Ausarbeitung Anwendungen 2
SS2011

Fabian Jäger

Scalable Videocodec in a Video Conference

Fakultät Technik und Informatik
Department Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Topic of this paper

Scalable Videocodec in a Video Conference

Keywords

Videoconferencing, SVC, scalable video codec, bandwidth estimation, adaptive codec, iPhone, mobile video conference, Daviko, Placecam, H.264

Abstract

This work focuses on the design of an adaptive, bandwidth aware streaming strategy for real-time video data in a conferencing system called PlaceCam by Daviko [1]. In this paper we analyze and discuss the research that has already been done in this area. We hope to find suggestions and approaches for our work.

Contents

List of Figures

1 Introduction

Compressing video data in a videoconference is an important task to improve its efficiency. In a videoconference with multiple participants, the hardware and available bandwidth of each participant may differ. Therefore it is an important task to scale the amount and complexity of the videodata for each participant individually. This can be done with the scalable video codec (SVC), which is an extension to H.264 [11]. The ability to scale the video data is not a new feature, but it is rarely used. The SVC could be a solution for the slow Internet connection many smartphones that are able to deliver and receive high quality video have to cope with [2] [3].

With these codecs it is possible to reduce the amount of data for clients with low bandwidths by dropping certain packages with additional video data that are not necessary to decode the video stream. This allows us to send each client as much data as possible without jamming the connection, if we know the available bandwidth. In this paper we will discuss various research results concerning this topic.

This paper is organized as follows: Section 1 gives an introduction and an overview of the research area. Section 2 describes the topics we are interested in. In section 3 papers which seems important to us, are presented and discussed. Concluding remarks are given in section 4.

2 Research area

There are several people and groups who have already done some research in this area. Even though their work is partly different from ours, there are certain aspects which could be used as helpful reference. We are especially interested in research about bandwidth estimation, video scaling and the combination of the two, so we tried to find parts of them in different projects and use the experience others have already made. We are going to discuss a lot of the prior research on this topic, and take a closer look on some of the work that has already been done.

The main topics we are interested in:

1. We have to find a way to determine the currently available bandwidth, which is a difficult task. Therefore we need a solution which is very fast (because we want real-time multimedia streaming), stable and if possible without any data overhead. As sustained, bandwidth may change rapidly, so we need to refresh the bandwidth information periodically.

2. H.264 has multiple options for scaling the video data. We must figure out which of them are the best in Quality of Experience (QoE) for a video conference and whether this decision depends on the participants or not. For example a mobile participant with UMTS connection could prefer a mode which is not optimal for a mobile participant with a WLAN link. Unfortunately, the scalable video codec enhances processing requirements, so we must be careful to not overload the source of the videostream. The first thing here is to compare the performance, the amount of extra data and the suitability for our purpose in each mode.
3. The scale of the video codec depends on the observed bandwidth, so we must find a way to adapt these two parts. Therefore we are looking for projects where this has already been done and try to find an approach which is adaptable to the requirements of a videoconference.

3 Related Work

In this section we present interesting papers which focus on the same research area as we do. We will give a short introduction to the first two paper and discuss their relevance for our project. The last paper will be discussed in detail, because it examines some very interesting aspects of our project.

3.1 Adaptive Bandwidth Measurement and Monitoring for Multimedia Transmission in Home Networks

Adaptive Bandwidth Measurement and Monitoring for Multimedia Transmission in Home Networks [6] is a paper by Ruchao Gao, Chaorui Zhang, Jiekai Zhang, Deyuan Li,Rong Yu and Shengli Xie from the University of Technology in Guangzhou, China. They proposed an idea on how to implement a scalable video transmission based on a bandwidth estimation.

3.1.1 Basic idea

Goa and his colleagues implemented a framework which basically consists of two modules. The Available Bandwidth Measurement (ABWM) Module constantly measures the currently available bandwidth and provides this information to the SVC module, which is shown in figure 1. Figure 2 shows the SVC module, which determines the number of enhancement layers for every client. Every enhancement layer increases the quality of the video unless it jams the network and causes packet loss. The balance between the quality and amount of video data is very important to guarantee a good Quality of Experience (QoE).

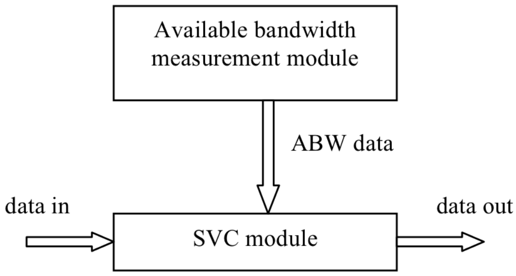

Figure 1: The available bandwidth is constantly measured and provides the SVC Module with information about the network conditions

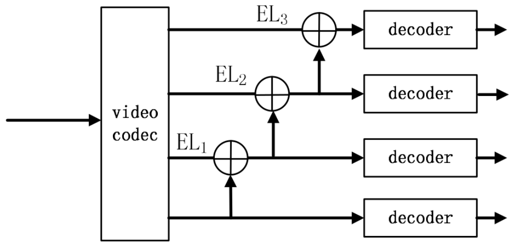

Figure 2: The number of Enhancement Layers (EL) depends on the available bandwidth

3.1.2 Bandwidth estimation

To estimate the available bandwidth, a relatively new Packet Rate Model (PRM) algorithm named PathChirp is used. The main idea is to send probing packets with a certain initial rate to the receiver. The receiver compares the rate of the probing packets he receives to their initial rate. If they do not differ, it usually means that the maximum available bandwidth has not been reached and that the initial rate will be increased by the sender. This is done until the rate on the receiver side is lower than the initial rate, which means that the maximum available bandwidth is exceeded.

The measurement must be constantly repeated in order to recognize changes of the network conditions. The refresh intervals depend on the network and the amount and type of the competing traffic, for example if the competing traffic is caused by a music streaming application the traffic should be fairly constant. If the competing traffic is HTTP traffic, the amount of data is rather low until a new webpage is ordered. This causes a short peak of data [9] therefore the refresh time should be increased if the measurement results are often changing.

3.1.3 SVC

The number of layers is an important factor for the scalability of the SVC. The more layers we have, the more possibilities we have to add or remove layers for certain clients, which increases or reduces the amount of network traffic and quality of the video. Every extra layer costs CPU power on the sender side [4]. The size of an enhancement layer packet can change rapidly, depending on the characteristics of the video. The maximum size of an enhancement layer packet is well known and could be used for a save bandwidth calculation, but most of the enhancement layer packets are far smaller than the maximum size. This approach would waste bandwidth. Therefore it is proposed to measure the size of each enhancement layer packet and use the average for the bandwidth calculation.

3.1.4 Conclusion and relevance

In this paper a very clear suggestion is made about how to design a bandwidth aware video transmission. Measurements are made to test the precision of the available bandwidth estimation. We will use a similar design for our project, but a different bandwidth estimation. The problem with PathChirp is, that it might not react fast enough to any changes on the network conditions and that it produces too much overhead due to the probing packets.

The SVC usage and the combination of the enhancement layers with the available bandwidth is an interesting approach and we will try to use it for our project. Since video conference stream characteristics differ from those normal videos have, we have to see how they inflict the size of the enhancement layer packets. The average of the video packet size will be very small until the person on the video conference makes fast movements which will cause a peak. Therefore, we must find out whether we should use an average function or find another approach to determine the needed bandwidth. Also, the hardware of smartphones is much slower than normal PCs and therefore we can just encode a few enhancement layers, which reduce the scalability.

3.2 Optimization Bandwidth Sharing for Multimedia Transmission Supporting Scalable Video Coding

Optimization Bandwidth Sharing for Multimedia Transmission Supporting Scalable Video Coding [12] is a paper by Talebi, M.S. and Khonsari, A. and Hajiesmaili, M.H. from the University of Tehran, Iran where the behavior of multiple SVC video streams on a network is examined. They developed an algorithm to optimally distribute the available bandwidth to several sources of video streams.

3.2.1 Approach

The idea is to distribute several video streams through a network as best as possible. The paper is very theoretical and complex, which is why we cannot to explain the algorithm in detail. The basic idea behind the algorithm is to calculate a shadow price for each link. The shadow price is used as a factor to determine the bandwidth for each stream and is calculated for each link using the method of Lagrangian multipliers based on the available bandwidth, the number of stream on the link and the scalability factor of the video streams.

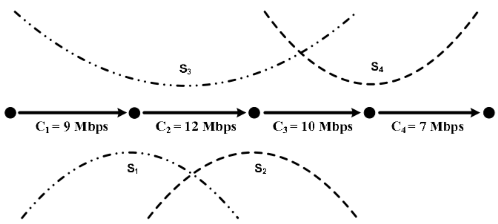

Figure 3: *Network with 4 videos streams*

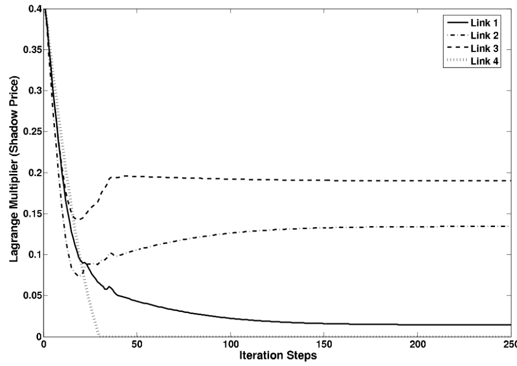

Figure 4: *Shadow price results for each link*

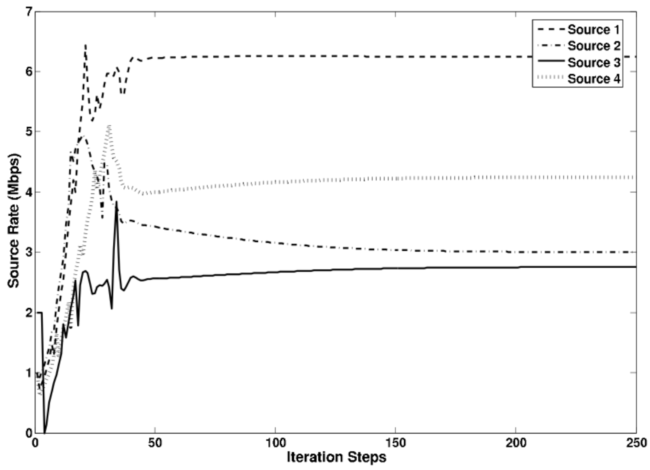

Figure 5: *Source Rate (Mbps) for each sender*

Video streams normally have different characteristics regarding their scalability. Every enhancement layer needs a certain amount of extra bandwidth, which depends on the encoder settings of the sender. If free bandwidth is available, it will allocate to a sender which is able to add an additional enhancement layer with the extra amount of bandwidth. This factor is also considered in the algorithm. The result is an indicator about the load factor of the link. This calculation is repeated a few times and with every iteration, a new shadow price is calculated based on the old shadow price, an updated bandwidth estimation and the amount of transmitted video data since the last calculation.

The shadow price result for the network in figure 3 is shown in figure 4. As we see, the link 4 with just one stream has a negative shadow price (because the link contains only one stream and therefore must not been shared) while link 2 and 3 have the highest shadow prices because they both contain three video streams.

Figure 5 shows the results for the bandwidth distribution on this network. The available bandwidth of each link is shared by the senders based on the amount of stream on the links, the available bandwidth and the scalability factor of the video streams.

3.2.2 Conclusion and relevance

This paper includes the development of an algorithm which splits an available bandwidth between several senders. The theoretical aspect of the algorithm is precisely explained and gives a good insight of bandwidth sharing.

This work could be relevant for us, since it focuses a lot of problems we usually have during a multiple user video conference. When several senders send data to one receiver, they might jam its network. In that case, every sender would recognize the jam and try to reduce the network traffic. For that matter, the jam would disappear and the senders would start to increase the video quality until the receiver is jammed again, but since we have a video conference software which is aware of each participant we could use this information to calculate the maximum bandwidth for each sender. This way we reduce the possibility of a buffer overflow on the receiver. The overhead is acceptable since we only need to share the information at the beginning of a session and they will not change during the conference.

3.3 Adaption Strategies for Streaming SVC Video

Adaption Strategies for Streaming SVC Video [7] is a paper by Burak Görkemli and A. Murat Tekalp from the Koç University in Instanbul, Turkey. In this paper they analyze the combination of the available bandwidth on a network with a scalable video stream. They focus on the behavior of the video quality in consideration of available bandwidth, retransmission of lost video packets and congestion control mechanism.

For their measurement, a relatively new protocol named DCCP [8] is used, which is a transport protocol for real-time applications. It provides a congestion control and allows flow-based semantics like TCP, but abandons a reliable in-order delivery, which slows real-time traffic down. Due to this characteristics, DCCP is a protocol which can be used for real-time data and is therefore often used in multimedia applications.

DCCP supports several congestion control mechanisms which have an identifier named Congestion Control Identifier (CCID). Görkemli and Tekalp compare the TCP-like congestion control (CCID2) and the TCP Friendly Rate Control (TFRC) [5] (CCID3) for their research. A TFRC ensures competing TCP traffic a certain amount of bandwidth unlike UDP for example, which does not consider any other traffic and always uses the maximum bandwidth. The results of their measurements are compared to standard TCP.

3.3.1 Testbed

For the measurement a simple network with one sender and one receiver is used. Also simulated HTTP packets are used as competing traffic. The sender extracts a Group of Pictures (GoP) [10] and sends the extracted video packets to the receiver. This occurs with a certain rate that depends on the network conditions and the extraction rate of the video. The latter is an information the sender already has, while the available bandwidth of the network needs to be estimated by the sender. The receiver inserts the packets in a buffer and starts decoding as soon as the buffer is half-filled. If the sending rate is higher than the decoding rate of the video, the buffer fills-up and could result in a buffer overflow. The sender can react to buffer overflows by decreasing the sending-rate, but it would be better to avoid buffer overflows in advance.

To prevent buffer overflows on the receiver side and also to provide competing flows with as much bandwidth as possible, they implement an application layer rate control. After extracting the video packets the dispatch to the sending queue can be delayed by a certain time. They implemented 3 different modes of rate control in the application:

1. Sending the video packets at the available bandwidth rate. This might be a waste of resources if the available bandwidth is higher than the video rate.
2. Sending the video packets at the video extraction rate, which can jam the network if the video rate is higher than the available bandwidth.
3. Sending the video packets at the available bandwidth rate, but never faster than the maximum video extraction rate.

Two video sequences with different characteristics and 2880 frames each, are used for the measurement. The size of the GoP is 16 frames and the Byte-limit slice mode is set to 1400 Bytes to guarantee that they will not fragment in the IP-layer. The SVC is used with 1 base layer and 2 enhancement layers.

3.3.2 Retransmission

An important point in real-time communication via networks is the retransmission of lost packets. If the receiver recognizes a packet loss, it will requests these packets again using an Automatic Repeat reQuest (ARQ) scheme depending on whether there is enough time for retransmission, which can be estimated by the size of the decode queue.

They use three different request schemes for their application:

1. Request all missing packets.
2. Request only base layer packets, but never an enhancement layer packet.

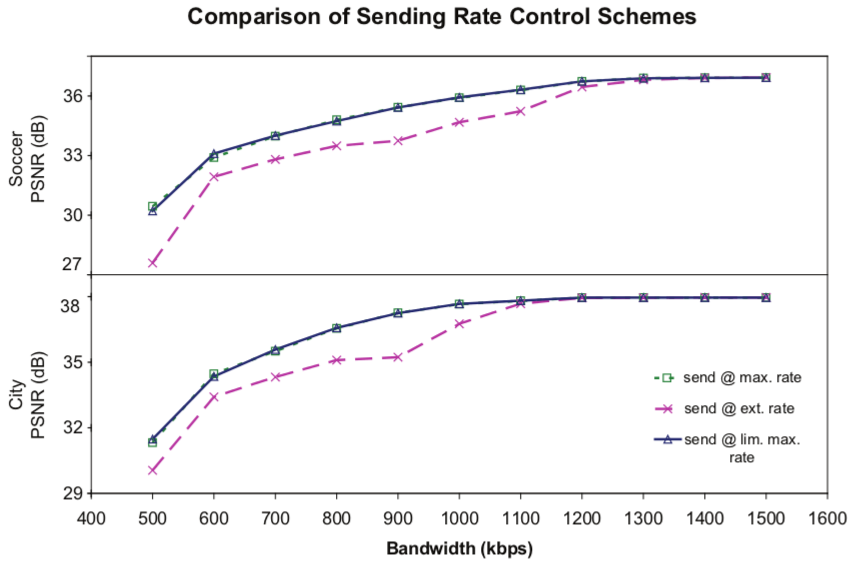

Figure 6: *Comparison of rate control schemes using DCCP CCID3*

3. Request all base layer packets and the enhancement layer packets only adaptively. The decision if a lost enhancement layer packet should be retransmitted or not depends on the decoder buffer. Unless the retransmission could cause an overflow, the lost enhancement layer packets are requested.

3.3.3 Bandwidth estimation

DCCP CCID3 continuously calculates the available bandwidth based on the packet loss and the delay statistic of the network. In this case no further work is necessary to get the available bandwidth. TCP and DCCP CCID2 do not offer this feature and therefore the calculation needs to be done on the application layer. The available bandwidth is estimated by the maximal sending rate of the video packets, which is not an optimal but at least a simple solution.

3.3.4 Results

Several measurements have been performed, but we are especially interested in the rate control measurements. In figure 6 the quality of the two videos is shown for three different rate strategies. The green line resembles the result for the video quality if the sending rate is set to the maximum available bandwidth, the purple line shows the results when sent

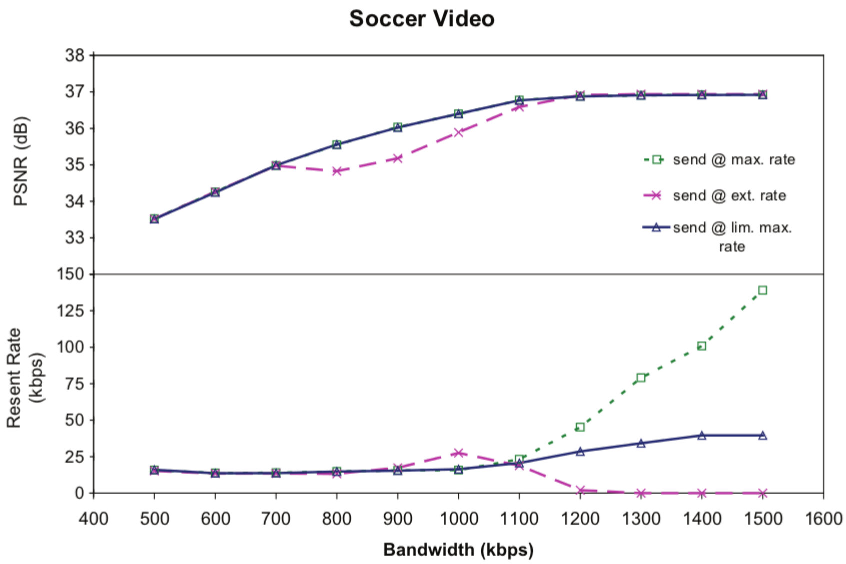

Figure 7: *Comparison of resent rate of DCCP CCID3*

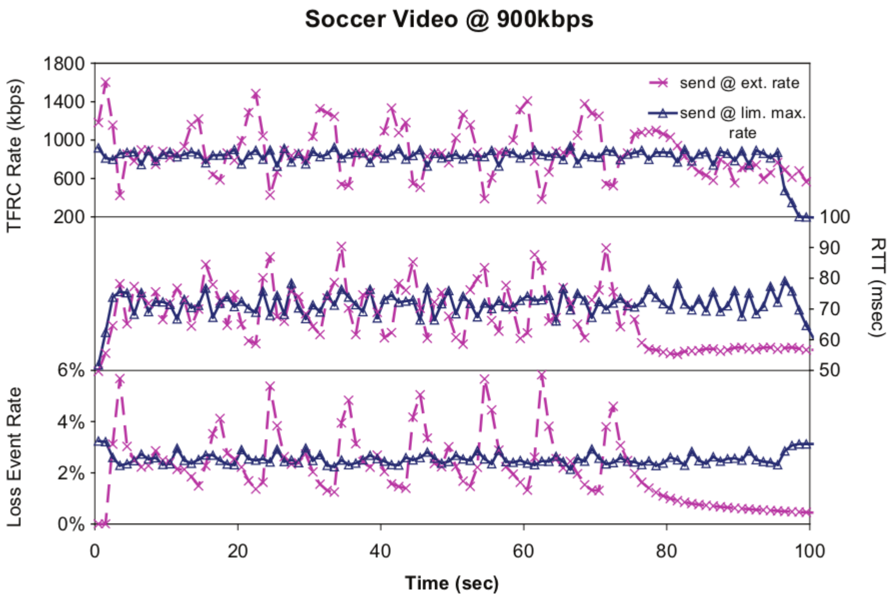

Figure 8: *TFRC, RTT and Packet Loss at 900 kbps over DCCP CCID3*

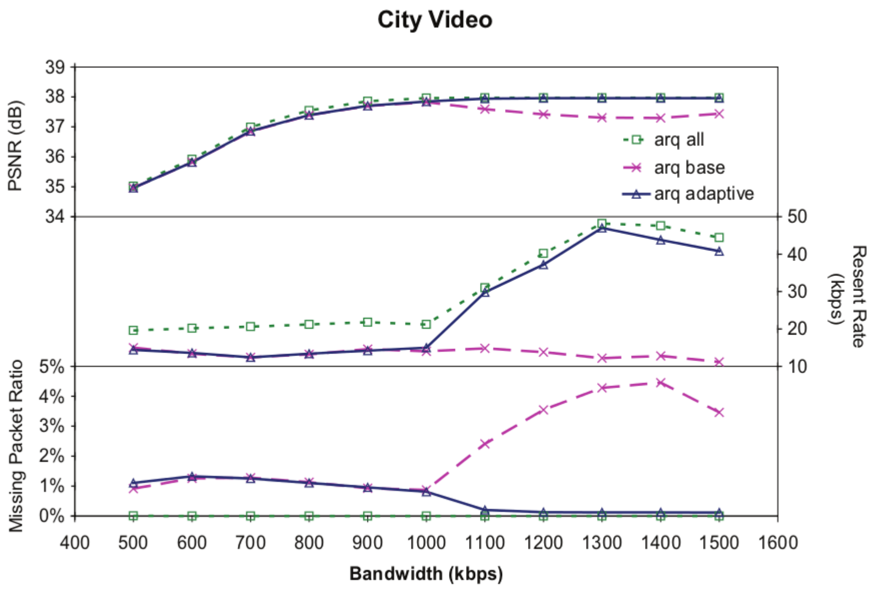

Figure 9: *ARQ schemes using DCCP CCID3*

with the video rate and the blue line is the result when sent with the estimated available bandwidth, but never higher than the maximal video extraction rate.

When sending with the rate of the extracted video, the quality is the worst. The reason is the extraction rate too low to guarantee a good quality. If one video packet is delayed by some competing traffic, the quality instantly gets worse. Both other strategies have better quality results. The maximum available bandwidth strategy and the limited bandwidth strategy differ in the amount of retransmitted packets. This is shown in figure 7. In this measurement, no competing traffic is used. The resent-graph shows a significant difference between the strategies. The video has 900 kbps considering all enhancement layers, which is why all strategies work the same until they reach the 900 kbps bandwidth. The resent rate of the extraction rate strategy decreases after 900 kbps. It has enough bandwidth for the video packets, because only the base layer packets are resent if they get lost, which rarely ever happens. The maximum bandwidth strategy on the other hand increases, because the high sending rate causes a buffer overflow and leads to more retransmissions. The limited bandwidth strategy produces the best result. It has a good video quality and only just a bit overhead caused by the retransmissions.

The measurement results of the extraction rate in figure 7 do not behave as expected, because of the divergence at the 900 kbps mark. Figure 8 shows the TFRC, RTT and the Loss Event Rate at this certain rate. When the sending rate matches the extraction rate the TFRC rate begins to oscillate. This happens due to the calculation of the TFRC, which uses the RTT and the Loss Event Rate to determine the available bandwidth. If the calculated TFRC is lower than the available bandwidth, the next estimation will be increased. The consequence is a rate which is too high and the packet loss and RTT increases. In the next estimation step, the rate is decreased again, because of the packet loss. This behavior repeats and causes the oscillation, which leads to a decreased video quality.

Figure 9 shows the results of the ARQ measurements for the City video, using DCCP CCID3 without any competing traffic. The quality of the video is almost equal for the retransmission of all lost packets (green) and the retransmission of all lost base layer packets and the enhancement layer adaptively (blue). The retransmission of the base layer only (purple) results in a decreased video quality.

The second graph resembles the resent rate, which increases after the 900 kbps mark caused by overflows on the receiver side. The adaptive ARQ has a lower resent rate than the ARQ where all packets are retransmitted, which reduces the overhead. The resent rate of the base ARQ does not change since only the base layer is retransmitted if it gets lost, which does not happen very often (as mentioned above).

The last graph shows the ratio of the missing packets, which is only high when the base ARQ is used due to the missing enhancement packets, which will never be retransmitted if they get lost.

3.3.5 Conclusion and relevance

Burak Görkemli and A. Murat Tekalp work is basically uses the same approach we had in mind. SVC is used for the video scaling instead of a multiple video stream approach with different quality. The bandwidth estimation is simple, fast and done by the sender. Also competing HTTP is considered, which is an important factor and can inflect the video transmission rate heavily. Altogether, this work is very useful for our project, but some differences still exist. First of all we are operating with much smaller devices than they do. Smartphones do not have as much bandwidth and their hardware is much slower. This does not necessary mean that the results would be completely different for smartphones, but it should be considered for our project.

They use an existing video for their measurement and stream it over a network. Therefore they are able to adjust the extraction rate of a video. This is not possible in our case, since our focus is a real-time video conference with a camera and we do not have a video file which need to be extracted first.

The scalability of the video is not the way we would use it. In this work the scalability is achieved due to an increased or decreased extraction rate. The sender always sends the whole video stream with all enhancement layers. The potential of the removable enhancement layer SVC is only used if they get lost and the receiver has to decide whether it should request the packet again or not. For our project the sender decides which enhancement layers a receiver should get depending on the available bandwidth.

4 Outlook and Conclusion

We discussed three papers in this work which gave us an overview of different, already existing approaches to our topic as well as ideas of how to develop a bandwidth aware multimedia streaming application. Every paper focused on a different issue and discussed a specific problem we may be facing during our research. The third paper was especially interesting and helpful, because its topic is closely related to ours and also because it discussed the problems and solutions very accurately. We hope to be able to improve the quality of our project because of the experience we got from the related work. Nevertheless, we still have to make our own measurements and experiences.

References

[1] Placecam. http://www.placecam.com.

[2] Hans L. Cycon, Valeri George, Gabriel Hege, Detlev Marpe, Mark Palkow, Thomas C. Schmidt, and Matthias Wählisch. Adaptive temporal scalability of h.264-compliant video conferencing in heterogeneous mobile environments, 2010.

[3] Hans L. Cycon, Thomas C. Schmidt, Gabriel Hege, Matthias Wählisch, Detlev Marpe, and Mark Palkow. Peer-to-Peer Videoconferencing with H.264 Software Codec for Mobiles. In Ramesh Jain and Mohan Kumar, editors, *WoWMoM08 – The 9th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks – Workshop on Mobile Video Delivery (MoViD)*, pages 1–6, Piscataway, NJ, USA, June 2008. IEEE, IEEE Press.

[4] Lino Ferreira, Luis Cruz, and Pedro Assunção. H.264/svc roi encoding with spatial scalability. *SigMap*, 2007.

[5] S. Floyd, M. Handley, J. Padhye, and J. Widmer. TCP Friendly Rate Control (TFRC): Protocol Specification. RFC 5348 (Proposed Standard), September 2008.

[6] Ruchao Gao, Chaorui Zhang, Jiekai Zhang, Deyuan Li, Rong Yu, and Shengli Xie. Adaptive bandwidth measurement and monitoring for multimedia transmission in home networks. In *E-Business and E-Government (ICEE), 2010 International Conference on*, pages 1473 –1476, may 2010.

[7] Burak Gorkemli and A. Murat Tekalp. Adaptation strategies for streaming svc video. In *ICIP'10*, pages 2913–2916, 2010.

[8] E. Kohler, M. Handley, and S. Floyd. Datagram Congestion Control Protocol (DCCP). RFC 4340 (Proposed Standard), March 2006. Updated by RFCs 5595, 5596.

[9] Qiang Liu and Jeng-Neng Hwang. End-to-end available bandwidth estimation and time measurement adjustment for multimedia QOS. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, volume 3, pages III–373–6 vol.3, 2003.

[10] Iain Richardson. *Video Codec Design: Developing Image and Video Compression Systems*. Wiley, May 2002.

[11] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1120, September 2007.

[12] M.S. Talebi, A. Khonsari, and M.H. Hajiesmaili. Optimization bandwidth sharing for multimedia transmission supporting scalable video coding. In *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*, pages 185 –192, oct. 2009.