Investigation of Scalable Video Delivery using H.264 SVC on an LTE Network.

Patrick McDonagh[1], Carlo Vallati[2], Amit Pande[3], Prasant Mohapatra[3], Philip Perry[1] and Enzo Mingozzi[2]
[1]Performance Eng. Lab, School of Comp. Sci. and Informatics, University College Dublin, Ireland.
Email: patrick.mcdonagh@ucd.ie, philip.perry@ucd.ie
[2]Dipartimento di Ingegneria dell'Informazione, University of Pisa, Italy.
Email: c.vallati@iet.unipi.it, e.mingozzi@iet.unipi.it
[3]Department of Computer Science, University of California, Davis, CA 95616, USA
Email: amit@cs.ucdavis.edu, prasant@cs.ucdavis.edu

***Abstract*—The combination of increased data rates in 4G cellular networks (such as LTE Advanced), dedicated multicast/ broadcast services (e-MBMS) and the emergence of scalable video coding standards (H.264 SVC) allow mobile operators to offer multimedia-based services with a high quality of experience to end users. H.264 SVC offers three dimensions of scalability v.i.z. Quality (SNR), Temporal and Spatial.**

In this paper we investigate the use of Scalable Video Coding (SVC) for video delivery over an LTE network. In particular, we carried on a two step performance evaluation: first, we perform a static analysis on how different types of scalability influences the video quality, then, we analyze through simulations how the transmission over a wireless means further affects the quality.

In our analysis, we adopted a wide range of metrics: 2 full-reference metrics, namely PSNR and SSIM, along with 2 no-reference metrics, MSU Blocking and Blurring.

Our results show that no-reference evaluation metrics could be employed alongside a frame-drop metric in the place of full-reference metrics. Moreover, we show that the video scalability alone is not sufficient to avoid a degradation of quality, in some cases in the order of seconds, when caused by packet loss.

I. Introduction

Cellular data networks are experiencing an increased demand for multimedia-based communications made viable by increasing bandwidth in evolving cellular wireless technologies such as LTE and WiMAX. Service and network providers are exploring the opportunity to further enhance their current offerings and to increase revenues by catering for the demand in rich multimedia services to both mobile and fixed users using cellular networks such as LTE.

There are two important factors to be considered by providers aiming to deliver video services over cellular networks. The first being the heterogeneity of user equipment - the equipment will range from power-constrained cellphones to home users requiring high definition video. Obviously, these devices will require video streams of different qualities, resolutions and decoding complexities. The second factor deals with the constant changes in delivery parameters in the network, this can occur due to congestion or the inherent variability of the wireless links. In this case, some levels of loss will occur in the network leading to an overall degradation in service quality. For video services, these degradations could manifest themselves in terms of macro-blocking of the video stream, temporary playback pauses due to buffering of the video stream or total loss of playback. An extension of the H.264 AVC (Advanced Video Coding) standard known as Scalable Video Coding (SVC) provides a solution to both of the above factors.

H.264 SVC [1] allows for the transmission of a variety of different quality layers, (in terms of spatial, temporal and picture quality) for a video sequence. In the presence of congestion, this layered approach helps avoid blocking, pausing or losses in playback by replacing the video sequence with a lower quality version with a reduced bandwidth requirement. This is achieved by reducing the signal-to-noise ratio of the sequence (greater compression), the frame rate or the spatial resolution. However, the decision as to which dimension(s) to scale depends of the nature of the video content being viewed and each content type is expected to have one or more optimum trajectories through the adaptation space [2].

Taking the above factors into account this paper presents an initial analysis regarding the use of H.264 SVC for delivery of video services over LTE networks. We perform a twofold analysis: first a static assessment to evaluate the impact of different type of scalability in the compressed video, than an experimental analysis on how the loss of packets caused by the transmission of the video over an LTE network influences the quality.

In detail, we first provide analysis of video quality when variations in all 3 dimensions (spatial, temporal and SNR) are considered, independently and simultaneously. Four image quality metrics are considered in our experiments: two no-reference (blocking, blurring) and two full-reference (Peak Signal to Noise Ratio or PSNR, and Structural Similarity Index Metric or SSIM).

On the top of the results of this static analysis, we simulated the transmission of an SVC video over an LTE network. For this latter analysis, we used the OPNET network simulator's LTE model [3] adopting the methodology presented in [4].

This work has received support from Science Foundation Ireland via the Federated, Autonomic Management of End-to-End Communications Services (grant no. 08/SRC/I1403)

Our goal is to evaluate how the loss of packets influences the video quality with different types of scalability.

The paper is organized as follows: Section 2 provides a discussion of H.264 SVC, along with the motivation for quality monitoring for IP-based video services. Section 3 gives details of simulation framework while Section 4 details our experimental analysis. Section 5 provides the results of this analysis. Section 6 provides some conclusions with directions for future work.

II. Background and Related Work

A. H.264 - Scalable Video Coding

In order to support scalability, H.264 SVC allows for the creation of "layers" within a single video file allowing for the transmission of different layers of a video sequence from the same file. The most basic representation of the video sequence is contained within the "base layer" which consists of the lowest quality representation in each of the temporal, spatial and quality dimensions. A series of "enhancement" layers are then encoded, each of these layers represent a point in the 3-dimensional (temporal, spatial and quality) space. Each enhancement layer is seen as an improvement in terms of one or more of the 3 dimensions and requires that all of the lower layers have been received and decoded successfully in order for itself to be decoded successfully. Using this approach the visual quality of a particular sequence can be tailored to suit the devices decoding complexity, as well as to satisfy bandwidth restrictions during periods of congestion.

There are three orthogonal dimensions along which scalability can be achieved. Spatial scalability refers to scalability with respect to resolution of decoded video. Quality scalability refers to scaling in terms of the level of compression applied to the source video during encoding. This is primarily controlled using the quantization parameter (QP). Temporal scalability refers to scaling a video in terms of frames displayed per second. To generate a H.264 SVC stream, we can use one of these scalable dimensions independently or scale along multiple dimensions. The selection of the layer parameters to scale up/ down is decided prior to the encoding phase and consequently during playback we need to scale up/ down along same path chosen before encoding. For example - if we encode using temporal and then spatial scalability (two layers), we have to upscale the video first along temporal and then over spatial dimension, we must follow the reverse path for the case where we wish to change to a lower layer during playback.

B. Quality Assessment for Video Services

Any issues that degrade a network's ability to deliver packets will, as a consequence, degrade the quality of any real-time services of customers currently connected to the network. In the case of video services this degradation is likely to take on the following forms: pausing of playback due to buffer starvation, macroblocking in the case of lost (bi-) predictive frames or full loss of picture in the case of lost Intra-frames.

In the face of varying network conditions, it is possible for the service provider to perform adaptation of their delivered stream [5]. This needs no-reference video evaluation on the client side, which is then provided to service provider as a feedback who then coordinates with the network provider to provide real-time adaptation.

Ksentini et al. [6] use a priority based cross layer architecture where they prioritize the I frames transmission of H.264 video over a wireless network to improve the overall performance. However, the number of priority classes in H.264 is restricted to 2 only, against SVC which gives a range of scalability options [4]. Lee et al. [7] present a subjective performance evaluation of H.264 SVC but they don't consider network losses or evaluation with no-reference metrics. Seling et al. [8] present a comparison of H.264 SVC and VP8 but don't consider the quality issues. This paper has two significant differences in the approach: we study the effect of different scalability options and with respect to full and no-reference quality metrics.

Video quality measurement using objective metrics is concerned with performing analysis of network and/or video stream data (typically, as close to the user as possible) in order to extract data which can ascertain the quality of the received video. The input data for these metrics can range from data which analyses a video in a pixel-by-pixel fashion to data from network QoS measurements. Two popular objective metrics are peak signal-to-noise ratio (PSNR) and Structural Similarity Index Metric (SSIM). They are full reference metrics, in the sense that they require the original video sequence for evaluation purposes.

Previous work [9] has shown that the scalability offered by the SVC can provide a graceful degradation of service in an MBMS scenario to increase the number of customers served in areas where the radio signal quality is variable. Our work in this paper extends some of these concepts to a multidimensional adaptation regime which can yield greater flexibility in the bit rate to give improved control of the user-perceived quality.

III. Simulation Framework

A reference H.264 SVC encoder [10] was used to encode the videos. Due to the relatively recent emergence of H.264 SVC, the transmission of videos of this type has not fully been implemented in the majority of network simulators. In order to overcome this fact, JSVM allows for the generation of a "packet trace" for a video sequence. This contains information about the output video file such as, length of each Network Abstraction Layer (NAL) unit, a pointer to the location of a slice in the H.264 bitstream, as well as other data regarding to which level in each of the 3-dimensions this slice belongs.

To simulate the transmission of the video sequence over an LTE network the OPNET [3] network simulator was used. The properties of the physical channel were configured in order to provide a highly dynamic transmission rate along with bursts of packet loss occurring throughout the simulation. In order to obtain a realistic simulation of the streaming of an H.264 SVC video we enable the generation of network packets at the eNB according to the trace of encoded video (V_{enc}) obtained

from JSVM. The video receiver at UE is modified as well in order to save the trace of the received packets along with the delay values to simulate a playout buffer. Figure 1 gives

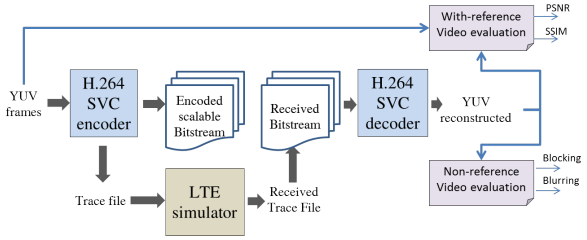

Fig. 1. Experimental setup to evaluate SVC video performance

a block diagram of the simulation process. In the case of a total loss of an I-Frame, we replace the lost I frame with the last correctly decoded frame. The decoded video at UE (V_{dec}) is used to assess the level of distortion in terms of full-reference (PSNR, SSIM), and no-reference (blurring and blocking) metrics.

IV. Experimental Analysis

The first step is to perform a static analysis of the effect on video quality when employing all 3 different modes of scalability. In order to ascertain the effect on quality, 2 video sequences ("City" and "Harbour") [1] were used. These sequences were in the raw YUV format in the 4:2:0 chroma subsampling format at a resolution of 704x576 pixels and a framerate of 30fps.

These sequences were then encoded using the JSVM SVC reference encoder [10] to a H.264 bitstream for a collection of different output resolutions, frame rates and fidelities. Note, that in order to perform quality analysis at each point in the 3 dimensional space (i.e. a single combination of a spatial, temporal and quality values) a separate H.264 video file was created for each point in the space. Table IV provides the parameters used during encoding for the low, medium and high settings for each dimension. These parameters were chosen to

TABLE IV
H.264 SVC Encoding Parameters

Quality Level	Spatial Resolution	FPS	QP
High	704x576	30	32
Medium	352x288	15	38
Low	176x144	7.5	44

allow for in depth analysis of how video degradation increases for both single and multi-dimensional reductions in quality.

For the purpose of our analysis, as detailed above, the following video quality metrics were used, 2 full reference metrics: PSNR [11], SSIM [12], as well as two non-reference metrics: Blocking and Blurring [11]. In all the figures below, these are referred as (a), (b), (c) and (d) respectively. We use the MSU VQM tool [11] to evaluate the SVC videos using these metrics (using their implementation of blocking and blurring, against multiple choices presented in previous works).

[1] source: ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/

In the case of a H.264 video where the spatial resolution and/or the temporal resolution was decreased, it was necessary to upscale this video to both the same resolution (temporal and/or spatial) as the source YUV to allow for comparison. For spatial upsampling, a procedure known as dyadic upsampling is used. In the case of temporal upsampling, the frames that are lost as a result of the temporal downsampling stage are simply replaced by repeating the previous frame a requisite number of times.

Our second step was to investigate the quality of video delivery using H.264 SVC over an LTE network, we used an OPNET LTE model and simulated packet losses due to factors in the wireless network and overflow in the playout buffer at the user equipment. The simulation parameters used were:

Duplex mode - FDD, PHY profile - 5 MHz, HARQ Retransmissions - 1, PHY loss probability - 0.01, Competing traffic - VBR (40-80 Mbps uniform distribution), Playout buffer size - 0.3 sec, Video length - 10s.

The transmission of the video is performed in multicast as an MBMS transmission. In order to vary the bandwidth available to the video transmission over the time, competing traffic with a variable bit rate is transmitted in the downlink channel with higher priority than video flow. The instantaneous rate of the competing traffic randomly changes over the time following a uniform distribution.

V. Results

A. SVC Static Analysis

In this section we provide an overview of the results obtained from the static analysis of the video with different types of scalability. In the figures presented below, the horizontal axis represents the frame number. Figure 2 gives the performance of SVC with different scalability options. For the purpose of this experiment, the chosen encoded sequences were taken from 3 points in the 3 dimensional space and were designated as "high", "medium" and "low" versions of the video sequence (see Table IV). The four video evaluation metrics follow the same trend in the sense that there is substantial loss in numerical value of PSNR, SSIM and Blurring values. The blocking metric works inversely (a lower value indicates lower levels of blocking) and follow inverse trend. It is observed that the blocking value for spatial downsampling gives inaccurate results - this requires further investigation but was included for completeness. On further subjective investigation, we observed the blocking metric value to be similarly affected by quality degradations as with other metrics. The zig-zag behavior of Figure 2 (A-B) is due to full-reference evaluation of temporally different video streams. The lower scalability video streams need to be scaled up to allow full-reference evaluation which leads to problems.

Our next experiment involved the analysis of the degradation in video quality when only one of the scalable dimensions is varied. Figure 3 presents the results of this analysis. All the conclusions drawn for the previous scenario still hold. The gap between the different layered videos is reduced and it confirms that the spatial scalability, as expected, is the

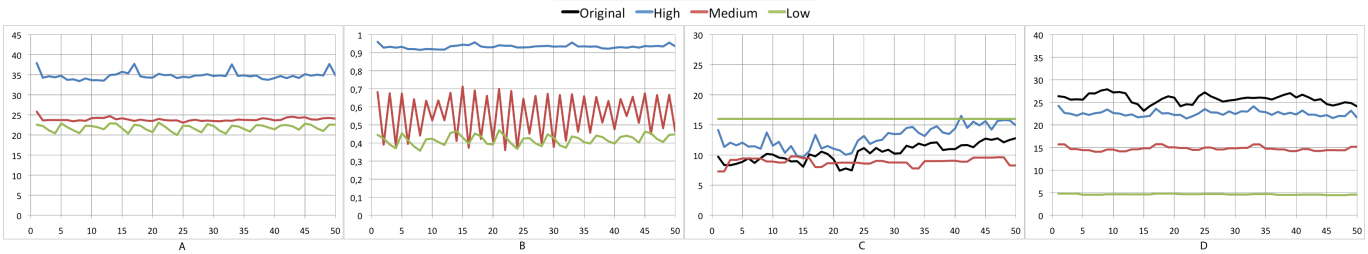

Fig. 2. Video Quality Degradation for *"City"* Video Sequence. High, Medium and Low represent the corresponding values of all scalability options in H.264 SVC as mentioned in Table IV. (A-D) represent measures for PSNR, SSIM, Blocking and Blurring metrics respectively in Y axis plotted against frame number in X axis in all figures .

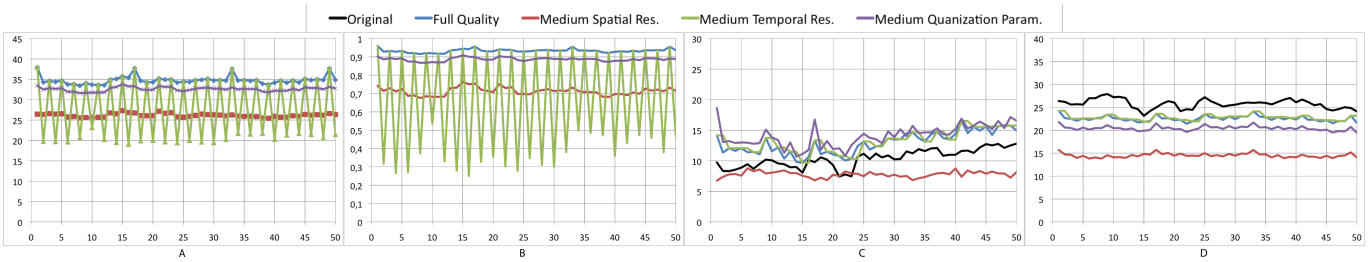

Fig. 3. Video Quality Degradation for *"City"* Video Sequence. Full (High) quality video is compared with results when there is degradation along only one scalability dimension (spatial or temporal or Quality).

main cause of quality degradation and its effects are the most noticeable. Apart from this, we can observe that blocking and blurring metrics have a graceful degradation when we move to lower quality streams. This is the same trend as in PSNR and SSIM. The peaks in PSNR and SSIM values visible at different quality levels (both Figure 2, Figure 3 and Figures 4 and 5, explained later) indicate location of I-frames. There is a smoothness associated with blocking and blurring metrics especially at low bitrates and in case of detecting network losses this makes them more suitable for real-time feedback to content provider and base station. This result corroborates recent work [13] focused on using a no-reference metric for video evaluation.

B. LTE simulations

This section presents the results obtained from LTE simulations. Figures 4 and 5 illustrate the quality of video resulting from simulations. As detailed in Section 4, network losses are caused through the introduction of competing VBR traffic (with a higher priority), this limits the bandwidth available to video transmission and will, at times cause packet loss. We can observe that even if the channel conditions and the load of the competing traffic are the same for all the scenarios, the quality degradation of the video is not always synchronized among all the traces. This is attributed to the fact that the packet loss over the LTE network depends on the size of the transmitted packets which varies amongst the different coding configurations.

As can be seen in almost all the cases, packet loss causes a temporary degradation of the quality which results in steep declines for PSNR and SSIM values. This loss in quality lasts for a short time if just P or B frames are lost or the loss in quality can last for a long time if an I-frame is lost or corrupted. The simulated interfering traffic had peak bandwidth utilisation periods equivalent to the duration of 70-100 frames. It can be observed that different quality levels of video observed different packet losses and that there was the lowest quality degradation in case of medium temporal resolution, this is due to the fact that the loss in network bandwidth resulted in an I-frame loss in other resolutions/scalability schemes.

Looking at the quality degradation of the different configurations, we can see that "medium" temporal scalability is subjected to higher degradations in quality caused by packet loss, in particular with respect to PSNR. This is due to the fact that when a whole frame is corrupted the last correctly decoded frame freezes, and therefore the display at decoder is not refreshed.

Comparing temporal to spatial and SNR, we can see that temporal is subject to similar levels of degradation but appears to suffer more frequent degradations in quality. This is due to the nature of the temporal layering, where each frame is attached to a single layer and the loss of a layer is a total frame loss. In the case of SNR and spatial degradations the loss of a single layer's data will still allow a frame to be reconstructed. Furthermore in the case of SNR and spatial, the loss of a frame results in a longer degradation because of the cross references between frames within the same GOP.

To explain further, the more frequent degradations in quality visible in the case of temporal scalability can be attributed to the fact that a single layer represents an entire I-frame, whereas in the case of SNR and spatial, the I-frame data is spread across multiple layers. Thus, when loss of a layer containing I-frame data occurs, the entire frame is lost in the case of temporal scalability, whereas in the case of SNR and spatial degradations the frame can be partially reconstructed as explained above.

Another interesting observation is the plateau observed in this region (Frame 70-100) for no-reference metrics. This is attributed to the repetition of previously decoded frame in the video. Thus, the no-reference metrics fail to observe drop of

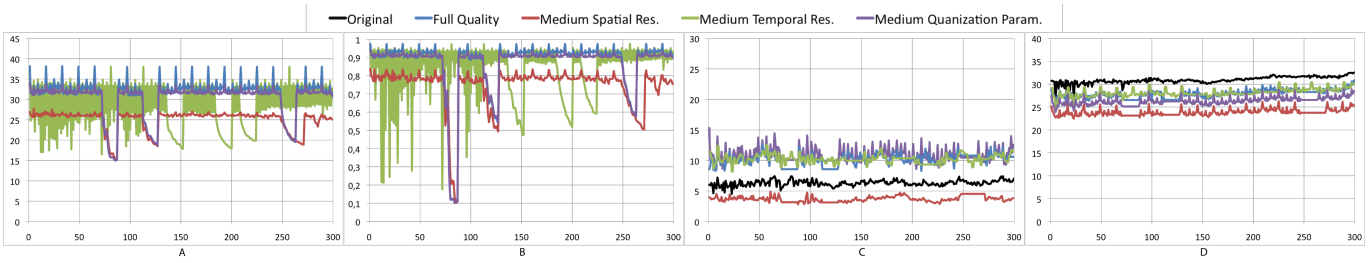

Fig. 4. SVC performance with network losses for *"Harbour"* Video Sequence. The steep decline in performance is associated with scenario with loss of I-frames.

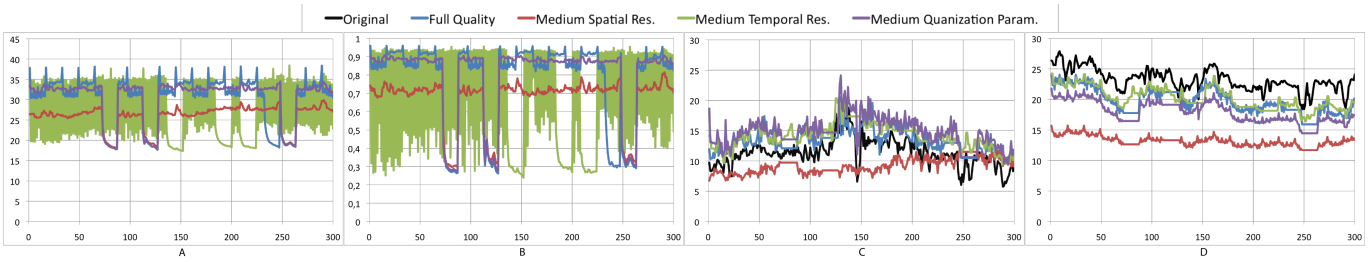

Fig. 5. SVC performance with network losses for *"City"* Video Sequence.

frames due to bandwidth congestion, which must be separately monitored by the application. This is easy to report for the decoder and requires no extra computation.

Both of the above points are somewhat linked. This link can be partially attributed to the reconstruction step of received video from the transmitted trace file. As a result of performing a simulation using a video trace file, the behaviour of the H.264 decoder at times of (partial) frame loss is not captured.

The typical effects of loss results in the decoder displaying increased blocking of the image, partial (visible) losses of macroblock data and other errors such as loss of smooth playback / motion. However as stated before, due to the limitations of the simulation setup, other methods must be employed to handle the loss of video data, such as freezing or re-use of the previous frame.

VI. Conclusion

In this work we studied the effect of scalability dimensions on the video quality by simulating scalable video transmission with and without network packet losses in LTE. The simulations show that the loss of packets causes different quality degradations over the time according to which packet has been lost. Current IPTV standards are primarily for fixed line communications systems where Packet Loss Rates (PLR) are of the order of one loss event in one to four hours ($PLR \leq 10^{-5}$) while in cellular systems the raw error rate can be as high as 1%. As illustrated by simulations, this higher packet corruption can lead to application layer loss events on the time scale of seconds or minutes. Further, we illustrate how no-reference metrics like blocking and blurring can serve as useful substitute for full-reference metrics for real-time video adaptation, which can be used to assess the quality of the received video while it is transmitted. We reserve as future work, a cross layer approach, aware of the structure of the video, which can help to decrease the packet loss exploiting its scalable structure.

References

[1] 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.

[2] N. Cranley, L. Murphy, and P. Perry, "User-perceived quality-aware adaptive delivery of mpeg-4 content," in *Proceedings of the 13th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, ser. NOSSDAV '03. New York, NY, USA: ACM, 2003, pp. 42–49.

[3] (2011, Jan.) Opnet network simulator. [Online]. Available: http://www.opnet.com/

[4] D. Migliorini, E. Mingozzi, and C. Vallati, "QoE-Oriented Performance Evaluation of Video Streaming over WiMAX," *Wired/Wireless Internet Communications*, pp. 240–251, 2010.

[5] G.-M. Muntean, P. Perry, and L. Murphy, "A new adaptive multimedia streaming system for all-ip multi-service networks," *Broadcasting, IEEE Transactions on*, vol. 50, no. 1, pp. 1 – 10, 2004.

[6] A. Ksentini, M. Naimi, and A. Gueroui, "Toward an Improvement of H. 264 Video Transmission over IEEE 802.11e Through a Cross-Layer Architecture," *Communications Magazine, IEEE*, vol. 44, no. 1, pp. 107–114, 2006.

[7] J. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurutepe, T. Sikora, J. Ostermann, E. Izquierdo, and T. Ebrahimi, "Subjective Evaluation of Scalable Video Coding for Content Distribution," in *Proceedings of the International Conference on Multimedia*. ACM, 2010, pp. 65–72.

[8] P. Seeling, F. H. P. Fitzek, G. Ertli, A. Pulipaka, and M. Reisslein, "Video network traffic and quality comparison of vp8 and h.264 svc," in *Proceedings of the 3rd Workshop on Mobile Video Delivery*, ser. MoViD '10. New York, NY, USA: ACM, 2010, pp. 33–38.

[9] C. Hellge, T. Schierl, J. Huschke, T. Rusert, M. Kampmann, and T. Wiegand, "Temporal scalability and layered transmission," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 2048 –2051.

[10] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-8," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q. 6, JVT- U*, 2006.

[11] (2011, Jan.) Video quality metric (vqm) software. [Online]. Available: http://compression.ru/video/quality_measure/video_measurement_tool_en.html

[12] Z. Wang, L. Lu, and A. Bovik, "Video Quality Assessment based on Structural Distortion Measurement," *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, 2004.

[13] C. Keimel, T. Oelbaum, and K. Diepold, "No-reference video quality evaluation for high-definition video," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 1145 –1148.