

A Comparison of Request and Measurement Based End to End Video Target Rate Adaptation Methods

Arto Mahkonen

Senior Research Scientist, NomadicLab, Ericsson Research

P.O. Box 58, FIN-02131 Espoo, Finland
+358 40 769 3903

arto.mahkonen@ericsson.com

Toomas Wigell

Research Scientist, NomadicLab, Ericsson Research

P.O. Box 58, FIN-02131 Espoo, Finland
+358 40 507 1332

toomas.wigell@ericsson.com

ABSTRACT

In this paper, we compare a request based video target rate adaptation method to a measurement based adaptation method. It is shown that by using methods, that use both estimation of the transport delay and its variation and packet loss jointly, it is possible to reduce packet losses and transport delay variation, when the transport channel is temporarily congested or suffering from other bad conditions (like fading radio access). The lower packet loss and transport delay variation consequently improves the perceptual quality of the video signal significantly. Further it is shown by simulated experiments, that request based video adaptation method is less sensitive to the nominal RTCP sender report (SR) interval than a measurement based adaptation, when same mathematical algorithms for target rate estimation are used.

Categories and Subject Descriptors

G.4 [Mathematical Software]: - *Algorithm design and analysis, Verification*; H.1.1 [Models and Principles]: Systems and Information Theory - *Information theory*; H.4.1 [Information Systems Applications]: Communications Applications - *Computer conferencing, teleconferencing, and videoconferencing*; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems - *Video (e.g., tape, disk, DVI)*; I.6 [Simulation and Modeling]: - *Applications, Model Validation and Analysis*

General Terms

Algorithms, Measurement, Documentation, Performance, Experimentation, Standardization, Theory, Verification.

Keywords

Video, target rate, adaptation, packet loss, transport delay variation.

1. INTRODUCTION

In conversational video telephony services the video encoder

often applies a rate control function, which tries to keep the average source bit-rate constant and as close to a given target bit-rate as possible. This is usually called a constant bit-rate (CBR) video.

In packet switched networks like Video Telephony over IP (VToIP) the transport delay of video packets will vary due to several reasons. One of the reasons is the varying load on the transport links, which may cause low frequency fluctuation of transport delay and even bursty congestion peaks. These may last up to several seconds and may cause degradation of perceptual quality like annoying delay variation (or jerkiness), problems with audio synchronization and even packet losses.

The integrity of decoded video is usually damaged due to packet losses. Video compression is heavily based on motion compensation by vectors (usually called motion vectors). It uses previously decoded frames as references for the future ones and thus the loss of quality is also progressing in correctly received subsequent frames via the earlier corrupted reference frames. How long these disturbances last depends on the used INTRA refreshing method, which means encoding of a frame or parts of it without motion compensation (i.e. without references to the previous frames).

There are several methods to cope with this kind of disturbances. One of them is to apply an end-to-end video target rate adaptation. It means that, when congestion or disturbance begins, the receiving end tries to influence on the sending end to reduce the used target rate of the CBR video encoder as soon as possible. When the congestion or disturbance is over the receiver influences to the sender to let the target rate slowly recover back up to its original value. This paper concentrates to the video target rate adaptation as the method of coping with this kind of situation.

In the video telephony service the rate adaption is usually enabled in both directions, which means that both ends serve as video sender and receiver simultaneously. However this paper describes everything from the view of one way sender-receiver model in order to avoid unnecessary complexity in the description, because the other direction is an asymmetric replica. However it is assumed, that the service in this context is a two way video telephony service (like for instance IMS Multimedia Telephony [5]), and not a one way streaming service.

There are two different principles to accomplish an end-to-end target rate adaptation. In this paper the principles are called request based and measurement based video target rate adaptation methods.

Figure 1 shows the principles of a request based video target rate adaptation method. In this the video receiver measures certain indicators based on the characteristics of the received packet flow. Our investigation has included the transport delay variation and packet loss in the indicators.

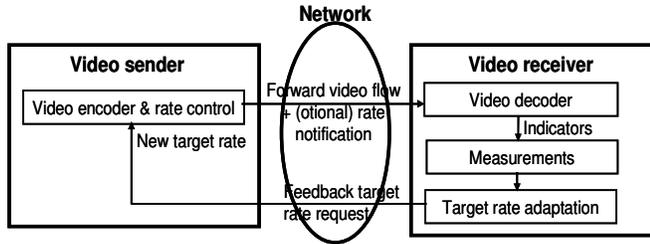


Figure 1. Request based video target rate adaptation.

Based on these measurements the receiver calculates a proper target rate, and if it is changed compared to the previous value of the target rate, then it will be sent back to the video sender as a request.

Figure 2 shows the principles of a measurement based target rate adaptation method. This deviates from the request based method, in that it sends the measurements as often as possible back to the video sender, who then calculates with the help of the received measurements a proper target rate to be used by the video encoder. In other words the video receiver does not know in this alternative, whether the measurements will cause the target rate to be changed or not.

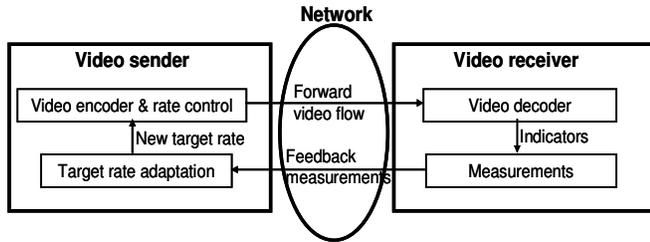


Figure 2. Measurement based video target rate adaptation.

This paper compares the performance of the request based and measurement based video target rate adaptation methods to each other. Further we assume that media is transported on RTP over UDP and feedback uses RTCP over UDP. Also it is assumed that the request based algorithm is able to use the early RTCP SR according to RFC 4585 ([3]) for downwards rate requests and regular RTCP SR according to RFC 3550 ([2]) for both up- and downwards requests. The reason for this is that target rate must react downwards quickly, but upwards the changing speed can be relaxed. It is assumed that either RTCP-APP ([2]) or Temporary Maximum Media Bit-rate Request or TMMBR ([4]) is used as the carrier of the feedback in the request based adaptation method. The measurement based method is using only the regular compound RTCP SR ([2]), because the new target rate is not calculated in the video receiver but first in the video sender, and the video receiver does not know, whether fast reduction of the target rate is required. In measurement based algorithm the feedback is based on the current information carried by the compound RTCP SR and no extra carrier is required.

The paper is organized as follows. Section 2 is an overview of the related work, section 3 describes the outlines of the model setup,

section 4 describes the experimental results and finally section 5 provides conclusions.

2. RELATED WORK

For multimedia applications transported over RTP the most important alternatives from congestion control point of view are either replacement of non-congestion-controlled UDP with congestion-controlled DCCP protocol (see for instance [1]), or accomplish the congestion control by RTP/RTCP protocol itself (see for instance [2], [3] and [4]).

There are not many widely agreed and standardized congestion control algorithms. Besides congestion control algorithms intended for the TCP, the most known algorithms are TCP-like congestion control (RFC 4341, specified for DCCP, ref. [3]), and TCP-friendly Rate Control (RFC 3448, ref. [6]).

As a general principle, the TCP-friendly algorithms model the TCP congestion control behavior, so that they are TCP-friendly over longer term but behaving different in short term. Usually the intention is to smooth the rate change intensity.

Unfortunately, the TCP-friendly algorithms emphasize the packet loss as a main congestion event. As a result, the algorithms allow quite high packet loss during congestion - behavior that is not quite suitable for real-time video applications.

Real-time media friendly congestion control mechanisms should consider the QoS requirements for the real-time media applications. This means that the real-time media friendly congestion control should rather be proactive – by sensing coming congestion and performing adaptation before packet loss ever happens. Where the packet loss indicates more a persistent congestion, a sudden increase of delay predicts a coming congestion.

A coming congestion on a bottleneck link can also be handled by means of cross-layer signaling locally or by indicating it to the traffic sender by some marking in a feedback signal, like ECN ([9], [10]).

ECN-based methods require ECN-compatible nodes along the connection and are depending on ECN marking algorithms in the nodes. Currently the ECN marking is defined for TCP and DCCP traffic only and ECN marking algorithms in routers assume TCP compatible congestion behavior in hosts.

3. MODEL SETUP

Experiments were made by using a simulator that consisted of a channel model and a probe user model with real video and audio source containing encoders and destination containing decoders as well as both ends consisting of required parts of request and measurement based video target rate adaptation algorithms as alternative options.

The channel model consisted of cascaded components that approximate UMTS enhanced uplink (EUL), core network (CN) and UMTS high speed downlink packet access (HSDPA). The crossing (background) traffic was simulated inside the channel models with the help of dummy user models.

3.1 Video target rate adaptation algorithm

Mathematically the video target rate adaptation algorithm used in this experiment is the same in both the request and measurement

based cases. It uses both transport delay variations and packet losses jointly as indicators.

In request based adaptation the transport delay is estimated by the equation (1) in the video receiver.

$$d = t_{arr} - t_{send} + x \quad (1)$$

where t_{arr} = arrival time set by the receiver's clock, t_{send} = sending time derived from timestamp of received video RTP-packet and set by the sender's clock and x = unknown offset due to for instance the clock skew (or offset) between sender's and receiver's clocks including also other possible unknown constant offsets.

In measurement based adaptation the transport delay estimation is a bit more complicated, but can be estimated by the equation (2) in the video sender. This equation is actually giving the transport delay of the most recent forward RTCP SR from video sender to video receiver. Note, that in measurement based method it is assumed that RTCP SR uses the same flow (or bearer) as the corresponding RTP, so that it experiences the same characteristics.

$$d = (t_{RTCP} - d_{dlsr}) - t_{lsr} + x \quad (2)$$

where $t_{RTCP} - d_{dlsr}$ represents the arrival time of the last forward RTCP SR packet from video sender, t_{RTCP} = sending time of the backward RTCP SR packet carrying these measurements from the video receiver to the sender, t_{RTCP} is derived from the RTP timestamp set by the clock of the video receiver and carried by the sender info field of a RTCP SR packet, d_{dlsr} = the delay since the reception of the last forward RTCP SR message from video sender until the sending time of the next backward RTCP SR message, d_{dlsr} is derived from the dlsr-field of the backward RTCP SR message, t_{lsr} represents the sending time of the most recent forward RTCP SR message, which is derived from the lsr-field of the backward RTCP SR report and which is a 32 bit integer carrying the middle bits of the 64 bit NTP-timestamp of the most recent forward RTCP SR packet and x = unknown offset due to for instance the clock skew (or offset) between sender's and receiver's clocks including also other possible unknown constant offsets ([2]).

In the request based adaptation method the packet loss is measured in the video receiver by monitoring, whether the sequence number (SN) of the received RTP packets increases smoothly by one or not. In the measurement based adaptation method the packet loss is indicated in the video sender by monitoring the cumulative-number-of-packets-lost field of the received backwards RTCP SR packet

This paper does not describe the algorithm in any details. Only the overall behavior is presented in the following steps. The steps are performed per each received video frame (or packet).

- 1) Wait for the next frame (in request based method) or RTCP SR (in measurement based method) and measure the individual transport delay d_i - with the help of equation (1) or (2) respectively - and packet loss indicators
- 2) Estimate the low pass (LP) filtered fixed delay $d_{fix,i}$ by an asymmetrically weighted (AWLP) infinite impulse response (IIR) filter by the equation (3).

$$d_{fix,i} = (1 - a) * d_{fix,i-1} + a * d_i \quad (3)$$

where $a = \max(1/i, 0.0005)$, if $d_i > d_{fix,i-1}$ else $a = \max(1/i, 0.005)$ and i = sample index, where 0.0005 and 0.005 were defined experimentally and max-operator is used to get faster settling time for the coefficient a at the beginning. So $d_{fix,i}$ adapts downwards 10 times faster than upwards, and it is in that sense representing a kind of LP-filtered minimum detector. Note, that $d_{fix,i}$ includes the unknown offset x . The meaning of the equation (3) is to have a smooth enough adaptive delay reference, that is approaching the minimum delay, but is smoother than just simple minimum operator. This is the basis for the thresholds $th_{low,i}$ and $th_{high,i}$ defined in the step 4.

- 3) Estimate the signal power level (p_i) of the short term jitter ripple by LP-filtering the absolute value (h_i) of the high pass (HP) filtered delay. The HP-filter is a simple finite impulse response (FIR) filter. Because the sum of its coefficients is 0, the unknown offset x is canceled out. This step is handled by the equation set (4).

$$h_i = |1 * d_i - 2 * d_{i-1} + 1 * d_{i-2}|$$

$$p_i = (1 - b) * p_{i-1} + b * h_i \quad (4)$$

$$b = \max(1/i, 0.001)$$

where 0.001 was defined experimentally and max-operator is used for faster settling time at the beginning.

The purpose of this step is to get margins for the thresholds $th_{low,i}$ and $th_{high,i}$ defined in the step 4, and the margins are proportional to the amplitude of the short term variation (i.e. the ripple) of the delay.

- 4) Based on the estimated AWLP-filtered fixed delay ($d_{fix,i}$) and the power level of the short term jitter ripple (p_i), calculate low (th_{low}) and high (th_{high}) thresholds according to equations (5). Note that these include the unknown offset x .

$$th_{low,i} = d_{fix,i} + 1.0 * p_i \quad (5)$$

$$th_{high,i} = d_{fix,i} + 3.5 * p_i$$

where the coefficients 1.0 and 3.5 were set experimentally.

- 5) Based on the comparison and analysis of individual delay values d_i and $th_{low,i}$ and $th_{high,i}$ as well as packet loss, decide whether video target rate should be changed up or down. Note that here the unknown x is canceled out, because it is included in both the individual delay and the thresholds, so that compensation of clock skewing between sender and receiver is not needed. Optionally some timing constraints (like not before the previous request is acknowledged etc.) can be included in this decision.

If the delay seems to clearly exceed $th_{high,i}$, which is the minimum delay increased by a certain margin, that is proportional to the level of the short term delay variation, or packets are lost, then target rate should be

reduced. On the other hand, if the delay seems to mainly and persistently be inside the current short term delay variation ($th_{low,i}$) and no packets are lost, then the target rate is allowed to slowly increase until a configured maximum rate is reached.

- 6) If the target rate is requested to be changed smaller, then go to the step 8, else continue to step 7.
- 7) If target rate is requested to be changed bigger, then go to step 9, else go back to the starting step 1 and no feedback is needed in the request based algorithm.
- 8) Change the target rate smaller by a certain function of the individual delay d_i , packet loss, thresholds and the current target rate. The function is chosen so, that the rate is decreased quite quickly in order to respond fast to a possible congestion or a degradation in radio conditions. In request based method send the decreased target rate to video sender either by an early non-compound RTCP SR, if previous early was sent before the previous regular, or else postpone the sending to the next regular compound RTCP SR packet. Finally go back to the starting step 1.
- 9) Change the target rate bigger by a certain function of the individual delay d_i , packet loss, thresholds and the current target rate. The function is chosen so, that the rate is increased quite slowly in order to peek, whether a possible congestion or a degradation in radio conditions is really over. In request based method send the increased target rate to video sender only by a regular compound RTCP SR message, because the changing pace of the target is allowed to be relaxed in the upwards direction. Finally go back to the starting step 1.

Note, that the target rate calculation functions in steps 8 and 9 are also insensitive to the unknown offset x .

3.2 Used test case

The used test case for the probe user consisted of a video clip. Its duration was about 270 s. A congestion period of 150 s was simulated to start at 20 s and stop at 170 s. It was accomplished by increasing the number of dummy crossing users at 20 s in the channel model, so that it was congested, and releasing them at 170 s. In addition to this, one of the dummy users was switched on and off repeatedly during the congestion period to create short delay peaks.

H.263 baseline video codec ([11]) was used for the probe user's video and AMR NB 12.2 kbps mode ([12]) for audio. Used test video clip was recorded with QCIF (176 x 144) frame resolution and with 12.5 frames per second. The maximum target rate for video was configured to 128 kbps and minimum to 64 kbps. The target rate was set to 128 kbps at the beginning and it was allowed to be adapted between these extremes.

Figure 3 shows the delay profiles of the video packets, when no target rate adaptation method was used. The delay profile on the top shows the behavior, when no packets were lost due to buffer overflow. Usually packets that have been pending too long time in buffers or in scheduling queues are dropped, which causes packet losses. The profile on the bottom shows the behavior as an example, when packets were dropped (i.e. lost), if they were pending in the transport channel model more than about 400 ms.

This caused a packet loss rate (PLR) of 1.4 % during the congestion period.

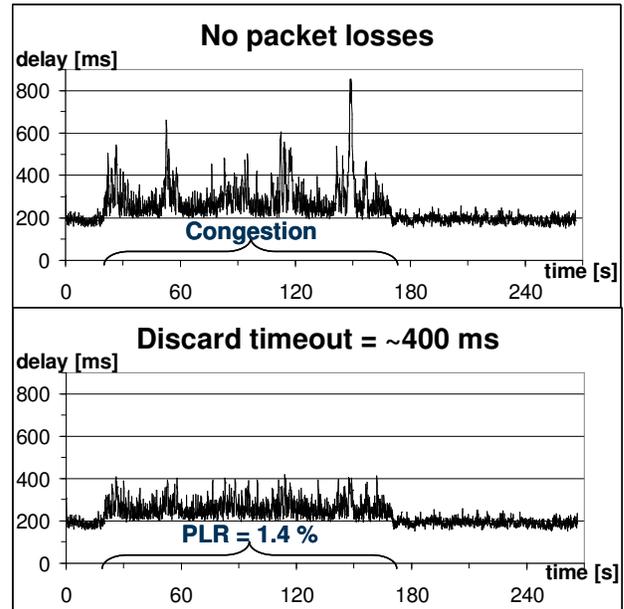


Figure 3. Delay profiles without rate adaptation.

4. EXPERIMENTAL RESULTS

When the request based target rate adaptation is applied to the lower alternative of the Figure 3, then the delays will behave as shown in the Figure 4 as a function of the regular RTCP SR interval. The figure on the top shows the profile, when the interval is 0.2 s, the figure in the middle corresponds to a case, when the interval is 0.5 s, and the figure in the bottom shows the delay, when the interval is 1 s. It can be seen, that the delay behavior is quite insensitive to the RTCP SR interval, when request based target rate adaptation is used. It also shows that $PLR = 0$ independently of the interval.

When the measurement based target rate adaptation is applied in the same alternative, then the delays will behave as shown in the Figure 5 as a function of the regular RTCP SR interval. The figure on the top shows the profile, when the interval is 0.2 s, the figure in the middle corresponds to a case, when the interval is 0.5 s, and the figure in the bottom shows the delay, when the interval is 1 s. It can be seen, that the delay behavior depends on the RTCP SR interval, when measurement based target rate adaptation is used. It also shows that $PLR = 0$ only with the shortest interval of 0.2 s.

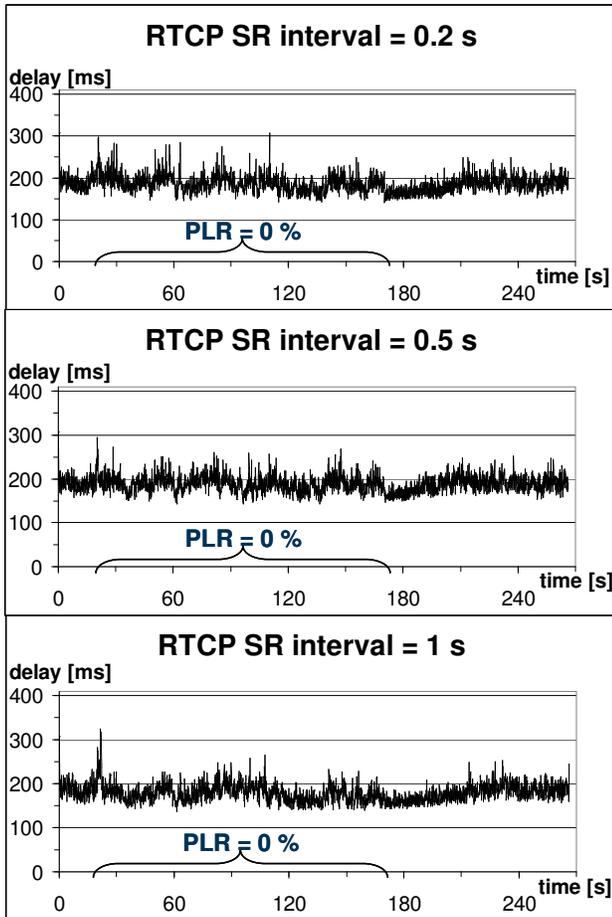


Figure 4. Delay profiles with the request based rate adaptation.

Figures 6 and 7 show the behavior of the adapted target rate with different RTCP SR intervals, when the request and measurement based adaptations are applied respectively. They show that with the request based adaptation the behavior of the target rate is less sensitive to the RTCP SR interval than with the measurement based method.

Figure 8 shows the complementary cumulative distribution functions (CCDF) of the delays with target rate adaptation. The figure on the top shows CCDF:s as a function of RTCP SR interval, when the request based method is used, and the figure at the bottom shows CCDF:s, when the measurement based adaptation is used. With request based adaptation 99 % of the delays are below ~230 ms almost independently from the RTCP SR interval. This can be seen as the x-values of the crossing points of the line $y = 0.01$ with the respective CCDF curves.

With measurement based adaptation the crossing point depends more on the used RTCP SR interval so, that with the interval of 0.2 s 99 % of the delays are below 270 ms and with intervals 0.5 s and 1 s the crossing point is at about 350 ms.

A peak signal to noise ratio (PSNR) is a well known method to compare different video processing methods to each other. However it does not always correlate well with the perceptual quality especially, when disturbances are bursty like in this case. However it seems that by looking the behavior of the cumulative

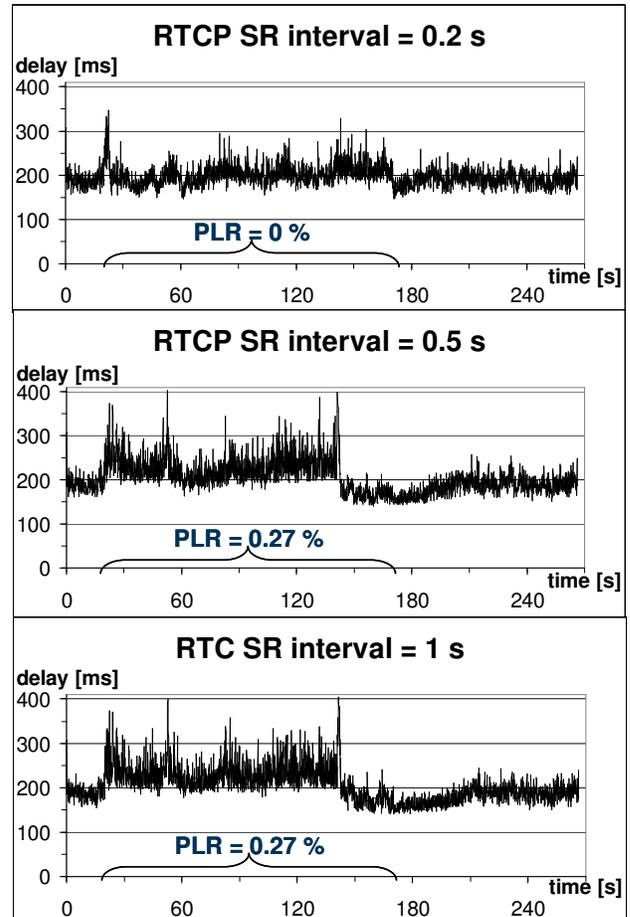


Figure 5. Delay profiles with the measurement based rate adaptation.

distribution functions (CDF) of PSNR-values a better correlation with perceptual quality can be achieved. The leading edge of the CDF reaches smaller values with distortions than without them, but the PSNR averages are still almost equal. Human eyes are usually sensitive to this kind of extreme quality degradations, which do not so much reveal themselves in the averages.

Figure 9 shows the CDF:s of the PSNR:s. The figure on the top shows CDF:s, when the request based method was used, and the figure at the bottom corresponds to the cases when the measurement based adaptation was used. In both figures the CDF corresponding to the simulation without rate adaptation is also included as a reference.

With the request based adaptation the quality is very independent from the RTCP SR interval and about 99 % of the frames will give a PSNR value that is bigger than 27 dB.

With the measurement based adaptation the quality is about the same with RTC SR interval of 0.2 s, whereas with the intervals of 0.5 s and 1 s the quality is reduced mainly due to lost packets. Jerkiness due to jitter is not revealed by the PSNR, because long enough static jitter buffering has been applied, so that rendering does not lose any frames due to too late arrivals.

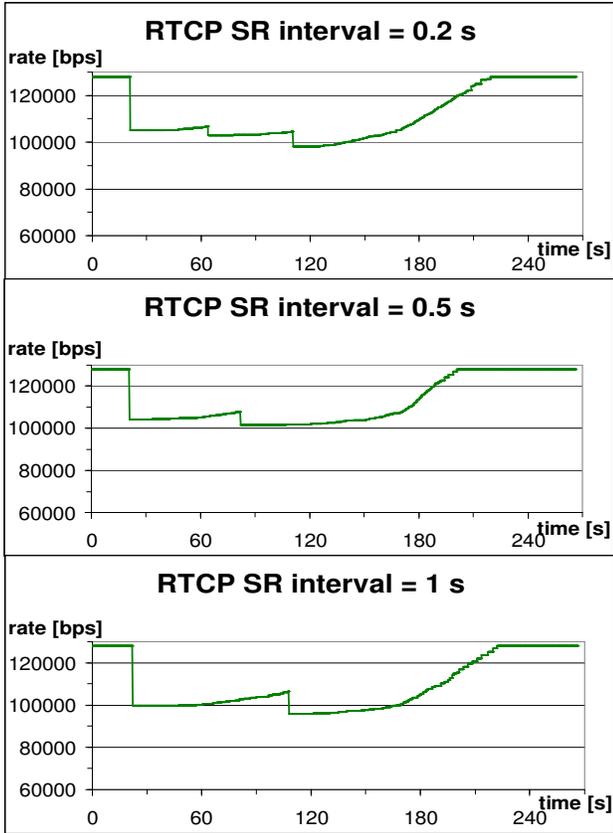


Figure 6. Target rate profiles with request based adaptation.

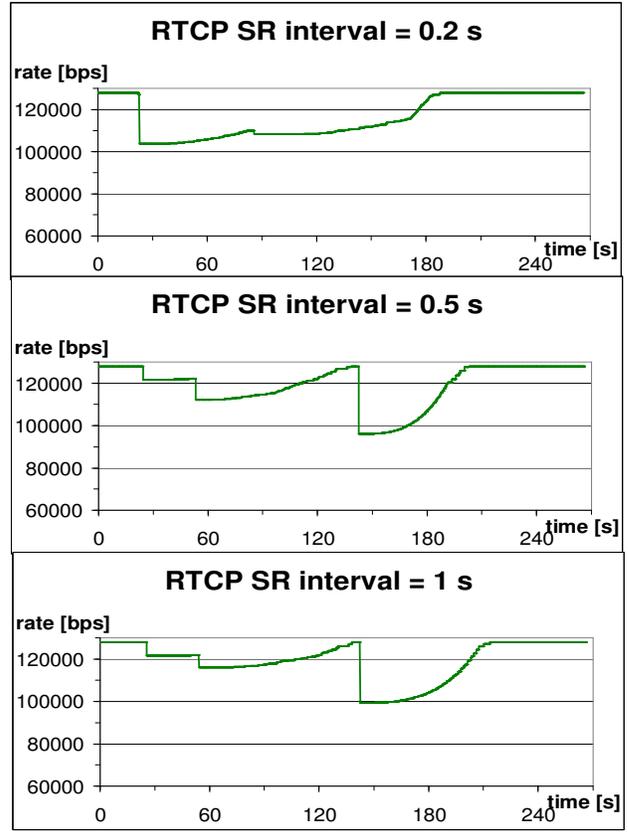


Figure 7. Target rate profiles with measurement based adaptation.

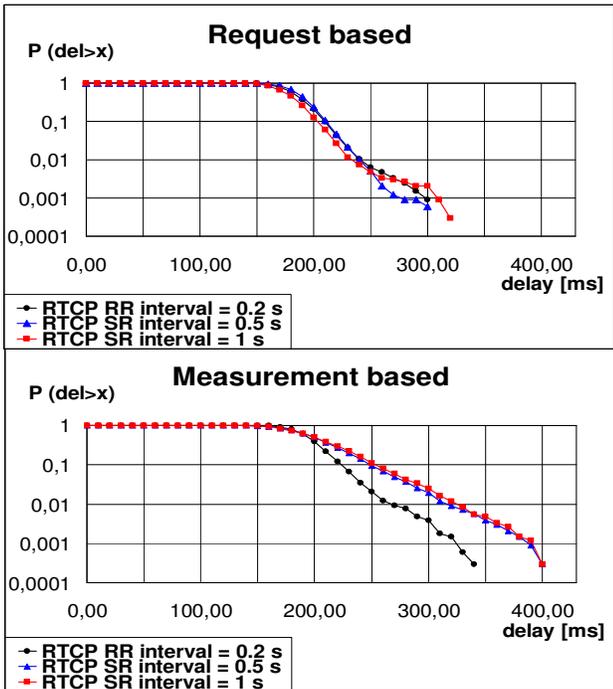


Figure 8. CCDF:s of the delays with rate adaptation.

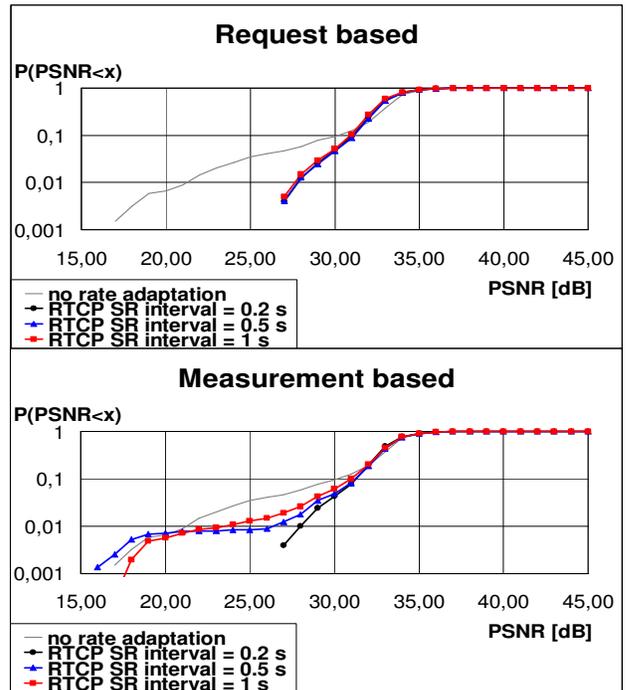


Figure 9. CDF:s of PSNR:s with rate adaptation.

The amount of needed bit-rate for the regular compound RTCP SR packets with different intervals is given in the Table 1. The assumption is that an RTC SR packet requires 140 bytes. As we noticed above the performance of the measurement based video target rate adaption is almost equal to the request based adaption only with the shortest RTCP SR interval of 0.2 s. It means, that with request based algorithm it is possible to have a longer RTCP SR interval (e.g. 1 s) than with measurement based adaptation, which requires a short interval (e.g. 0.2 s) in order to have about the same performance. If we further take into account that early RTCP SR is used only for downwards rate requests, allowed only once between regular RTCP SR messages ([3]) and assume that the early RTCP SR could be transported as a non-compound message ([8]), then it could be possible to save feedback signaling capacity by a factor of about 5 by using request based target rate adaptation instead of measurement based.

Table 1. bitrates needed for RTCP SR

RTCP SR interval [s]	Bitrate [bps]
0.2	5600
0.5	2240
1	1120

5. CONCLUSIONS

This paper illustrates that the end-to-end video target rate adaptation method reduces packet losses and transport delay variation significantly with constant bit-rate (CBR) video services, when they are suffering from congestion or bad transport conditions (like fading in radio access) compared to the case without rate adaptation. Consequently the perceptual quality is improved.

Further it illustrates that using the measured delay variation as one of the indicators for the target rate adaptation acts as a predictor of the congestion or bad transport conditions, so that they are detected already before packets are lost.

Further this paper shows that the request based video target rate adaptation is optimized with respect to time response and transport capacity, when compared to the corresponding measurement based adaptation using the same indicators and same mathematical algorithm. The saving of transport capacity is true with the conditions, that early RTCP SR is used only for downwards rate requests, its transmission is not allowed more

often than once between two consecutive regular compound RTCP SR messages ([3]) and that non-compound RTCP ([8]) is available for the early rate request messages.

6. REFERENCES

- [1] RFC 4336. S. Floyd, M. Handley, and E. Kohler. Problem Statement of the Datagram Congestion Control Protocol (DCCP). March 2006
- [2] RFC 3550. H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. July 2003
- [3] RFC 4585. J. Ott, S. Wenger, N. Sato, C. Burmeister, and J. Rey. Extended RTP Profile for Real-time Transport Protocol (RTCP)-Based Feedback (RTP/AVPF). July 2006.
- [4] IETF Internet-Draft: draft-ietf-avt-avpf-ccm-04. S. Wenger, U. Chandra, M. Westerlund, and B. Burman. Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF). March 5, 2007 (exp: May 2007).
- [5] RFC 4341. S. Floyd, and E. Kohler. Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control. March 2006.
- [6] RFC 3448. Sally Floyd, Mark Handley, Jitendra Padhye, Jörg Widmer. TCP Friendly Rate Control (TFRC): Protocol Specification. January 2003
- [7] 3GPP TS 26.114, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction. March 2007.
- [8] IETF Internet-Draft: draft-ietf-avt-rtcp-non-compound-02. I. Johansson and M. Westerlund. Non-compound RTCP in RTCP profile.
- [9] RFC 3168. K. Ramakrishnan, S. Floyd, and D. Black. The Addition of Explicit Congestion Notification (ECN) to IP. September 2001.
- [10] RFC 3540. N. Spring, D. Wetherall, and D. Ely. Robust Explicit Congestion Notification (ECN) Signaling with Nonces. June 2003.
- [11] ITU-T H.263. Video coding for low bit rate communication.
- [12] 3GPP TS 26.090. Adaptive Multi-Rate (AMR) speech codec; Transcoding functions