Advanced Rate Adaption for Unicast Streaming of Scalable Video

Chenghao Liu
Tampere University of Technology
chenghao{dot}liu{at}tut{dot}fi

Imed Bouazizi
Nokia Research Center, Tampere
imed{dot}bouazizi{at}nokia{dot}com

Moncef Gabbouj
Tampere University of Technology
moncef{dot}gabbouj{at} tut{dot}fi

***Abstract*-Recent advances in scalable video coding have paved the way for the development of flexible and adaptive media streaming applications. In order to support rate adaptation, 3GPP Packet-Switched Streaming Service (PSS) has specified a feedback mechanism that includes transmitting information about the status of the client buffer. This paper presents a novel Multiple Virtual Client Buffer Feedback (MVCBF) mechanism, which includes information about multiple media times each of which corresponding to a different set of sub-streams in scalable media streaming. Moreover, a new rate adaptation method is presented that a) assigns different requirements for the buffering time to the different sets of sub-streams, b) observes the media time variation at sub-stream level, and c) efficiently changes the operation point to maintain the required buffering time in priority order. Simulation results show that the proposed MVCBF-based rate adaptation algorithm outperforms the traditional PSS compliant rate adaptation method in the overall throughput by quickly adapting the operation point to the varying network resources and avoiding unnecessary bouncing between operation points.**

I. INTRODUCTION

Scalable video coding has been widely investigated in academia and industry during the past twenty years. However, before the standardization of the scalable video coding (SVC) extension of H.264/AVC [1], scalable video coding was rarely used in commercial applications due to the increased complexity and the significant drop in coding efficiency when compared to non-scalable video coding. Alternatively simulcast, which transmits two or more independent single layer streams simultaneously, were preferred. However, simulcast causes significant increases in the resulting total bit rate. Recently scalable video coding (SVC) [1] [2] has been finalized as an extension of H.264/AVC. SVC introduces single loop decoding and drift control to address earlier shortcomings of scalable video coding. As a result, SVC became an attractive candidate as the media codec of choice in video streaming applications.

To improve the quality of experience for the users, 3GPP Packet-Switched Streaming Service (PSS) [3] supports the adaptation of continuous media. When rate adaptation is performed at the server, the feedback from the client to the server is used at the server to adjust the transmission bit rate and the quality in an end-to-end scenario. In order to signal the client feedback information from the client to the server, 3GPP introduced a new Application-defined Real Time Control Protocol (RTCP) called Next Application Data Unit (NADU) that is delivered as a report block together with other RTCP feedback information.

In the literature, a comprehensive discussion of adaptive streaming within the 3GPP packet-switched streaming service has been presented in [4]. Kampmann *et al.* [5] proposed a 3GPP PSS compliant stream switching solution. In the proposed solution, stream switching is carried out based on certain thresholds of the total media time in the client buffer and using NADU feedback at the server. Schierl *et al.* [6] presented a 3GPP PSS compliant adaptive video streaming strategy for H.264/AVC encoded stream. Here, a transmission rate estimation algorithm was proposed based on NADU and RTCP Receiver Report (RR) from the client to the server. Based on the estimated transmission rate, adapting the video bit rate is realized by the combination of Bit-Stream Switching and Temporal Scalability. However, all those adaptive video streaming solutions are carried out based on a single layer client buffer feedback.

In the streaming adaptation using SVC, using switching flexibility between different medium-grain quality scalability (MGS) layers is essential for achieving accurate adaptation of the transmission rate to varying wireless link conditions. To fully utilize the quality switching flexibility in SVC, knowledge of the reception status at sub-stream level is required at the server side. The client buffer feedback mechanism provided in 3GPP PSS is a good candidate to provide feedback to the server for monitoring the reception status. However, 3GPP PSS lacks sub-stream based client buffer feedback, particularly when the different layers are transmitted over the same RTP stream. When monitoring the client buffer status using the client buffer feedback specified in 3GPP PSS, a straightforward way is that the server estimates the client buffer status on a stream level and applies the estimation to the entire set of sub-streams. However, it is worthwhile to assign different target protection levels to the different layers of the stream, as this will ensure playback continuity with graceful degradation in case of significant packet loss. Thus, the straightforward way results in intolerable inaccuracies in the client buffer estimation at sub-stream level. A more advanced way to estimate the client buffer status for each sub-stream is presented in the sequel. Basically the server has the entire view on the transmission order of packets, their decoding order and their playout timestamp. Hence, the server can estimate the client buffer status such as playout media time on client buffer for each sub-stream under the assumption that first transmitted packets from the server arrived firstly at the client (called first transmitted first arrived in this paper). However, first transmitted first arrived is not guaranteed in

978-1-4244-6404-3/10/$26.00

SVC video transmission, since the packets of different layers are scheduled with different prioritization during transmission. Hence, current solutions provided in 3GPP PSS have limitations on monitoring the client buffer status on a sub-stream level which limits the stream adaptation accuracy. In addition without strategy for assigning different target buffer protection time for different layer of a sub-stream such as in 3GPP PSS, re-buffering happens more frequently as the more important sub-stream of the base layer does not get appropriate high protection.

Contrary to previous works, MGS switching flexibility is exploited when adapting a scalable stream to varying wireless link conditions in the proposed rate adaptation mechanism. A sub-stream level of a client buffer feedback mechanism is proposed in this paper. Based on the proposed feedback mechanism, the server precisely monitors the reception status of each sub-stream as well as adapts the transmission rate to the varying wireless link by switching between different sub-streams of different quality refinement layers. In addition, the sub-stream level of the target buffer protection time is proposed to provide the prioritized protection for the sub-stream of important base layers compared to enhanced layers.

The rest of the paper is organized as follows. Section 2 gives a flexible bit stream adaptation in H.264/SVC. The proposed scalable streaming adaptation based on Multiple Virtual Client Buffer Feedback (MVCBF) and Target Virtual Buffer Protection Time (TVBPT) is presented in Section 3. Simulation results and conclusions and future work are presented in Sections 4 and 5, respectively.

II. Flexible rate adaptation with SVC

The SVC extension to H.264/AVC, called SVC also in this paper, is one of the most promising scalable video coding standards nowadays. SVC enables to produce a single scalable bit-stream which contains a base layer and several enhancement layers. A scalable video bit-stream is structured in such a way that it allows the extraction of more than one sub-stream, each representing a different operation point (quality, temporal resolution, and spatial resolution). An overview of SVC extension of H.264/AVC is presented in [2].

For quality scalability, SVC supports coarse-grain quality scalability (CGS) as well as medium-grain quality scalability (MGS). CGS achieves quality refinement by decreasing the quantization step when encoding the residual data in the spatial domain. For encoding a refinement quality layer, CGS supports inter-layer prediction tools (used in spatial scalability) including motion prediction, residual prediction and intra-prediction. In addition, inter-layer prediction can be directly performed in the transform domain for intra prediction and residual prediction. The number of bit-stream switching points in CGS is restricted by the frequency of IDR access units as switching layers in CGS is only possible at random access points. In contrast to CGS, MGS enables packet-based quality scalability by allowing bit-stream switching between different MGS layers at any access unit. In SVC, the transform coefficients of the enhancement layer may be split over multiple slices. The coefficient scan index specifies the specific transform coefficients that are included in a given slice. Hence, the transform coefficients that correspond to the same quantization step may be distributed into several network abstraction layer (NAL) units [2].

III. Proposed Fine Granular Rate Adaptation

We propose a mechanism based on MVCBF and TVBPT. Based on the above mentioned feedback and metrics, a novel rate adaptation technique using SVC is proposed.

A. *Definitions and Terminology*

In SVC rate adaptation, a set of operation points (OP) is used to adapt the transmission rate to the varying network resources. An OP consists of a set of layers of the SVC bit stream that enables the decoding of the video at certain media-specific characteristics, i.e. the spatial resolution, the temporal resolution, and the quality level. Switching to a higher/lower OP is performed by adding/dropping a portion of the bit stream for transmission, which is denoted as sub-stream in this paper. Note that a sub-stream only consists of exactly one single scalability layer. Furthermore, each sub-stream is provided with a sub-stream identification (ID) to describe the highest scalability layer of each sub-stream. The sub-stream ID of value 1 identifies the portion of the bit stream that represents OP 1. Otherwise, (for a sub-stream ID larger than 1), the sub-stream ID identifies the portion of the bit stream that needs to be added to the next lower OP resulting in the OP with the corresponding ID. The scalability information which characterizes each operation point may be signaled from the server to the client using appropriate extensions of RTSP [7] or SDP [8].

A Multiple Virtual Client Buffer (MVCB) is used to store the media data of each sub-stream, although, in reality, the media data from all sub-streams are stored in the same physical client buffer. The MVCBF proposed in this paper includes the feedback messages for each of the virtual client buffers.

The Target Buffer Protection Time (TBPT) specifies the pre-buffering time to provide interruption-free play-out. The TVBPT proposed in this paper specifies the sub-stream level of pre-buffering time. The TVBPT decreases along with the increase of sub-stream ID.

B. *SVC Bit Stream Transmission System Based on MVCBF and TVBPT*

A block diagram of SVC bit stream transmission system based on MVCBF and TVBPT is shown in Fig. 1. At the SVC server side, the server adapts the transmission rate to the varying network recourses by changing OP. Based on the TVBPT, the server monitors each sub-stream reception status, e.g. deduces the media time of each sub-stream in the client buffer. The SVC stream adaptation is carried out as explained in Section 3.D.

At the client side, the client generates and transmits the MVCBF message, which contains the feedback messages for each virtual client buffer (VCB). To create the MVCBF, the

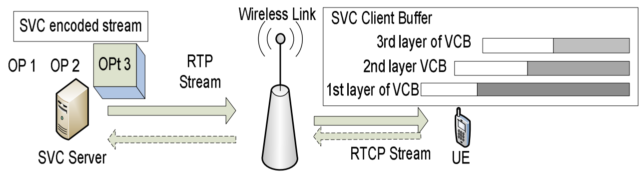

Fig. 1. SVC bit stream transmission system based on MVCB feedback

client sorts the incoming media packets to the corresponding VCB based on the scalability information, i.e. dependency_id, quality_id and temporal_id as extracted from the Network Abstraction Layer (NAL) header of each packet and the VCB ID.

C. Multiple Virtual Client Buffer (MVCB)

For providing MVCBF, a new RTCP application feedback message named Multiple Virtual Client Buffer (MVCB) is proposed in this paper. MVCBF is encapsulated into RTCP packet as an application-dependant report block. The Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF) can be used for the transmission of RTCP [9]. Fig. 2 shows the Report block format of MVCBF and the semantics are defined as follows. In MVCBF, the definition of SSRC, playout delay, NSN, reserved, NUN and FBS are identical with the definitions for the client buffer feedback in 3GPP PSS. These fields are not described here and the reader is referred to the 3GPP PSS specification [3].

The new fields, introduced by MVCBF, are defined as follows:

NVCB (8 bit): indicates the number of VCB report blocks, which corresponds to the number of sub-streams at the current OP.

VCB ID (8 bits): identifies the sub-stream that corresponds to the virtual client buffer. It starts from 1 and ends at NVCB.

VNUN (5 bits): indicates the unit number of the next ADU to be decoded from the corresponding virtual client buffer that is identified by VCB ID.

VHSNR (16 bits): indicates the highest sequence number received in an RTP data packet for the corresponding VCB.

VLSNR (16 bits): indicates the lowest sequence number received in an RTP data packet for the corresponding VCB.

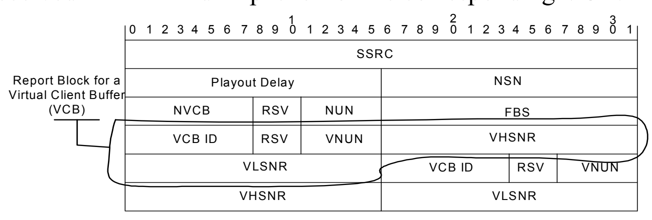

Fig. 2. Report block format for the MVCBF

D. Rate Adaptation Based on MVCBF and TVBPT

The flowchart for the rate adaptation is presented in Fig. 3. Upon receiving the MVCBF, the media time level for the VCB with VCB ID i is calculated according to the following formula

$$MT(i) = TS_{VHSNR}(i) - TS_{VLSNR}(i) + PD \quad (1)$$

where $MT(i)$ denotes the media time, $TS_{VHSNR}(i)$ denotes the timestamp of the $VHSNR$ and $TS_{VLSNR}(i)$ denotes the timestamp of the $VLSNR$, PD denotes the playout delay and i identifies the VCB ID. The expression presented in [5] to calculate the media time, where i is equal to 1, is a special case of (1).

The VCB media time comparison module compares the media times of each VCB against the 5 thresholds to decide appropriate adaptation actions as described in the following paragraph. This operation is repeated for each VCB starting with the VCB with lowest VCB ID as shown Fig. 3.

The thresholds differ from one sub-stream to another, in order to ensure a higher protection time for more important SVC layers. For each sub-stream, the adaptation includes increasing the current OP, dropping all higher sub-streams, accelerating the transmission rate of the current sub-stream, and finally dropping the current sub-stream as shown in Fig. 4. The VCB initial buffering (VIB) module checks whether the media time of the VCB of the lately added sub-stream has accumulated to the TVBPT and increases the OP to the next higher one when true. The TVBPT decreases gradually along with increasing VCB ID. The VIB module drops the lately added sub-stream as soon as the media time of the VCB of the lower sub-stream is observed to have decreased beyond a specific threshold as shown in Fig. 5.

Additionally, the OP is increased after certain time intervals to check and make use of any recently available additional bandwidth as shown by the left side blocks of Fig. 3.

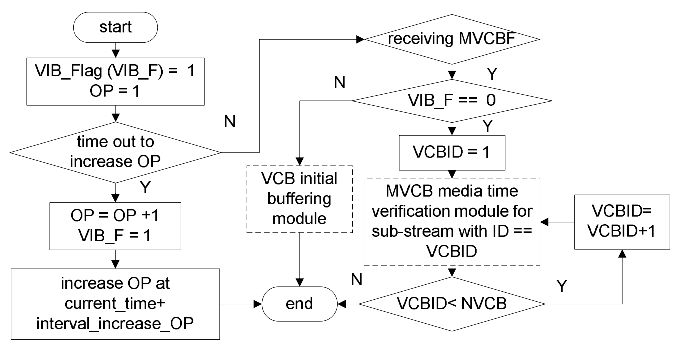

Fig. 3. Flowchart of scalable stream adaptation based on proposed method

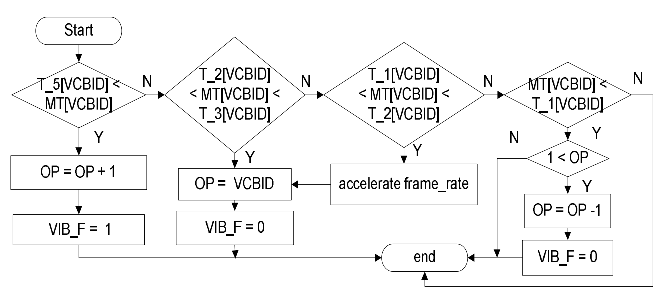

Fig. 4. MVCB media time verification module

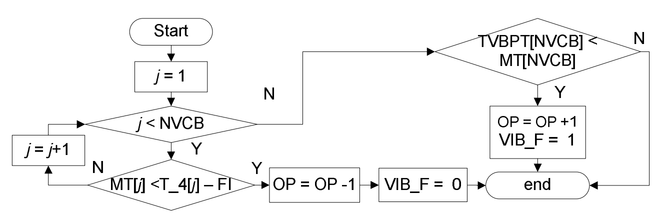

Fig. 5. VCB initial buffering module

IV. SIMULATION RESULTS

The JSVM software version 9.6 is used as the reference software of H.264/SVC. QVGA-resolution (320 × 240 luma samples) test sequences including *City* and *Crew* were encoded at a fixed frame rate 12.5 fps. A base layer and two quality enhancement layers achieved by setting QP equals 33 as the base layer and QP equals 32 and 31 as the two enhancement layers. Each enhancement layer is further divided into 3 MGS layers. Hence 7 sub-streams are used to provide 7 OPs. The PSNR and the bitrate for each OP of the two encoded sequences are shown in Fig. 6. Two different scalable rate adaptation algorithms, namely the proposed method and the PSS Client Buffer Feedback (CBF) based method are implemented in NS2. The proposed method carries out the streaming adaptation as defined in section 3.D. The 5 thresholds (denoted as T_i) for each of the 7 VCBs, each corresponding to an MGS encoded sub-stream, are shown in Table I. The large TVBPT is assigned to the lower layer of VCB to ensure playback continuity. In order to switch to a lower OP as smooth as possible, the distance between T_4 and T_3 as well as the distance between T_2 and T_1 decrease linearly with the increase of VCB layer. The fluctuation_interval (FI) in Fig. 5 is set to 250ms. Whereas PSS CBF based method carries out the rate adaptation based on the client buffer feedback specified in 3GPP PSS. Fig. 7 shows the flowchart for scalable stream rate adaptation based on PSS single CBF. The PSS CBF based method calculates a single media time using (1) which is compared against the predefined 4 thresholds. For the CBF approach, the rate adaptation includes increase of the OP, decrease of the current OP to the next lower one, transmission of the current sub-stream at the accelerated frame rate, and decrease of the current OP to the next lower one. During the initial buffering phase, which ends after elapsing TBPT since receiving the first packets, the rate adaptation is not performed. In the CBF based method, the 4 thresholds, T_1 to T_4, are set to 8.0s, 6.5s, 6.0s, and 5.5s, respectively. TBPT is set to 7.0.

The network topology of our simulation is shown in Fig. 8. Node 0 and node 4 are the SVC streaming server and client, respectively. Node 1 and node 5 are additionally inserted to the network as UDP source and sink with traffic rate 150Kbps in order to simulate background traffic. Nodes 0, 1 and nodes 4, 5 are connected through node 2 and node 3, respectively. All network links have a fixed bandwidth, except the link between node 3 and node 4 as shown in Fig. 8.

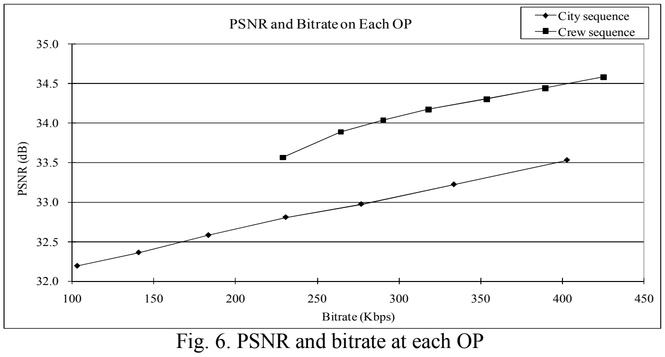

Fig. 6. PSNR and bitrate at each OP

TABLE I THRESHOLDS (T) AT EACH VCB LAYER FOR THE MVCBF METHOD

VCB Layer	T_1 (s)	T_2 (s)	T_3 (s)	T_4 (s)	T_5 (s)	TVBPT (s)
1	1.2	5.0	5.7	7	8	7.2
2	1.5	4.8	5.6	6.7	7.7	6.9
3	1.8	4.6	5.5	6.4	7.4	6.6
4	2.1	4.4	5.4	6.1	7.1	6.3
5	2.4	4.2	5.3	5.8	6.8	6.0
6	2.7	4.0	5.2	5.5	6.5	5.7
7	3.0	3.8	5.1	5.2	6.2	5.4

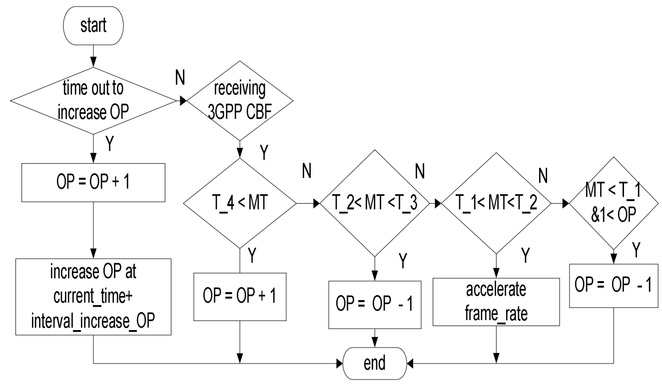

Fig. 7. Flowchart of scalable stream rate adaptation based on PSS CBF

The RTCP reporting interval is set to 500ms. Fig. 9 and Fig. 10 shows the links between node 3 and node 4 which varies between 170 Kbps and 430Kbps for the *City* sequence and between 260 Kbps and 460Kbps for the *Crew* sequence, respectively. In addition, Fig. 9 and Fig. 10 depict the transmission rates (averaged over 10 second intervals) for the proposed method and for the PSS CBF based method for the *City* and *Crew* sequences, respectively. The simulation results show that the transmission rate of the proposed method approaches the link rate better than the PSS CBF based method. The average deviations between the transmission rate and the link capacity for the proposed method and for the PSS CBF based method are 38.96 Kbps and 113 Kbps, respectively, for *City* and 79 Kbps and 99Kbps, respectively, for *Crew*. Fig. 11 and Fig. 12 show the actual OP for the *City* and *Crew* sequences, respectively. The OP of the proposed method remains stable upon reaching the optimal layer whereas the OP for the PSS CBF based method fluctuates dramatically. Both methods increase the OP to probe for the availability of additional end-to-end bandwidth that may then be used for improving the video quality. After a failed probing, the proposed method decreases the OP back to the optimal level. In contrast, the PSS CBF based method decreases the OP that stabilizes at a level lower than the optimal level which results in increased oscillation between OPs and sub-optimal rate adaptation performance.

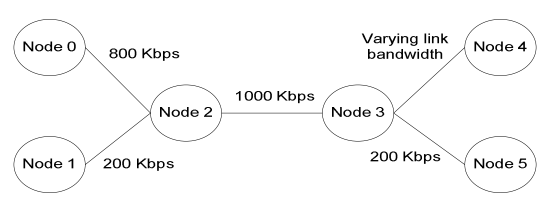

Fig. 8. Network topology

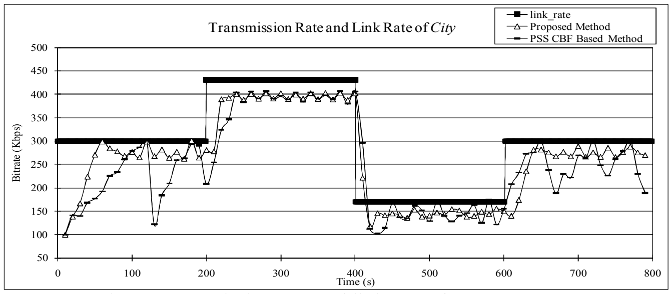

Fig. 9. Transmission rate and link rate of *City* sequence

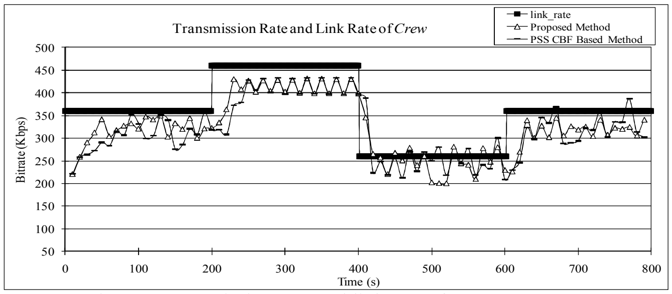

Fig. 10. Transmission rate and link rate of *Crew* sequence

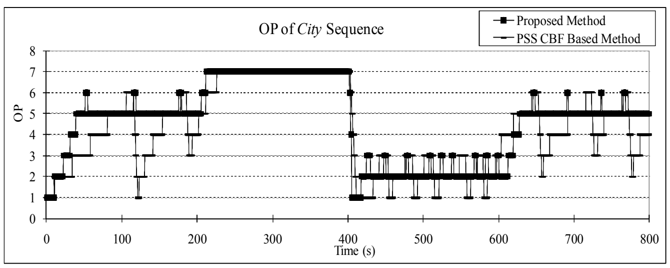

Fig. 11. OP of *City* sequence

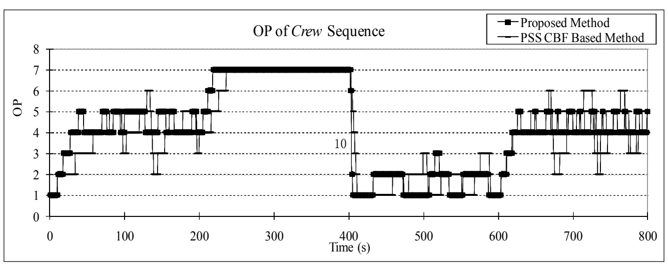

Fig. 12. OP of *Crew* sequence

Fig. 13 shows the media time level in the client buffer for the C*ity* sequence. VCB1 to VCB7 denote the media time for the 7 virtual client buffers based on the proposed method; the single Client Buffer (CB) denotes the media time level for the single CBF as specified in 3GPP PSS. Eq. (1) is used to calculate the media time, where i is set to a value from 1 to 7 for the proposed method, and i is set to 1 for the PSS CBF based method. The media time decreases as the VCB layer increases from 1 to 7 as shown in Fig. 13. So the size of the client buffer in the proposed method is smaller than that of the client buffer in the PSS CBF based method, which assigns the same target buffer protection time for all VCB layers. Furthermore, the media time of the PSS CBF based method drops significantly to 0 at time 418s due to the late rate adaptation, which causes re-buffering and playback interruption.

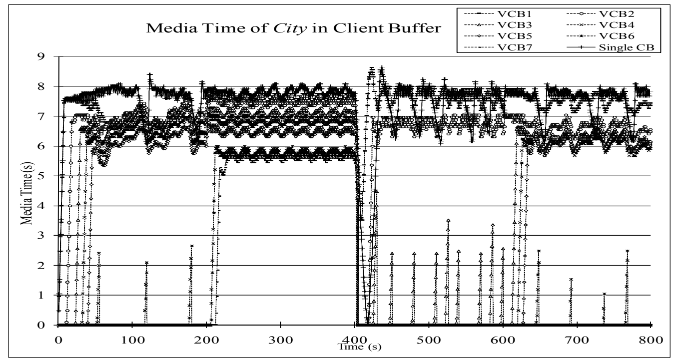

Fig. 13. Media time of *City* sequence in client buffer

V. CONCLUSIONS AND FUTURE WORK

A scalable rate adaptation method based on Multiple Virtual Client Buffer Feedback (MVCBF) and Target Virtual Buffer Protection Time (TVBPT) was proposed in this paper. MVCBF reports the multiple virtual client buffer feedback to the server to enable monitoring of the available media time in the client buffer at sub-stream level. TVBPT specifies high pre-buffering time for the base layer compared to enhancement layers. By monitoring the packet reception status and client buffer at a sub-stream level, precise and priority-based rate adaptation is realized at the server in order to cope with varying network recourses. Simulation results show that the proposed method based on MVCBF and TVBPT outperforms the scalable rate adaptation method based on client buffer feedback as specified in 3GPP PSS. In future work, we will investigate the theoretical aspects of the MVCBF based rate adaptation method.

ACKNOWLEDGMENT

This work was supported by the Academy of Finland, (application number 129657, Finnish Programme for Centres of Excellence in Research 2006-2011).

REFERENCES

[1] *Advanced Video Coding for Generic Audiovisual Services,* ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, v3: 2005, Amendment 3: Scalable Video Coding.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, 2007.

[3] 3GPP TS 26.234, “Transparent End-To-End Packet-Switched Streaming Service (PSS): Protocols and Codecs (Release 6

[4] P. Fröjdh, U. Horn, and M. Kampmann, “Adaptive Streaming within the 3GPP Packet-Switched Streaming Service,” *IEEE Network*, vol 20, no 2, pp. 34-40, March-April 2006.

[5] M. Kampmann, and C. Plum, “Stream switching for 3GPP PSS compliant adaptive wireless video streaming,” *IEEE Consumer Communications and Networking Conference, 2006 (CCNC 2006)*, Jan. 2006.

[6] T. Schierl, M. Kampmann, and T. Wiegand, “3GPP Compliant Adaptive Wireless Video Streaming Using H.264/AVC,” in IEEE Int. Conf. Image Proc., Genova, Italy, Sept. 2005.

[7] H. Schulzrinne, A. Rao, R. Lanphier, "Real Time Streaming Protocol (RTSP)", RFC 2326, April 1998.l

[8] Handley, M., Jacobson, V, and C. Perkins, "SDP: Session Description Protocol", RFC 4566, July 2006.

[9] Ott, J., Wenger, S., Sato, N., Burmeister, C., and J. Rey, “Extended RTP profile for real-time transport control protocol (RTCP)-based feedback (RTP/AVPF),”IETF RFC 4585, July 2006.