AN ADAPTIVE SYSTEM FOR REAL-TIME SCALABLE VIDEO STREAMING WITH END-TO-END QOS CONTROL

Beilu Shao[1], *Daniele Renzi*[2], *Peter Amon*[3], *George Xilouris*[4], *Nikolaos Zotos*[4], *Stefano Battista*[2], *Anastasios Kourtis*[4], *Marco Mattavelli*[1]

[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015, Lausanne, Switzerland
[2]bSoft Ltd, 62100, Macerata, Italy
[3]Siemens Corporate Technology, Real-time Vision & Industrial Imaging, 81730 Munich, Germany
[4]National Center of Scientific Research "Demokritos", 15310, Agia Paraskevi, Greece

ABSTRACT

This paper presents a real-time adaptive video streaming system based on the latest standardized video codec H.264/MPEG-4 AVC scalable extension (SVC). The system provides a full MPEG-21 media access framework over heterogeneous networks and terminals with end-to-end QoS control and multimedia adaptation based on SVC. This adaptive streaming system is composed of a server with a real-time SVC encoder, an adaptive network node, and a terminal with appropriate feedback of perceptual quality, network conditions and user preferences for adaptation support. The system facilitates a general content adaptation solution to achieve the end-to-end QoS control.

Index Terms- Multimedia adaptation, multimedia system, real-time streaming, H.264/SVC, MANE, QoS

1. INTRODUCTION

With the latest development of technologies, real-time multimedia applications have gained much attention and contribution from both academia and industry. Many different techniques have been developed, e.g., video coding with high compression efficiency and scalability, multimedia adaptation, etc. Different transmission systems have been deployed, such as IP networks, DVB-T/H, UMTS and the terminal devices have been evolving from straightforward TVs, Set-Top Boxes, PCs, to mobile devices such as PDAs, smart phones. The capabilities and working scenarios of these devices vary widely in terms of network bandwidth, computing power, screen display, decoding capabilities and user interactions. At the same time, a variety of new media content formats, standard-compliant or proprietary, have emerged resulting in a diversity of media itself. In summary, multimedia delivery has been growing from single format video transmission over the monolithic data service network to the adaptive delivery of the complex multimedia across heterogeneous networks, terminals, and users [1].

Particularly, the recent advances in video coding have led to the new standard SVC, which enables the scalability in spatial, temporal, and SNR quality, while keeping compression at high efficiency [2]. The scalable video coding allows efficient adaptation directly in the bit stream level without the need of transcoding or re-encoding. Such benefits remove the restriction of the conventional rate adaptation architecture, i.e., bit rate adaptation on the media server or the real-time encoder for appropriate channel rate, possibly with feedback supports from the receivers. A more practical and valuable system topology, as first introduced in [3] for H.264/MPEG-4 AVC and later for SVC [4], is the Media Aware Network Element (MANE) which performs adaptation operations in the middle. In this scenario, the terminal can request the adaptation by means of feedback, carrying QoS metrics, which can be processed by the MANE to make adaptation decisions. Another advantage of this architecture applies to overlay networks. In practice, it is unlikely that every content server has the adaptation capability; therefore, adaptation-capable overlay nodes arguably provide a good approach.

A number of prior efforts have been taken on the multi-dimensional video adaptation under heterogeneous network and terminal conditions [5-8]. [5] formulated the adaptation decision-taking with terminal and network constraints as a multi-criteria optimization problem and formulated it in MPEG-21 context. [6] proposed the similar approach and provided more specific investigation on the SVC adaptation. [7] extended further to incorporate content-based machine learning methods for predicting the relation between adaptation operations and subjective quality. The goal of a comprehensive framework for scalable video streaming with adaptation under MANE topology for heterogeneous delivery was only shared by the lately work [8]. In this paper, we advanced further and presented some more complete results in terms of the system design. The adaptive streaming system consists in a server with a real-time SVC

encoder, an adaptive network node assisted with an MPEG-21 adaptation decision taking engine (ADTE), and a terminal with appropriate perceptual quality, network conditions and user preferences for adaptation support.

The paper is organized as follows. Section 2 presents the adaptive video streaming architecture based on SVC. Section 3 highlights mechanisms for adaptation. We present experimental results in Section 4. Finally, Section 5 draws the conclusion.

2. ADAPTIVE SCALABLE STREAMING SYSTEM

2.1. Overall streaming architecture

To achieve the adaptive streaming with scalable video, the entire system includes three major components, as illustrated in Fig.1, the Server, the Adaptation Node (MANE), and the Client.

- The Server, as a content producer, encodes and streams the live content to the Adaptation Node;
- The Adaptation Node (MANE) adapts the scalable video and relays it to the client, based on the network and terminal constraints.
- The Client decodes the video stream, and feedbacks the end user's perceptual quality for the Adaptation Node.

For simplicity, we assume the connection between the Adaptation Node and the Client to be more problematic, compared to the pair of the Server and the Adaptation Node. This assumption usually holds since the server devices are more professional and well maintained in service.

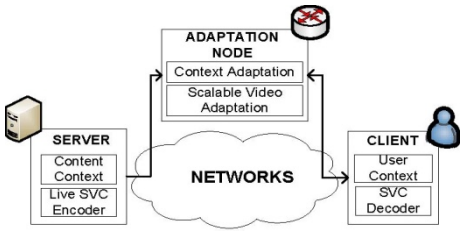

Fig. 1. Adaptive streaming architecture

2.2. Server: real-time SVC encoder

The latest standard Scalable Video Coding (SVC) provides an advanced scalable approach to video coding, by inheriting from H.264/MPEG-4 AVC its highly efficient compression algorithms, such as multiple reference pictures, hierarchical B pictures and the high motion estimation accuracy. This makes SVC much more attractive than its various ancestors, such as MPEG-2 Scalable Extensions, which did not gain much success, due to the high sacrifice in compression efficiency and the significant increase in decoder complexity, compared to the simulcast solution.

The Server is composed of three distinct modules, as shown in Fig.2:

- The real-time SVC encoder, performing high run-time SVC encoding of content captured by a video camera;
- The RTP streamer, streaming SVC bit stream over IP network;
- The mediation service module, for configuration and control.

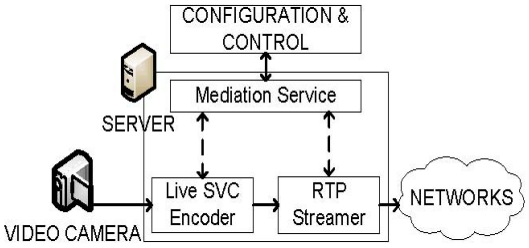

Fig. 2. SVC video server block diagram

The video content is encoded to produce a stream including a base layer and enhancement layers in a hierarchical structure. Each layer, together with all the layers it depends on, forms an operation point, i.e., a representation of the video content at a certain spatial (frame format), temporal (frame rate) and SNR (quality) resolution. The operation points can be extracted from original scalable streams by means of straightforward operations, without need for trans-coding or trans-moding.

To achieve high real-time performance for applications like video-conference, great efforts have been made to design and optimize the SVC encoder. These include both encoding level strategies (e.g., simplified mode decision, reduced search window for motion estimation, etc.) and code level strategies (e.g., using assembly code for computationally intensive code sections). The real-time SVC encoder has been designed to support up to 5 temporal layers (with a hierarchical prediction), 2 spatial layers and 2 CGS layers.

The SVC layers are encapsulated into NAL units, which are the fundamental units of the H.264/MPEG-4 AVC-SVC bit stream and represent packets with an integer number of bytes. A NAL unit starts with a one-byte header, which signals the type of the contained data. The remaining bytes represent payload data and scalability information for SVC. The NAL units are then packetized in RTP packets, in either single mode or non-interleaved mode as in Section 4.2, and transmitted over the network, towards either the Adaptation Node or the Terminal.

An important role in the Server is played by the mediation service: it allows configuring and controlling the Server in the MPEG-21 framework, e.g., to automatically provide the server with several configuration parameters, such as the IP address of the MANE, if present in the chain, the number of layers to be encoded for any scalability dimension, the target bit rate, etc. This provides the SVC encoder itself with the adaptation capability and allows the ADTE to refine and improve the adaptation process and guarantee to the users the best media experience.

2.3. Node in the middle: adaptive SVC MANE system

The SVC MANE, a typical adaptation node in the middle, is depicted in Fig. 3. The SVC adaption module is a passive module controlled by a controller, the Configuration

Manager in Fig. 3, which configures the module, provides media packets from an RTP stream, requests the adaptation to take place and finally retrieves the modified RTP packets.

The functionality of the SVC Header Analyzer consists of selecting the packets from the RTP stream that are required for a successful stream adaptation to network and terminal constraints given by the Configuration Manager.

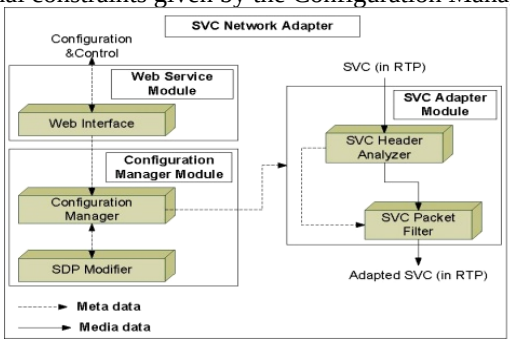

Fig. 3. The Adaptation Node: SVC MANE

An SVC RTP Extractor is instantiated for each pair of destination internet address and port. Each destination describes unicast, multicast or broadcast session. A typical MPEG-21 ADTE computes the best combination of temporal, spatial, and quality layers for a client (i.e., unicast) or set of clients (i.e., multicast, broadcast). The result for each session is then delivered to the SVC MANE at session setup and updated throughout the lifetime of the session. The SVC RTP Extractor is then configured via the Configuration Manager with a spatial layer, a temporal layer, and a quality layer which correspond to the highest media layers that should be delivered in this session. All packets to decode higher layers are not forwarded.

In an SVC stream all packets not belonging to the base layer encapsulate information about the spatial, quality and temporal layers it belongs to. For the base layer, this information is provided by the so-called prefix NAL units. Decoding this field, the SVC Header Analyzer figures out, if the packet needs to be forwarded. In this case, the sequence number field of the RTP packet needs to be modified to hide any packet dropping in the MANE from the Client. Other RTP fields, such as the packet timestamp, are preserved.

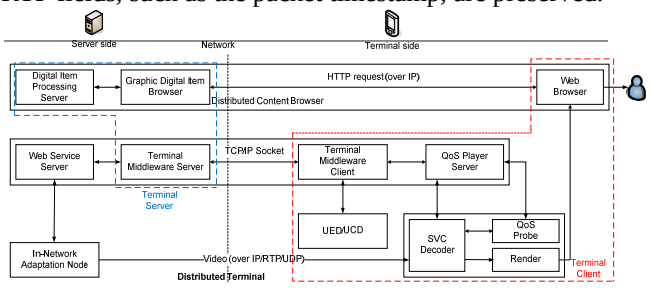

Fig. 4. Distributed terminal architecture

2.4. End user: terminal with PQoS feedback

To support adaptation, a terminal comprising metadata processor, usage environment characteristics, QoS monitor and SVC player has been implemented based on previous result [9]. The above modules are coordinated under a distributed middleware, as illustrated in Fig. 4. Such middleware provides the following responsibilities: 1) externally interacting with server or intermediate service, to feedback QoS for content adaptation or to contact content server for content browsing; and 2) locally coordinating the convergence of media player, usage environment, and video QoS probes.

3. SVC STREAM ADAPTATION METHODS

3.1. Perceptual QoS-application QoS mapping

During the adaptation, a series of QoS mapping are performed, which allows a correspondence between perceived quality levels and practical adaptation parameters.

1) Subjective QoS to Objective QoS mapping: for real-time audiovisual service, such perceptual mapping is required for perceptual criteria. The pQoS probe, which is integrated in the terminal, provides this mapping: it extracts the objective quality by combining objective parameters, such as bit rate b, frame rate f, packet loss rate l, jitter j, quantization parameter q, etc. Such parameters are combined through a non-linear formula defined below, by using specific weighting factors and exponents, which can be adapted to the service and the terminal targeted by the application.

$$MOS \approx B \times b^{bb} + F \times f^{ff} + L \times l^{ll} + J \times j^{jj} + Q \times q^{qq} + \ldots$$

Notably, video quality mapping for adaptation is a classic topic in video processing [10], and it is not our goal to develop a specific or more advanced quality model for our in-network adaptation system. However, the proposed system can be readily modified to accommodate any QoS models.

2) Objective perceptual QoS to Application QoS mapping: the adaptation need to be provided with the correspondence between perceived QoS and the application parameters, such as frame rate, frame size, quantization parameter etc. We use three levels of Application QoS (bronze, silver, gold), and each of them is mapped to a specific range of pQoS (MOS) values, through specific tables, strictly depending on the application. e.g., the frame rate, expressed for the scalable video by the number of temporal layers, can be mapped to such aQoS levels by means as the Table 1.

aQoS	SVC Temporal Layer	Frame Rate(fps)	MOS
GOLD	2	25.000	4.4
SILVER	1	12.500	3.7
BRONZE	0	6.250	3.2

Table 1. Objective QoS to Application QoS mapping

3.2. Adaptation with RTP packets for SVC payload

The Real-time Transport Protocol (RTP) [11] defines a variable-length extension for the RTP header. Such an extension allows particular implementations extending the use of the RTP header with specifically conceived data fields. These fields could be used for SVC adaptation and

layer based error resilience methods. Moreover, a more specific approach for SVC transport over RTP has been established lately [12]. The information in the SVC NAL unit header extension can be re-used in RTP signaling, since the NAL unit header can serve as a payload header for some (non-aggregating) packetization modes. For aggregations, the SVC payload format also defines a new NAL unit in order to provide information for adaptation, Payload Content Scalability Information (PACSI) NAL unit. Our system adopted the original approach but could readily follow this new method.

As described in [2], among the three most common SVC distribution models, the framework proposed in this paper is based on the unicast model. This model permits to get over the problems aroused by the other two, especially in scenario of multiple firewall pinholes. However, the unicast model requires specific methods to guarantee the error resilience of terminals, in scenarios where adaptation is performed, because the RTP sequence number, being related to the entire SVC stream, cannot help identifying losses at a layer-basis. We proposed solutions to detect and signal packet losses, by preserving the compliance with all the involved standards in [13].

4. EXPERIMENTAL RESULTS

We analyze the benefits of integrating adaptation and SVC in streaming scenario. To set up a good evaluation, we use a sufficiently long video sequence with maximum bit rate of 700kbps and the average bit rate of 300kbps.

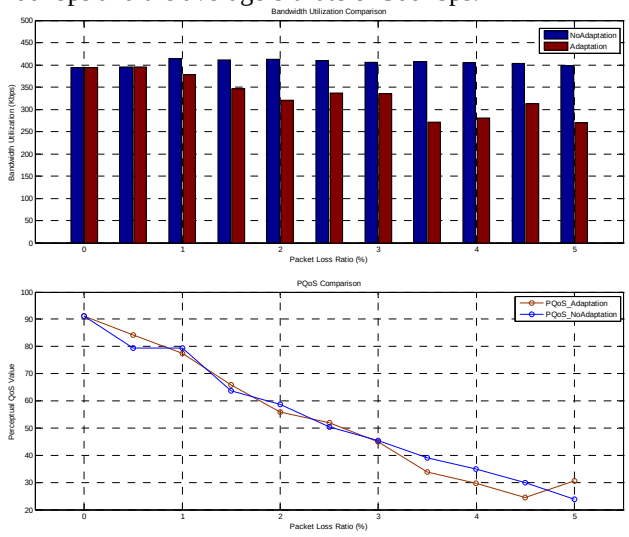

Fig. 5. Bandwidth utilization comparison v.s. packet loss

This evaluation applies to the SVC delivery in adaptation with QoS control. Fig. 5 illustrates the bandwidth utilization and the corresponding perceptual QoS of the two approaches of the SVC transmission without adaptation and our proposed SVC transmission with adaptation under the same packet loss conditions. Bandwidth saving is achieved with comparably little degradation of quality, which is acceptable. We have to notify that these two cases are both using the SVC. If we compare with non-scalable video, the results of bandwidth saving performance will be more.

5. CONCLUSION

A real-time scalable video streaming system with adaptive End-to-End QoS control for multimedia content delivery over heterogeneous networks has been investigated. We present a general adaptive multimedia delivery platform based on SVC and MANE topology. The scheme comprises a real-time SVC encoder, an SVC network adaptation node, and a terminal with probes for adaptation feedback. The experimental results indicate the advantage of such an adaptation system that facilitates bandwidth utility to obtain user's fruitful experience of video quality.

6. REFERENCES

[1] I. Burnett, R. Van de Walle, K. Hill, J. Bormans, and F. Pereira, "MPEG-21: goals and achievements," *IEEE MultiMedia*, vol.10, no.6, pp-60-70, 2003.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. CSVT.*, vol.17, no.9, pp.1103-1120, Sept. 2007.

[3] S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP payload format for H.264 video," *IETF RFC 3984*, Feb. 2005.

[4] S. Wenger, Y.-K. Wang, and T. Schierl "Transport and signaling of SVC in IP networks," *IEEE Trans. CSVT.*, vol.17, no.9, pp.1164-1173, Sept. 2007.

[5] D. Mukherjee, E. Delfosse, J. Kim, and Y. Wang, "Optimal adaptation decision-taking for terminal and network quality-of-service," *IEEE Trans. Multimedia*, vol. 7, no. 9, pp.454–462, Jun. 2005.

[6] T. C. Thang, Y. S. Kim, Y.M. Ro, J. W. Kang, and J.-G. Kim, "SVC bitstream adaptation in MPEG-21 multimedia framework," *Proc. Packet Video*, Apr. 2006.

[7] Y. Wang, M. van der Schaar, S.-F. Chang, and A. Loui, "Classification-based multi-dimensional adaptation prediction for scalable video coding using subjective quality evaluation", *IEEE Trans. CSVT.*, vol 15, no. 10, pp. 1270-1179, Oct. 2005.

[8] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real-time system for adaptive video streaming based on SVC", *IEEE Trans. CSVT*, vol. 17, no. 9, pp. 1227-1137, Sept. 2007.

[9] B. Shao, M. Mattavelli, D. Renzi, M. Andrade, S. Battista, S. Keller, G. Ciobanu, and P. Carvalho, "A multimedia terminal for Adaptation and end-to-end QoS Control," *Proc. ICME*, Jun. 2008.

[10] H. R. Wu and K. R. Rao. "Digital video image quality and perceptual coding", CRC Press, Boca Raton, FL, USA, 2005.

[11] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson (eds.), "RTP: a transport protocol for real-time applications," *IETF/STD 0064, RFC3550*, Jul. 2003.

[12] S. Wenger, Y.-K. Wang, T. Schierl, and A. Eleftheriadis (eds.), "RTP payload format for SVC video," *IETF Internet Draft draft-ietf-avt-rtp-svc-16.txt*, Dec.2008.

[13] D. Renzi, P. Amon, and S. Battista, "Video content adaptation based on SVC and associated RTP packet loss detection and signaling," *Proc. 9th Intl. Workshop of Image Analysis for Multimedia Interactive Services (WIAMIS)*, May 2008.