A Parametric QoE Video Quality Estimator for Wireless Networks

E. Cerqueira[1], Augusto Neto[2,5], M. Curado[3], A. Riker[3], R. Immich[3], H. Barros[4]
E. Aguiar, A. Abelém, [1]Federal University of Para (UFPA), Belém-PA, Brazil
[2]Federal University of Ceara, Department of Informatics Engineering (DETI), Fortaleza-CE, Brazil
[3]University of Coimbra, Department of Informatics Engineering (DEI), Coimbra, Portugal
[4]Federal University of Rio Grande do Norte, Post Graduation Program in Computer Systems (PPgSC), Natal-RN, Brazil
[5]Institute of Telecommunications, Aveiro, Portugal
{cerqueira, eaguiar, abelem}@ufpa.br, augusto.deti@ufc.br, {marilia, ariker, immich}@dei.uc.pt, hgb@ppgsc.ufrn.br

Abstract— **The development of real-time quality estimator schemes for emerging Internet videos with different content types remains a significant challenge and is crucial for the success of wireless multimedia systems. However, currently in-service assessment schemes fail in capturing subjective aspects of multimedia content related to the user perception. Therefore, this paper proposes an on-the-fly parametric video quality estimator approach (called MultiQoE) for real-time video streaming applications. Experiments in a Wireless Mesh Network (WMN) scenario were carried out to show the accuracy, benefit, and impact of MultiQoE compared to widely used Quality of Experience (QoE) subjective, objective and parametric methods.**

Keywords – Quality of Experience; Video Streaming; Wireless Mesh Networks; Service Assessment.

I. INTRODUCTION

Recent advances in wireless communications and real-time multimedia applications, as well as, the proliferation of wireless users are changing the Internet and creating a next generation content-aware multimedia era. Regarding real-time multimedia, thousands of new users and providers are sharing their content ubiquitously, where quality level assurance is the main requirement for the success of next generation networks.

In traditional wireless networking multimedia quality level measurement techniques, only network/packet level parameters, such as bandwidth, loss and delay are used to evaluate Quality of Service (QoS) issues of applications [1]. However, in future wireless networks, where users will create, access and share thousands of real-time multimedia types of content, QoS metrics (in-service schemes) by themselves are not enough to assess the quality level of real-time multimedia applications, because they fail to capture subjective aspects of multimedia content related to user perception/satisfaction. In the light of this, many researchers have been studying Quality of Experience (QoE) assessment methods/metrics [2], which from the user's perspective, can be considered to be the overall performance of a system.

In order to measure the user experience, subjective experiments are usually conducted under well-defined test plans and controlled conditions, using quality scores rated by human observers [3]. This approach provides the most accurate QoE evaluation of multimedia content, but is costly solutions which become impractical for real-time analysis. On the other hand, objective assessment solutions [4] [5] [6] are designed to estimate QoE using explicit functions with measurable parameters related to the encoder or to the network. Existing objective QoE methods and metrics generally show low performance due to poor reference models [7] [8].

To cope with the above limitations, this paper proposes a parametric QoE Video Quality Estimator for Wireless Networks Estimator (MultiQoE) for real-time video streaming. MultiQoE correlates well with the results obtained from subjective and objective tests. This tool was built to assess the QoE of the video perceived by end-users, which took as its input the values of a set of parameters related to the video characteristics of an encoder, and which correspondingly quantified the video quality. MultiQoE is based on statistical learning using Multiple Artificial Neural Network (MANN) and was set up with real videos generated over a Wireless Mesh Networks (WMN) scenario.

Performance evaluation results show the benefits of MultiQoE in carrying out an in-service assessment of the quality level of multimedia content in WMNs. MultiQoE outperforms well-known objective metrics and also the subjective methods for evaluating the perceived video quality.

The remainder of the paper is organized as follows; Section II describes some related works. Section III explains how MultiQoE can be used for video streaming. The test environment, scenario, implementations, results of experiments and simulations are described in Section IV. Finally, Section V summarizes the findings as well as suggesting future work in the area of QoE prediction.

II. RELATED WORKS

An accurate video quality estimator scheme must integrate comprehensive monitoring schemes and service metrics such as the nature of visual codec, Group of Picture (GoP) length, spatial-temporal video activity, network impairments and other relevant factors, such as the capacity of wireless systems. It is believed that accurate QoE-aware multimedia assessment models can be achieved through efficient, network cross-layer agnostic, content-awareness, QoE monitoring and Artificial Intelligent (AI) techniques along with the corresponding cognitive evaluation of user inputs [7] [9] [10].

978-1-4673-4941-3/12/$31.00 ©2012 IEEE

As an example of online approach, a Pseudo-Subjective Quality Assessment (PSQA) [8] has been created and developed based on the trial of a Random Neural Network (RNN). For example, a number of different applications have already been used by PSQA in the following areas: VoIP over WLANs [9], video streaming over p2p networks [11], voice [12], video streaming over DiffSev networks [13], video streaming over WLAN [14], video delivery network monitoring suite [15], Multiple Description Coding (MDC) video streaming, over multiple overlay paths in video distribution networks (VDN) [16], a scheme for network selection [17] and an admission control mechanism for IEEE 802.11 [18].

A discrete perceptual impact quality assessment framework (DEQA) was proposed in [19] which enabled a real-time, non-intrusive assessment service by efficiently recognizing and assessing the individual IPTV quality violation events in the distribution network. The discrete perceptual impacts on a media session are aggregated for the overall level of quality evaluation by the user.

A MintMOS framework was established in [7] as a loadable kernel module that is an accurate, lightweight, non-reference framework for capturing video QoE inside the network core. It consists of an inference engine to infer QoE, a network sniffer to snoop traffic, and a QoE space. A QoE space is a well-known characterization of perceptual quality for the various parameters that affect it. QoS parameters, such as bit rate, delay and loss, are also employed, together with a comparison of objective metrics (Peak Signal to Noise Ratio (PSNR) [6] and Video Quality Metric (VQM) [4]) and subjective metric (Mean Opinion Score (MOS) [3]).

Based on a state-of-the-art analysis, it is clear that the accuracy and performance of QoE-multimedia video estimation or prediction is strongly dependent on the video characteristics, GoP length [20] [21], frame importance [22], AI technique and MOS experiments. PSQA can be regarded as a good assessment solution, although it does not consider explicit parameters to reflect the spatial-temporal activity of the videos. Compared to the PSQA [8] and its extensions, for example [16], MultiQoE considers videos with different levels of spatial-temporal activity to improve the system accuracy in predicting the quality level of emerging videos. MutiQoE also assures that packets are treated according to their relevance from the user perception, which is very important to determine the extent of the video impairment as analyzed and proven in [22]. Moreover, the number of observers in MultiQoE is higher than in PSQA, which increases the system accuracy. Another key difference is the use of MANNs.

MultiQoE takes into account the current network conditions and different video parameters (percentage of losses in I, P and B frames, total of losses, GoP length, and motion and complexity levels) that directly impact the quality of the video in terms of MOS prediction.

III. MultiQoE Proposal

A. MPEG Video Structure

The MPEG standard [23] defines three frame types for the compressed video streams, namely *I* (Intra-coded), *P* (Predictive-coded) and *B* (Bi-directionally predictive-coded). The frame classification is mainly based on the procedure, through which each frame type has been generated and encoded. The successive frames between two succeeding I frames define a GoP.

This is shown in Fig. 1 where the arrows indicate the encoding/decoding correlation between the frames and more specifically, the fact that the B and P frames depend on the respective preceding and succeeding I or P frames. Thus, in a GoP an I frame is the main reference-point of a P and B frames, and the I frame is coded without any reference to any other frame.

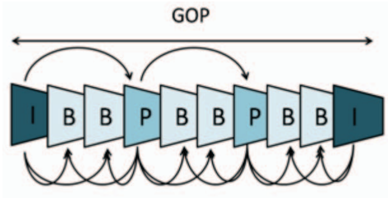


Figure1. MPEG GoP Structure

In addition, P frames are predicted from I frames and from another P frame, although only in the forward time manner. Each P frame within the GoP is predicted from the frame immediately preceding it (an I or a P frame). The B frame uses forward/backward-interpolated prediction. B frames have references from the previous I or P frames, as well as from the succeeding I or P frames and B frames are not used to predict other frames, as shown in Fig. 1.

B. Implementation of MultiQoE in a WMN

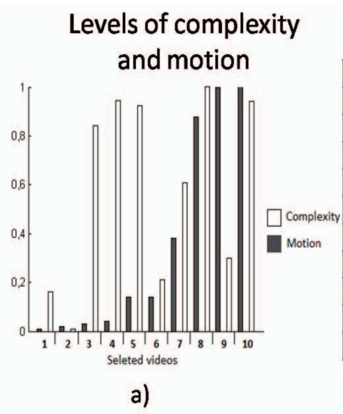
Four successive stages were required to implement the MultiQoE approach in a networking system (including WMNs): (1) Quality-affecting factors; (2) Distorted video database generation; (3) Subjective Quality Assessment and; (4) Learning the quality of the behavior with MANN.

In the first stage, a set of factors affecting quality is selected that have an impact on quality. The factors selected are parameters that can differentiate between the levels of impact that can affect the video quality when a packet is dropped, including the percentage losses of I, P and B frames, video content characteristics (video motion and complexity) and GoP length as quality estimators. We consider that the I, P and B frames loss rate is important due to the hierarchical structure of encoding methods (e.g., MPEG). Also, the GoP length is used as an input factor because it changes dramatically the amount of I (and other) frames in the video sequence.

For this paper, ten videos were selected from the Video Trace Library [24] to generate a distorted video database. The selected videos have high, medium and low levels of motion and complexity, which are depicted in Fig. 2.

Fig. 2.a shows that four videos have low motion, two videos have medium motion and four videos have high motion. In Fig. 2.a it is possible to see that the relation between motion and complexity is not proportional. In four videos (3, 4, 5 and 9) the difference between motion-complexity was more than 70%. Both the name and the number of frames selected for each video is shown in Figure 2.b. The motion and complexity

values shown in Fig. 2.a were obtained by using an algorithm to calculate the estimated complexity that was adapted from [25]. Single images were recommended since this is a measurement for image activity that is derived from the amount of edges in an image.



Selected videos

Number	Video name	Frames
1	Akiyo	252
2	Mother and daughter	252
3	Hall monitor	252
4	Container	211
5	News	211
6	Silent	252
7	Coastguard	252
8	Mobile	252
9	Football	218
10	Flower	210

b)

Figure 2. (a) estimates of motion and complexity of videos and (b) selected Videos

To optimize the performance of the model for scenarios that reflect network environments in practice, it was necessary to transmit those videos through one existing scenario. Then, a study case of MultiQoE was conducted over the WMN backbone of the Federal University of Para (UFPA), which is located in Amazon/Brazil, as presented in Fig. 4. To ensure that the losses suffered in each video transmission were not arbitrary, an exponential loss model was applied.

In the third stage, the video database was evaluated subjectively. This involved asking a panel of human observers to evaluate the distorted videos. The *Mean Opinion Score* (MOS) [2] which was based on an average score obtained from all the observers and corresponding MOS, was put into two separate databases called 'training' and 'validation', as shown in Fig. 3.

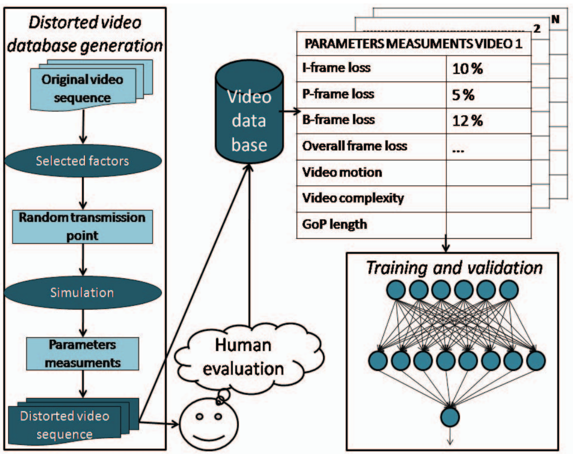


Figure 3. Methodology of MultiQoE

As Fig. 3 makes clear, after humans had evaluated each distorted video in specific parameter conditions, the training process was conducted with the video database for training to obtain the mapping between the selected parameters and human scores. Once the tool has been trained, it was proposed that a function *f* could map the selected parameters into MOS. After the training process, the validation task is carried out with a validated video database to ensure that the training is acceptable.

Once MANN has been trained and validated, MultiQoE can be used for real-time QoE prediction without any intervention by real viewers. It is necessary to measure the quality-affecting parameters at time *t* and evaluate these values with the MANN to obtain the perceived quality instantaneously. MultiQoE gives scores in terms of MOS and which are as close as possible to those of a human MOS.

IV. MultiQoE: Use Case

A. Test Environment, Scenario and Implementation

Advances in WMNs are essential for the future of next generation wireless systems. Thus, IEE 802.11s networks were selected to implement, evaluate and validate MultiQoE. As shown in Fig. 2.b, ten videos were encoded with H.264 and all of them have different patterns (duration, complexity and motion) and three different GoP length, namely 10, 20 and 30.

The selected scenario is based on the mesh backbone used by UFPA, which is formed by six mesh routers and two gateways as depicted in Fig. 4. As well as the mesh backbone, a mesh client was simulated to receive video streaming from gateway 1 or 2. The client experiences a different loss rate because, in each simulation, the location of the client (different wireless conditions) was chosen at random.

Figure 4. UFPA mesh backbone

The simulated experiments were carried out by using Network Simulator 2.34 [26], Evalvid tool [27], MSU Video Quality Measurement Tool (VQMT) [28] and the MANN was built using MATLAB. This was evaluated together with a well-known MOS so that a comparison could be made between MultiQoE and PSQA [16].

When Evalvid was employed, the simulations were able to generate the real simulated video. Each selected video was simulated 90 times to provide a large enough video database, and a total of 900 videos with GoP length 10, 20 and 30, were obtained. 810 videos were selected from this database for the training base and 90 videos selected for the validation database. This selection was carried out at random.

An MOS system based on the [2, 25, 31] recommendations with a total of 55 observers was employed for subjective evaluation. They had normal vision and their age ranged from 18 to 45 years old. The observers include undergraduate students, postgraduate students, and, university staff. After each observer had given its own score, the MOS average score was calculated for each video. The possible MOS scores range from 1 to 5, as shown in Table I.

TABLE 1. MOS SCORES

MOS	QUALITY	IMPAIRMENT
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

B. Performance Evaluation and Results

The performance results obtained with MultiQoE are presented in Fig. 5a. The results reveal that, on average, MultiQoE gives the same scores as MOS in 90% of the experiments, while for PSQA only in 41%. The advantages of MultiQoE over PSQA are clear and demonstrate its accuracy and efficiency in predicting the quality level of Internet-based video in emerging wireless multimedia systems.

Compared to main QoE objective metrics, MultiQoE presents yields much better results in measuring the quality level of videos as shown in Figs. 5b, 5c and 5d. All of the objective schemes show low performance for all experiments, where the PSNR, SSIM and VQM only provides good estimation for 18.88%, 15.55% and 27.77% of the tests, respectively. It is fair to say that these metrics do not correlate well with human video system perception.

To summarize the performance of MultiQoE, Figs. 6a and 6b present the absolute accuracy of MultiQoE, PSQA, and objective (PSNR, SSIM and VQM) models against MOS scores for each cluster type and GoP length. The absolute accuracy is the number of time that the estimated QoE error is zero (e.g., MultiQoE MOS is equal to human MOS).

Moreover, MultiQoE is a more stable solution for QoE estimation for videos with different content characteristics compared to objective QoE metrics and PSQA.

Fig. 6 illustrates the benefits of the QoE assessment mechanisms for each GoP length. In the case of GoP 10, the MultiQoE and PSQA absolute accuracy is 90% and 32%, respectively. For GoP 20, MultiQoE presents an accuracy of 85%, while PSQA only 35%. Finally, for video with GoP 30, the MultiQoE accuracy is 85%, while PSQA 70%, where PSQA shows better performance because the I frame loss has a higher impact on the video quality but it yields low performance for GoP lengths 10 and 20. On average, MultiQoE improves the system accuracy by 41% when compared to PSQA.

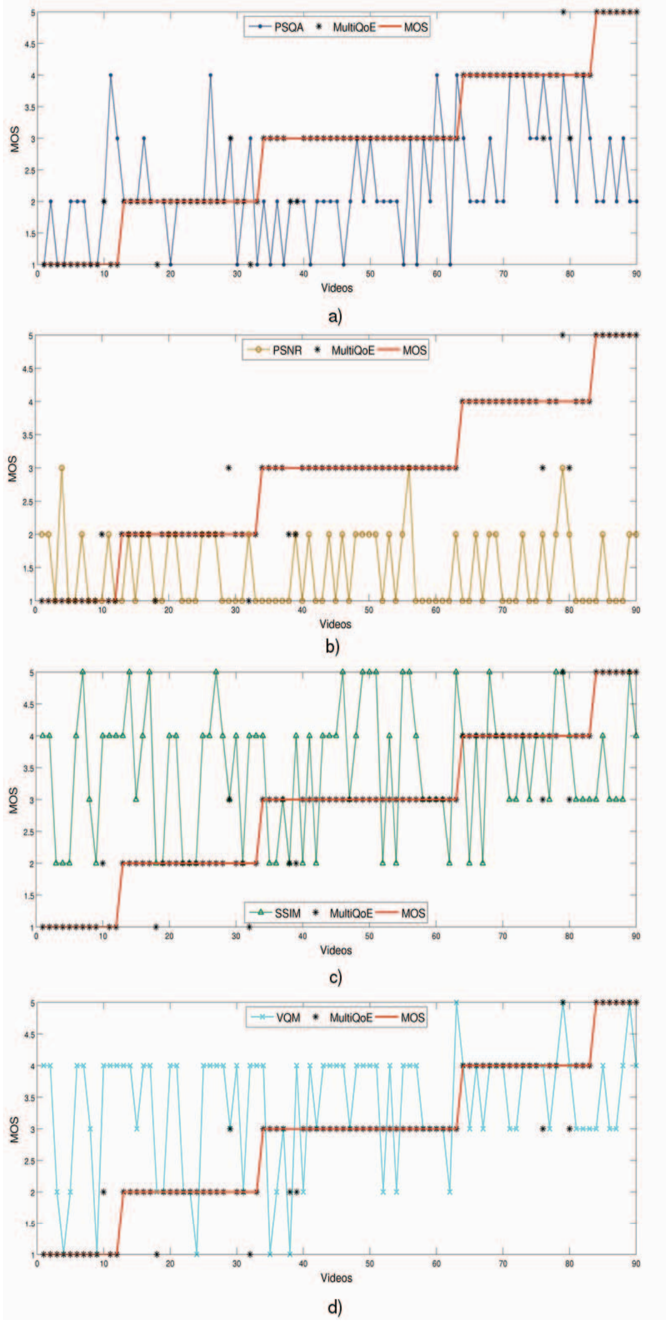


Figure 5. Use Case Results: a) MultiQoE vs. PSQA vs. MOS; b) MultiQoE vs. PSNR vs. MOS; c) MultiQoE vs. SSIM vs. MOS; d) MultiQoE vs. VQM vs. MOS.

V. CONCLUSIONS

Current QoS video measurement approaches implemented by network and content providers are not enough to match user experience. Therefore, understanding the user perception in real-time without the need for any decoding processing is becoming a key strategy for multimedia/network providers and the best way to keep and attract new clients, while offering new value-added services with QoE-awareness. Thus, customers can be notified about the real quality level of their multimedia applications, while operators improve their profits, optimize the usage of scarce network resources, and increase the satisfaction of customers.

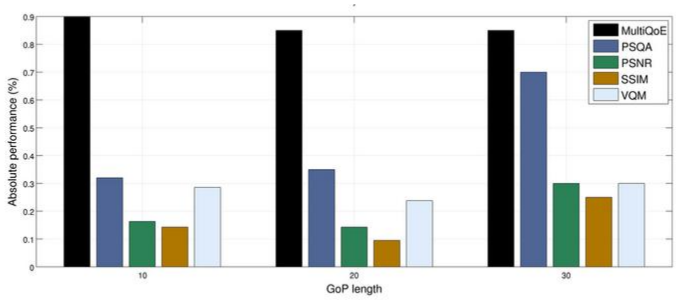


Figure 6. Absolute accuracy for each GoP length

This paper introduced MultiQoE to monitor/predict the experience of the user with emerging video applications. A use case in a WMN was implemented to demonstrate the impact and benefits of MultiQoE in a wireless network. The results obtained show that MultiQoE provides results equal to MOS in approximately 87% of the tests, while for PSQA only in 45%. The benefits of MultiQoE are also highlighted when compared to PSNR, VQM and SSIM which do not correlate well with human visual perception. Based on the results obtained, we conclude that MultiQoE is a suitable solution to assess the quality level of emerging real-time videos in multmedia-aware networks. For future works, MultiQoE will be further evaluated with gaming and 3D videos.

Acknowledgment

This work was funded by Coordination for the Improvement of Higher Level Personnel (CAPES), National Council for Scientific and Technological Development (CNPq), Fundação Amazônia Paraense de Amparo à Pesquisa (FAPESPA) and FCT UBIQUIMESH project (PTDC/EEATEL/105472/ 2008).

References

[1] Stankiewicz, R.; Cholda, P.; Jajszczyk, A.; “QoX: What is it really?”, Communications Magazine, IEEE , vol 49, no 4, pp. 148-158, 2011.

[2] Mu M *et all*; “Framework for the integrated video quality assessment”, Springer Multimedia Tools and Applications, pp. 1-31, 2011.

[3] ITU-R Recommendation BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” International Telecommunication Union, Geneva, Switzerland, 2002.

[4] M. H. Pinson and S.Wolf, “A new standardized method for objectively measuring video quality,” IEEE Trans. Broadcasting, vol. 50, no. 3, pp. 312–322, September 2004.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing,, vol. 13, pp. 600-612, 2004.

[6] Y. Park, Y. Seok, N. Choi, Y. Choi, and J.-M. Bonnin, “Rate-adaptive multimedia multicasting over IEEE 802.11 wireless LANs,” Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE, vol. 1, pp. 178–182, 8-10 Jan. 2006.

[7] M. Venkataraman and M. Chatterjee, “Inferring video QoE in real time,” IEEE Network, vol. 25, no. 1, pp. 4–13, 2011.

[8] S. Mohamed and G. Rubino, “A study of real-time packet video quality using random neural networks,” Circuits and Systems for Video Technology, IEEE Transactions on, vol. 12, no. 12, pp. 1071–1083, 2002.

[9] G. Rubino, M. Varela and J. Bonnin “Controlling multimedia QoS in the future home network using the PSQA metric”, The Computer Journal, vol. 49,issue 2,pp. 137–155,2006.

[10] J. Gross, J. Klaue, H. Karl, and A. Wolisz, “Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming,” Computer Communications, vol. 27, no. 11, pp. 1044–1055.

[11] H. Cancela, P. Rodriguez-Bocca, G. Rubino and A. Sabiguero. “A robust P2P streaming architecture and its application to a high quality live-video service,” The IV Latin-American Algorithms, Graphs, and Optimization Symposium, pp. 219 – 224, nov. 2008.

[12] A. Silva, M. Varela, E. Silva, E. Leão, and G. Rubino, “Quality assessment of interactive voice applications,” Computer Networks: The International Journal of Computer and Telecommunications Networkingm, vol. 52, issue 6, 2008.

[13] H. Cancela, P. Rodriguez-Bocca, G. Rubino and A. Sabiguero, “Perceptual quality in P2P multi-source video streaming policies”, GLOBECOM, Washington, DC, pp. 2780-2785, nov. 2007.

[14] K. Piamrat, C. Viho, J. Bonnin and A. Ksentini, “Quality of Experience Measurements for Video Streaming over Wireless Networks”, International Conference on Information Technology: New Generations, p.1184-1189, Las Vegas, Nevada, USA ,april 2009.

[15] D. De Vera, P. Rodr´ıguez-Bocca, and G. Rubino, “QoE monitoring platform for video delivery networks“. In Proceedings of the 7th IEEE International Conference on IP Operations and Management . IPOM' 2007, pp. 131-142.

[16] M. Ghareeb and C. Viho, “Hybrid qoe assessment is well-suited for multiple description coding video streaming in overlay networks,” in 8th Annual Communication Networks and Services Research Conference. IEEE, 2010, pp. 327–333.

[17] K. Piamrat, “QoE-based network selection for multimedia users in IEEE 802.11 wireless networks,” IEEE Conference on Local Computer Networks, pp.388-394, out. 2008.

[18] K. Piamrat, “QoE-Aware Admission Control for Multimedia Applications in IEEE 802.11 Wireless Networks,” Vehicular Technology Conference, Calgary, pp.1-5, set. 2008.

[19] M. Mu, “A Discrete Perceptual Impact Evaluation Quality Assessment Framework for IPTV Services,” IEEE International Conference on Multimedia & Expo, Singapore, july 2010.

[20] R. Sousa *et all*, “GoP size influence in high resolution video streaming over wireless mesh network”, in Computers and Communications (ISCC), 2010 IEEE Symposium on, 2010, pp. 787–789.

[21] A´. Husza´k and S. Imre, “Analysing GoP structure and packet loss effects on error propagation in mpeg-4 video streams,” in Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on. IEEE, 2010, pp. 1–5.

[22] J. Greengrass, J. Evans and A. Begen, “Not All Packets Are Equal, Part I and II,” IEEE Internet Computing, 2009.

[23] J. Mitchell, W. Pennebaker, C. Fogg, D. LeGall, “MPEG Video Compression Standard,” Chapman & Hall, New York, 1996.

[24] Video Trace -Video Trace Library. http://trace.eas.asu.edu/

[25] ITU-T Recommendation P.910, “SubjectiveVideo Quality Assessment Methods for Multimedia Applications,” International Telecommunication Union, Geneva, Switzerland, 1999.

[26] NS2, “Network simulator 2.34,” http://www.isi.edu/nsnam/ns/, 2010.

[27] J. Klaue, B. Rathke, and A. Wolisz, “Evalvid - a framework for video transmission and quality evaluation,” Computer Performance Evaluation. Modelling Techniques and Tools, pp. 255–272, 2003.

[28] D. Vatolin, A. Moskvin, O. Petrov, and N. Trunichkin, “Msu video quality measurement tool,” http://www.download3k.com/Install-MSU-Video-Quality-Measurement-Tool.html.

[29] ITU-T Recommendation P.911, Subjective Audiovisual Quality Assessment Methods for Multimedia Applications International Telecommunication Union, Geneva, Switzerland, 1998.