Towards QoE Management for Scalable Video Streaming

Thomas Zinner*, Osama Abboud†, Oliver Hohlfeld‡, Tobias Hossfeld*, Phuoc Tran-Gia*

* University of Wuerzburg, Institute of Computer Science, Chair of Distributed Systems, Wuerzburg, Germany
eMail: [zinner|hossfeld|trangia]@informtik.uni-wuerzburg.de

†Multimedia Communications Lab, Technische Universitaet Darmstadt, Germany
eMail: osama.abboud@kom.tu-darmstadt.de

‡ Chair Intelligent Networks, Technische Universitaet Berlin / Deutsche Telekom Laboratories, Germany
eMail: oliver@net.t-labs.tu-berlin.de

***Abstract*—Video streaming applications are a major driver for the evolution of the future Internet. In this paper we introduce a framework for QoE management for video streaming systems based on H.264/SVC codec, the scalable extension of H.264/AVC. A relevant feature is to control the user perceived quality of experience (QoE) by exploiting parameters offered by SVC. A proper design of a control mechanisms requires the quantification of the main influence parameters on the QoE. For this purpose, we conducted an extensive measurement study and quantified the influence of i) video resolution, ii) scaling method, iii) network conditions in terms of packet loss and iv) video content types on the QoE by means of the SSIM and PSNR full-reference metrics. Further, we discuss the trade-off between these different control knobs and their influence on the QoE.**

I. Introduction

In Next Generation Networks (NGNs), video streaming is expected to be the killer application dominating the traffic share worldwide. According to [1], Internet video will account for over 60 % of all consumer Internet traffic in 2013 and will generate over 18 exabytes per month. The user will demand high-quality image resolutions that may require bandwidths greater than what is supported in the current Internet architecture. Massive investments by network and service providers are one pathway to cope with the emerging challenges. In this paper, we propose an alternative approach referred to as *QoE management* which will lead to much more economic and efficient use of the available resource while improving the quality of experience (QoE) for end users.

In particular, QoE management includes a) the monitoring of the current situation from the network's and the user's point of view as well as b) control mechanisms to dynamically adapt the video system to deliver the optimal QoE. The *monitoring of the current situation* aims at retrieving information about (i) the network environment of the user, like fixed FTTH or DSL access, or wireless WiMAX, WLAN or 3G access; (ii) the current network conditions like available end-to-end throughput or packet loss; (iii) user related topics, like the capabilities of the end device (CPU power, resolution) or SLAs with the network or service operator; (iv) service and application specific information, like used video codec or type of content (sports clip, music clip, news, etc.). Similar investigations for the mapping from QoS to QoE have been conducted for voice and web traffic in [2], [3].

The work has been conducted in the project G-Lab, funded by the German Ministry of Educations and Research (Förderkennzeichen 01 BK 0800, G-Lab). The authors alone are responsible for the content of the paper.

The *QoE control mechanism* takes into account the monitoring information and adjusts corresponding influence factors. For video streaming systems, the dynamic adapation of the video quality according to the current situation can be smartly realized with the de facto state of the art video codec H.264 and its scalable extension (H.264/SVC). This extension provides an integrated solution for different temporal, spatial and quality scalabilities and a seamless switching between, e.g., different resolutions or frame rates. The concept enables an adoption of the delivered video quality to the available bandwidth. In case of network problems like congestion, the resolution, and thereby the necessary bandwidth, could be reduced in order to avoid packet loss and the emerging video quality degradation. In this context, the question occurs how the end user perceives the actual quality of the delivered video. In particular, is a user more satisfied with i) a low resolution, but a smooth video playout or ii) a high resolution at the cost of quality degradations due to packet loss in the network.

Figure 1 illustrates the acceptable area of QoE control knob settings for SVC in a spider plot. The different axes denote the influence of the control knob settings on the QoE. A highly sophisticated QoE control mechanims determines the optimal settings in terms of best QoE with minimal costs. The focus of the paper is the quantification of the acceptable area, i.e. of different influence factors and control knob settings on the QoE and the required resources (bandwidth, CPU). This is mandatory in order to design appropriate QoE control mechanisms. We rely on objective QoE metrics like PSNR and SSIM which allow to conduct extensive measurement studies and to derive simple relationships applicable in QoE control. In particular, we take a closer look at the influence of packet loss, the content type of the video and its resolution, as well as the

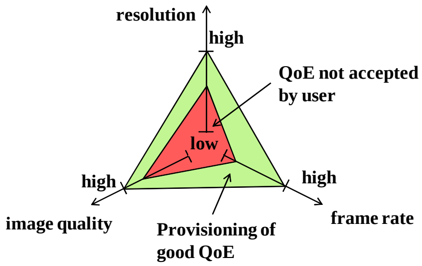


Fig. 1. Acceptable area of QoE control knob settings

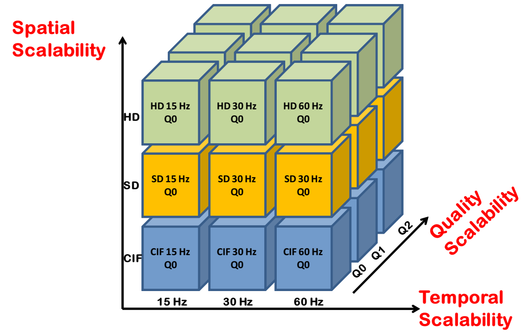


Fig. 2. SVC Cube, illustrating the possible scalability dimensions for a video file

scaling method. The scaling method is an additional control knob on application layer. User prefer to watch a video clip in an adequate size [4], that means, they will scale up the video, if possible, to be displayed on full screen. For resizing, the various interpolation algorithms differ in their computational complexity at the user device as well as the achieved video quality.

The main contribution of the paper is twofold. First, we conduct extensive measurement studies for the quantification of the QoE for different control knobs (video resolution and scaling method), network problems (packet loss) and video content types. The measurement results support results well-known from literature [4]–[6] and confirm similar conclusions for the de facto video standard H.264. Second, we compose the investigated control knobs for a common QoE control mechanisms. The measurement results are a first step to define thresholds fo a highly sophisticated QoE control.

The remainder of this paper is structured as follows. Section II gives comprehensive background on scalable video coding and existing work which links main influence factors of different video dimensions on the user perceived quality. In addition, metrics for quantifying QoE are briefly introduced. By means of laboratory measurement studies, we quantify the influence factors on QoE which allow to design control mechanism in the QoE management framework. The measurement methodology is discussed in Section III, while the measurement results are presented in Section IV. Finally, Section V concludes this work and gives an outlook on future work.

II. Background

A. Scalable Video Coding

The video codec H.264/SVC, cf. [7], [8], is based on H.264/AVC, a video codec used widely in the Internet, for instance by video platforms (e.g., YouTube, GoogleVideo) or video streaming applications (e.g., Zattoo). H.264/AVC is a so called single-layer codec, which means that different encoded video files are needed to to support heterogeneous end user devices. The Scalable Video Coding (SVC) extension of H.264/AVC enables the encoding of a video file at different qualities within the same layered bit stream. This includes besides different resolutions also different frequencies (frames displayed per second) and different qualities w.r.t. Signal-to-Noise Ratio (SNR). Different qualities can be considered as a special case of spatial scalability with identical picture size for base and enhancement layers. These three dimensions are denoted to as spatial, temporal and quality scalability. Figure 2 gives an example of different possible scalabilities for a video file. The scalable video file can be watched in three different temporal resolutions (15Hz, 30Hz, 60Hz), three different spatial resolutions (CIF, SD, HD) and three different quality resolutions (Q0, Q1, Q2). The left bottom "subcube", CIF resolution with 15 Hz and quality Q0, is the base layer which is necessary to play the video file. Based on this layer different enhancement layers permit a better video experience with a higher resolution, better SNR or higher frame rate, respectively. The more subcubes along any of the three axes are available the higher the quality in this respect is. If all subcubes are available the video can be played back in highest quality. If all subcubes within quality Q0 are available, the video can be played back in HD-resolution with 60 Hz, but only with a low SNR. This concepts allows an adaptation of the video quality to the service parameters, for instance to the connection throughput. Due to the integration of different layers within one video file a seamless switch between different layers is possible. Thus, the bandwidth of a video stream may be adjusted to the network conditions. If the offered end-to-end throughput is not sufficient for playing back the video file in maximum quality it is possible to reduce the delivered frame rate, image quality or resolution. Therefore the bandwidth of the video stream can be reduced, that means the bandwidth is adopted to the offered network quality of service parameters. The influence of a bandwidth reduction on the user perceived quality is discussed in the next subsection. In order to investigate the capabilities of SVC we emulated the codec behavior with H.264/AVC. It has to be noted that we consider UDP-based video streaming in this paper. Thus, packet loss results in artifacts or missing frames in the video. Transmitting the video with TCP faces a different kind of QoE degradation, that is stalling of the video.

B. Main Influence Factors of Different Video Dimensions on the User Perceived Quality

When using a video codec like H.264, bandwidth reduction is usually achieved by one of the following ways (i) reduce the image resolution of the video, (ii) decrease the image quality due to higher image compression rates (larger quantization), or (iii) reduce the frame rate (fps). Recently user surveys have been conducted investigating the impact of these influence factors on the subjective quality of digital video, especially in the context of mobile environments. In [5] Buchinger et al. described the interconnection between the compression rate and the frame rate in mobile environments. It turns out, that, for a given resolution, users prefer a video higher image quality, i.e. lower compression rate, and low frame rate instead of a video with medium picture quality and high frame rate. Similar investigations have been carried out by McCarthy et al., cf. [6]. For their experiments they showed test videos on desktop computers and palmtops in two different resolutions, 352x244 for the desktop experiments and 176x144 for the palmtop experiments. The conducted surveys confirm, that users tend to neglect a reduction of the frame rate, but that a decrease of picture quality leads to dissatisfied users. Our work differs from the approaches mentioned above, since we investigate H.264 encoded video sequences with higher resolutions. Further, we use objective metrics for determining the Quality of Experience instead of subjective ones like MOS or acceptability.

The issue how a video clip with given resolution should be displayed on the screen is discussed by Knoche, cf. [4]. He discovered that the video has to be displayed at an adequate size. This includes that users prefer to resize the video picture to the highest possible size with still a sufficient image quality. His work does not include an investigation of resizing mechanisms. Usually, this is done by the player either with simple mechanisms like nearest neighborhood interpolation or more sophistic mechanisms like cubic or bicubic polynomial interpolation. The more complex the resizing algorithm, the more expensive is it in terms of CPU and energy consumption. On the other hand a more complex algorithm might increase the user perceived quality. An investigation of this issue is performed with full reference models in Section IV-A. Yamagishi [9] discuss the influence of the coded bit rate on the video quality. This differs from our contribution since we detail the different scalability mechanisms provided by H.264/SVC.

C. Quantifying Quality of Experience

Quality of Experience is defined as the subjectively perceived acceptability of a service [10]. The perceived quality can be investigated in subjective tests, where presented stimuli—such as impaired video sequences—are rated by subjects under controlled conditions. The obtained rating expresses the *subjective Quality of Experience* (sQoE), typically described by the Mean Opinion Score (MOS).

However, subjective tests are time-consuming, expensive and have to be undertaken manually, which does not allow for automatic quality ratings by software. This aspect motivates objective metrics, which are designed to correlate with human perception, and, thus avoid cost and time intensive empirical evaluations. Estimates for the quality obtained by metrics are called *objective Quality of Experience* (oQoE). A more comprehensive discussion on this subject can be found in [11].

Quality metrics can be classified into three categories by the required amount of reference information [12]: *Full-reference* (FR) metrics are based on frame-by-frame comparison between a reference video and the video to be evaluated; *No-reference* (NR) metrics have to make assumptions about the video content and distortions, e.g. by evaluating the blockiness of a frame, as a common artifact in block-based compression algorithms such as MPEG; *Reduced-reference* (RR) metrics evaluate the test video based on a subset of features previously extracted from the reference video. Based on the complex nature of cognitive aspects and the human visual system, objective quality metrics do not capture its entire complexity and focus on aspects, which have been shown to correlate well with human perception in subjective tests. Thus, they are biased by model limitations and limited in their performance.

In this paper, we focus on two full reference metrics, PSNR and SSIM, due to the availability of the unimpaired reference video in laboratory conditions. The peak signal-to-noise ratio (PSNR) [13] as a FR metric has been found not to correlate well with human perception [14], as it is defined as a binary comparision of images and thus neglects the complex human perception. The PSNR is appealing due to its easy computation and physical meaning, and, therefore often used as a reference.

In principle, the subjective judgement of video quality depends on factors such as content (e.g. interview, soccer match, movie) or context (e.g. viewed on a mobile device, HDTV capable screen). Subjective tests showed a measureable content dependency [15], which is not reflected by SSIM and PSNR. However, Huynh-Thu and Ghanbari [16] argue that PSNR is a valid quality measure as long as the video content and its codec is not changed.

The Structural Similarity Index Metric (SSIM) [17] introduced by Wang et al. is motivated by the assumption that human visual perception is highly adapted for extracting structural information. It has been shown to have a high correlation with image [17] and video quality [18].

Based on results obtained for still images in [17], we introduce a mapping of PSNR and SSIM (oQoE) to a nominal 5-point MOS scale (sQoE) according to Table I for expressing an approximation of sQoE.

III. Measurement Setup

This section deals with the measurement setup, the investigated parameters and the used measurement methodology.

A. Aim of the Measurements

There are three main parts we investigate with the measurements. First, we want to identify whether full reference models

TABLE I
Mapping of oQoE to sQoE

MOS	PSNR	SSIM
5 (excellent)	≥ 45	> 0.99
4 (good)	≥ 33 & < 45	≥ 0.95 & < 0.99
3 (fair)	≥ 27.4 & < 33	≥ 0.88 & < 0.95
2 (poor)	≥ 18.7 & < 27.4	≥ 0.5 & < 0.88
1 (bad)	< 18.7	< 0.5

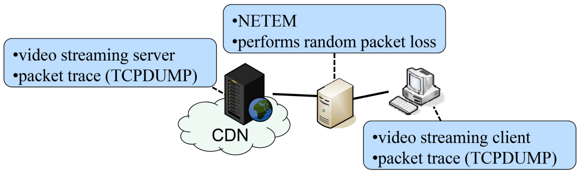


Fig. 3. Measurement Setup

can be used to classify the decrease of the QoE for different resolutions. This is motivated by the fact, that users scale up the provided resolution, if possible to full screen mode. For that, video players provide different interpolation mechanisms leading to larger images with less quality. This influence is examined in Subsection IV-A. Second, we want to investigate the case of UDP video streaming over lossy links. For that the testbed described in the next section is used. The results are presented in Subsection IV-B with additional focus on the behavior of the used full reference models in case of different content. The last issue we investigate is the influence of quality degradation in case of lossy links for different resolutions. Here we want to discuss which of the following cases users prefer:

1) best available video with high bandwidth requirements which is disturbed due to network congestion *or*
2) video in a lower resolution which is not disturbed.

Results of this study are provided in Subsection IV-C.

B. Measurement Setup

This subsection discusses the used measurement setup and describes the tools and video sequences used for the transmissions. For the conducted measurements we used the setup depicted in Fig. 3. As operation systems Debian Sid with Kernel 2.6.26-2-686 was used for all three hosts. One host, a Pentium IV equipt with a 2.4 GHz processor and 1 GB RAM acted as video streaming server, and another, a Dual Pentium III with 2x 1.2 GHz and 512 MB RAM acted as client. The experiments were traced using tcpdump [19]. For investigating the influence of network conditions, i.e. packet loss, on the video degradation we used a software based network emulation approach. This is performed by the network entity located in the middle of our testbed, a Pentium III with a 0.8 GHz processor and 256 MB RAM, running Netem [20] as network emulation software. On this entity we adjusted the desired random packet loss $p_l \in [0, 5\%]$. The video sequences were transmitted using the Evalvid framework, cf. [21]. The framework provides an approach computing the received video clip of the client with the recorded packet traces and the original video. The original and the computed video were used as input for computations with the used full reference models.

TABLE II
PROPERTIES OF REFERENCE SEQUENCES

Name	blue sky	crowd run	park joy
# Frames	216	499	499
Frame rate	30	30	30
Average bandwidth (Mbyte/s)	0.82	1.54	1.85
Length (sec)	7.2	16.63	16.63
Motion type	low-medium	medium	medium

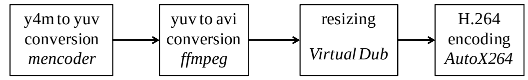


Fig. 4. Encoding methodology

C. Measurement Methodology

This subsection describes the used video sequences, how this video sequences were encoded and which full reference models we used for the evaluation. As video clips we used *blue sky*, *crowd run* and *park joy*, cf. Table II, in y4m format with a resolution of 1080p, provided by xiph.org [22]. As models for evaluating the user perceived quality for the resized and disturbed video sequences we used PSNR and SSIM metric. An efficient implementation of these metrics is provided by the MSU Video Quality Measurement Tool [23]. The encoding steps of the video clips are depicted in Fig. 4. The starting format of the video clips was uncompressed y4m format, i.e. each single image was available in uncompressed raw format. With mencoder [24] we computed uncompressed yuv files and embedded the images in an avi container with ffmpeg. For the conducted experiments we resized the video files with Virtual Dub [25] to the resolutions depicted in Table III. The encoding of the video clips to H.264/AVC was done with AutoX264 [26], an application mainly using mencoder and the x264 [27] codec. For the encoding process we assured, that every 30 frame is an I frame. Since we used a frame rate of 30 frames per second we assure that image failures spread over one second at maximum.

IV. RESULTS

Next, we discuss results obtained with by measurements which were conducted at the G-Lab local site in Wuerzburg.

A. Objective Quality of Experience for Different Resolutions

Before we discuss the influence of different video resolutions compared with the the best available resolution, we take a look at the bandwidth constraints. This is depicted in Fig. 5 for the different video clips blue sky, crowd run and park joy. On the x-axis the different resolutions are displayed, cf. Table III, whereas the scale of the axis is proportional to the number of pixels of each resolution. The y-axis shows the average bandwidth of the video sequence in Mbyte/sec. We observe that there is a strong influence of the content on the required average bandwidth. Further it can be seen that, regardless of the content type, huge bandwidth savings can be achieved by lower resolutions. Now we want to discuss the influence of lower resolutions on the user perceived quality of experience. As additional influence we investigate this behavior for different interpolation mechanisms, nearest neighborhood and

TABLE III
PROPERTIES OF REFERENCE SEQUENCES

Descriptor	1	2	3	4	5	6
Width	1216	1024	960	640	448	320
Height	684	576	540	360	252	180

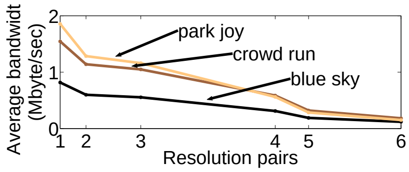
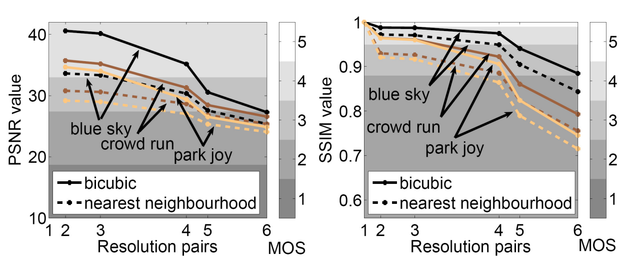


Fig. 5. Required average bandwidth for different resolutions

bicubic interpolation. The results are illustrated in Fig. 6 for the objective quality estimators PSNR and SSIM mapped to the subjective MOS scale. For both figures, the x-axis shows the different resolution pairs, assigned with the axis just as well as in Fig. 5. The y-axis denotes the used objective metric, and the gray colored areas illustrate the corresponding MOS values. A darker gray scale indicates a worse MOS value. The dashed lines show the results for nearest neighborhood interpolation, the solid lines for bicubic interpolation. Fig. 6(a) depicts PSNR and the corresponding MOS values. It can be seen that, regardless of the content, a decrease of the resolution yields in a decrease of the PSNR metric and accordingly to the MOS value. Further nearest neighborhood interpolation is perceived always worse than bicubic interpolation. The same holds for the SSIM metric which is illustrated in Fig. 6(b). Both figures show that for the three video sequences, resolution 4 (640x360) is still classified as MOS value fair (3), if bicubic interpolation is used. Taking the results of Fig. 5 into account, this means that a reduction of the required bandwidth to less than a half of the original bandwidth results still in a viewable video clip. We can conclude, that both objective metrics can be used to classify the degradation of video content in case of lower resolutions. Further both metrics estimate different content unequal and generate approximately equal results for the conducted study. Further we have seen that there is a strong influence of the content on the oQoE metrics, i.e. if we want to control the user perceived quality we have to take the content into account.

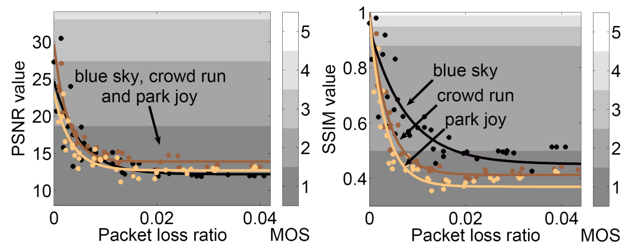
B. Influence of Packet Loss on Objective Metrics

This subsection deals with the influence of packet loss on video transmission. For that we conducted experiments with different content types and varying packet loss ratios, as explained in Section III-C. The results for the objective metrics and the corresponding subjective MOS values are depicted in Fig. 7. In both subfigures, the y-axis denotes the objective QoE value, while the different MOS values are illustrated by the areas with different gray colors. The PSNR value, depicted in Fig. 7(a), decreases with increasing packet loss ratio. The same behavior is investigated for the SSIM metric, shown in Fig. 7(b). Further, it can be seen that the PSNR values decrease equally for different content. That means, that PSNR can not differ between different content. This inability of PSNR are in accordance with the results obtained by Huynh-Thu [16], showing that PSNR cannot be a reliable method for assessing the video quality across different video contents. On the other hand, SSIM is able to distinguish different content as can be seen in Fig. 7(b). Here, blue sky is affected least by packet loss, followed by the crowd run sequence. We can conclude that the gradient of the SSIM curves for increasing packet loss differs for the investigated content types, i.e. QoE control has to take the type of content into account. Both metrics indicate, that already a small packet loss probability decreases the user perceived quality highly. Referring to SSIM, a packet loss rate $p_l = 0.007\%$ yields to a MOS value of 3, and a MOS value bigger than 2 is obtained for $p_l < 0.5\%$. Thus, we can conclude that for the given encoded video clips and scenarios, packet loss should be avoided in any case.



(a) PSNR (b) SSIM

Fig. 6. Objective comparison of different resolution pairs

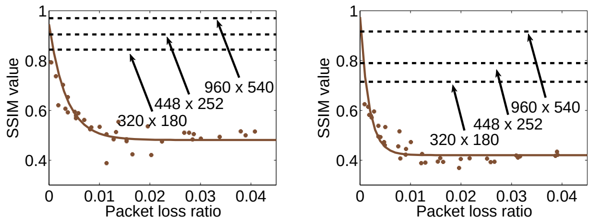


(a) PSNR (b) SSIM

Fig. 7. Objective comparison for different packet loss ratios

C. Trade-off between Packet Loss and Content Quality

In this section we want to investigate the following scenario. A network is under heavy load and users want to watch a high resolution video stream via UDP. In this case, the already heavily loaded network would become congested, which results in packet loss. As we concluded from the previous subsection, packet loss has a strong negative impact on the user perceived quality. Thus, it might be better to relieve from the congestion state by throttling the bandwidth constraints of the video stream by reducing the video resolution. This is discussed for the content type blue sky and park joy in Fig. 8. For this scenarios, the x-axis denotes the random packet loss probability, and the y-axis the corresponding SSIM values. The dashed lines indicate the corresponding SSIM values for content in lower resolution, respectively 960x540, 448x252 and 320x180 pixels, compared to the best investigated quality, 1216x684. Since the used interpolation method is the nearest neighborhood mechanism we assure a worst case approximation of the oQoE of the lower resolution content. The results for blue sky are depicted in Fig. 8(a). It can be seen that the SSIM thresholds are close to each other. A resolution of 960x540 is still rated with a SSIM value of 0.97 resulting in a MOS value of 4 and reduces the bandwidth requirements by



(a) SSIM values for blue sky

(b) SSIM values for park joy

Fig. 8. Objective comparison of different packet loss ratios and resolutions

more than 25%. The other resolutions still result in a MOS value of 3 by reducing bandwidth requirements of more than 50%. In order to cope with this SSIM values, a packet loss ratio $p_l \leq 0.04\%$ has to be assured. The same results can be deduced from Fig. 8(b) which illustrates this investigation for the sequence park joy. This content type differs from blue sky, since the SSIM thresholds for other resolutions yield in a lower SSIM value. Nevertheless the distertion of the high resolution content is even worse in case of packet loss; a packet loss ratio $p_l \leq 0.01\%$ would effectuate a higher SSIM value. We can conclude that in case of network congestion it is better to reduce the resolution of a UDP video stream if the congestion is relieved and packet loss can be avoided. This conclusion also holds for burst and other types of packet losses, but the threshold when to reduce bandwidth requirements changes. Future work will deeply investigate packet loss patterns and determine their impact on the user perceived experience.

V. Conclusion

In this paper we proposed a framework for QoE mamagement for content distribution systems based on H.264 SVC. Bandwidth adaption with SVC can be achieved seamlessly in three ways: (i) reducing the video resolution, (ii) decreasing the image quality, or (iii) reducing the frame rate. Additionally to these three control knobs the scaling method on application layer has to be taken into account as control knob. We conducted an extensive measurement study to quantify the *objective Quality of Experience* and the resulting *subjective Quality of Experience*. For that we focused on the scaling method for different resolutions and on the impact network conditions, respectively packet loss. We showed that the gain in terms of user perceived quality for higher resolutions can be estimated with objective metrics. Second, our results show, that SSIM can be used to quantify the behavior of different content and the influence of packet loss on the QoE. This does not hold for PSNR. Further we showed that video sequences with lower resolution perform better than disturbed high resolution content with respect to SSIM. Future work will investigate the influence of the frame rate and image quality on the user perceived quality. This results will be combined with the present results in an overall QoE control mechanism for SVC based on the integrated SVC scalabilities. This control mechanisms will permit an adaption of the video content to the network conditions and allow a maximization of the QoE.

Acknowledgments

We deeply want to thank Hans Henjes for his efforts and assistance by setting up the measurement infrastructure. Special thanks go to Dirk Rauscher and Michael Lang for performing the measurements.

References

[1] Cisco Systems Inc., "Cisco Visual Networking Index: Forecast and Methodology, 2008-2013," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, June 2009.

[2] T. Hoßfeld, P. Tran-Gia, and M. Fiedler, "Quantification of Quality of Experience for Edge-Based Applications," in *20th International Teletraffic Congress (ITC20)*, Ottawa, Canada, jun 2007.

[3] M. Fiedler, T. Hoßfeld, and P. Tran-Gia, "A Generic Quantitative Relationship between Quality of Experience and Quality of Service," Submitted, 2009.

[4] H. Knoche and M. Sasse, "The Big Picture on Small Screens Delivering Acceptable Video Quality in Mobile TV," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 5, no. 3, pp. 1–27, 2009.

[5] S. Buchinger and H. Hlavacs, "Subjective quality of mobile mpeg-4 videos with different frame rates," *Journal of Mobile Multimedia (JMM)*, vol. 1, no. 4, pp. 327–341, 2006.

[6] J. McCarthy, M. Sasse, and D. Miras, "Sharp or Smooth?: Comparing the Effects of Quantization vs. Frame Rate for Streamed Video," in *CHI '04: Proceedings of the SIGCHI conference*, New York, NY, USA, 2004, pp. 535–542, ACM.

[7] 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. 1103–1129, 2007.

[8] ITU-T Rec. & ISO/IEC 14496-10 AVC, , "Advanced Video Coding for Generic Audiovisual Services," .

[9] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of iptv services," in *ICC*, 2008, pp. 110–114.

[10] ITU-T Delayed Contribution D.197, "Definition of Quality of Experience," 2004.

[11] B. Staehle, A. Binzenhöfer, D.Schlosser, and B. Boder, "Quantifying the influence of network conditions on the service quality experienced by a thin client user," in *14. GI/ITG MMB*, Dortmund, Germany, apr 2008.

[12] S. Winkler, "Video Quality and Beyond," in *Proc. European Signal Processing Conference*, 2007.

[13] B. Girod, "What's Wrong with Mean-Squared Error?," *Digital images and human vision*, pp. 207–220, 1993.

[14] E. Ong, X. Yang, W. Lin, Z. Lu, and S. Yao, "Video Quality Metric for Low Bitrate Compressed Videos," in *IEEE International Conference on Image Processing (ICIP)*, 2004, vol. 5.

[15] A. Raake, M. Garcia, S. Moeller, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann, "T-V-Model: Parameter-based Prediction of IPTV Quality," in *Proceedings of ICASSP*, 2008, pp. 1149–1152.

[16] Q. Huynoh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, 2008.

[17] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[18] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 1, Jan. 2004.

[19] "tcpdump/libcap," http://www.tcpdump.org/.

[20] The Linux Foundation, "Compression project," http://www.compression.ru/video/index.htm.

[21] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid - A Framework for Video Transmission and Quality Evaluation," in *Proc. of 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, Urbana, Illinois, USA, Sept. 2003.

[22] "Xiph.org test media," http://media.xiph.org/video/derf/.

[23] Graphics and Moscow State University Media Lab, CMC department, "Net:netem," http://www.linuxfoundation.org/en/Net:Netem.

[24] "Mplayer - the movie player," http://www.mplayerhq.hu/.

[25] "Virtual dub," http://www.virtualdub.org/.

[26] "Autox264," http://www.videohelp.com/tools/AutoX264.

[27] "x264 - h264/avc encoder," http://www.videolan.org/developers/x264.html.