Scalable Video Coding Based Video Streaming on JXTA Peer-to-Peer Network

Youssef Lahbabi[1], Ahmed Hammouch[2], El Hassan IBN ELHAJ[3]
[1]ENSIAS Mohammed V Souissi University
[2]ENSET
[3]National Institute of Posts and Telecommunications (INPT)
[1, 2, 3]Rabat, Morocco
lahbabi_youssef@yahoo.fr, hammouch_a@yahoo.com, ibnelhaj@inpt.ac.ma

ABSTRACT: *Video streaming through peer-to-peer (P2P) networks using scalable video coding (SVC) has been an active research. In this way, we propose to design and implement such solution by using a JXTA peer-to-peer framework to share and stream scalable videos. This new architecture use jointly the scalable video coding H.264/SVC and JXTA P2P network in term of video streaming especially Video-On-Demand (VoD). Most importantly, in this architecture we include some relevant metrics that reflect the key features of perceived video quality adaptation for different resources consumption. For this purpose, we have defined some objective metrics Quality of Experience (oQoE) by exploiting parameters offered by SVC (SVC level, Bitrate and SNR) which are designed to correlate with subjective Quality of Experience (sQoE), typically described by the Mean Opinion Score (MOS), and, thus avoid cost and time intensive empirical evaluations. This paper proposes design and implement video streaming through JXTA P2P networks using SVC.The problem is well motivated. The experimental results show that the proposed schemes based on the strong relationship between sQoE and oQoE provide a good control for studies factors relating to QoE.*

Keywords: H.264/SVC, Video Streaming, VoD, P2P, JXTA, QoE, sQoE, oQoE

Received: 2 June 2013, Revised 8 July 2013, Accepted 17 July 2013

1. Introduction

Live video streaming over networks, especially over the Internet, is more and more used in large applications, such as Video-On Demand and Internet TV. The broadcast is usually insured by deploying several streaming servers .These services are accessible to a large number of receivers, which is increasing and can result in a server link bottleneck. Furthermore, these clients are connected to the network via various network accesses (LAN,ADSL,Wi-Fi...)and using diverse access terminals such as PC, PDA, and cell phone. These heterogeneous conditions make the stream adaptation a hard task to achieve in order to guarantee a proper Quality of Service (QoS) to the client. Stream parameter negotiation can be an effective solution to the heterogeneous conditions and be achieved by using the scalable video coding .Such can provide several sub-streams, each one providing certain fidelity of the original video, so the client can choose the one adapted to this capacities. On the other hand, peer-to-peer applications have become popular in recent years and offer an effective solution to overcome the limitations of centralized servers. In the audiovisual streaming area, the popularity of P2P real-time and Video onDemand (VoD) streaming applications such as PPLive [1], PPStream [2], UUSee [3], Pando[4], Zattoo [5] has been demonstrated. As an example, PPLive has registered over 110 million users, 2 million users concurrently connected, offers more than 600 channels and has users in more than 200 countries. In addition, Youtube [6], a worldwide well-known web 2.0 streaming applications, is preparing to use P2P computing in order to improve its download rate while reducing transmission costs (especially interesting in current financial crisis time). Kawashima and Ma [7] developed a collaborative tool called TOMSCOP, which is based on the elementary peer group services offered by the JXTA general framework. TOMSCOP provides four types of services: synchronous message transportation, peer room administration, peer communication support and application space management. However, TOMSCOP suffers from inefficient pipe advertisement problems and it takes a long time. On the other, STARCast [8] ,which is a JXTA based platform that offers streaming services. STARCast uses an ALM (Application Layer Multicast) structure for streaming tasks. But, opposite to platform CIMSLive (Collaborative and Interactive Media Streaming Platform) which is a hybrid solution based on the P2P Java API of JXTA , DONET (Data-driven Overlay Network) [9] and ALM structure, it treats all the streams in the same manner, that is, it does not distinguish between low and high bitrate streams. In addition, there is no adjusted the level of SVC layer and no optimized by QoE.Nowadays, P2P supposes the major part of current Internet traffic (more than 60% of Internet traffic in 2006 was consumed by P2P file-sharing applications [10]). In addition, regarding to high-consuming P2P video applications, statistics in one of the biggest Chinese Internet Service Providers (ISP) show that PPLive accounts 10% of the total Internet backbone traffic, even more than file-sharing (Bittorrent [11] represents the 8% of the traffic). Some studies [12] manifest that streaming is taking over P2P users for video content. So, video and peer-to-peer contents are both rapidly increasing Internet bandwidth demands. Recent reports predict an exaflood [13] from advances in video over the Internet, rich media content, and User Generated Content (UGC). Moreover, it is expected that by 2013, the sum of all forms of video (TV, VoD, Internet video, and P2P) will exceed 90 percent of global consumer IP traffic. In that sense, new systems and studies to optimize future P2P and video traffic may have a very high impact on the future of the Internet. In this paper we propose to design and implement such solution by using a JXTA peer-to-peer architecture to share and stream scalable videos and we define relevant metrics that reflect the key features of perceived video quality. The SVC extension of the H.264/AVC standard strives to overcome the problem of coding the underlying videos. The communication between a client and the sender peers is ensured by establishing several Real Time Streaming Protocol (RTSP) sessions and transporting the streams using RTP sessions. This paper also aims to throw light on the differences between user experience and the objective measurements using a video quality assessment tool. The rest of the paper is organized in the following way. In section 2, we propose architecture for scalable P2P video streaming. In section 3 and 4, we present our simulative analysis and then conclude the paper in section 5.

2. Proposed Architecture

Consider distributed video peers for storing and streaming videos where single video processing machines are connected to a network of distributed peers connected via switches and routers. This peer interconnection provides a virtual video server to the player application.The management for the virtual server can be implemented by a JXTA-based middleware which ensures on the one hand Quality of Service for video streaming and manages on the other hand the storage and distribution of recorded videos. Each peer offers video content and may play videos from other peers by streaming simultaneously. The basic architecture of P2P video streaming SVC is presented in Figure 1. Based on combination of scalable video coding SVC and Peer-to-Peer networks. This system is composed of several modules as follows:

2.1. Graphical user interface

The client needs the GUI module to display the GUI software and to facilitate communication between users of peers.

2.2 Mplayer

For the actual playback, we use a modified version of the MPlayer [14], which supports SVC Encoding [15].

2.3 Quality of Experience

Quality of Experience is defined as the subjectively perceived acceptability of a service [16]. 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 [17][18].

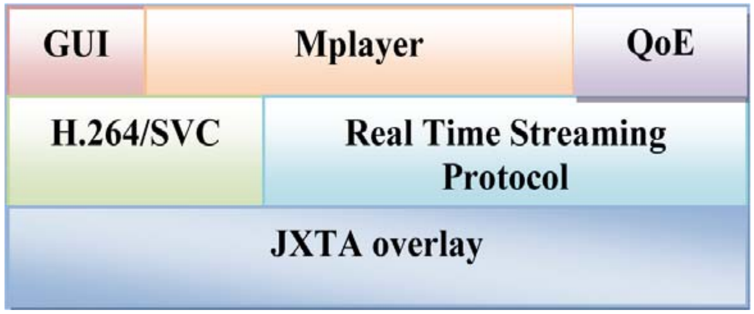

Figure 1. Architecture of scalable P2P video streaming

2.4 Scalable Video Coding (SVC)

The MPEG-4 Scalable Video Coding standard is an extension of the H.264/AVC standard and provides a number of different layers within one encoded bitstream. While the H.264/AVC compliant base layer of a scalable bitstream provides the minimum quality, the enhancement layers are used to further increase the quality, resolution or frame rate of the bitstream [19][20]. Thus, a client only needs to receive a small part of the scalable bitstream to consume the data in low quality, while it has to receive and decode the complete scalable bitstream to consume the data in best quality. The usage of scalable codecs simplifies the adaptation of bitstream significantly, as an adaptation of such a bitstream can be performed by simply skipping some or all of the data related with enhancement layers. The SVC base layer may be enhanced in three dimensions: the temporal dimension (frame rate), the spatial dimension (resolution) and the quality dimension (SNR). When considering networks with fluctuating bandwidth, especially temporal and SNR scalabilities enable powerful adaptation by diminishing the video bitrate. However, when several terminals with unique device capabilities exist also spatial scalability is a considerable alternative for saving the encoding time of various different types of sequences. An essential feature of the design of the SVC extension is that the majority of the components of the H.264/AVC standard were adopted. This implies that transform coding, entropy coding, motion compensation, intra-prediction, the deblocking filter or the structure of the NAL units (NALU) are used as intended for the H.264/AVC standard . One advantage of this approach is that the base layer of an SVC encoded bitstream can generally be processed by a H.264/AVC compatible decoder, as the extensions of the H.264/AVC standard are only used to support spatial and signal-to-noise ratio (SNR) scalability. The levels are the same as the H.264/AVC levels. However, the characteristic number of macroblocks per second in an SVC stream is calculated according to the number of layers in the stream (see formula below) [1].

$$N = \begin{cases} \frac{1}{2} * \sum_{i=1}^{L-2} NL_i ; L > 2 \\ NL_2 ; \text{where } L = 2 \end{cases} \tag{1}$$

With, N = Total number of Macroblocks.

L = Total number of layers in the stream.

NL_i = Number of Macroblocks on layer i.

2.5 Real Time Streaming Protocol

The Real Time Streaming Protocol (RTSP) is a network control protocol designed for use in entertainment and communications systems to control streaming media servers. The protocol is used for establishing and controlling media sessions between end points. Clients of media servers issue VCR-like commands, such as play and pause, to facilitate real-time control of playback of media files from the server. The transmission of streaming data itself is not a task of the RTSP protocol. Most RTSP servers use the Real-time Transport Protocol (RTP) in conjunction with Real-time Control Protocol (RTCP) for media stream delivery; however some vendors implement proprietary transport protocols.

2.6. JXTA Platform

JXTA [21] is a programming language and platform independent Open Source protocol started by Sun Microsystems for peer-to-peer (P2P) networking in 2001(See Figure 2).You can use JXTA technology to create peer-to-peer (P2P) applications based on Java technology. The JXTA technology is a set of open protocols that enable any connected device on the network, ranging from cell phones and wireless PDAs to PCs and servers, to communicate and collaborate in a P2P manner. JXTA peers create a

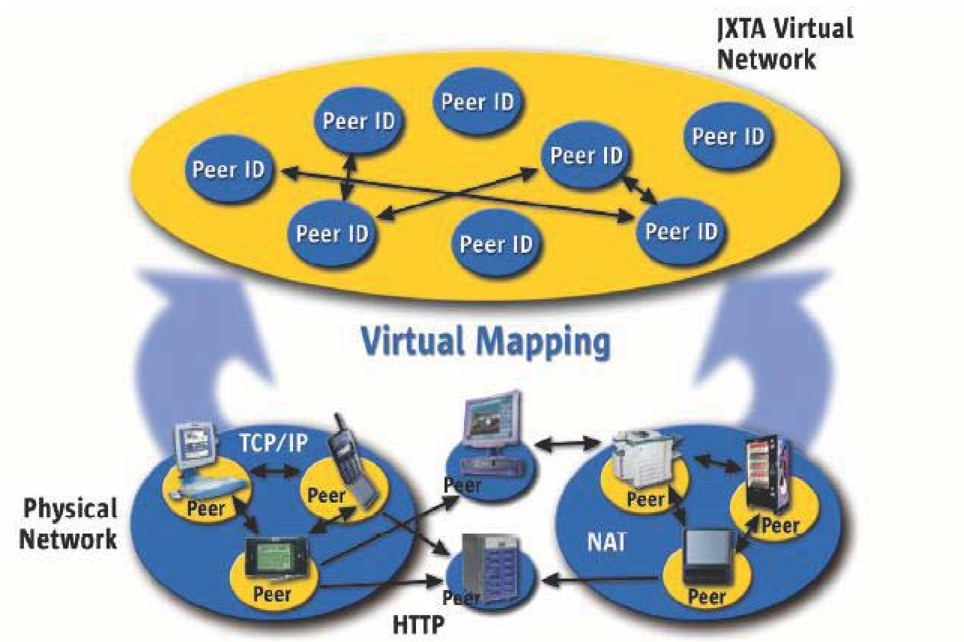

Figure 2. JXTA Technology Virtual Network

virtual network where any peer can interact with other peers and resources directly, even when some of the peers and resources are behind firewalls and network address translations (NATs) or on different network transports.

3. Evaluation

3.1 Simulation setup

The video traffic trace used for the experiment is "*Highway*" video that uses H.264/SVC video coding with Joint Scalable Video Model (JSVM 9.16) codec [22]. This video is collected from the standard test media [23].The encoding parameters can be seen in Table 1. In order to have a variation to PSNRs and bitrates we encoded the test sequences without enabling the rate control. Table 2 presents the encoded PSNR average for each layers level. From Table 3 the quality scalable sequence takes on three values namely spatial, temporal and quality identifiers denoted by (D, T, Q) respectively. All the layers corresponding to enhancement layer has an additional quality level it is denoted by identifiers 1 and 2. The layer extraction in this case is done on bitrate allocations. Each layer has an extraction point and hence there are eleven extraction points for eleven different layers. In this case, the extraction is done on bitrate allocation for each scalable layer, where as the layer extraction was done based on the D, T, Q identifiers in combined scalability. For this video to extract layer 11 with the D,T, Q values (0, 3, 2) from the Table 3, the extraction was based on the bitrate that is equal to 1412.2 kb/s. The comparisons were done between the layers (0, x, 0) and (0, x, 2) where x is the temporal level ranging from 0 to 3.

3.2 Demonstration scenario

The demonstration will make use of the Scalable Video for Peer-to-Peer Streaming. It will start by presenting how the streaming system is built over this JXTA Virtual Peer to Peer Networks. After a stream has been selected, data transmission starts and, after some startup time, MPlayer is automatically launched and playback starts. The GUI of the Scalable Video for Peer-to-Peer Streaming application is presented in Figure 3.

4. Results and Analysis

The sample test video that is subjected to the experiment is downloaded from the standard test media. Scalable video coding software is built to encode the test video. The test video is encoded using the software called as Joint Scalable Video Model JSVM). Necessary settings in JSVM are made to enable scalable mode. The resulting streams are considered as the scalable streams and are scalable in spatial and temporal aspects. The scalable video coding structure is analyzed and a systematic study of scalable stream obtained from a single video stream is done. The scalable stream is extracted using an extraction tool in a

Number of Layers	3
Base Layer & Enhancement layer resolution	352×288
Frames to be encoded	2000
GoP size	8
Frame Rate	30
Quantization parameters	60; 30; 15
Base Layer bitrate	61.3; 81.4; 109.1;143.2 kb/s
Enhancement layer 1 bitrate	90.7; 250.8; 324; 405.2 kb/s
Enhancement layer 2 bitrate	790; 964.3;1177.7;1412.2kb/s

Table 1. SVC Encoding Parameters

Layers level	**PSNR average of Layers level**
0; 4; 8	43.75
1; 5; 9	40.04
2; 6; 10	37.58
3; 7; 11	37.58

Table 2. Encoded PSNRS Average for each layers Level

Base Layer	(0, 0, 0) L_0	(0, 1,0) L_1	(0, 2, 0) L_2	(0, 3, 0) L_3
Enhancement layer 1	(0, 0, 1) L_4	(0, 1, 1) L_5	(0, 2, 1) L_6	(0, 3, 1) L_7
Enhancement layer 2	(0, 0, 2) L_8	(0,1, 2) L_9	(0, 2, 2) L_{10}	(0, 3, 2) L_{11}

Table 3. Layer Structure For Quality Scalable Sequence

structural manner starting from the higher dependency layers (Enhancement Layers) and ending up with the lower dependency layer (Base layer). Each scalable stream forming a valid bit stream known as the substream, is decoded to investigate the impact of scaling on a single video stream. The obtained substreams that correspond to a specific layer have varying spatial and temporal details compared to the original stream. As the substreams are parts of the original stream, bit streams that are decoded faced challenges with spatial and temporal scalability are not compatible with the original encoded video.In order to make the substreams compatible with the original stream in spatial and temporal domains for the purpose of spatial, temporal and quality scalability have been employed. To measure the quality of video SVC when the layers are varied from a scalable stream, objective and subjective metrics Quality of Experience are been employed .The reconstructed substreams that contain few layers are compared with the original stream containing all the layers. In this section we represent the plots for objective video quality assessments performed over video streaming for all the scalable layers. The test sequences that are subjected to both objective and subjective video quality tests yielded results that characterize the effect of scaling on objective video quality. Objective tests are categorized into three plots, namely MOS versus SVC level of Layers, MOS versus Bitrate and MOS versus PSNR. The test criteria for the objective measurements include test streams that are reconstructed spatially, temporally and quality. Table 4 shows the terminology used for formulating subjective quality experience, typically described by 5-point MOS scale versus some characteristics of scalable video coding (Table 5).

These Metrics are defined by the general form of an arithmetic sequence as follow:

$$\begin{cases} U_0 = 0 \\ U_{n+1} = U_n + \lambda; \quad \forall \lambda \in \mathbb{R} \end{cases} \tag{2}$$

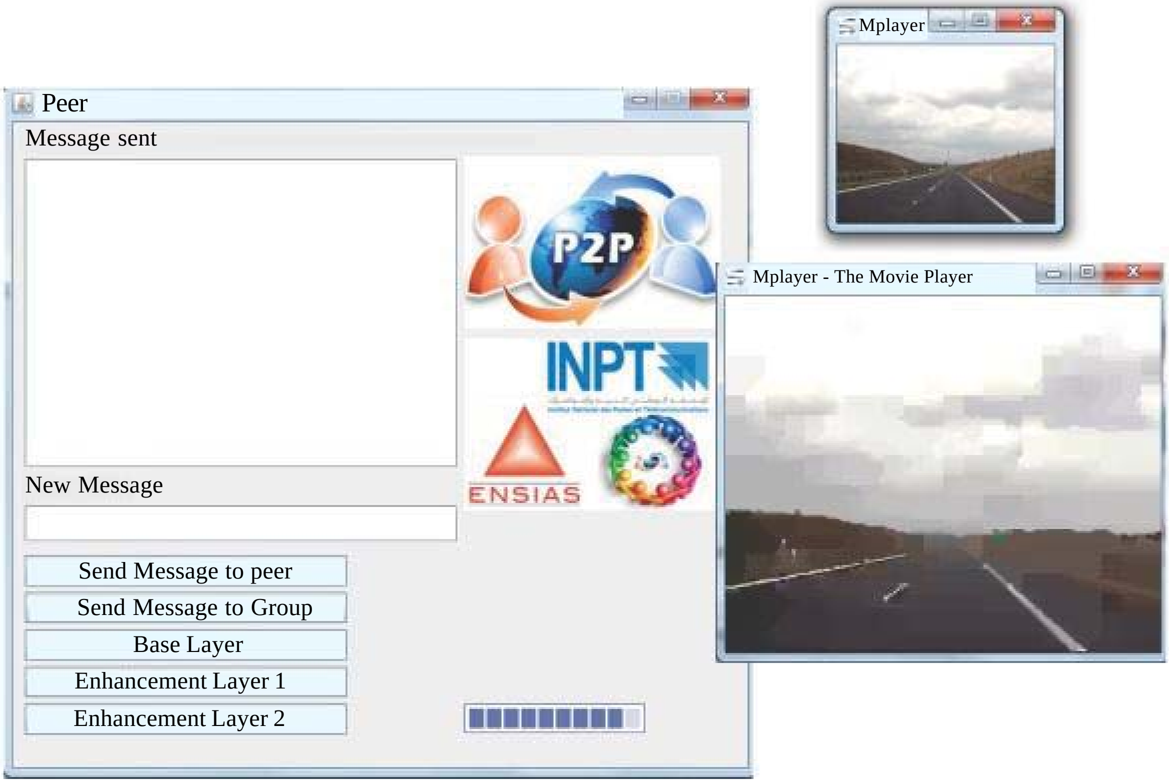

Figure 3. A screen shot of the Scalable Video for Peer-to-Peer Streaming

Variable	Description
MOSmax	Maximum Mean Opinion Score (MOS) is 5.
Lmax	Identifier of maximum SVC level for maximum enhancement layer.
Bmax	Maximum bitrate for maximum enhancement layer.

Table 4. Terminology Used For Formula

Scaling Method	Metric
MOS vs. SVC level	$\begin{cases} U_0 = 0 \\ U_{n+1} = U_n + \lambda_{MOS_{max}}^{L_{max}} \\ \text{Where : } \lambda_{MOS_{max}}^{L_{max}} = \frac{L_{max}}{MOS_{max}} \end{cases}$
MOS vs. Bitrate	$\begin{cases} U_0 = 0 \\ U_{n+1} = U_n + \lambda_{MOS_{max}}^{B_{max}} \\ \text{Where : } \lambda_{MOS_{max}}^{B_{max}} = \frac{B_{max}}{MOS_{max}} \end{cases}$

Table 5. Formula

With the initial term of an arithmetic progression is U_0 and the common difference of successive members is λ.

While in general, PSNR in given in dB, it can be simpler to map the dB scale into the MOS one, to make comparisons more clear. A possible MOS 5-point scale is as explain more in Section IV. C.

4.1 MOS versus SVC level of Layer

Table 6 shows the mapping from MOS scores to SVC level of layer and Figure 4 shows plot of the MOS (Mean Opinion Score) versus various identifiers for Highway sequence. Each point on the x-axis denotes the spatial, temporal and quality identifiers represented in three coordinates. The y-axis denotes the MOS obtained from score reference on the perceived quality of the received quality and the curve indicates the MOS as a function of different spatial, temporal and quality identifiers for this sequence. From this figure, it can be observed that the curve exhibit a linear relationship between the SVC level of layers and MOS scale. The MOS drops down from (0, 3, 2) of enhancement regions to (0, 0, 0) of base regions and vice versa. This is due to reducing or expanding in quality scalability that has yielded for lower or higher MOS.

SVC Level	MOS
> 8.8	5 (Excellent)
6.6 - 8.8	4 (Good)
4.4 - 6.6	3 (Fair)
2.2 - 4.4	2 (Poor)
< 2.2	1 (bad)

Table 6. Possible SVC level of layer to MOS conversion

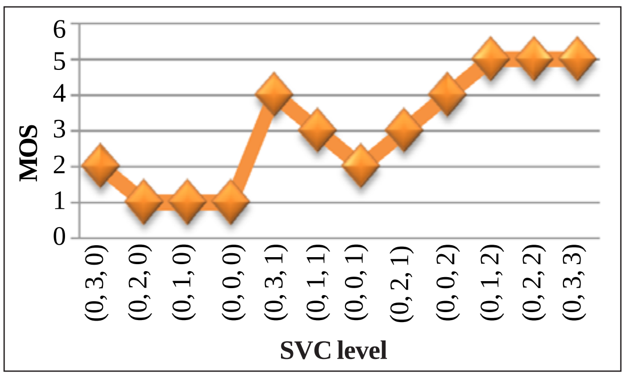

Figure 4. Plot of MOS versus SVC level of Layers

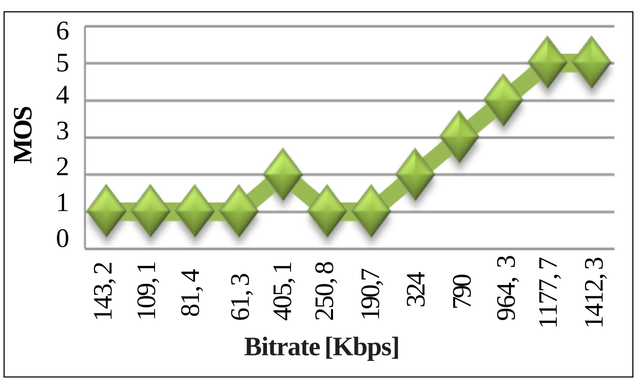

Figure 5. Plot of MOS versus Bitrate

4.2 MOS versus Bitrate

Table 7 shows the mapping from MOS scores to Bitrate and Figure 5 shows plot of MOS versus bitrate (Kbps) for this sequence. The *x*-axis denotes the bitrate in Kilobits per second and the y-axis denotes the MOS obtained. The curve gives the MOS as a function of bitrate for both Enhancement and Base Layers. It can be observed that there is a strong correlation between the bitrates corresponding to quality identifiers and the MOS. Layers containing higher quality identifiers from (0, 0, 2) to (0, 3, 2) achieve higher MOS compared to lower quality identifiers. Layers containing higher bitrate may not necessary produce higher quality unless the temporal levels are high.

Bitrate (Kbps)	MOS
> 1129.76	5 (Excellent)
847.32 – 1129.76	4 (Good)
564.88 – 847.32	3 (Fair)
282.44 – 564.88	2 (Poor)
< 282.44	1 (bad)

Table 7. Possible Bitrate To MOS Conversion

4.3 MOS versus PSNR

MOS and PSNR Analysis. There are basically two approaches to measure the digital video quality, namely subjective quality measure and objective quality measure. Subjective quality metrics always grasp crucial factors which are the impression of the user when watching the video. The human quality impression is usually given in a scale from 5 (best) to 1 (worst).

This scale is called Mean Opinion Score (MOS) (Table 8) [24], [25]. While the objective quality is measured by taking the average PSNR (Peak-Signal-to-Noise-Ratio) over all the decoded frames. PSNR computes the maximum possible signal energy to noise energy. PSNR measures the difference between the reconstructed video file and the original video file. As long as the video content and the codec type are not changed, PSNR is a valid quality measure [26], [27]. Figure 6 indicates the plot of Peak Signal to Noise (PSNR) versus MOS for this sequence. The x-axis denotes PSNR and the y-axis denotes MOS obtained. This indicates that the fidelity values for PSNR have high correlation with the MOS scores observed. The MOS score drops as the quality or temporal identifiers are reduced and are increased when expanding their identifiers. This observation is due to the lower or higher quality and temporal identifiers, which affect the video quality.

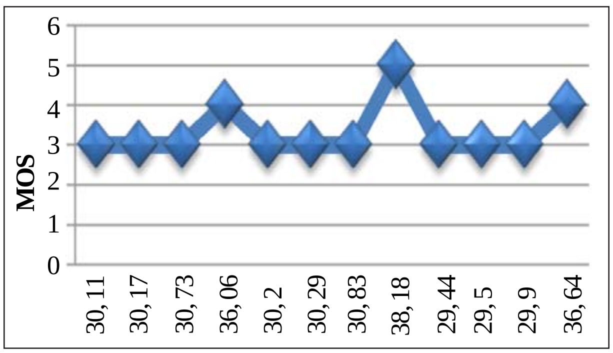

Figure 6. Plot of MOS versus PSNR

PSNR	MOS
> 37	5 (Excellent)
31 - 37	4 (Good)
25 - 31	3 (Fair)
20 - 25	2 (Poor)
< 20	1 (bad)

Table 8. Possible PSNR To MOS Conversion

5. Conclusion

In this paper, the objective is to enhance P2P video streaming by using advanced video coding techniques, mainly layered video and to design appropriate video streaming techniques based on the Scalable Video Coding over JXTA Peer-to-Peer Network. On the other hand, investigate the behavior of scaling a video stream and to provide a systematic view of the scalable video coding structure. Further, evaluation a new the metrics in order to make the comparison between oQoE and sQoE. Finally, we believe that our results show that the use of multi-layer video coding, such as SVC, is an important step towards the future generation of P2P VoD systems. Also, our metrics are available to measure the performance of video SVC by conversion between sQoE and oQoE. For the future perspective, we aim to perform real test-bed evaluation for the more personalized IPTV services delivery over P2P network and another parameters quality of experience QoE.

References

[1] PPLive official site, www.pplive.com (last visit July 2013).

[2] PPStream official site, www.ppstream.com (last visit July 2013).

[3] UUSee official site, www.uusee.com (last visit July 2013).

[4] Pando official site, www.pando.com (last visit July 2013).

[5] Zattoo official site, www.zattoo.com (last visit July 2013).

[6] Youtube official site, www.youtube.com (last visit July 2013).

[7] Kawashima, T., Ma, J. (2004). TOMSCOP—A synchronous P2P collaboration platform over JXTA. *In*: Proceeding of the international workshop on multimedia network systems and applications (MNSA'2004), In Conjunction with the 24th international conference on distributed computing systems. Tokyo, Japan, March, p. 85–90.

[8] Tsuchiya, T., Yoshinaga, H., Koyanagi, K., Honda, A., Minami, A. (2006). STARCast:Streaming Collaboration Platform using the Overlay Technology , *In*: Proc. SAINTW'06, IEEE Computer Society, January.

[9] Zhang, X., Liuy, J., Liz, B., Yum, T. P. (2005). CoolStreaming/DONet: A Data-Driven Overlay Network for Efficient Live Media Streaming, *In*: Proc. IEEE INFOCOM, March.

[10] Cachelogic official site, www.cachelogic.com/ (last visit July 2013).

[11] Bitttorrent official site, www.bittorrent.com/ (last visit July 2013).

[12] Internet Study 2008-2009 by Ipoque. (2013). http://www.ipoque.com/resources/internet-studies.

[13] CISCO IP traffic forecast. (2013). http://newsroom.cisco.com/dlls/2009/prod_060909.html.

[14] MPlayer. (2013). The Movie Player. Online: http://www.mplayerhq.hu/.

[15] Mederic Blestel and Mickael Raulet. Open SVC Decoder. available at http://sourceforge.net/projects/opensvcdecoder/.

[16] ITU-T Delayed Contribution D.197. (2004). Definition of Quality of Experience.

[17] Staehle, B., Binzenh¨ofer, A., Schlosser, D., Boder, B. (2008) . Quantifying the influence of network conditions on the service quality experienced by a thin client user, in 14. GI/ITG MMB, Dortmund, Germany, Apr.

[18] Thomas Zinne, Osama Abboudy, Oliver Hohlfeldz, Tobias Hossfeld, Phuoc Tran-Gia. (2010). Towards QoE Management for Scalable Video Streaming. *In*: 21st ITC Specialist Seminar on Multimedia Applications -Traffic, Performance and QoE, March.

[19] Schwarz, H., Marpe, D., Wiegand, T. (2007). Overview of the Scalable Video Coding Extension of the H.264/AVC Standard, IEEE Trans. on CSVT, 17 (9) 1103- 1120, September.

[20] Kouadio, A., Clare, M., Noblet, L., Bottreau, V. (2008). SVC a highly-scalable version of H.264/AVC, EBU TECHNICAL REVIEW, V. 22.

[21] JXTA , https://java.net/.

[22] SVC Reference Software (JSVM software). [Online]. Available: http://ip.hhi.de/imagecom_G1/savce/downloads/SVCReference-Software.htm, accessed on 2013.

[23] Xiph.org Test Media [online, Verified February 2013]. Available from: http://media.xiph.org/video/derf/

[24] Klaue, J., Tathke, B., Wolisz, A. (2003). Evalvid – A framework for video transmission and quality evaluation , *In*: Proc. Of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Urbana, Illinois, USA, p. 255-272.

[25] Korhonen, J., You, J. (2010). Improving objective video quality assessment with content analysis, *In*: Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM'10), Scottsdale, AZ, USA, Jan.

[26] ITU-T Rec. (2008). J. 247, Objective perceptual multimedia video qualitymeasurement in the presence of a full reference, *International Telecommunication Union*, Geneva.

[27] Hyunh-Thu, Q., Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment, *Electronics Letters*, 44 (13) 800-801, June.