Scalable Video Coding Transmission over Heterogeneous Networks

Reuben A. Farrugia, Lucianne Cutajar
Department of Communications and Computer Engineering
University of Malta
Msida, MSD 2080, Malta
reuben.farrugia@um.edu.mt, lucycut88@hotmail.com

***Abstract*—Video streaming is currently occupying a huge chunk of the Internet bandwidth. This is mainly attributed to the wide variety of applications that are being transmitted over current Internet infrastructure, such as videoconferencing, mobile television (TV), Internet video streaming, and Internet Protocol TV (IPTV). These applications are generally encoded using the H.264/AVC codec which encodes the video content into a single layer stream with a fixed spatio-temporal video resolution. This poses a limitation for such applications since the same video content must be encoded into different streams in order to cater for heterogeneous devices demanding different spatio-temporal resolutions. This paper presents the performance evaluation of the recent H.264/SVC standard. The H.264/SVC encodes the video into different layers and the receiving device can decide to drop some layers in order to meet the required spatio-temporal resolution. This work shows that transmission of H.264/SVC using multicasting provides a substantial reduction in bandwidth requirement over traditional H.264/AVC. Simulation results further demonstrate that the H.264/SVC provides less congestion and is thus provides better Quality of Experience (QoE).**

***Keywords*-Computer Networks; H.264/SVC; Quality of Service; Scalable Video Coding; Video Streaming**

I. INTRODUCTION

Internet video is expected to consume 91% of the global consumer Internet traffic by 2014 [1]. The increase in popularity of multimedia content is accredited to the wide range of devices which make multimedia content available on several devices. Typical video streaming applications adopt the H.264/AVC standard [2] to deliver video content over the Internet. It achieves high compression efficiency relative to other standards and encodes the video content into a unique bitstream. Therefore, the generated bitstream is only suitable for a particular spatio-temporal resolution.

However, as shown in Fig. 1, different devices provide different requirements in terms of frame rate and image resolution. Therefore, the traditional H.264/AVC must generate different streams for different devices, thus becoming inefficient in terms of bandwidth utilization. For example, consider that the Main Video Server in Fig. 1 needs to transmit the same video content to two different devices; a mobile device and a High Definition (HD) Client. The standard H.264/AVC must encode two different streams, a lower resolution stream to mobile devices and a higher resolution stream to HD Clients.

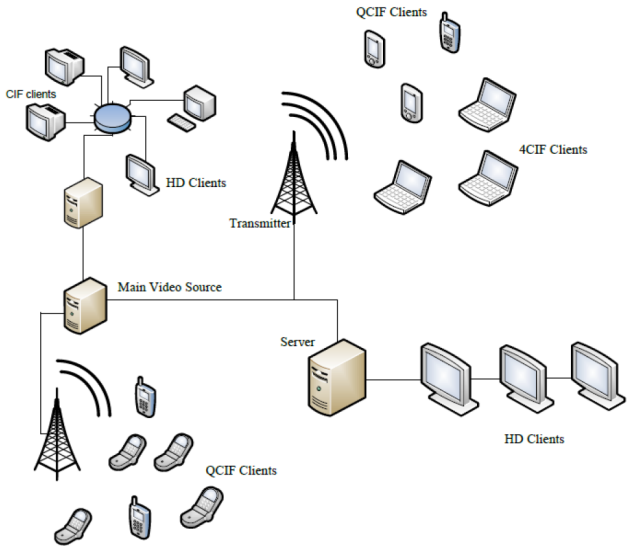

Figure 1. Typical Heterogeneous Network

Scalable Video Coding (SVC) [3] poses an attractive solution to the above mentioned problems encountered by the standard H.264/AVC codec and was recently introduced as an extension to the same standard. The H.264/SVC offers scalability by allowing the removal of parts of the video bitstream in order to comply with the various needs or preferences of the end user and to adhere to the network/device capabilities. Taking the above mentioned example, the H.264/SVC generates a unique bitstream that will be received by both devices. The HD client will decode the whole stream and thus recover the HD content, while the mobile device will drop part of the bitstream to recover a lower resolution version.

The authors in [5] have proposed a rate adaptation mechanism for H.264/SVC. On the other hand, this paper is aimed to analyze the performance of the H.264/SVC standard relative to traditional H.264/AVC codec in both unicast and multicast scenarios. Simulation results show that the H.264/SVC multicast is the most promising solution since it poses a bitrate reduction of 72 % relative to the traditional H.264/AVC unicast. This work further demonstrates that the H.264/SVC multicast transmission generates less packet loss

 ISBN: 978-1-61208-129-8

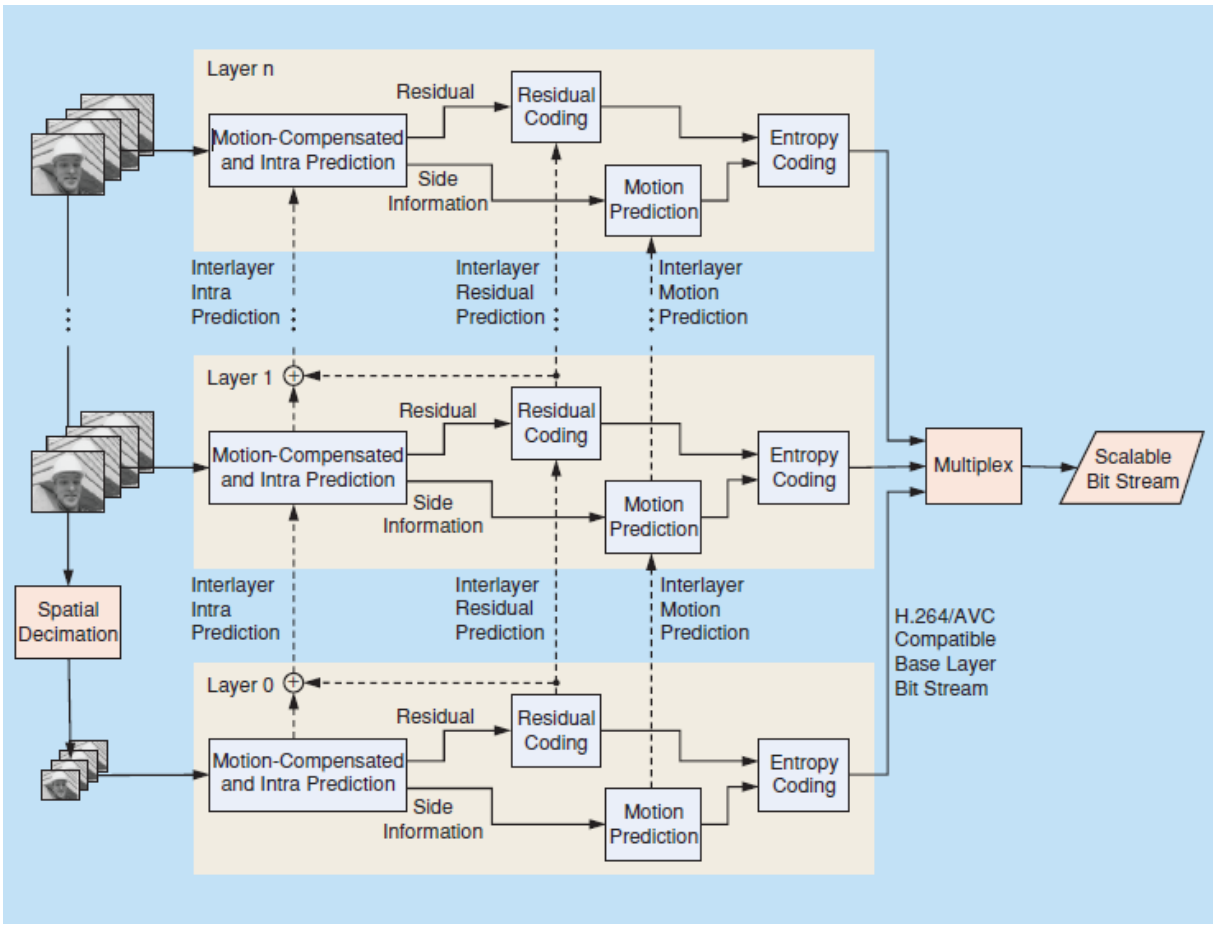

Figure 2. Simplified H.264/SVC Encoder Structure (from [4])

due to congestion and thus provides a higher level of Quality of Experience (QoE).

This paper is organized as follows. The Scalable Video Coding paradigm is presented in Section II. Section III presents the methods and protocols adopted in order to transmit H.264/SVC content over the Internet. The simulation environment is described in some detail in Section IV followed by the simulation results in Section V. This paper is concluded with the comments and conclusion in Section VI.

II. Scalable Video Coding

The H.264/SVC is encoded using a layered structure that allows the user/device to derive the most appropriate spatio-temporal resolution. Therefore, scalable video coding enables the encoder to encode only once while decoding many times at different spatio-temporal resolutions. The same bitsteam is delivered to all devices (mobile devices, HDTV etc.). However, the required spatio-temporal resolution is achieved by dropping part of the bitstream. Fig. 2 outlines the basic encoding steps taken by the H.264/AVC encoder. Each representation of the same video content can be altered to various spatial and temporal resolutions. The number of layers utilized for decoder depends on the needs of the application i.e. higher spatio-temporal resolutions are achieved by increasing the number of enhancement layers. The following sub-sections introduce the theory after which the scalable video coding paradigm is based on. More information can be found in [3].

A. Temporal Scalability

Temporal scalability refers to the frame rate of the video representation. A higher temporal layer would imply a higher frame rate. The H.264/SVC is encoded to achieve the highest frame rate *T*. Applications which need a frame rate of *N*, where $N \leq T$, drop the temporal enhancement layers *M* within the range $N < M \leq T$.

The H.264/SVC employs the Hierarchical B-picture [6] for temporal scalability. Fig. 3 illustrates the three separate bitstreams which can be extracted and independently decoded to give three temporal layers: layer 0 $T0$ and enhancement layers $T1$ and $T2$. Devices which decode only the base layer achieves one ninth the full rate, while if layers $T0$ and $T1$ are decoded one third the full rate is achieved.

On the other hand, in order to achieve full rate the base layer and all enhancement layers must be decoded.

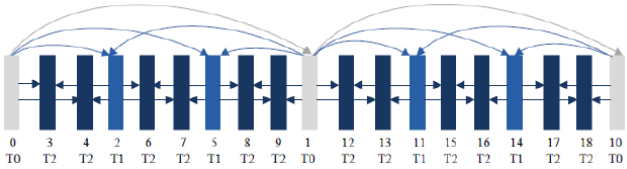

Figure 3. Hierarchical B-Picture structures for enabling temporal scalability (from [3])

The coding efficiency of H.264/SVC is dependent on the quantization parameters of each layer. This is because the motion-compensation prediction process of one layer is dependent on the other succeeding it. Therefore, the quantization parameters for the lower layers must not be very large in magnitude since this would result in reducing the image quality. Thus, the quantization parameters must be in increasing order of magnitude, with the uppermost layer having the largest quantization parameters.

B. Spatial Scalability

The Spatial Scalability process adopted by H.264/SVC employs the multilayer coding concept where each layer supports a particular spatial resolution. Similar to temporal scalability, the base layer provides enough information in order to reconstruct a low resolution video. Each enhancement layer enhances the video to a higher resolution. Therefore, the decoder can drop a number of enhancement layers in order to achieve the required spatial resolution. Fig. 4 illustrates typical spatial-scalability architecture. If only the base layer is decided the resulting frame will have an image resolution of 176×144. Increasing the number of layers will increase the spatial resolution to a maximum of 1704×576, which is the largest resolution supported by this system.

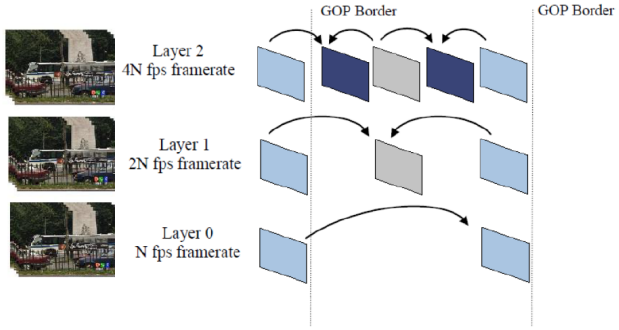

Figure 4. Inter-Layer Prediction for Spatial Scalability

In order to maximize coding efficiency, each spatial layer adopts both inter and intra predictions as for H.264/AVC. To improve coding efficiency, the inter-layer prediction is used to encode the enhancement layers. The lower resolution frames are upsampled and the similarities between the upsambled reference and the current frame are exploited using Inter-Layer prediction.

C. Spatio-Temporal Scalability

Both Spatial and Temporal scalable coding can be combined to form the spatio-temporal scalability. Fig. 5 shows a typical example of spatio-temporal scalability with a GOP of 8. The key pictures at the GOP borders are intra–coded. Higher data rates can be achieved by decoding more temporal enhancement layers. The frame structure illustrated in Fig. 5 adopts two spatial resolutions (spatial layer 0 at *QCIF* resolution and spatial layer 1 at *CIF* resolution). The upper stream adopts four temporal layers, with the top most enhancement layer being at a frame rate eight times the frame rate of the lower most base temporal layer. The lower stream employs three temporal layers.

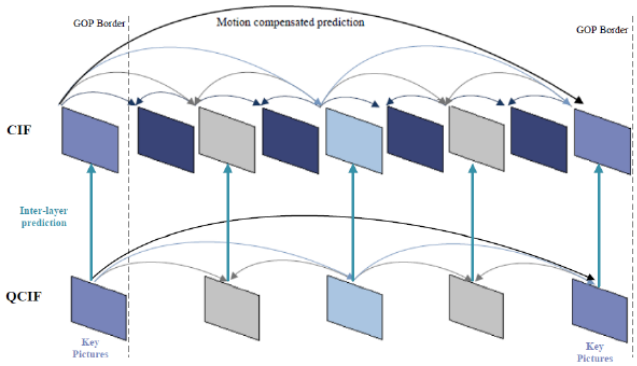

Figure 5. Spatio-Temporal Scalability (from [7])

III. Scalable Video Transmission

The H.264/SVC video related information is encapsulated within Network Abstraction Layer Units (NALUs). As illustrated in Fig. 6, the H.264/SVC standard employs a 4-byte header where the first byte is similar to the one adopted by the H.264/AVC, while the remaining bytes in the header indicate SVC related information.

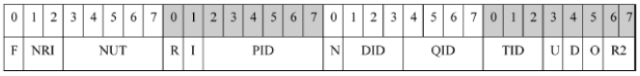

Figure 6. 4-byte SVC NALU header structure (from [8])

The *forbidden_zero_bit* (*F*) is used to indicate an error in the particular NALU while the *nal_ref_idc* (*NRI*) is used as an indication of the visual importance of the particular NALU. The *nal_unit_type* (*NUT*) field indicates which type of payload is being used for the particular NALU i.e. whether the unit is a Video Coding Layer (VCL) or non-VCL.

 ISBN: 978-1-61208-129-8

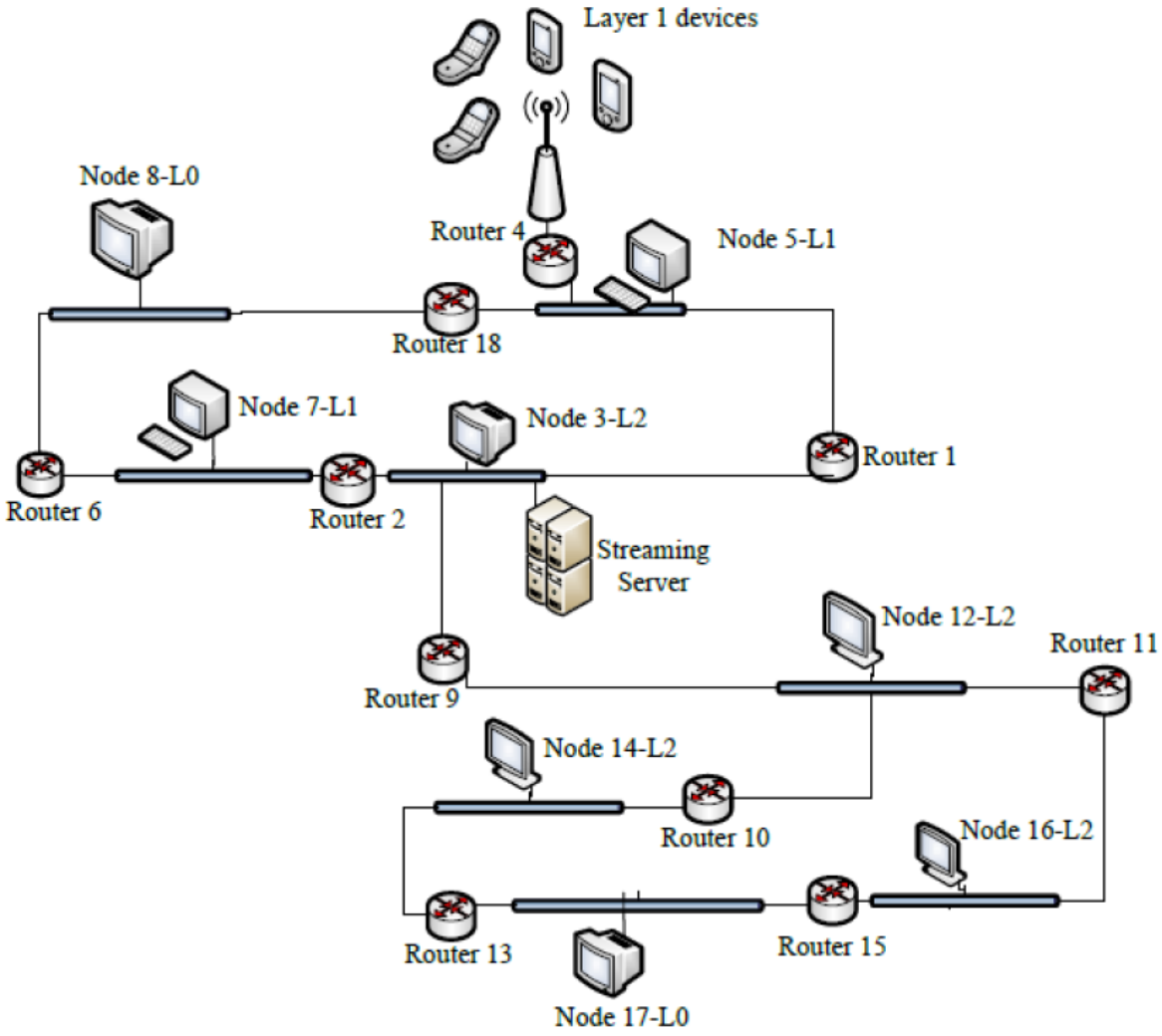

Figure 7. Network topology used in the following simulations

The remaining fields are used by the H.264/SVC codec. The *dependency_id* (*DID*) denotes the spatial scalability inter layer coding structure. The *temporal_id* (*TID*) indicates the temporal scalability hierarchically. The *quality_id* (*QID*) is used to define the quality scalability structure while the *priority_id* (*PID*) is used to assign priority to the stream. More information about each field is provided in [8].

Typical video streaming services are provided using the User Datagram Protocol (UDP) at the transport layer. UDP is a connectionless and unreliable protocol and therefore the sending node does not have a feedback channel. Therefore, the sender has no knowledge about the receiver nodes.

The UDP is a simple protocol which is convenient for real time applications, especially when using multicasting transmission. However, UDP does not ensure good Quality of Experience (QoE) and thus cannot be employed on its own. The Real–Time Transport Protocol (RTP) is a protocol that stands in between the transport and the application layers and provides additional functionalities such as timestamping and sequencing. These methods reduce the effect of jittering and enable the reordering of the received packets thus making transmission of real-time multimedia content feasible.

IV. Simulation Environment

The JSVM [7] software model was used as a reference for both the H.264/AVC and H.264/SVC. This software package contains libraries that can be used to generate both single and multiple layer video streams according to the respective standard. Every NALU is encapsulated within RTP/UDP/IP packets, where the single NALU packetization mode is adopted [9]. Unless otherwise specified, 100 frames at 60fps of the *City.YUV* sequence were encoded and transmitted

 ISBN: 978-1-61208-129-8

using the JSVM reference model. The SVC stream was composed of three spatial layers and six temporal layers.

The packetized video content are simulated to be transmitted over the Internet using the Network Simulator 3 (NS-3) [10], which is a discrete-event network simulator. It allows the study of Internet-protocols and monitoring of data flow of large scale systems in a controlled environment. For this work, the network topology illustrated in Fig. 7 is used. The video stream was transmitted using both unicast and multicast transmissions.

V. Simulation Results

The ns-3 was combined with the JSVM codec to simulate the transmission of both H.264/AVC and H.264/SVC over heterogeneous networks. Fig. 8 shows the throughput at every router for both unicast and multicast transmission modes of H.264/SVC streams. The simulation results clearly demonstrate that under multicast transmission, a reduction of around 60% was achieved relative to unicast transmission. This can be easily explained since multicast transmission sends one stream to a group while unicast transmits a stream for each receiving node. Fig. 9 shows the bit rates at the video server when using different encoding and transmission modes. SVC multicast transmission results in a 92% decrease in bandwidth over the generated H.264/SVC unicast and a 72% decrease over .AVC unicast. This confirms that SVM multicast is the most appropriate mode for video streaming applications.

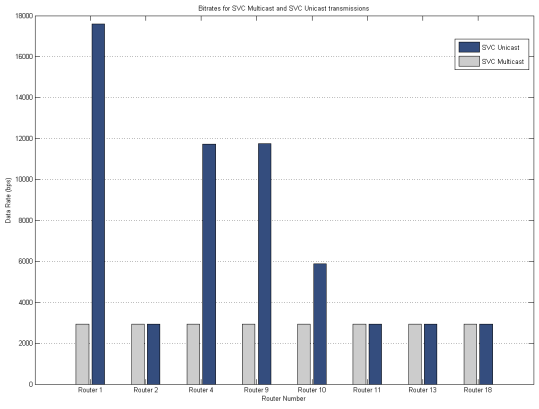

Figure 8. Throughput Analysis at routers for unicast and multicast transmissions

The H.264/SVC is inherently more robust to packet loss, especially when adopting multicasting. This is attributed to the fact that since the H.264/SVC multicast requires lower bitrates, thus reducing the probability of packet loss due to congestion. This can be easily observed from Fig. 10 where the packet loss for certain devices is much higher for H.264/AVC unicast than for H.264/SVC multicast. Furthermore, H.264/SVC is more robust to transmission errors

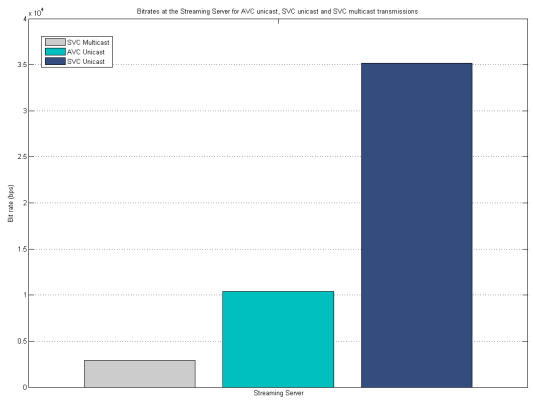

Figure 9. Bitrate of the Video Streaming Server at different modes of operation

Figure 10. Number of lost packets by each receiver for AVC unicast and SVC multicast transmissions

relative to the H.264/AVC since the H.264/SVC sequence will either reduce the image resolution or else reduce the frame rate, thus providing minimal distortion. On the other hand, H.264/AVC has only one stream and therefore the only option is to conceal the damaged region of the frame which generally results in lower video quality.

VI. Conclusion and Future Works

This paper has presented a detailed analysis of the transmission of H.264/SVC over heterogeneous networks. It was shown that the best solution is to transmit the H.264/SVC stream using multicast transmission mode. This is mainly attributed to the fact that in multicast transmission the video stream is transmitted to a group of devices opposed to unicast transmission. Moreover, it was shown that H.264/SVC multicast encounters less congestion mainly due to the smaller data-rate required opposed to H.264/AVC unicast. Simulation results have shown that the congestion level using H.264/SVC is 75% smaller than when using

 ISBN: 978-1-61208-129-8

H.264/AVC. Furthermore, H.264/SVC is inherently more robust to transmission errors and thus making it ideal in packet loss scenarios such as IPTV. Future work involves the application of Overlay networks for H.264/AVC.

REFERENCES

[1] Cisco. (2010, Jun) Cisco visual networking index: Forecast and methodology, 2009-2014 ONLINE. [Online]. Available: http://www.cisco.com/

[2] *Advamced Video Coding for Generic Applications*, ITU-T Std. H.264, 2005.

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

[4] H. Schwarz and M. Wien, "The scalable video coding extension of the h.264/avc standard [standards in a nutshell]," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 135 –141, March 2008.

[5] B. Zhang, X. Li, M. Wien, and J.-R. Ohm, "Optimized channel rate allocation for h.264/avc scalable video multicast streaming over heterogeneous networks," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 2917 –2920.

[6] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical b pictures and mctf," in *Multimedia and Expo, 2006 IEEE International Conference on*, Dec. 2006, pp. 1929 – 1932.

[7] *JSVM Software Manual*, JVT, 2009.

[8] W. Ye-Kui, M. Hannuksela, S. Pateux, A. Eleftheriadis, and S. Wenger, "System and transport interface of svc," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1149 –1163, Sept. 2007.

[9] D. Wenger, T. Stockhammer, M. Hannuksela, M. Westerlund, and D. Singer, "Rtp payload format for h.264 video," *IETF RFC3984*, Feb. 2005.

[10] *ns-3 Reference Manual*, ns-3 Project, 2009.

 ISBN: 978-1-61208-129-8