10] I. Martinez-Yelmo, I.Seoane, C.Guerrero, "Fair Quality of Experience (QoE) Measurements Related with Networking Technologies", 2010.

[11] Kamal Deep Singh, Adlen Ksentini ,Baptiste Marienval, Quality of Experience measurement tool for SVC video coding, IEEE Communications, 2010.

[12] I. Politis, M. Tsagkaropoulos, T. Dagiuklas, S. Kotsopoulos and P. Stavroulakis, "On the QoS Assessment of Video Sessions in Heterogeneous 3G-WLAN Networks with Seamless and Secure Mobility Support ", China Communications Magazine, Special Issue on Communications and Information Security, vol. 4, no.1, pp. 105-119, February 2007.

Mamadou Tourad Diallo received his M.Sc. from Paris University 6 in 2011. He is currently working on his Ph.D. within Orange Labs. His thesis work concerns new video content networks and context adaptation.

Hassnaa Moustafa received her B.Sc. from Alexandria University and Ph.D. from ENST Paris. She is now with Intel Labs USA.

Hossam Afifi is professor at Institut Mines Télécom, Saclay, France. He works on multimedia services and applications.

Nicolas Marechal received his Ph.D. from Rennes University. He works as senior engineer at Orange Labs Issy les Moulineaux.

No-reference IPTV Video Quality Assessment Based on End-to-End Visual Distortion Estimation

Ning Liao and Zhibo Chen
Technicolor Research & Innovation, Beijing, China
ning.liao@technicolor.com, chenzb@ieee.org

1. Introduction
No-reference H.264 video quality assessment models for IPTV scenario and mobile streaming scenario are studied in ITU-T P.NBAMS [1] (Non-intrusive Bitstream model for the Assessment of performance of Multimedia Streaming) work group. The parsing-mode P.NBAMS models do not completely decode the H.264 video stream; any kind of analysis of the bitstream, without using pixel information, can be done. On-line video quality monitoring, e.g. at gateway or setup box, is a major target application of P.NBAMS model. Thus, both prediction accuracy and algorithm complexity are important aspects to evaluate a model. Test conditions of P.NBAMS databases [2] are designed to reflect the realistic application situations, which includes compression artifacts introduced due to lossy video compression and slicing/freezing artifacts [3] introduced due to lossy transmission and different Packet Loss Concealment (PLC) method used.

We developed a complete solution for parsing-mode P.NBAMS in IPTV scenario and Mobile video streaming scenario, which demonstrated the best performance in ITU-T P.NBAMS standard competition. In this model, the visual distortions of three types of above-mentioned artifacts are modeled separately, then the mutual influence of perceptible compression artifacts and slicing/freezing artifacts is modeled by linearly combining the output of the two worst quality levels of the individual artifact models. The idea is that the overall quality is determined mainly by the worst artifact type, regardless of specific artifact types.

In this paper, we only present our quality assessment proposal named E2EVD (End-to-End Visual Distortion) for slicing artifacts, which is more challenging to estimate compared with coding and freezing artifacts. One challenge is to evaluate the concealed artifacts without the explicit knowledge of the pixel signal of the artifacts, which is available only after decoding and applying PLC strategy to the lost macro blocks (MBs) at decoder. The effectiveness of a PLC strategy depends heavily on video content characteristics.

In our previous packet-layer model [4], we demonstrated that visibility estimation of lost frames based on video complexity can significantly improved prediction accuracy of video quality, as compared with PLR. We also demonstrated that considering error propagation effects can further improve model performance. In E2EVD proposal, the visibility level of lost packets is estimated at MB level with more accurate content features extracted from video bitstream, and scene change detection methods are proposed to improve the video quality prediction accuracy. The E2EVD scheme demonstrates statistically significantly better performance for both IPTV (i.e. High Resolution, HR) scenario and mobile video stream (i.e. Low Resolution, LR) scenario.

2. Description of E2EVD
E2EVD model is shown in Figure 1. Generally, the goal of PLC is to estimate lost MBs in order to minimize perceptual quality degradation. Visual artifacts may still be perceived after PLC, because PLC may be not effective therein. Such visual artifacts caused by lost MB are denoted as initial visible artifacts. If a block having initial visible artifacts is used as a reference, for example, for intra prediction or inter prediction, the initial visible artifacts may propagate spatially or temporally to other macro blocks in the same or other frames through prediction. The overall visible artifact of a MB is caused by initial and/or propagated visible artifacts. Finally, all MBs' visual artifacts in the sequence are aggregated and mapped to a numeric video quality index in 1-5 scale.

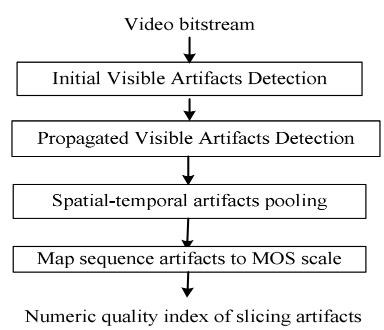


Fig. 1. Block diagram of E2EVD model.

Initial visible artifact estimation in a scene
The perceived strength of artifacts produced by transmission errors depends heavily on the employed PLC techniques. For example, if a frame far away from a current frame is used to conceal a current macro block, the concealed macro block is more likely to have visible artifacts. So, the distance $ecdist$, in

display order, between the to-be-concealed frame and the concealing frame is a parameter for modeling. In addition, the artifact strength is also related to the video content. For example, a slow moving video is easier to be concealed. Thus, parameters, such as motion vectors and error concealment distance, can be used to assess the error concealment effectiveness and the quality of concealed video at a bitstream level. In HR scenario, the initial artifacts visibility level for a lost MB indexed by (i ,j) of frame n is given by

$$LoVA_{init}(n,i,j) = f(MV_{n,i,j} * ecdist/4.0)$$

$$f(x) = \begin{cases} v_1, & ,x < T_1 \\ \frac{(v_2-v_1)}{T_2-T_1} * (x - T_1) & ,T_1 \le x \le T_2 \\ v_2 & ,x > T_2 \end{cases}$$

$v_1 = 0, v_2 = 100; T_1 = 1$ and $T_2 = 8$ in the unit of pixel.

In LR scenario, in smooth areas of some video sequences, for example, in sky and grassland which are usually easy to be concealed, unlike in HR scenario, the estimated motion vectors in H.264 encoding may be large even the movement between pictures are small. Consequently, a video quality measurement based on motion vectors may falsely estimate strong visible artifacts even though the concealed areas have good visual quality. By contrast, the energy of prediction residual signal in the smooth areas may be relatively small and may provide better indication about the perceived visual quality. Thus, residual energy $rsd_{n,i,j}$ is used as another parameter in estimating the artifact level, i.e.,

$$LoVA_{init}(n,i,j) = \min\{f\left(MV_{n,i,j} * \frac{ecdist}{4.0}\right), f\left(rsd_{n,i,j}\right)\}$$

For scaling of residual signal, $T_1 = 1$, and $T_2 = 64$.

Scene cut artifact estimation

When there is a significant scene change between two adjacent pictures and packet loss occurs in the second picture of the two adjacent pictures, the concealed second picture will have very strong visible artifacts. Scene cut artifacts occur at partially received scene cut frames or at frames referring to lost scene cut frames. The idea behind detecting scene cut frame is that the prediction residual energy change or motion change around a scene change is often greater. For the lost MBs in a detected scene cut frame, set its initial visible artifact to a larger value, i.e. 100.

Note that, when one scene changes gradually to another scene, and if packet loss occurs in a gradual transition picture, the artifacts in the error concealed picture are less visible. This is quite contrary to scene cut artifacts. However, the energy or motion difference of two gradually changed scenes may also be great. Thus, it is also important to differentiate the gradual scene change from significant scene change.

Propagated visible artifacts

How the artifact level propagates can be traced through motion vectors. Experiment shows that it is sufficient to use zero motion vectors instead of accurate motion vectors to roughly track the temporal propagation of visible artifacts. The overall artifacts in a MB should consider both initial artifacts and propagated artifacts:

$$LoVA(n,i,j) = max(LoVA(n-k,i,j),\ LoVA_{init}(n,i,j))$$

where $LoVA(n-k,i,j)$ is the propagated visible artifact from reference frame n-k.

Spatio-temporal pooling

Finally, sequence-level artifacts are aggregated from frames' artifacts LoVA(n) by a log function, and mapped to a MOS by 2nd-order polynomial fitting.

$$LoVA_{seq} = log_{10}\left(\left(\sum_n LoVA(n)\right)/F_{fps} + 1\right)$$

$$Q_s = C_1 \times LoVA_{seq}^2 + C_2 \times LoVA_{seq} + C_3$$

where F_{fps} is frame rate, the parameters C_1, C_2, C_3 are trained on selected samples that are dominated by perceptible slicing artifacts, i.e., the influence of coding artifacts on perceptual quality can almost be ignored.

3. Performance analysis

In HR case, compared with the full-reference metric MSE on five standard defined video databases, no-reference E2EVD scheme achieved better performance with an average correlation of 0.83 with subjective score, when tested on slicing samples without perceptible coding artifacts. Interestingly, the MSE considering slicing artifacts only outperforms the MSE equally counting in both coding and slicing artifacts by a large margin on some databases. This illustrates that, the overall quality is determined by the dominant visual distortion; discriminative treatment of signal difference caused by different types of artifacts may bring significant performance gain.

Using our complete solution, the average RMSE for video packet loss conditions causing slicing artifact is about 0.4 on 5 scales in HR case, and around 0.5 in LR case.

References

[1] ITU-T Q14/12, "P.NBAMS Terms of Reference," Jan. 2011.
[2] ITU-T Q14/12, "P.NAMS and P.NBAMS Test Plan," Sept. 2011.
[3] ITU-T Q14/12, "Draft recommendation P。1202", Sept. 2012.
[4] N. Liao, ZB Chen, "A Packet-Layer Video Quality Assessment Model with Spatiotemporal Complexity Estimation," EURASIP Journal on Image and Video Processing, 2011
[5] N. Liao, ZB Chen, "No-reference IPTV video quality assessment based on end-to-end visual distortion estimation", ICASSP, 2013, submitted.

Ning Liao is with Research & Innovation Dept. of Technicolor, Beijing, China. She received the B.E. degree in wireless communication and the Ph.D. degree in Telecommunication Engineering from the Beijing University of Posts & Telecommunications, China, in 1998, and in 2007 respectively. From 1998 to 2001, she was a network engineer in China Telecom. She worked on scalable video compression, error resilient video coding and transmission, and power-efficient service scheduling in WiMAX network from 2003 to 2008. Her research interests are in the areas of video quality assessment, video coding, and wireless multimedia communications.

Zhibo Chen (M'01–SM'11, chenzb@ieee.org) received his B. Sc., and Ph.D. from EE Tsinghua University. He has been with Technicolor since 2004 and worked in Sony Research before that. He is principal scientist in Technicolor Research & Innovation Department, Distinguished Fellow of Technicolor fellowship program, manager of media processing lab in Technicolor. His areas of expertise and interests include: media processing and coding, media Quality of Experience analysis and management for content delivery, perceptual based rendering, etc. He was the Media QoE lead for Technicolor Research & Innovation, contributed to the design of media QoE assessment platforms for CE applications and corresponding ITU-T standardization. He has more than 100 granted and filed EU and US patent applications, more than 60 publications and standard proposals. He is IEEE senior member, member of IEEE Visual Signal Processing and Communications Committee, and member of IEEE Multimedia Communication Committee. He was RC member of ISCAS meetings from 2007 to 2013, key member in IEEE MMTC special interest group on QoE. He was TPC member of PCS and VCIP, Chair of ICME 2011 Multimedia track. He also served as co-Editor of IEEE Journal on Selected Areas in Communications QoE-Aware Wireless Multimedia Systems 2011, member of best paper selection committee in IEEE VCIP 2012.

Video Quality as a Driver for Traffic Management with Multiple Subscriber Classes

Martín Varela and Janne Seppänen
VTT Technical Research Centre of Finland
{martin.varela, janne.seppanen}@vtt.fi

1. Introduction

Video services currently account for a very large portion of the total traffic on the Internet, and this portion is foreseen to keep rising [*1*]. This trend, coupled with the resource-hungry nature of video services, poses significant problems for network management, if good perceptual quality levels are to be achieved. In mobile networks, in particular, this can be a problem when a cell contains several users streaming video concurrently. In this paper we present a short overview of a multi-faceted mechanism for cross-layer quality-driven traffic management for video services at the last hop, which we have proposed in [*2*]. We consider over-the-top (OTT) services, where the network operator does neither control the content nor profit directly from it. Despite ever-more-efficient encoding schemes, mobile video traffic is poised to keep increasing its need for resources, as high-resolution displays appear in mobile devices, and users become accustomed to HD video on their TV and desktop/laptop systems. Since bad quality might lead to user churn, solutions in the form of access control, or *Differentiated Services*, have been explored, which may allow implementing network QoS mechanisms that result in better QoE for the end users. An immediate problem that appears in this context is that of identifying the traffic to mark as high-priority. In the case of RTP-based streams, simply looking at packet streams might be sufficient, but with the majority of OTT services being HTTP-based, the problem becomes non-trivial.

Research on quality-driven traffic management for video services has been done for IPTV (e.g. in [*3*] [*4*]), and to a lesser extent on OTT services [*5*] [*6*] in wireless contexts.

2. A multi-faceted approach

In this work, we propose a composite approach to managing the traffic in order to provide adequate QoE to the users. We propose a subscriber-based differentiation scheme (implemented, without loss of generality, with two classes of users, namely *premium* and *normal*), and a traffic management scheme based on both access control and application-based traffic differentiation.

Overall, our solution works as follows. New flows arriving at a generic *Access Point* are classified both by their subscriber class and by their application type (the latter classification is done by using the two-stage statistical classifier described in [*7*]). Regardless of subscriber class, inelastic flows are only admitted if the average Mean Opinion Score (MOS) of other video streams is above a set threshold (note that premium users cannot preempt normal users, and so a premium user's stream might be dropped even if normal users are currently streaming video). If a flow is admitted, then it is assigned to a queue with the adequate priority based on its application type, subscriber class, and current estimated QoE. The objective of this process is to ensure that a) admitted video streams provide acceptable quality, b) premium users' streams achieve better quality when congestion arises, and c) the system is fair to normal users as well (not preempting them, and interleaving the priority of premium and normal users' application classes).

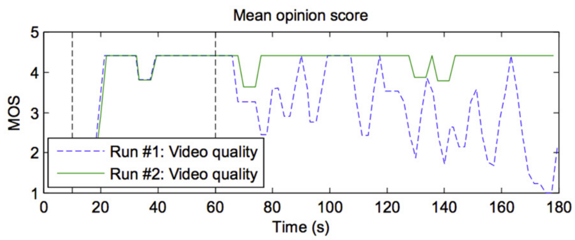
All flows enter a single FIFO queue, and individual flows are promoted on an as-needed basis to higher priority queues depending on their current quality, subscriber class, and application class. In particular, a threshold of 3.0 in the usual 5-point MOS DC scale is set, so that actions are taken when a flow's quality estimation drops below this value. A hysteresis mechanism is then implemented, ensuring that the improved quality achieved by promoting the flow is stable for a set period of time over a second threshold (4.0 points in the case of the reported results) before returning the flow to a lower-priority class (if possible). Queues of higher priority are emptied before those of lower priority (up to a certain limit), and traffic within each queue is handled with stochastic fairness queuing [*8*].

The quality estimations were performed using a PSQA [*9*] based model for IPTV-like services. The traffic management was implemented on top of a Linux-based router, by using the Hierarchical Token Bucket (HTB) queuing discipline [*10*]. HTB is rather complex, but provides a very flexible approach to handling different traffic classes. The system uses both the priority of a class and a set limit for each class to achieve fairness. Within each class' queue, stochastic fairness queuing (as implemented in Linux [*11*]) is used.

3. Performance evaluation

The performance of the proposed approach was tested as a proof-of-concept in a laboratory environment. Four different aspects of the proposed system's performance were tested, namely 1) responsiveness in the case of congestion (application differentiation), 2) subscriber priority handling, 3) reaction times, and 4) admission control. All tests were done between 15 and 40 times, and the results presented herein are representative of the average behavior of the system.

In the first set of tests, a test video stream was subjected to contention by a large bulk transfer. Figure 1 shows the results of the first test set.



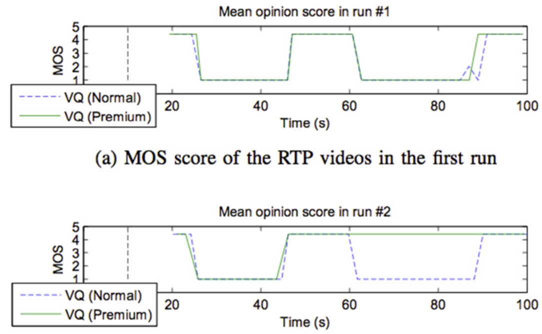
MOS score of the RTP video in both runs

Figure 1 - Application differentiation test

Run #1 was performed without the application-based priority handling, whereas run #2 was performed with it. The dashed vertical lines indicate the start of the video and bulk streams, respectively at 10s and 60s. It is clear from Figure 1 that with the traffic control off, the video quality quickly turns unacceptable, while when it is on, the quality remains at acceptable levels.

The second set of tests involved two video streams belonging to different subscriber classes. The link bandwidth was set so that one flow could be served without problems, but two flows would congest it. **Figure** 2 shows the results obtained. In the first run, the quality of both streams suffers, as expected, since the link cannot support both at their peak rates (Figure 2a). **Figure** 2b shows the effect of the subscriber class differentiation at work, and it is easy to see that the premium user enjoys a significantly better quality than the normal user. The reader may notice that there is a period in which the premium user will also suffer from a lower quality in a first instance in run #2. This is due to a trade-off between the size of the time window over which the MOS is estimated, and the estimation's accuracy. In practical use, a smaller window would probably be useful to avoid the user stopping the streaming due to the lowered quality.

This leads us to the thirds test set, regarding the reaction times of the system. The fastest performance achieved resulted in flows being promoted to a higher-priority class in 2.8s on average over 40 test runs (recall that all flows start out in the same FIFO queue by default if there's no contention). The total reaction time was of 4.0s. As mentioned before, however, these smaller values impose a trade-off in the QoS calculations, which become noisier as a consequence of having fewer samples.



(a) MOS score of the RTP videos in the first run

(b) MOS score of the RTP videos in the second run

Figure 2 - Subscriber differentiation test

The final set of tests was related to the admission control. The QoE of the flows in the system was averaged over a 30s sliding window, in order to avoid noise in the measurements. In this setup, two streams were started at different times, and the link bandwidth was set low, so that the quality of the first stream was acceptable, but not good.

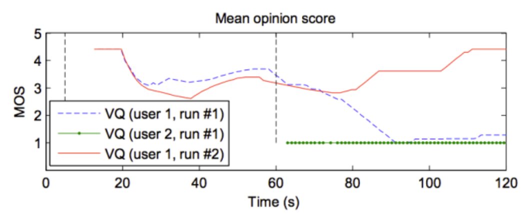


Figure 3 – Admission control test

In Figure 3, we can see that in run #1, when the admission control is not enabled, the start of the second stream results in a completely unacceptable quality for both streams. Note that actual degradation is sharper than it appears in the plot, as the plot is smoothed by the 30s averaging window. In run #2, with admission control enabled, when the second flow starts it is immediately dropped, as the quality of the first flow is below the activation threshold. Thus, the user watching that stream attains an acceptable quality throughout the whole period.

4. Conclusions

We have proposed a multi-faceted approach to QoE-based traffic control by considering different subscriber and application types and using them to perform admission control and traffic differentiation. The results obtained show a clear QoE improvement for OTT video streams when the proposed mechanisms are in place instead of a simple best-effort policy. Further work on this subject includes the extension and refinement of the traffic classification mechanism used to work on adaptive HTTP-based video streaming schemes, as well as the development of suitable parametric QoE models for them.

References

[1] Cisco Systems, Inc., "Cisco Visual Networking Index: Forecast and Methodology, 2011-2016," May 2012.

[2] Janne Seppänen and Martín Varela, "QoE-Driven Network Management for Real-Time Over-the-Top Multimedia Services," in *IEEE WCNC*, Shanghai, 2013, Accepted for publication.

[3] D. Waring, G. Lapiotis, J. Lyles, and R. Vaidyanathan K. Kerpez, "IPTV Service Assurance," *IEEE Communications Magazine*, vol. 44, no. 9, pp. 166-172, September 2006.

[4] M. Garcia, M. Atenas, and A. Canovas J. Lloret, "A QoE management system to improve the IPTV network," *International Journal of Communication Systems*, vol. 24, no. 1, pp. 118-138, 2011.

[5] K. Haugene and A. Jacobsen, "Network based QoE optimization for ”over the top” services," *Norwegian University of Science and Technology*, Department of Telematics, 2011.

[6] A. Ksentini, C. Viho, and J.-M. Bonnin K. Piamrat, "QoE-aware admission control for multimedia applications in IEEE 802.11 wireless networks," in *IEEE VTS - Vehicular Technology Conference (Fall)*, 2008, pp. 1-5.

[7] M. Hirvonen, "Two-phased network traffic classification method for quality of service management," University of Oulu, Dpt. of Electrical and Information Engineering, 2009.

[8] P. E. McKenney, "Stochastic Fairness Queueing," in *IEEE INFOCOM'90*, vol. 2, 1990, pp. 733-740.

[9] Martín Varela, "Pseudo-subjective Quality Assessment of Multimedia Streams and its Applications in Control," *Université de Rennes 1*, Doctoral Thesis Nov. 2005.

[10] M. Devera. (2003) HTB Home. [Online]. http://luxik.cdi.cz/~devik/qos/htb/

[11] B. Hubert. Linux Advanced Routing & Traffic Control HOWTO. [Online]. http://lartc.org/lartc.html#LARTC.SFQ

Dr. Martín Varela received his PhD and MSc from the University of Rennes 1 (Rennes, France), in 2005 and 2002 respectively. He has been an ERCIM fellow, and spent time at SICS and VTT, where he has been a Senior Scientist since 2007. His research interests lie in the QoE domain, both in QoE for real-time multimedia services and for generic services.

Janne Seppänen received his Bachelor's Degree in University of Oulu in 2010 and his Master's Degree two years later. In his studies, he specialized in computer engineering and signal processing. He currently works at Technical Research Centre of Finland (VTT) as a research scientist, covering topics such as Quality of Experience, network management, traffic classification and network traffic measurement.

Cross-layer Design for Quality-Driven Multi-user Multimedia Transmission in Mobile Networks

Maria G. Martini
Kingston University London
{m.martini}@kingston.ac.uk

1. Introduction

A critical problem in next generation wireless multimedia networks is how to efficiently ensure good quality video streaming over a multiple access wireless channel with shared communications resources. The main aim of achieving a satisfactory Quality of Experience (QoE) for the users of the system can be afforded at different layers of the protocol stack.

Dynamic rate control strategies optimized across the users can be considered at the application layer in order to allocate the available resources according to users' requirements and transmission conditions. Rate control was originally adopted with the goal of achieving a constant bit-rate, then with the goal to adapt the source data to the available bandwidth [1]. Dynamic adaptation to variable channel and network conditions, i.e., by exploiting the time-varying information about lower layers, can be adopted.

Packet scheduling schemes across multiple users can be considered below in the protocol stack (MAC layer) in order to adapt each stream to the available resources [2]. Content-aware scheduling can also be considered, as in [3].

At the physical layer, adaptive modulation and coding (AMC) can be exploited to improve the system performance, by adapting the relevant parameters to both the channel and the source characteristics.

2. Cross-layer design strategies

Cross-layer design (CLD) solutions should be investigated in order to optimize the global system based on a Quality of Experience criterion. As an example, in [4] a cross layer design approach is considered with multiuser diversity which explores source and channel heterogeneity for different users.

Typically cross-layer design is performed by jointly designing two layers [5] - [9]. In [9] cross layer design takes the form of a network-aware joint source and channel coding approach, where source coding - at the application layer - and channel coding and modulation - at the physical layer - are jointly designed by taking the impact of the network into account. In [7] cross-layer optimization also involves two layers, application layer and MAC layer of a radio communications system. The proposed model for the MAC layer is suitable for a transmitter without instantaneous channel state information (CSI). A way of reducing the amount of exchanged control information is considered, by emulating the layer behavior in the optimizer based on a few model parameters to be exchanged. The parameters of the model are determined at the corresponding layer and only these model parameters are transmitted as control information to the optimizer. The latter can tune the model to investigate several layer states without the need of exchanging further control information with the layer. A significant reduction of control information to be transmitted from a layer to the optimizer is achieved, at the expense of the control information from the optimizer to the layers that might slightly increase.

The work in [6] includes in the analysis MAC-PHY and APP layers, presenting as an example a MAC/application-layer optimization strategy for video transmission over 802.11a wireless LANs based on classification

3. Cross-layer signaling

Very few contributions consider jointly all the layers of the protocol stack. Furthermore, most of the cross-layer approaches presented in the literature do not address signaling across layers, necessary to pass the needed side information and control messages among layers even of different network devices. Some mechanisms are in use, for instance for the exchange of information about resource reservation or prioritisation among the different system layers, such as those proposed by IETF for QoS provisioning, namely differentiated services (DiffServ) and integrated services (IntServ). Their aim was to allow an application to reserve resources or a specific service level from the interconnecting IP network by mapping the user requirements at network protocol level.

Another example of inter-layer signaling can be found in the IEEE 802.11e standard where the QoS provisioning is performed between the application and the medium access layers. The QoS information consisting of the priorities of IP packets, to drop them selectively, is not sufficient however as an optimization method for multimedia transmission. More detailed cross-layer information needs to be delivered in order to fully optimize the end-to-end transmission.

In [11] the authors focus on cross-layer feedback, i.e., making information from one layer available to another

layer of the stack. They highlight the need for a cross-layer feedback architecture and identify key design goals for such architecture.
A review of existing relevant solutions can be found in [10]. One solution proposed for transferring the required controlling information is to extend the current protocols such as Internet Protocol version 6 (IPv6) or Internet Control Message Protocol version 6 (ICMPv6) through the definition of new options and message types, respectively.
This concept of transmitting cross-layer information can be referred as "network transparency" [9]; this includes the abstract idea of making the underlying network infrastructure almost invisible to all the entities involved in the joint optimization. The mentioned transparency solutions are potential candidates for transferring the control information through both wired and wireless network but they do not solve fully the problem of transferring control information through the protocol layers from application to physical layer and vice versa. Furthermore, they do not propose solutions to use this information for end-to-end optimization, which requires taking into account all protocol layers and particularly applications. In addition, the QoS information consisting of the IP packet priorities alone is not sufficient for delivering optimization information between the layers of source and destination devices. Hence more detailed information needs to be delivered in order to fully optimize the end-to-end QoS of multimedia transmission systems across different system layers. Besides efforts in research, standardization is also needed in the area and very recently standardization groups started addressing this issue.

The CONCERTO and the OPTIMIX European address(ed) cross-layer design strategies, cross-information to be exchanged and the strategies to pass such information among the layers in mobile networks. In order to control the system parameters based on the observed data, two controller units were proposed in the OPTIMIX project: one at the application layer (APP) and one at the base station (BSC) to control lower layers parameters [14] and in particular resource allocation among the different users based on the (aggregated) multiple feedback. The two controllers operate at different time scales, since more frequent updates are possible at the base station controller, and rely on different sets of observed parameters. The goal of the proposed system is to provide a satisfactory quality of experience to video users, hence video quality is the major target and evaluation criteria, not neglecting efficient bandwidth use, real time constraints, robustness and backward compatibility.

This system has been implemented in a realistic simulation platform based on OMNET++ to test its performance: while most cross-layer methodologies do not realistically consider the impact on the performance of the signaling overhead and complexity, our approach allows performance evaluation and comparison of cross layer-methodologies in a realistic environment.

4. Quality assessment and utility design

Quality assessment of the received multimedia information is crucial for two main purposes:

- Final system performance evaluation;
- "On the fly" system adaptation.

In the first case, the goal is to assess the final quality reflecting the subjective quality experienced by the users (through subjective tests or objective metrics well matching subjective results). In the second case, while matching subjective results is also of importance, the main requirements are the possibility to calculate the video quality metric in real-time and without reference to the original transmitted signal. An example of real-time reduced-reference metrics is [15].
Utility design is also a key issue in this framework: when transmitting multimedia signals to multiple users, the trade-off between resource utilization and fairness among users has to be addressed. In [13] this was addressed by focusing on fairness, targeting at the maximization of the minimum weighted quality among the different users. In [12] we addressed quality fairness by relying on the Nash bargaining solution.

5. Conclusion

In this paper we have briefly presented the main aspects of quality-driven design for multimedia transmission over mobile networks, highlighting design issues and open points. Some recent results have been also presented. For further recent works on the topic, the reader may also refer to [16].

Acknowledgement

The author acknowledges EU FP7 for the support provided in the framework of the projects OPTIMIX and CONCERTO. The colleagues involved in the projects are also acknowledged.

References

[1] M. Chen and A. Zakhor, "Rate control for streaming video over wireless," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 32–41, 2005.

[2] D. Jurca and P. Frossard, “Video packet selection and scheduling for multipath streaming,” *IEEE Trans. on Multimedia*, vol. 9, no. 3, pp. 629–641, 2007.

[3] P. V. Pahalawatta, R. Berry, T. N. Pappas, and A. K. Katsaggelos, “Content-aware resource allocation and packet scheduling for video transmission over wireless networks,” *IEEE JSAC*, vol. 25, no. 4, pp. 749–759, 2007.

[4] G.-M. Su, Z. Han, and K. Liu, “Multiuser cross-layer resource allocation for video transmission over wireless networks,” *IEEE Network*, March/April 2006.

[5] G. Dimic, N. D. Sidiropoulos, and R. Zhang, “Medium Access Control–Physical Cross-Layer Design”, *IEEE Signal Processing Magazine*, Sept. 2004.

[6] M. van der Schaar, D. S. Turaga and R. Wong, “Classification-Based System For Cross-Layer Optimized Wireless Video Transmission”, *IEEE Trans. on Multimedia*, Vol. 8, No. 5, October 2006.

[7] A. Saul, S. Khan, G. Auer, W. Kellerer and E. Steinbach, “Cross-Layer Optimization Using Model-Based Parameter Exchange”, *IEEE Intl. Conference on Communications (ICC 2007),* Glasgow, UK. June 24-28, 2007.

[8] Q. Liu, X. Wang, and G. B. Giannakis, “A Cross-Layer Scheduling Algorithm With QoS Support in Wireless Networks”, *IEEE Trans. on Veh. Technology*, Vol. 55, no. 3, pp. 839-847. 2007.

[9] M. G. Martini, M. Mazzotti, C. Lamy-Bergot, J. Huusko, and P. Amon, “Content adaptive network aware joint optimization of wireless video transmission,” *IEEE Communications Magazine*, vol. 45, no. 1, pp. 84-90, 2007.

[10] Q. Wang and M. A. Abu-Rgheff, “Cross-Layer Signalling for Next-Generation Wireless Systems”, *Proc. IEEE WCNC 2003*.

[11] V. T. Raisinghani and S. Iyer, “Cross-Layer Feedback Architecture for Mobile Device Protocol Stacks”, *IEEE Comms. Magazine*, Jan. 2006.

[12] N. Khan, M. G. Martini, and Z. Bharucha, "Quality-aware Fair Downlink scheduling for scalable video transmission over LTE systems*", IEEE SPAWC 2012*, Cesme, Turkey, June 2012.

[13] M.G. Martini and V. Tralli, “Video quality based adaptive wireless video streaming to multiple users”, *IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting*, 2008.

[14] C. Lamy–Bergot, M. G. Martini, P. Hammes, P. Amon, J. Vehkaperä, G. Panza, L. Hanzo, M. Chiani, G. Jeney, “Optimisation of Multimedia over wireless IP links via X-layer design”, *EUMOB 2008.*

[15] C. Hewage and M. G. Martini. "Edge-Based Reduced-Reference Quality Metric for 3-D Video Compression and Transmission." *IEEE Journal of Selected Topics in Signal Processing, vol.* 6, no. 5, pp. 471-482, 2012.

[16] M. G. Martini, C. W. Chen, Z. Chen, T. Dagiuklas, L. Sun, and X. Zhu, “Guest Editorial - QoE-Aware Wireless Multimedia Systems”, *IEEE Journal on Selected Areas in Communications,* vol. 30, no. 7, pp. 1153-1156, 2012.

[17] J. Huusko, J. Vehkapera, P. Amon, C. Lamy-Bergot, G. Panza, J. Peltola, and M.G. Martini, “Cross-layer architecture for scalable video transmission in wireless network”, *Signal Processing: Image Communication*, vol. 22, no. 3, pp. 317—330, 2007.

Maria G. Martini Maria G. Martini is a Reader (Assoc. Prof.) in the Faculty of Science, Engineering and Computing in Kingston University, London. She received the *Laurea* in electronic engineering (summa cum laude) from the University of Perugia (Italy) in 1998 and the Ph.D. in Electronics and Computer Science from the University of Bologna (Italy) in 2002. She has led the Wireless Multimedia Networking research group in Kingston University in a number of national and international research projects, funded by the European Commission, UK research councils and international industries.

An IEEE Senior Member, she has served as editor and reviewer for international journals (recently lead guest editor for the IEEE JSAC special issue on "QoE-aware wireless multimedia systems" (2012)) and she has been in the organising and programme committee of several international conferences (e.g., general chair of ICST/ACM MOBIMEDIA 2009 and EUMOB 2008, organizer of IEEE Streamcomm (ICME 2011) and of the First International Workshop on Cross-Layer Operation Aided Multimedia Streaming (IEEE VTC 2011)).

Her research interests include QoE-aware wireless multimedia networks, cross-layer design, joint source and channel coding, 2D/3D error resilient video, 2D/3D video quality assessment, decision theory, and medical applications. She has published extensively and she is the inventor of several patents on wireless video.

INDUSTRIAL COLUMN: SPEICAL ISSUE ON "DYNAMIC ADAPTIVE STREAMING OVER HTTP"

Guest Editors: Alex Giladi, Zhenyu Wu, Futurewei Technologies, U.S.A and Guosen Yue, NEC Laboratories America

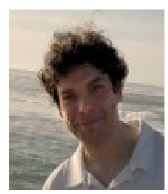
Several market trends and technology developments have resulted in the emergence of "over-the-top" (OTT) streaming, which utilizes the Internet as a delivery medium. Hardware capabilities have evolved enough to create a wide range of video-capable devices, ranging from mobile devices to connected TVs, while broadband penetration made high-quality Internet streaming viable.

As opposed to the traditional "closed" networks, which are completely controlled by the multi-system operator (MSO), Internet is a "best effort" environment, where bandwidth and latency are constantly changing. In particular, network conditions are highly volatile in mobile networks. Such volatility makes dynamic adaptation to network changes a necessity in order to provide a tolerable user experience.

Adaptive streaming has become synonymous with HTTP streaming. At first glance, HTTP does not seem a good fit for a video transport protocol. After all, real-time UDP-based streaming has been widely used for more than a decade. However, ubiquity and scalability of HTTP infrastructure, makes use of HTTP for Internet video streaming significantly more attractive and more scalable. Firewall penetration is yet another factor increasing the attractiveness of HTTP streaming. Among others, these factors made HTTP streaming the technology of choice for rate-adaptive streaming. The ubiquity of HTTP infrastructure made HTTP streaming a technology of choice for multiplatform and multi-screen applications even in operator-owned non-OTT scenarios.

Several proprietary technologies, such as Apple HTTP Live Streaming, Microsoft Smooth Streaming, have emerged as popular adaptive streaming solutions. MPEG DASH is a newcomer in this family. It is versatile and interoperable standard, developed in MPEG and 3GPP with the participation of most major players in the market and drawing extensively on the experience with existing technologies. DASH is backed by a large amount of vendors, and adopted by multiple standardization organizations and consortia. In parallel, it has been steadily gaining attention in the academia.

This special issue of e-letter starts with a brief technical introduction to DASH. The second paper by O. Oyman describes experimentation with use of DASH over LTE networks. As DASH does not define adaptation logic, this naturally is one of the most active research areas, and the last three papers cover different aspects of it. The paper by C. Mueller et. al. addresses the issue of fairness in bandwidth allocation -- a problem inherent in a client-driven adaptation scheme. The paper by S. Zhang et al. describes a different shifting paradigm, where both media bit rate and quality are taken into account. Lastly, the paper by Y. Reznik adds a completely new dimension by using information available to a mobile device via its sensors to optimize the rate adaptation logic



Alex Giladi (M'06) is a senior architect at FutureWei Technologies (Bridgewater NJ), where he works on adaptive streaming ecosystem and its standardization. Previously he worked at Avail-TVN, Digital Fountain, and Harmonic Lightwaves. He is actively involved in MPEG DASH standardization since 2010, currently serving as the editor of two MPEG standards, Part 4 of MPEG DASH and Common Encryption for MPEG-2 TS. Alex holds a MSEE degree from Stanford University and BSc degree from the Technion, Haifa.

Zhenyu Wu (S'99 – M'05) is a senior researcher and project manager from Media Labs, Futurewei Technologies, Bridgewater, NJ. Prior to that, he was with Thomson Corporate Research, Princeton, NJ from 2005 to 2010, and a consultant for Siemens Corporate Research, Princeton, NJ in the summers of 2002 and 2003. He has been working in the general areas of video/image coding, processing and media delivery, and has participated in multiple related standards. He is author or co-author of over 30 technical publications. He received Ph.D degree in electrical engineering from University of Arizona in 2005, and B.S. and M.S. degrees from Shanghai

University, Shanghai, China in 1996 and 1998 respectively.

Guosen Yue (S'03-M'04-SM'09) received the B.S. degree in physics and the M.S. degree in electrical engineering from Nanjing University, Nanjing, China in 1994 and 1997, respectively, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, in 2004. Since August 2004, he has been a Staff Member with the Mobile Communications and Networking Research Department, NEC Laboratories America, Princeton, New Jersey, conducting research for broadband wireless systems and mobile networks. His research interests are in the general areas of wireless communications and signal processing. Dr. Yue now serves as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has served as the Associate Editor for RESEARCH LETTERS IN COMMUNICATIONS, the Guest Editors for EURASIP JOURNAL OF WIRELESS COMMUNICATION AND NETWORKING special issue on interference management and ELSEVIER PHYCOM special issue on signal processing and coding. He was the Symposium Co-Chair of IEEE ICC 2010, the Track Co-Chair of IEEE ICCCN 2008, the steering committee member of IEEE RWS 2009, and the subcommittee chair of IEEE RWS 2008. He is a senior member of the IEEE.

MPEG DASH: A Brief Introduction

Alex Giladi FutureWei Technologies
400 Crossing Blvd., Bridgewater NJ, 08807
Email: alex.giladi@huawei.com

Abstract

The Dynamic Adaptive Streaming over HTTP (DASH) specification, ISO/IEC 23009-1:2012, is the newest addition to the growing number of adaptive HTTP streaming systems. Openness, feature richness and efficiency are what sets DASH apart, and make migration from first generation proprietary adaptive systems viable and valuable.

In this letter, we provide a consise technical overview of the DASH standard and its emerging extensions. An excellent review of the standard and its background is provided in [11], while an introduction specific to MPEG-2 TS is provided in [12]. An in-depth review of recommendeded implementation practices is provided by MPEG in [2].

1 Introduction

DASH [1] defines a manifest format, *Media Presentation Description* (MPD), and segment formats for ISO Base Media File Format (ISO-BMFF) [7] and MPEG-2 Transport Streams [6].

A *segment* is the minimal individually addressable unit of data: it is the entity that can be downloaded using URLs advertised via the MPD. One example of a media segment is a 4-second part of a live broadcast, which starts at playout time 0:42:38 , ends at 0:42:42, and is available within a 3-min time window. Another one is a complete on-demand movie, which is available for the whole period this movie is licensed.

A *representation* is one of the core concepts of DASH. It is defined as a single encoded version of the complete asset, or of a subset of its components. A typical representation would be e.g. ISO-BMFF containing unmultiplexed 2.5 Mbps 720p AVC video, and separate ISO-BMFF representations for 96 Kbps MPEG-4 AAC audio in different languages. This is the structure recommended in DASH264 [10]. Conversely, a single transport stream containing video, audio and subtitles can be a single multiplexed representation. A combined structure is possible: video and English audio may be a single multiplexed representation, while Spanish and Chinese audio tracks are separate unmultiplexed representations.

MPD is an XML document, which advertises the available media and provides information needed by the client in order to select a representation, make adaptation decisions, and retrieve segments from the network. MPD is completely independent of segment, and only signals the properties needed to determine whether a representation can be successfully played and its functional properties (e.g., whether segments start at random access points). MPD uses a hierarchical data model to describe the complete presentation.

2 MPD

Representations are the lowest conceptual level of the hierarchical data model. At this level, MPD signals information such as bandwidth and codecs required for successful presentation, as well as ways of constructing URLs for accessing segments. Additional information can be provided at this level, starting from trick mode and random access information to layer and view information for scalable and multiview codecs to generic schemes which should be supported by a client wishing to play a given representation (the latter added in [5]).

DASH provides a very rich and flexible URL construction functionality. As opposed to a single monolithic per-segment URL list (also possible in DASH), it allows dynamic construction of URLs, by combining parts of the URL (*base URLs*) that appear at different levels of the hierarchy. As multiple base URLs can be used, segments can be requested from more than one location. This way, DASH allows path diversity, improving performance and fault tolerance.

If short segments are used, an explicit list of URLs and byte ranges can reach thousands of elements per representation. This is inefficient and wasteful, especially in case there is a larger amount of representations. DASH allows using predefined variables (such as segment number, segment time, etc.) and printf-style syntax for on-the-fly construction of URLs using *templates*. Instead of listing all segments (e.g., seg_00001.ts, seg_00002.ts, ... , seg_03600.ts), its enough to write a single line, seg_$Index%05$.ts, to express any number of segments, even if they cannot be retrieved at the time the MPD is fetched. Due efficiency of templates, DASH264 [10] multi-segment representations are required to use templates.

Different representations of the same asset (or same component, in the un-multiplexed case) are grouped into *adaptation sets*. All representations within an adaptation set will render the same content, and a client can switch between them, if it wishes to do so.

An example of an adaptation set would be a collection of 10 representations with video encoded in different bitrates and resolutions. It is possible to switch between each one of these at a segment (or even a subsegment) granularity, while presenting same content to the viewer. Under some segment-level restrictions and time alignment, seamless representation switch is possible.

A *period*, is a time-limited subset of the presentation. All adaptation sets are only valid within the period, and there is no guarantee that adaptation sets in different periods will contain similar representations (in terms of codecs, bitrates, etc.). An MPD may contain a single period for the whole duration of the asset. It is possible to use periods for ad markup, where separate periods are dedicated to parts of the asset itself and to each advertisement.

DASH uses a simplified version of XLink in order to allow loading parts of the MPD (e.g., periods) in real time from a remote location. A simple use case for this can be ad insertion, when precise timing of ad breaks is known ahead of time, whereas ad servers determine the exact ad in real time.

A *dynamic* MPD can change and will be periodically reloaded by the client, while a *static* MPD is valid for the whole presentation. Static MPD's are a good fit for VoD applications, whereas dynamic MPD's are used for live and PVR applications.

3 Segments

3.1 Media segments

Media segments are time-bounded parts of a representation, and approximate segment durations appear in the MPD. Segment duration does not have to be the same for all segments, though in practice segment durations will probably be close to constant (e.g., DASH264 [10] uses segments with durations within a 25% tolerance margin).

MPD can contain information regarding media segments that are unavailable at the time it is read by the client, and a client needs to calculate when a media segment will be available.

3.2 Index segments

Another segment type of major importance is the *index segment*. Index segments may appear either as side files, or within the media segments, and contain timing and random access information. Indexes make efficient implementation of random access and trick modes, but the concept is useful beyond that â€" index segments can be used for more efficient bitstream switching. While indispensible for VoD and PVR type of applications, indexing less useful in live cases.

3.3 Bitstream switching

Several segment-level and representation-level properties are necessary to implement efficient bitstream switching. DASH provides explicit functional requirements for these, which are expressed in the MPD in a format-independent way. Each segment format specification has to contain the format-level restrictions that correspond to these generic requirements.

Let us denote media segment i of representation R as $S_R(i)$, its duration as $D(S_R(i))$. Furthermore, let its earliest presentation time be $EPT(S_R(i))$. EPT corresponds to the earliest presentation time of the segment, rather than the time is not the time at which a segment can be successfully played out at random access.

The key to efficient switching is *time alingment* of segments for all representations within an adaptation set. This translates intor a requirement that for any pair of representation R_a and R_b and segment i, $EPT(S_{R_a}(i)) < EPT(S_{R_b}(i-1)) + D(S_{R_b}(i-1))$. This, combined with the requirement that a segment starts with a random access point of certain types, ensures the ability to switch at segment border w/o need for overlapped downloads and dual decoding. The full definition of random access points is described extensively in [7].

When indexing is used, it is possible to do bitstream switching at a subsegment level as well, if similar requirements hold for subsegments.

Most systems require time alignment and random access point placement restrictions. In terms of video encoding, these restrictions typically translate into encodings with matching IDR frames at segment borders and closed GOP's.

4 System model

A DASH client conceptually consists of an access client, which is an HTTP client, a media engine, which decodes and presents media provided to it, and an application, to which the access client passes events. The only interfaces defined are the on-the-wire formats of the MPD and segments, the rest is left to the implementers' disgression.

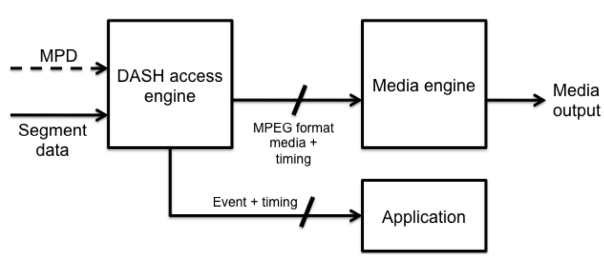


Figure 1: DASH system model, as defined in [4]

Timing behavior of a DASH client is slightly more complex than that of earlier technologies. While in Apple HLS all segments mentioned in a manifest are valid, and a client is always polling for new manifests, DASH MPD reduces the polling behavior by defining

MPD update frequency and allowing explicit calculation of segment availability.

A static MPD is always valid, whereas a dynamic MPD is valid from the time it was fetched by the client, for an explicitly stated refresh period. An MPD also has a notion of versioning – it may explicitly expose its publication time.

MPD provides the easy means for calculating availability time of the earliest segment of a period, $T_A(0)$. Media segment n is available starting from time $T_A(n) = T_A(0) + \sum_{i=0}^{n-1} D(S_R(i))$, for the duration of the timeshift buffer $T_t s$, the latter being explicitly stated in the MPD. Availability window size has a direct impact on the catch-up TV functionality of a DASH deployment. Segment availability time can be relied upon by the access client as long as it falls within the MPD validity period.

For any representation R MPD declares bandwidth B_R. MPD also defines a global minimum buffering time, BT_{min}. An access client will be able to pass a segment to the media engine after $B_R \times BT_{min}$ bits were downloaded, thus, given a segment starts with a random access point, the earliest time segment n can be passed to the media engine is $T_A(n) + T_d(n) + BT_{min}$, where $T_d(n)$ stands for the download time of segment n. In order to minimize the delay, a DASH client may want to start the playout immediately, however MPD may propose a presentation delay (as an offset from $T_A(n)$) in order to ensure synchronization between different clients. Note that tight synchronization of segment HTTP GET requests may create a thundering herd effect, severely taxing the infrastructure.

MPD validity and segment availability are calculated using absolute (i.e., wall-clock) time. Media time is expressed within the segments themselves, and in the live case drift can develop between the encoder and client clocks. This is addressed at the container level, where both MPEG-2 TS and ISO-BMFF provide synchronization functionality.

The definitions above are somewhat simplified, and an excellent in-depth overview of DASH timing behavior is provided in [13].

5 Events

Events [4] are a very recent extension to DASH, added in Amendment 1 [4]. As HTTP is stateless and client-driven, "push"-style events can be emulated using frequent polls. In current ad insertion practice in cable/IPTV systems, upcoming ad breaks are signaled 3-8 sec. before their start. Thus a straightforward poll-based implementation would be inefficient, and events were designed to address such use cases.

Events are "blobs" with explicit time and duration information and application-specific payloads. Inband events are small message boxes appearing at the beginning of media segments, while MPD events are a period-level list of timed elements. DASH defines an MPD validity expiration event [4], which identifies the earliest MPD version valid after a given presentation time.

Events are a powerful new tool, which is unique to DASH. New event types and schemes are actively explored now, and we would expect some amount of upcoming standardization activity in this area.

6 Content Protection

DASH is agnostic to digital rights management (DRM), and supports signaling DRM scheme and its properties within the MPD. A DRM scheme can be signaled via the ContentProtection descriptor, and an opaque value can be passed within it. In order to signal a DRM scheme, it is enough to have a unique identifier for a given scheme and define the meaning of the opaque value (or use a scheme-specific namespace instead).

MPEG developed two content protection standards, Common Encryption for ISO-BMFF (CENC) [8] and Segment Encryption and Authentication [3]. Common encryption standardizes which parts of a sample are encrypted, and how encryption metadata is signaled within a track. This means that the DRM module is responsible for delivering the keys to the client, given the encryption metadata in the segment, while decryption itself uses standard AES-CTR or AES-CBC modes. The CENC framework is extensible and can use other encryption algorithms beyond these two, if defined. Common Encryption is used with several commercial DRM systems, and is the system used in DASH264 [10].

DASH Segment Encryption and Authentication (DASH-SEA) [3] is agnostic to the segment format – encryption metadata is passed via the MPD, as opposed to the inband mechanisms of CENC [8] and traditional MPEG-2 Conditional Access [6]. For example, MPD contains information on which key is used for decryption of a given segment, and how to obtain this key. The baseline system is equivalent to the one defined in HTTP Live Streaming (HLS) [14], with AES-CBC encryption and HTTPS-based key transport. This has a side effect of making MPEG-2 TS media segments compatible with encrypted HLS segments. The standard itself is very extensible, and allows other encryption algorithms and more DRM systems, similarly to CENC.

DASH-SEA also offers a segment authenticity framework. This frameworks ensures that the segment received by the client is same as the one the MPD author intended the client to receive. This is being done using MAC or digest algorithms, and the intent is to prevent content modification within the network (e.g., ad replacement, altering inband events, etc.)

7 Current Standardization Activities

MPEG is actively working on exploration in areas related to DASH. Several core experiments were established in 2012-2013, exploring areas such as quality-driven streaming, non-HTTP distribution, issues related to live and low-delay services, advanced uses of events, and more. Activities related to reference software and an extensive set of implementation guidelines are at their advanced stage. Several standardization organizations and consortia adopted DASH or are in process of adopting it.

DASH industry forum was established in 2012. This forum is not an SDO -- it provides a set of interoperability guidelines and precisely defined interoperability points, DASH264 [10], as well as corresponding test material and software tools. The first version of DASH264 is at the public review stage, with more extensions coming.

8 Summary

In this letter we have provided a technical overview of the MPEG DASH standard and its newer extensions. During the 14 months that passed from the time it finalized, DASH gaining momentum in the industry. More than 50 companies joined the newly established DASH Industry Forum, and client implementations are already available. In the next 14 months we hope this interest will result in commercial deployments.

References

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

[2] ISO/IEC PDTR 23009-3 *Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 3: Implementation Guidelines*

[3] ISO/IEC FDIS 23009-4 *Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 4: Segment Encryption and Authentication*

[4] ISO/IEC Study DAM 23009-1 *Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, Amendment 1: Support for Event Messages and Extended Audio Channel Configuration*

[5] ISO/IEC COR 23009-1 *Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, Technical Corrigendum 1*

[6] ITU-T Rec. H.222.0 | ISO/IEC 13818-1:2012, *Information technology " Generic coding of moving pictures and associated audio information: Systems*

[7] 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)*

[8] ISO/IEC 23001-7, *Information technology " MPEG systems technologies" Part 7: Common encryption in ISO base media file format files*

[9] WebM Project, "WebM Dash Specification", http://wiki.webmproject.org/adaptive-streaming/webm-dash-specification

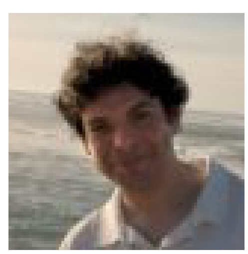
[10] DASH Industry Forum, "DASH264 Implementation Guidelines v0.9", January 2013, http://dashif.org/guidelines/DASH264-base-v09.pdf

[11] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE Multimedia*, Oct-Nov, 2011.

[12] M. Kar, A. Giladi, "Using DASH and MPEG-2 TS for Adaptive Multiplatform Delivery," *SCTE Cable-Tec Expo,* ," November 15-17, 2011, Atlanta GA.

[13] T. Stockhammer, "Live Timing for DASH v0.56", DASH-IF, January 14, 2013

[14] R. Pantos, W. May, "HTTP Live Streaming", http://tools.ietf.org/html/draft-pantos-http-live-streaming (work in progress)



Alex Giladi (M'06) is a senior architect at FutureWei Technologies (Bridgewater NJ), where he works on adaptive streaming ecosystem and its standardization. Previously he worked at Avail-TVN, Digital Fountain, and Harmonic Lightwaves. He is actively involved in MPEG DASH standardization since 2010, currently serving as the editor of two MPEG standards, Part 4 of MPEG DASH and Common Encryption for MPEG-2 TS. Alex holds a MSEE degree from Stanford University and BSc degree from the Technion, Haifa.

Optimizing DASH Delivery Services over Wireless Networks

Ozgur Oyman, Intel Labs, Santa Clara, CA 95054 USA
ozgur.oyman@intel.com

1- Introduction on DASH: The growing consumer demand for mobile video services is one of the key drivers of the evolution of wireless multimedia solutions requiring exploration of new ways to optimize future wireless networks for video services towards delivering enhanced capacity and quality of experience (QoE). One of these key video enhancing solutions is HTTP adaptive streaming (HAS), which has recently been spreading as a form of internet video delivery with the recent deployments of proprietary solutions such as Apple HTTP Live Streaming, Microsoft Smooth Streaming and Adobe HTTP Dynamic Streaming, and is expected to be deployed more broadly over the next few years.

In the meantime, the standardization of HTTP Adaptive Streaming has also made great progress with the recent completion of technical specifications by various standards bodies. More specifically, the Dynamic Adaptive Streaming over HTTP (DASH) has recently been standardized by Moving Picture Experts Group (MPEG) and Third Generation Partnership Project (3GPP) as a converged format for video streaming [1]-[2], and the standard has been adopted by other organizations including Digital Living Network Alliance (DLNA), Open IPTV Forum (OIPF), Digital Entertainment Content Ecosystem (DECE), and Hybrid Broadcast Broadband TV (HbbTV). DASH today is endorsed by an ecosystem of over 50 member companies at the DASH Industry Forum (DASH IF).

2- Research Challenges for Optimizing DASH Deployment in Wireless Networks: As a relatively new technology in comparison with traditional streaming techniques such as Real-Time Streaming Protocol (RTSP) and HTTP progressive download, deployment of DASH services presents new technical challenges. In particular, enabling optimized end-to-end delivery of DASH services over wireless networks requires developing new algorithms, architectures, and signaling protocols for efficiently managing the limited network resources and enhancing service capacity and user QoE.

Such development must also ensure access-specific (e.g., 3GPP-specific) optimizations for DASH services, which clearly require different approaches and methods compared to those for traditional streaming techniques, observing the client-driven nature of DASH and presence of the TCP layer, and that QoE for DASH is measured via different performance metrics. The rich set of research vectors in this space include:

i) Development of evaluation methodologies and performance metrics to accurately assess user QoE for DASH services (e.g., those adopted as part of MPEG and 3GPP's DASH specifications [1, 2]), and utilization of these metrics for service provisioning and optimizing network adaptation.

ii) DASH-specific QoS delivery and service adaptation at the network level, that involves developing new policy and charging control (PCC) guidelines, QoS mapping rules and resource management techniques over radio access network and core IP network architectures,

iii) QoE/QoS-based adaptation schemes for DASH at the client, network and server (potentially assisted by QoE feedback reporting from clients), to jointly determine the best video, transport, network and radio configurations toward realizing the highest possible service capacity and end user QoE. The broad range of QoE-aware DASH optimization problems emerging from this kind of a cross-layer cooperation framework includes investigation topics such as QoE-aware radio resource management and scheduling, QoE-aware service differentiation, admission control, and QoS prioritization, and QoE-aware server/proxy and metadata adaptation.

iv) DASH-specific transport optimizations over heterogeneous network environments, where content is delivered over multiple access networks such as WWAN unicast (e.g., 3GPP packet-switched streaming [3]), WWAN broadcast (e.g., 3GPP multimedia broadcast and multicast service [4]) and WLAN (e.g., WiFi) technologies.

3- Evaluation of Service Capacity and QoE for DASH: This section addresses the first and partly third research vectors above for optimizing DASH delivery in wireless networks. More specifically, we summarize our proposed capacity and QoE evaluation methodology for DASH services based on the notion of rebuffering percentage as the central indicator of user QoE, and associated empirical data based on simulations conducted over 3GPP Long Term Evolution (LTE) networks. Further details on our work can be found in the papers listed in [5]-[6].

For our capacity evaluation, we use a dynamic system-level simulator for the LTE air-interface based on a MATLAB-based software platform with suitable abstractions of application, transport, MAC and physical layers (details in [5]-[6]). We measure video capacity in terms of the number of unicast video streams that can be simultaneously supported for a given target QoE. The QoE metric of interest here is the *rebuffering percentage*, which is defined as the percentage of the total presentation time in which the user experiences rebuffering due to buffer starvation which has a significant impact on the end user quality of experience. It is worth noting here that in a recent study conducted by Conviva, rebuffering has been identified the single most dominating QoE impairment. In particular, our LTE capacity evaluation counts the number of users simultaneously supported via HTTP-based unicast video streaming sessions where the users are "satisfied" A_{cov} percentile of the time, with a user being counted as satisfied if and only if the rebuffering percentage in its video streaming session is less than A_{out}.

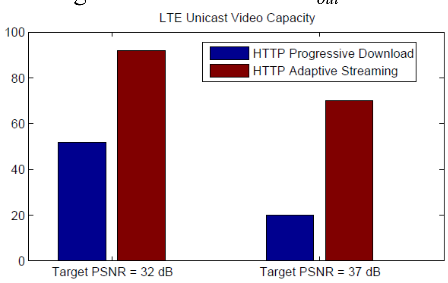


Fig. 1: LTE unicast video capacity comparison

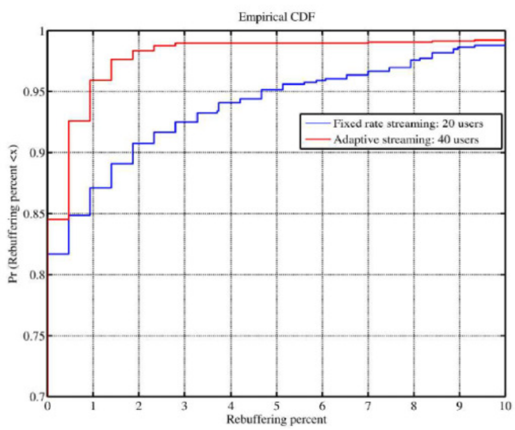


Fig. 2: Empirical CDF of rebuffering percentage for HTTP progressive download and DASH-based adaptive streaming

Fig. 1 shows the LTE unicast video capacities of HTTP-based fixed-rate streaming (i.e., progressive download) and DASH-based HTTP adaptive streaming, with A_{cov} = 95% and A_{out} = 5% subject to different target peak-to-signal ratio (PSNR) values of 32 dB and 37 dB, respectively. For the same set of streaming protocols, Fig. 2 shows the distribution of rebuffering percentage for the 37 dB target PSNR case. The capacity-quality tradeoff as a function of target PSNR is evident from the empirical data, i.e., the LTE system can support much higher number of users when the target PSNR is reduced. More importantly, the results clearly demonstrate that DASH-based HTTP adaptive streaming allows for supporting a significantly larger number of video users in comparison with HTTP-based progressive download techniques: With fixed rate streaming over LTE at a target PSNR of 37 dB and only with 20 users in the system, the 95-th percentile value of rebuffering percentage is 5% whereas the corresponding value for an LTE system with DASH-based adaptive streaming and twice as load (*i.e.*, with 40 users) is less than 1%. This is an intuitively expected outcome, given the significantly varying link quality among the users in the LTE network, leading to frequent occurrences of rebuffering with HTTP-based progressive download in the absence of any video quality/bitrate adaptation, especially when the network is unable to support the fixed bitrate during moments of low throughput caused by unfavorable link conditions. In contrast, with DASH-based HTTP adaptive streaming, each client device can dynamically select the quality/bitrate levels of the fetched videos to ensure continuous playback while also optimizing quality that could be achieved for the given link throughput, and such adaptation capability ensures finding the best possible compromise between high video quality and minimal occurrences of rebuffering events and delivering enhanced QoE to a larger number of LTE clients.

REFERENCES

[1] ISO/IEC DIS 23009-1: "Information technology — Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats"
[2] 3GPP TS 26.247: "Transparent end-to-end packet switched streaming service (PSS); Progressive download and dynamic adaptive streaming over HTTP (3GP-DASH)"
[3] 3GPP TS 26.234: "Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs"
[4] 3GPP TS 26.346: "Multimedia Broadcast Multicast Service (MBMS); Protocols and codecs"
[5] O. Oyman and S. Singh, "Quality of experience for HTTP adaptive streaming services," IEEE Commun. Mag., vol. 50, no:4, pp. 20-27, Apr. 2012.
[6] S. Singh, O. Oyman, A. Papathanassiou, D. Chatterjee and J. Andrews, "Video Capacity and QoE Enhancements over LTE", IEEE International Conference on Communications(ICC) Workshop on Realizing Advanced Video Optimized Wireless Networks, June 2012.

OZGUR OYMAN (ozgur.oyman@intel.com) is a senior research scientist and project leader in the Wireless Communications Lab of Intel Labs. He joined Intel in 2005. He is currently in charge of video over 3GPP Long Term Evolution (LTE) research and standardization, with the aim of developing end-to-end video delivery solutions enhancing network capacity and user quality of experience (QoE). He also serves as the principal member of the Intel delegation responsible for standardization at 3GPP SA4

Working Group (codecs). Prior to his current roles, he was principal investigator for exploratory research projects on wireless communications addressing topics such as client cooperation, relaying, heterogeneous networking, cognitive radios and polar codes. He is author or co-author of over 70 technical publications, and has won Best Paper Awards at IEEE GLOBECOM'07, ISSSTA'08 and CROWNCOM'08. His service includes Technical Program Committee Chair roles for technical symposia at IEEE WCNC'09, ICC'11, WCNC'12, ICCC'12 and WCNC'14. He also serves an editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He holds Ph.D. and M.S. degrees from Stanford University, and a B.S. degree from Cornell University.

Fair Share Dynamic Adaptive Streaming over HTTP

Christopher Müeller, Stefan Lederer, and Christian Timmerer
Multimedia Communication (MMC) Research Group, Institute of Information Technology (ITEC)
Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria
{Christopher.Mueller, Stefan.Lederer, Christian.Timmerer}@itec.aau.at

Abstract

Multimedia delivery over the Hypertext Transfer Protocol (HTTP) is currently very popular and with MPEGs' Dynamic Adaptive Streaming over HTTP (DASH) a standard is available to provide interoperability and enable large-scale deployments using existing infrastructures (servers, proxies, caches, etc.). This paper identifies some issue when multiple DASH clients compete for a bandwidth bottleneck when transparent proxy caches are deployed. Therefore, we propose a fair share adaptation scheme to be included within the client which – through experimental results – achieve a more efficient utilization of the bottleneck bandwidth and less quality switches.

Index Terms—Dynamic Adaptive Streaming over HTTP, DASH, Fair Adaptation, Proxy Cache, Multimedia

1. INTRODUCTION

The delivery of multimedia content over-the-top of existing infrastructures (servers, networks, caches, proxies) using the Hypertext Transfer Protocol (HTTP) is gaining more and more momentum despite its being designed for best effort and not for real-time multimedia transport. The Moving Picture Experts Group (MPEG) has recently ratified a standard for the Dynamic Adaptive Streaming over HTTP (DASH) [1] which is able to handle varying bandwidth conditions and allows for flexible deployments over existing infrastructures. The basic principle of DASH-based multimedia delivery is to *(i)* provide multiple versions of the same content – referred to as representations –, *(ii)* chop the content into time-aligned segments to enable seamless switching between different representations, and *(iii)* enable the client to request these segments individually based on its current conditions. This approach scales very well but it may introduce some new drawbacks, as clients are not aware of each other and when transparent proxy caches are deployed. In particular, consider the use case where multiple clients compete for a bottleneck bandwidth with a transparent proxy cache involved. Each client requests segments based on its own estimated throughput that may result in an uneven distribution of segments corresponding to different representations. Thus, clients may switch frequently between representations receiving segments alternatingly from the proxy cache and origin server respectively. The problem is further detailed in [4].

A similar problem has been identified in [2], TCP fairness has been addressed in [3] but without considering proxy caches, and a fair share adaptation scheme has been proposed in [4]. This paper aims to summarize the major findings from [4] and is organized as follows. Section 2 describes the fair share adaptation scheme for DASH. Section 3 provides experimental results while Section 4 concludes this paper and provides future work.

2. FAIR SHARE DYNAMIC ADAPTIVE STREAMING OVER HTTP

Our Fair Share Adaptation Scheme (FSAS) aims to address the problem identified in Section 1. Therefore, we introduce an exponential backoff within the adaptation logic of the DASH client. Although this approach decreases the number of switch ups to a higher representation (after prior switch down) but does not consider whether bandwidth fluctuations are either caused by the client itself or the network. Self-caused frequent switching between representations get introduced because the clients' adaptation logic is not aware whether segments are received from the proxy cache or the origin server. Note that these negative effects only occur when a client switches to a higher quality level due to a wrong interpretation of the throughput estimation. Therefore, we have incorporated a probe method to identify the effective available bandwidth. The following techniques have been identified:

1. The server provides a non-cacheable object which guarantees that the client will measure the bandwidth to the server.
2. The client downloads the first few bytes or a random byte range of the next segment to estimate the effective available bandwidth. Typically, most proxies do not cache byte range requests.
3. The proxy cache modifies the MPD and removes the qualities that could not be served due to bandwidth limitations.

4. The proxy cache offers a service that provides information about the effective available bandwidth.

We have decided to use method 2 for our system as it does not require any changes on the network side and, thus, can be easily deployed over existing infrastructures. However, any other method will lead to the same results.

Algorithm 1. Fair share adaptation algorithm using exponential backoff with probe.

```
if backoff > 0
 backoff := backoff - γ
endif
quality_level := find
(measured_bandwidth)
if quality_level > quality_last_segment
 if backoff <= 0
  if probe(quality_level)
   count := 0
  else
   backoff := (int) α * e^(β * count)
   count := count + δ
   quality_level :=
   quality_last_segment
  endelse
 else
  quality_level := quality_last_segment
 endelse
endif
return quality_level
```

Algorithm 1 depicts our adaptation logic that returns the quality level for the next segment. The backoff could be adjusted to the network characteristics with the parameters α and β. In our experiments we set them to 1 for simplicity reasons. Additionally, it is possible to accelerate or decelerate the backoff process with the parameters γ and δ. Furthermore, this algorithm uses the previously mentioned probe method to identify the effective available bandwidth for the next segment. This means that every adaptation decision which leads to a switch up will be verified.

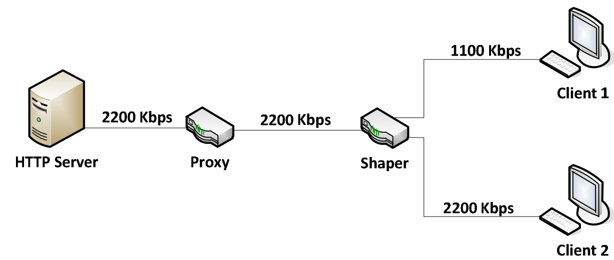
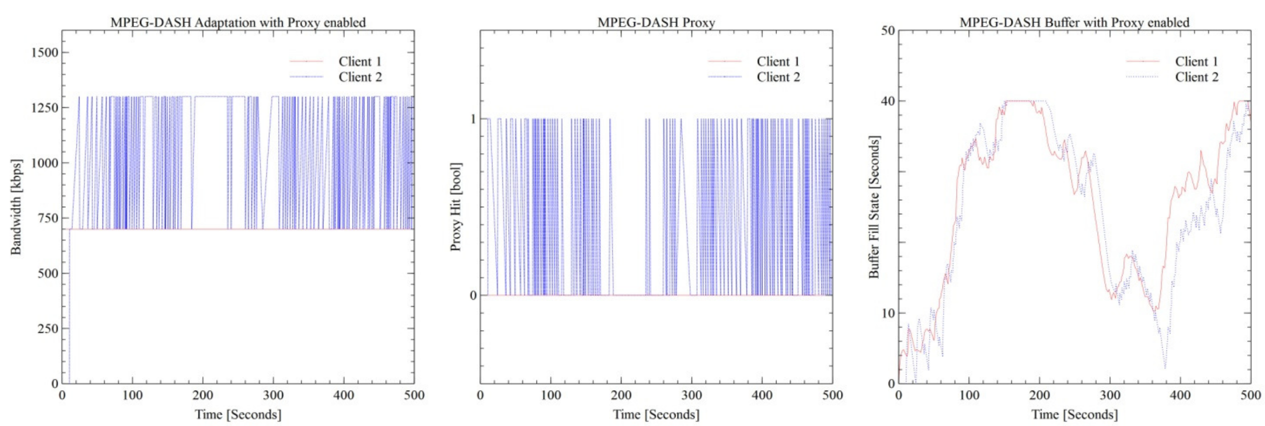


Figure 1. Experimental setup.

3. EXPERIMENTAL RESULTS

The architecture of our evaluation network is depicted in Figure 1. The proxy and the shaper are both based on Ubuntu 10.04. The shaper controls the bandwidth of the clients with the Linux traffic control system (tc) and the hierarchical token bucket (htb) has been used which is a classfull queuing discipline (qdisc). The available bandwidth for both clients remains static over the whole evaluation, i.e., 1100 Kbps for client 1 and 2200 Kbps for client 2. The proxy is based on the Squid [5] in transparent mode which also limits the bandwidth to the shaper with tc and htb. The evaluation has been performed with the Big Buck Bunny sequence at two representations with 700 and 1300 Kbps. Please note that for this experiment the available bandwidth will not change during the whole streaming session as dynamic bandwidth conditions may influence the negative effects even more. For example, the client makes an unfavorable adaptation decision when the network bandwidth drops. These evaluations under dynamic bandwidth conditions will be part of our future research.



(a) (b) (c)

Figure 2. MPEG-DASH clients without fair share adaptation scheme.

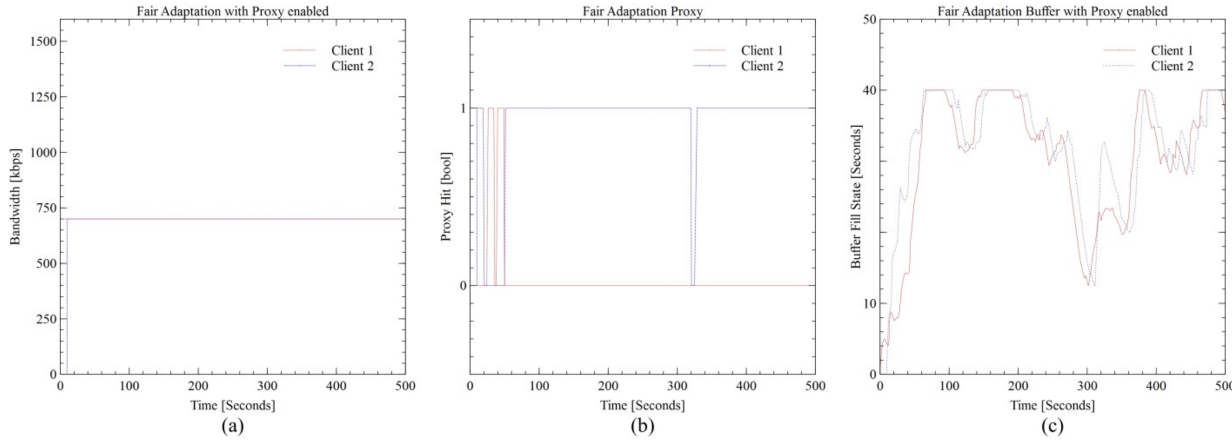


Figure 3. MPEG-DASH clients with fair share adaptation scheme.

The evaluation results for MPEG-DASH without our fair share adaptation scheme are depicted in Figure 2. Figure 2(a) shows the adaptation process, Figure 2(b) shows the behavior of the proxy cache and the cache hits for each request, and Figure 2(c) shows the buffer fill state. While client 1 does not switch to a higher representation (due to a max. bandwidth of 1100 Kbps), client 2 constantly tries to request a higher representation but fails due to the bottleneck bandwidth between proxy and server that is shared between both clients and, thus, results in frequent quality switches (cf. Figure 2(a)). This is behavior is also reflected by the cache hits in Figure 2(b). However, both clients ensure an almost smooth playback shown in Figure 2(c) except around second 380 where the buffer of client 2 is almost empty which resulted in stalling.

For MPEG-DASH clients employed with our fair share adaptation scheme the number of quality switches is reduced significantly as shown in Figure 3(a). The green lines indicate the probe points of the algorithm. Our approach also increases the cache performance depicted in Figure 3(b) and the buffer fill state of both clients is always far away from stalling as shown in Figure 3(c), thus, smooth playback is ensured by the client.

4. CONCLUSIONS AND FUTURE WORK

This paper identified some issues when multiple DASH clients compete for a bottleneck bandwidth with transparent proxy caches involved resulting in frequent quality switches which leads to a lower user experience including stalling. As a solution to this problem we proposed a fair share adaptation scheme using an exponential backoff algorithm with probing that reduces the number of quality switches, increases the cache performance, and ensures smooth playback without stalling. Our future work comprises the evaluation of the fair share adaptation scheme under dynamic bandwidth conditions with more than two clients and competing non-DASH traffic. We assume that the negative effects will be increased in such a scenario.

5. ACKNOWLEDGMENTS

This work was supported in part by the EC in the context of ALICANTE (FP7-ICT-248652), SocialSensor (FP7- ICT-287975), and partly performed in the Lakeside Labs research cluster at AAU.

REFERENCES

[1] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet", *IEEE MultiMedia*, vol.18, no.4, pp.62-67, Apr. 2011.

[2] S. Akhshabi, L. Anantakrishnan, C. Dovrolis, A. C. Begen, "What Happens when HTTP Adaptive Streaming Players Compete for Bandwidth?", *NOSSDAV 2012*, Toronto, Canada, Jun. 2012.

[3] R. Kuschnig, I. Kofler, and H. Hellwagner, "An evaluation of TCP-based rate-control algorithms for adaptive internet streaming of H.264/SVC", *ACM Multimedia systems (MMSys)*, Scottsdale, AZ, USA, Feb. 2010.

[4] C. Mueller, S. Lederer, C. Timmerer, "A proxy effect analyis and fair adatpation algorithm for multiple competing Dynamic Adaptive Streaming over HTTP clients", *IEEE Visual Communications and Image Processing (VCIP) 2012*, San Diego, CA, USA, Nov. 2012.

[5] Squid, http://www.squid-cache.org/, (last access: Feb. 2013).

Christopher Müller received his B.Sc. (Bakk.) in Mar'12 from the Alpen-Adria-Universität Klagenfurt. His research interests are multimedia streaming, networking, and multimedia adaptation; he has published more than 10 papers in these areas and currently holds one U.S. patent in the area of DASH. He gained practical expertise in various companies (Infineon, Dolby Laboratories Inc. LA, etc.) and participated in the MPEG-DASH standardization, contributed several open source tools (VLC plugin, libdash) and participated in several EC-funded projects (ALICANTE, SocialSensor).

Stefan Lederer holds the position of an Assistant Professor at the Institute of Information Technology (ITEC), Multimedia Communication Group. He received his B.Sc. (Bakk.) in Business Administration as well as his B.Sc. (Bakk.) in Information Management in Aug'10 and his M.Sc. (Dipl.-Ing.) in Computer Science in Mar'12, all from the Alpen-Adria-Universität Klagenfurt. His research topics include transport of modern/rich media, multimedia adaptation, QoS/QoE as well as future internet architectures and he has published more than 10 papers in these areas He participated in the MPEG-DASH standardization, contributed several open source tools (DASHEncoder, DASHoverCCN VLC plugin) and datasets. He also participated in several EC-funded projects (ALICANTE, SocialSensor).

Christian Timmerer is an assistant professor in the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria. His research interests include immersive multimedia communication, streaming, adaptation, and Quality of Experience. He was the general chair of WIAMIS'08, ISWM'09, EUMOB'09, AVSTP2P'10, WoMAN'11 and has participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, and SocialSensor. He also participated in ISO/MPEG work for several years, notably in the area of MPEG-21, MPEG-M, MPEG-V, and DASH/MMT. He received his PhD in 2006 from the Alpen-Adria-Universität Klagenfurt. Publications and MPEG contributions can be found under research.timmerer.com, follow him on twitter.com/timse7, and subscribe to his blog blog.timmerer.com.

Quality Driven Streaming Using MPEG-DASH

Shaobo Zhang, Yangpo Xu, Peiyun Di, Alex Giladi, Changquan Ai, Xin Wang

Huawei Technologies

Abstract

Video streaming is becoming more and more popular, with video traffic exceeding 50% of the total mobile traffic in 2012 according to Cisco VNI. Adaptive streaming over HTTP is responsible for a significant part of this traffic. Current HTTP streaming systems are client-driven – i.e., the client makes decision on how to adapt to the constantly changing network environment. Currently, such adaptation is based on bitrate of encoded content. In many cases, higher bitrates do not result in comparable increase in perceptual quality.

In this paper, we propose quality driven streaming – a paradigm where adaptation decisions are taken based on a combination of bitrate and bandwidth, with adaptation performance compared to the one of traditional bitrate-only algorithms.

In this paper, quality-driven streaming was implemented as an extension of MPEG DASH, as a part of the ongoing MPEG DASH Quality-Driven Streaming core experiment. With that said, the concept of quality-driven streaming is not DASH-specific and can be used with any of the modern HTTP streaming systems

Index Terms— DASH, streaming, adaptation, quality

1. INTRODUCTION

Video streaming is becoming more and more popular, with video traffic exceeding 50% of the total mobile traffic in 2012 according to Cisco VNI [11]. Adaptive streaming over HTTP is responsible for a significant part of this traffic.

Apple HTTP Live Streaming [2] and Microsoft Smooth Streaming [1] are the mostly widely deployed proprietary adaptive streaming solutions. A relative newcomer to the scene, MPEG Dynamic Adaptive Streaming over HTTP (DASH) [3] is an international standard which builds on the previous industry experience. DASH is gaining traction in the industry, and the authors expect it to be widely adopted and enable interoperability in multimedia streaming.

Capability to adapt to constantly changing network conditions, thus maintaining good user experience, is the most fundamental requirement for streaming over an open non-provisioned network. Adaptive streaming over HTTP is client-driven. An asset is encoded in multiple versions, *representations* in DASH terminology-- at multiple bitrates, resolutions, etc., and broken into discrete *segments* and (possibly) *subsegments* – addressable and playable parts of an asset, typically several seconds long.

Available representations are advertised to a client via the MPD (Media Presentation Description), an XML document describing representation properties, timing, and HTTP URL's for retrieving (sub)segments. MPD advertises representation properties such as bandwidth, codecs, resolutions, etc.

The client, on its part, selects most acceptable alternative advertised to it, and re-evaluates its decision as a response to changing network conditions.

More detailed introductions to DASH are provided in [7], [8], and [9], as well as in many other papers.

2. ADAPTATION PARAMETERS

Most representation properties provided by the MPD, such as resolution, aspect ratio, frame rate, etc., are used by the client for static adaptation – i.e., determining which representations can be played out successfully. The only parameter used for dynamic adaptation is bitrate.

The DASH buffering model defines buffering duration, with the physical buffer size being a product of this duration and the stated representation-level bandwidth. This buffer is assumed to be sufficient for normal playback of a representation. With that said, there is no strict bitrate constancy requirement in DASH.

While DASH segments are addressable via URL's that are derived directly from the MPD, finer-granularity addressing can be done via an index – a binary structure providing a mapping of playable byte ranges within segments to their playout times and random access properties. When DASH indexes are used, per-segment bitrate information is accessible to the client, whereas only per-representation bandwidth is known to the client otherwise. Use of indexes allows for more optimal adaptation decisions, as well as better random access and trick mode functionality, and is expected from on-demand applications.

In absence of information beyond bitrate, the client makes an implicit assumption that higher bitrate is equivalent to better perceived quality. Such greedy bitrate-driven rate adaptation behavior of a Microsoft Smooth Streaming [1] client is described in [10]. When indexes are used, more optimal decisions can be taken, as the client has per-segment information, however this does not change the greedy nature of the algorithm.

Greedy adaptation behavior may lead to amplified quality fluctuation or inefficient usage of bandwidth. This is due to the fact that complexity of content changes with time. In streaming, content is encoded either CBR (capped VBR, in most cases), or unconstrained VBR. For CBR representations, its bitrate is well controlled, but its quality may fluctuate significantly unless the bitrate is sufficiently high. Changing content complexity, such as switching between sports and static scenes in news channels makes it very difficult for video encoders to deliver consistent quality and at the same time produce bitstream that has a certain specified bitrate. In this case, bandwidth is not efficiently used. For VBR representation, it allows a higher bitrate to be allocated to the more complex scenes while fewer bits to less complex scenes. Its quality fluctuation is relatively small but far from constant.

If quality information is provided, either per-representation (as an additional attribute in the MPD), or per segment (in MPD, in an index, or in a separate resource), both quality and bitrate can be taken into account for the purposes. Many of the proposals to the MPEG Quality Driven Streaming core experiment [4] propose syntactical ways of providing such information.

3. EVALUATION OF QUALITY BASED ADAPTATION

In this section, we compare adaptation policies using and not using quality information. There are 4 different combinations on which level bitrate and/or quality are provided. For comparison, a typical algorithm for each kind policy is selected which uses the specified information in a simple and straightforward way. It is recognized that the selected algorithm may not be best one of its kind in performance, but it does not bias the comparison.

3.1. Adaptation Policies

1. Adaptation according to per-representation bitrate information (@bandwidth), denoted as RB.
   With only representation level bitrate information available, the only option is to match bitrate of a representation to the available bandwidth. At time when a new segment is requested, a representation with bitrate lower than but closest to the available bandwidth is selected. This algorithm is equivalent to the one used in the Microsoft Smooth Streaming client, as described in [10].
2. Adaptation according to per-representation bitrate and quality information, denoted as RBRQ.
   In addition to per-representation bitrate, average PSNR over the whole sequence is signaled per representation to represent its quality. A quality threshold qth is used in this algorithm to enhance RB. It works if there is enough bandwidth to deliver representation with quality higher than the threshold. Thus with quality information bandwidth waste is avoided.
3. Adaptation according to per-segment bitrate information, denoted as SB.
   If the Segment Index is present in representations, client can download it and obtain detailed bitrate/size information over time. It helps to better match (sub)segments from representations with the available bandwidth to achieve a good bandwidth usage. This essentially is a refinement of the algorithm from [10]
4. Adaptation according to per-segment bitrate and quality information, denoted as SBSQ
   It uses both bitrate and quality on (sub)segment level. Compared with algorithm RBRQ, the adaptation logic is the same but works on a finer granular level, segment instead of representation. Bandwidth adaptation and quality threshold play a bigger role.
5. Adaptation according to per-segment bitrate and quality information, denoted as SBSQ
   M-SBSQ improves on the basis of SBSQ to further reduce quality fluctuation. When quality threshold works, segments of lower bitrate than possible under available bandwidth are downloaded, and the client's buffer is then filled to a higher level. The higher level of buffered media data is used to buy time sometime later to download segments with higher quality than allowed by available bandwidth. To avoid overuse of buffered data, there is a buffer threshold - buf_{th2}, data in buffer lower the threshold is only used to counter bandwidth fluctuation, quality improvement, i.e. requesting segment of higher quality than allowed by available bandwidth, is only performed when buffer level is above the threshold and should not result in a lower buffer level than the threshold.

3.2. Experiment Setup

To compare the performance of adaptation policies, experiment is done via simulation. Figure 1 depicts the experiment setup: Content Server and Client are connected via 1Gbps Ethernet links. For simplicity reason, it is assumed that available bandwidth measured on application level, i.e. TCP throughput is directly used by adaptation algorithm at client, which eliminates the difference resulting from bandwidth measurement and estimation. As long as the link capacity between Content Server and Client is large enough, traffic between the two is only restricted by the available bandwidth.

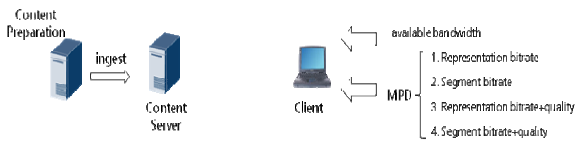


Figure 1 Experiment Setup

3.3. Result

Figure 2 depicts quality of the streamed content in PSNR. Though SB has the highest mean quality it exhibits large quality fluctuation. On quality fluctuation, M-SBSQ and SBSQ performs best with fairly small deviation. It is because M-SBSQ and SBSQ use quality information while SB not. RB and RBRQ do not perform well on both mean value and deviation in cases of constrained and unconstrained representations. The reason is that it is the maximum bitrate and averaged quality of all segments provided on representation level, which is too inaccurate to guide adaptation on segment basis.

As for bandwidth consumption (Figure 3), SB consumes more bandwidth than RB. As SB always request segments with highest possible bitrate. For RB, it uses representation level bitrate which is the maximum bitrate of a representation (it ensures any segments in the representation can be delivered in time when available bandwidth is equal to the value). Apparently, some, even most segments in the representation have lower bitrates than this value. As a result, RB "overestimates" bandwidth required for segments in the representation and consumes a low bandwidth.

For RBRQ and SBSQ, their performances depend on the value of the parameter quality threshold q_{th} in the algorithms. When the threshold is high enough it has no effect, RBRQ/SBSQ performs the same as RB/SB. When the threshold becomes smaller, it takes effect and segments of quality exceeding the threshold are replaced with those of lower quality, thus less bandwidth is consumed and quality of the streamed content becomes smoother.

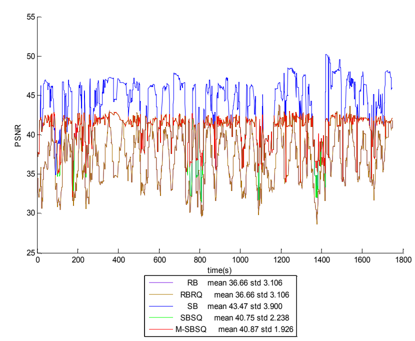
When q_{th}=41dB, compared with SB, SBSQ saves consumed bandwidth about 40%.

Additional quality information helps client to improve quality smoothness therefore quality of experience and at the same time reduces bandwidth consumption.

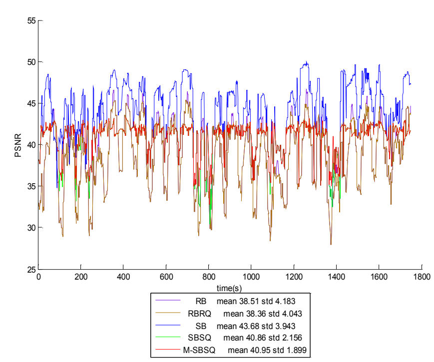
Simple use of quality threshold helps to reduce quality fluctuation and save bandwidth, as shown in SBSQ. It comes with no cost of buffer level decline.

Quality information can further guide quality improvement of streaming service, i.e. to prevent short term quality drop. Without quality information it is difficult if not impossible. M-SBSQ is an example which attempts to request segments with quality in a specific range, when quality of a segment exceeding an up bound, it requests a segment with lower quality and vice versa.

A comparison of results between unconstrained VBR and constrained VBR representation (sub figure(a) and sub figure (b) in Figure 4) shows constrained VBR benefits more from quality information and quality driven adaptation algorithms e.g. RBRQ, SBSQ, SBSQ-M, than unconstrained VBR in aspects, such as mean value of quality, quality deviation, bandwidth saving. This result can be extended to CBR, which is the extreme of constrained VBR with strict restriction on bitrate fluctuation. In a special case that available bandwidth remains unchanged over time and content is prepared in CBR, bitrate only based adaptation selects a single representation, however, quality driven adaptation performs better by selecting (sub)segments from different representations. The explanation is that from VBR to CBR, as the bitrate fluctuation reduced and quality fluctuation increased, quality information plays a more important role in adaptation. The conclusion is important, since in most cases of streaming service, content is offered as constrained VBR representations rather than of unconstrained ones.

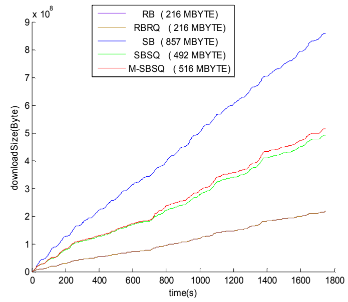


(a) unconstrained

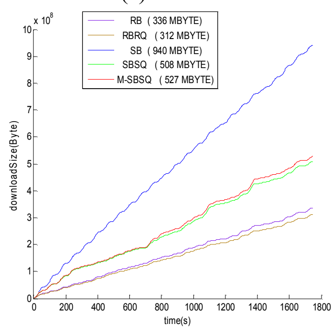


(b) unconstrained

Figure 2 quality of the streamed content (for SBSQ and M-SBSQ, qth= 41dB; for M-SBSQ, delta_q=6dB)

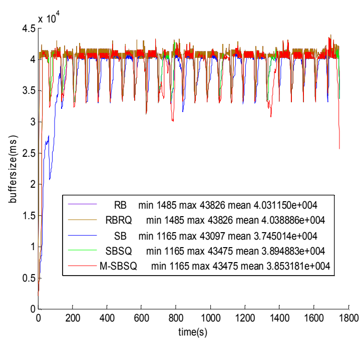


(a) unconstrained

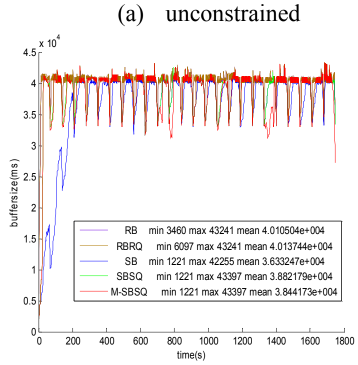


(b) unconstrained

Figure 3 bandwidth consumption (for SBSQ and M-SBSQ, qth= 41dB; for M-SBSQ, delta_q=6dB)



(a) unconstrained



(b) unconstrained

Figure 4 buffer level(for SBSQ and M-SBSQ, qth= 41dB; for M-SBSQ, delta_q=6dB)

REFERENCES

[1] Microsoft, IIS Smooth Streaming Technical Overview, Mar. 2009; http:// www.iis.net/learn/media/on-demand-smooth-streaming/smooth-streaming-technical-overview.

[2] R. Pantos and E.W. May, “HTTP Live Streaming”, IETF Internet draft, Oct. 2012

[3] ISO/IEC 23009-1:2012 Information technology — Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, Apr. 2012

[4] MPEG, Descriptions of Core Experiments on DASH amendment, w13082, Shanghai, Oct. 2012.

[5] C. Liu, I. Bouazizi, M. Gabbouj, “Rate Adaptation for Adaptive HTTP Streaming”, in Proceedings of the second annual ACM Conference on Multimedia Systems(MMSys), Feb. 2011.

[6] F.Z. Yang, S. Wan, Q. Xie and H.R. Wu, “No-reference Quality Assessment for Networked Video via Primary Analysis of Bit-stream,” IEEE Trans. CSVT, vol.20, no. 11, pp.1544-1554, Nov. 2010.

[7] I. Sodagar, “The MPEG-DASH Standard for Multimedia Streaming Over the Internet”, IEEE MultiMedia, vol.18, no.4, pp.62-67, Apr. 2011.

[8] I. Sodagar and H. Pyle, Reinventing multimedia delivery with MPEG-DASH, Proc. SPIE 8135, 81350R (2011)

[9] M. Kar, A. Giladi, Using DASH and MPEG-2 TS for Adaptive Multiplatform Delivery, SCTE Cable-Tec Expo, November 15-17, 2011, Atlanta GA

[10] S. Akhshabi, L. Anantakrishnan, C. Dovrolis, A. C. Begen, “What Happens when HTTP Adaptive Streaming Players Compete for Bandwidth?”, NOSSDAV 2012, Toronto, Canada, Jun. 2012.

[11] Cisco Visual Networking Index: Global Mobile Data, Traffic Forecast Update, 2012–2017,

[12] February 6, 2013, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

Shaobo Zhang received his B.E.degree in Radio Engineering in 1991, M.S. degree in Electronic and Communication in from Southeast University (Nanjing) in 1996. In the same year, he joined Huawei Technologies. He had been working in Wireless Department before 2006 where he engaged in development, design and planning of products for cellular communication. After then, he joined the Corporate Research Department. His research interest includes multimedia delivery, network optimization, system design and mobile communication.

Yangpo Xu received B.E. degree and M.S. degree from Tianjin university in 2003 and 2006 respectively. He joined Huawei Technologies in 2006 and works in the Corporate Research Department. His research interests are in video transmission, media file format and network protocols.

Peiyun Di received her B.S. degree in Applied Mathematics and the M.S. degree in electronic engineering from Xi Dian University, Xi'An, China, in 2004 and 2007, respectively. She is currently working at Huawei Tech Co.,Ltd.

Alex Giladi [M'06] is a senior architect at FutureWei Technologies (Bridgewater NJ), where he works on adaptive streaming ecosystem and its standardization. Previously he worked at Avail-TVN, Digital Fountain, and Harmonic Lightwaves. He is actively involved in MPEG DASH standardization since 2010, currently serving as the editor of two MPEG standards, Part 4 of MPEG DASH and Common Encryption for MPEG-2 TS. Alex holds a MSEE degree from Stanford University and BSc degree from the Technion, Haifa.

Changquan Ai received his M.S. degree in computer software, Xi'an, China. Currently He works as a research engineer in field of multimedia system.

Xin Wang (xwang@huawei.com) has been working as a chief scientist in multimedia systems at Corporate Research of Huawei Technologies, USA. Prior to joining Huawei, he was the chief scientist at ContentGuard, Inc. He is a chair of the MPEG-M and MPEG-21 CEL ad hoc groups, and actively participates the MPEG DASH standard development. He is also adjunct faculty of the Department of Computer Science at the University of Southern California, Los Angeles, USA.

User-Adaptive Mobile Video Streaming Using MPEG-DASH

Yuriy A. Reznik, InterDigital Communications, Inc.
9710 Scranton Road, San Diego, CA 92122
Email: yuriy.reznik@interdigital.com

Abstract

MPEG-DASH is a new international standard for dynamic adaptive streaming over HTTP. In this letter, we show how this standard can be used to design intelligent streaming applications adapting video delivery to user behavior and viewing conditions, resulting in better utilization of network and power resources.

1 Introduction

During last two decades Internet streaming has experienced a dramatic growth and transformation from an early concept into a mainstream technology used for delivery of multi-media content [1-3]. A recently issued MPEG-DASH standard [4] consolidates many advances achieved in the design of streaming media delivery systems, including full use of the existing HTTP infrastructure, bandwidth adaptation mechanisms, latest audio and video codecs, etc. Yet, some challenges in implementation and deployment of streaming systems still exist. In particular, they arise in delivery of streaming video content to mobile devices, such as smartphones and tablets.

On one hand, many mobile devices are already matching and surpassing HDTV sets in terms of graphics capabilities. They often feature high-density "retina" screens with 720p, 1080p, and even higher resolutions. They also come equipped with powerful processors, making it possible to receive, decode and play HD-resolution videos. On the other hand, network and battery/power resources in mobile devices remain limited. Wireless networks, including latest 4G/LTE networks, are fundamentally constrained by capacities of their cells. Each cell's capacity is shared between its users, and it can be saturated by as few as $5-10$ users simultaneously watching high-quality videos [5]. High data rates used to transmit video also cause high power consumption by the receiving devices, draining their batteries rapidly.

All these factors suggest that technologies for reducing bandwidth and power use in mobile video streaming are very much needed. In this letter we describe one such technology. It is based on an observation that in many cases, mobile phone users can see only a fraction of information projected on the screen.

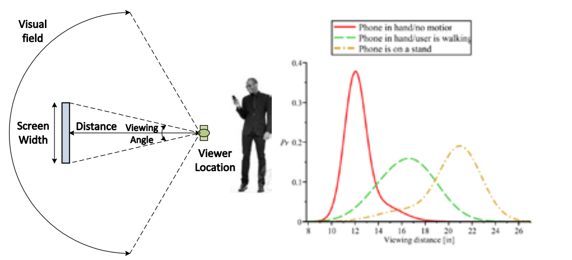


Figure 1: Characteristics of mobile viewing setup. The right sub-figure shows how viewing distance can be affected by user's activity.

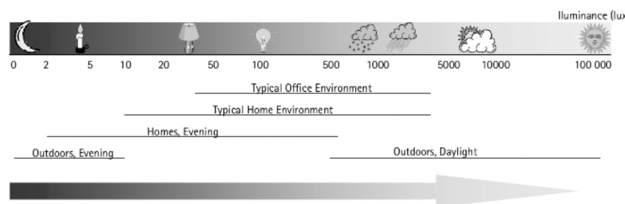


Figure 2: Ambient illuminance in different environments [6].

2 Factors affecting user ability to discern visual content

We illustrate some factors that affect user's ability to discern visual information Figures 1 and 2. For example, the user may hold a phone close to his eyes, or at arm's length. This affects viewing angle and density of information seen on the screen. Ambient illuminance may also change significantly. The user may be in the office, outside under direct sunlight, in a shadow, or in a completely dark area. Reflection of ambient light from the screen lowers the contrast of video or images seen by the user [6]. Finally, the user may pay full attention to visual content on the screen, or he could be distracted.

Together with characteristics of the mobile display and user vision, all these factors affect the capacity of the "visual channel", serving as the last link in a communication system delivering information to the user. The main idea of this letter, as well as several of our related publications [7, 8] is to show that characteristics of this last link can also be effectively measured and utilized in optimizing streaming video delivery. The recently developed MPEG-DASH standard offers an excellent framework using which this idea can be realized.

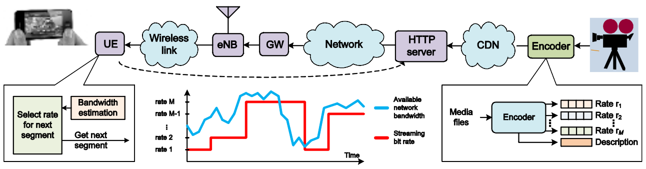


Figure 3: Illustration of functionality of mobile DASH-based streaming system. The multimedia content is encoded at encoded at multiple rates, and segmented in chunks allowing client to select portions that can be delivered in real-time, while also adopting to changing network bandwidth.

3 MPEG DASH-standard: the basics

We present a conceptual model of DASH-based mobile video streaming system in Figure 2. The original video content is captured, encoded, and placed on an HTTP server. To scale distribution, the content may also be pushed to many servers forming a Content Distribution Network (CDN). It is typically the web browser or a streaming client application running on a mobile phone (UE) that discovers this content, retrieves it, and shows it on a mobile device.

3.1 Content preparation

In order to support bandwidth adaptive streaming, the content is usually encoded at a plurality of bit rates. Such encodings are also prepared such that they consist of multiple segments with time-aligned boundaries, allowing switches between encodings at different rates. In MPEG DASH standard, points at which switching is allowed are called stream access points (SAP). In the simplest case, SAP may correspond to an I- or IDR- video frame, allowing sequential decoding of all frames that follow. In addition to producing encoded media streams the encoder also produces a file containing information about parameters of each of the encodings and URL links to them. This file is called media presentation description (.mpd) file.

3.2 Adaptation to bandwidth changes

The streaming session is controlled entirely by the DASH streaming client. It opens an HTTP connection to the server, retrieves the .mpd file, and learns about different encodings (representations) that are available on the server. Then it picks representation with most suitable bitrate, and start retrieving its segments by issuing HTTP GET requests. As bandwidth changes, the streaming client may request segments encoded at different bit rates, allowing uninterrupted playback of the content. We illustrate this in Figure 3.

3.3 Communication of encoding parameters

MPEG-DASH media presentation description file allows encoders to share specific parameters of each encoded version of the content. In case of video, these parameters include resolution (width × height), pixel aspect ratio, frame rate, and required bandwidth. When the content is prepared, the encoder may choose to use different combination of these parameters to produce encodings for each target bitrate. The encoder may also produce multiple encodings considering different screen resolutions and other specific capabilities of target devices, allowing streaming clients to pick versions that are optimized for each particular device.

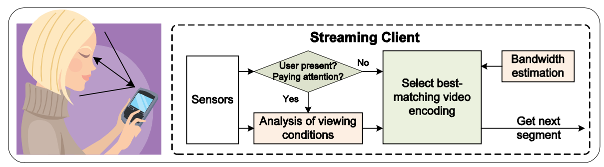


Figure 4: Illustration of functionality of DASH streaming client incorporating adaptation to user behavior and viewing conditions.

4 Enabling adaptation to user behavior and viewing conditions

We provide conceptual illustration of user-adaptive design of DASH streaming client in Figure 3. In order to adapt to viewing conditions, the client uses sensors of a mobile device, such as front-facing camera and accelerometer to detect the presence of the user, his proximity, pose, and viewing angle. The client also uses ambient illuminance sensor and information about brightness settings of the screen to estimate effective contrast ration of the screen.

Using these estimates, the client obtains minimum characteristics of encoded video, such as spatial resolution, framerate, and bitrate that are sufficient to achieve high level of visual quality. In finding such characteristics the client can use spatio-temporal contrast sensitivity functions [9] or other related results from studies on human vision and video coding. Once such characteristics are obtained, the client searches through a list of available video representations and selects one that is best suited for delivery.

As illustrated in Figure 4, adaptation to viewing distance and contrast can result in lowering bandwidth required to receive video. Increase in viewing distance lowers our ability to discern individual pixels and hence it becomes possible to select representations encoded using lower resolution and bitrate. Likewise increase of ambient illuminance lowers effective contrast of the screen and range of spatial frequencies that we can see. This also opens opportunity for lowering the resolution and required bitrate. For additional details the reader is referred to our related publications [7, 8].

In cases when client detects that user is not present next to the device, even more significant bandwidth savings are possible. For example, the client may stop receiving video while continuing playing only audio track.

As bandwidth usage is directly related to power consumption in mobile phones, the above described optimizations can also result in increased battery life.

5 Conclusion

In this letter we have shown that MPEG-DASH standard enables design of intelligent streaming systems adapting not only to bandwidth but also to factors affecting user ability to see visual information. Such adaptation can result in reduced bandwidth usage, increased battery life, and improved quality of user experience.

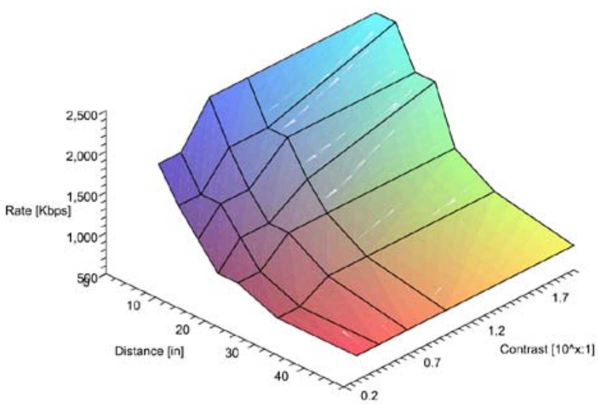


Figure 5: Example of rate allocation achieving approximately the same level of perceived quality under different viewing distances and contrast rates. This particular allocation was obtained assuming reproduction on a mobile device with 720 p-resolution screen, 340 dpi pixel density, and when using the H.264 (Main profile) video encoder.

References

[1] D. Wu, Y.T. Hou, W. Zhu, Y-Q. Zhang, and J.M. Peha, “Streaming video over the Internet: approaches and directions,” *IEEE Trans. Cir. Syst. Video Tech.*, Mar 2001, vol. 11, no. 3, pp. 282-300.

[2] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, “Video Coding for Streaming Media Delivery on the Internet,” *IEEE Trans. Cir. Syst. Video Tech.*, Mar 2001, vol. 11, no. 3, pp. 20-34.

[3] I. Sodagar, “The MPEG-DASH Standard for Multimedia Streaming Over the Internet,” *IEEE Multimedia*, Oct-Nov, 2011.

[4] ISO/IEC 23009-1 *Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats*, ISO/IEC, January 5, 2012.

[5] A. Talukdar, M. Cudak, and A. Ghosh, “Streaming Video Capacities of LTE Air Interface,” *Proc. IEEE Int. Conf. Comm. (ICC)*, 2010, pp. 1-5.

[6] J. Bergquist, “Resolution and contrast requirements on mobile displays for different applications in varying luminous environments,” *Proc. 2nd Int. Symp. Nanovision Science.*, 2005, pp. 143-145.

[7] Y. Reznik, et al., “User-adaptive mobile video streaming,” *Proc. of IEEE Visual Communication and Image Processing*, Aug 2012.

[8] R.Vanam, Y.Reznik, “Improving the Efficiency of Video Coding by using Perceptual Preprocessing Filter,” *Proc. Data Compression Conference*, March 2013.

[9] D. H. Kelly, “Motion and vision. II Stabilized spatio-temporal threshold surface,” *Journal of the Optical Society of America*, 1969, vol. 69, pp. 1340-1349.

Yuriy A. Reznik (M’97-SM’07) is a Director at InterDigital Communications (San Diego, CA), where he leads R&D in multimedia coding and delivery over wireless networks. He also serves as an editor of Part 3 of MPEG-DASH standard in MPEG. Previously, he worked at Qualcomm (2005-20011), and RealNetworks (1998-2005), and also stayed as Visiting Scholar at Stanford University (2008). He holds a Ph.D. degree in Computer Science from Kiev University. He has authored/co-authored over 80 conference and journal papers, and co-invented 14 issued (and over 50 pending) US patents.

CALL FOR PAPERS
COMMUNICATION SOFTWARE, SERVICES, AND MULTIMEDIA APPLICATION SYMPOSIUM

Symposium Co-Chairs

Vincent Wong, University of British Columbia, Canada
vincentw@ece.ubc.ca
Liang Zhou, Nanjing University of Posts and Telecommunications, China
liang.zhou@ieee.org

Scope and Topics of Interest

The Communications Software, Services and Multimedia Application Symposium will provide an international technical forum for discussing and presenting recent research results on any aspects of software, services, and multimedia communications. It aims at bringing together experts from industry and academia to exchange ideas and present results on advancing the state-of-the-art and overcoming research on the challenging issues related to the software design, system deployment of services, and multimedia applications over heterogeneous networks. Papers may present theories, techniques, applications, or practical experiences related to that. Topics of interest for this Symposium include, but are not limited to:

Multimedia Applications and Services
- Multimedia delivery and streaming over wired and wireless networks
- Cross-layer optimization for multimedia service support
- Multicast, broadcast and IPTV
- Multimedia computing systems and human-machine interaction
- Interactive media and immersive environments
- Multimedia content analysis and search
- Multimedia databases and digital libraries
- Converged application/communication servers and services
- Multimedia security and privacy
- Multimedia analysis and social media

Network and Service Management and Provisioning
- Multimedia QoS provisioning
- Multimedia streaming over mobile social networks and service overlay networks
- Service creation, delivery, management
- Virtual home environment and network management
- Charging, pricing, business models
- Security and privacy in network and service management
- Cooperative networking for streaming media content

Next Generation Services and Service Platforms
- Location-based services
- Social networking communication services
- Mobile services and service platforms

Home network service platforms
VoP2P and P2P-SIP services

Software and Protocol Technologies for Advanced Service Support
Ubiquitous computing services and applications
Networked autonomous systems
Communications software in vehicular communications
Web services and distributed software technology
Software for distributed systems and applications, including smart grid and cloud computing services
Peer-to-Peer technologies for communication services
Context awareness and personalization

Submission Guidelines

Prospective authors are invited to submit original technical papers by the deadline of **15 March 2013** for publication in the IEEE Globecom 2013 Conference Proceedings and for presentation at the conference. Submissions will be accepted through EDAS. All submissions must be written in English and be at most six (6) printed pages in length, including figures. For full details, please visit the following website:

http://www.ieee-globecom.org/2013/submguide.html

MMTC OFFICERS

E-LETTER BOARD MEMBERS