Empirical Study of Subjective Quality for Massive Multiplayer Games

Michal Ries, Philipp Svoboda, Markus Rupp
Vienna University of Technology
Institute of Communications and Radio-Frequency
Gusshausstasse. 25, A-1040 Vienna, Austria
Phone: (+43) 1-58801 38999 Fax: (+43) 1-58801 38931 E-mail: (mries, psvoboda, mrupp)@nt.tuwien.ac.at

Keywords: online gaming, QoS, quality metric, MMOG.

Abstract – Online gaming is a member of the real time interactive multimedia services. This service suffers from all impairments on the networking layer that increase the latency, respectively the responsiveness. The Quality of Service (QoS) of online gaming is becoming an important issue due to a massive increase of online games (Half-Life,War-Craft) with a significant player community. The purpose of this project is to investigate the impact of network impairments (delay, jitter) on the perceived QoS. We hereby focus on Massively Multiplayer Online Games (MMOG), due to their amount of active users and their popularity. The initial step is to develop a test setup for subjective quality evaluation for MMOG games. The final outcome of this paper is to propose a QoS metric for online gaming.

1. INTRODUCTION

Recently, significant growth of online games has taken place [1]. Players were attracted to these interactive multimedia services, due to the substantial increase of penetration of broadband Internet services, availability of gaming consoles as well as the introduction of game titles already known and established in the gaming community available in interactive Internet environments.

Online Gaming is an application that evolved from applications called multiplayer games. In multiplayer games two or more players enter the same game either in competition or in cooperation mode. The early adaptations were realized by the same machine, later the players anticipated by playing on separate computers connected via a network.

Recent developments already allow to define different classes of games: a First Person Shooter (FPS) for fast action games, Real Time Strategy (RTS) for slow interaction and MMOG for long lasting adventures. The most significant game class is MMOG. The ten most popular MMOG games have more than 35 million active subscribers world wide [2]. The most popular game of all is World of Warcraft (WoW) with more than nine million subscribers alone.

For the provisioning of online gaming services it is essential to provide a required level of customer satisfaction, given by the perceived gaming quality [15]. The subjective requirements vary according to the gaming class. FPS games have very strict limitations to RTT (round trip times) or delay for good service quality. The MMOG class is much less sensitive to network delays. In [3] we discovered online gaming traffic of an MMOG, namely *World of Warcraft*, within a live 3G cellular network. We decided in this paper to investigate the MMOG service class, in order to perform a QoS study regarding impact of delay and jitter on subjective quality.

The paper is organized as follows: Section 2 describes actual state of the art in area of empirical modelling for online games. In Section 3 we describe the measurement and test setup for subjective online gaming evaluation. Key Performance Indicators (KPI) results are analyzed and mapped, and for the subjective data in Section 4, the main focus is given on the subjective quality estimation. Section 5 contains conclusions.

2. RECENT WORK

The recent work in this area is mainly focused on determining KPI, e.g., delay, packet loss and jitter, and investigating their impact on the game play [3], [4], [5], [6], [7], [8], [9]. Moreover, subjective experiments for FPS were conducted in [11]. Earlier papers had investigated only on the maximum impairments that a user will tolerate. The first paper to introduce a model for the perceived QoS is [10]. In this work an MOS model is designed for an FPS game, namely Quake IV. The focus of our research group is towards mobile cellular networks. The network conditions of these systems are far above the minimum requirements of fast action games such as FPS (see [11]). However, with the introduction of HSDPA we assume that a good perceived quality may be obtained for MMOGs. Therefore, we decided to evaluate the MOS for mobile networks.

Our approach is triggered on subjective evaluation of the MMOG game class and the introduction of a model for automatic subjective quality estimation for MMOG games. Finally, a model is proposed for estimation of subjective quality in online gaming. To our knowledge this is the first paper which investigates MOS scores for MMOG games up to now.

3. MEASUREMENT AND TEST SETUP

The aim of this work was to develop test setups for subjective quality evaluation and to map network parameters on subjective data. As an initial step, it was necessary to develop a controlled measurement environment which allows us to determine and collect subjective and network parameters. Finally, we investigated mutual dependencies of the obtained data.

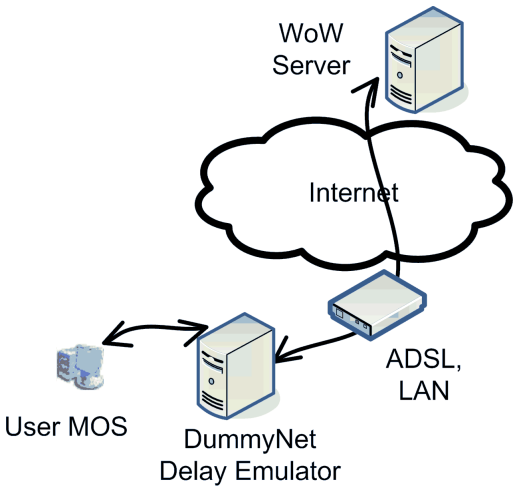

Figure 1: Measurement setup with additional delay.

3.1 Measurement setup

In contrast to fast action games, which are based on UDP (user datagram protocol) traffic, WoW runs on top of TCP (transport control protocol). The packet loss on the MAC layer is handled directly in the transport protocol causing additional delay. Therefore, we decided to test the user satisfaction concerning the input parameter delay.

The setup which we installed for the testing is based on active measurements. Figure 1 depicts the measurement setup following similar conditions as in video quality estimation [14]. We set up a standard PC to run the WoW client without any modifications. The PC was connected to the Internet via a `dummynet`[1] system. The `dummynet` software is a flexible tool based on FreeBSD which can be used for bandwidth management. In fact the software was designed to test protocols and can be reused to shape bandwidth. It introduces an intermediate layer into the protocol stack that can be configured by the user. The computer is set up as a bridge between the test device and the Internet access. The `dummynet` is then used to add an additional network delay to the Internet access. Therefore, this setup allows one to evaluate the user satisfaction for different delays and jitters.

Jitter or packet delay variation is the difference of the end-to-end delay between different packets within a flow. A flow is a TCP connection between two different hosts that may transport many packets. The term jitter itself is not well defined in terms of computer networks. We define jitter in this work as the variation of the packet delays. The jitter is realized with a number of pipes, each featuring a discrete delay value. For each packet another pipe is selected, based on a probability which was set as an input parameter. The resulting jitter is therefore only an approximation of a delay distribution.

As we have no control over the quality of the connection to the game servers and the reaction time we cannot directly use the delay values set via the `dummynet`. Therefore, we evaluated the TCP delay based on packet dumps from a `wireshark`[2] client running on the WoW client. The RTT is the time a message needs to travel to a remote host and back again. This time is important to data systems which adapt their throughput rate based on the delay. In TCP the RTT is measured between segment transmission and ACK receipt. It is then estimated using Karn's Algorithm, TCP timestamps (see RFC 1323) or Jacobson's algorithm. In this work we estimated the RTT according to the Karn's algorithm. The delay presented in the figures is the RTT divided by two, e.g., 40 ms of delay correspond to 80 ms of RTT.

3.2 The test setup for online gaming evaluation

Our online gaming test setup was designed in order to reflect a typical real world scenario. The online gaming scenario is specified by the game interactivity and the network performance. Moreover, the tested scenario followed ITU recommendations "Methods for subjective determination of transmission quality" [12] and "Subjective video quality assessment methods for multimedia applications" [13], although the online gaming scenario is strictly different in comparison with classical broadband video conference or IP-TV services. The ITU recommendations are not designed specifically for online gaming but they do provide a sufficient basis for subjective testing of interactive multimedia applications regarding also online games. Subjective testing was used for typical office desktop PCs with LCD screens. The screen resolution was set to 1280x768 which today represents a standard setting. The viewing distance from the LCD screen is not fixed, but selected by the test person. We have noticed that the users are comfortable viewing the LCD at a distance of 35-50 cm. The test was carried out in our video quality laboratory.

Figure 2: Snapshot of tested MMOG game: World of Warcraft

To obtain MOS values, we worked with 10 volunteers. The chosen group of test persons varied in age (between 20 and 30), gender, education and experience with image

[1] http://www.dummynet.com/

[2] http://www.wireshark.org/

processing. The test method was Absolute Category Rating (ACR) as it imitates marks provided by the user. ACR is a category judgment method. One test sequence is presented at a time and rated independently on a category scale. The subjects do not know the undistorted situation and rate the quality they perceived in their opinion on a scale from five to one (five is excellent, one is bad).

Thus, the subjects did not have the optimal network performance as a reference, resulting in a greater variance. The tested network settings were presented in arbitrary order and the test environment followed ITU recommendation [13]. People evaluated the subjective quality after each sequence using a five grade MOS scale (1-bad, 5-excellent) in a prepared form.

3.3 The test setup for online gaming evaluation

We already depicted the test setup in Figure 1. The ADSL line provided a downlink rate of 2 Mbit/s and an uplink rate of 512 kbit/s. According to our analysis in [3] the game should generate less than 128 kbit/s.

The parameters for the measurements are shown in Table 1. The delay added in the `dummynet` system was increased in steps of 20 ms up to 200 ms, for each step we ran all of the different jitter settings. The delay parameters for acceptable playing quality are well within the reach of modern mobile cellular networks [3]. These values are consistent with traces from a mobile core network that did show more than 1000 active gamers in a 3G network in Austria. The jitter was modified in steps of 10 ms. There were 15 different settings for each users. The test duration lasted between three and five minutes. The actual parameters remained unknown to the test subjects themselves, in order to avoid a bias in the MOS.

Delay (ms)	40, ... , 200
Jitter (ms)	0, 10, 30
ADSL data rate	2/0.5 Mbit/s

Table 1: Setup parameters.

4. TEST RESULTS

The obtained MOS data are depicted in Figure 3. From visual inspection we can already assume that delay has stronger influence on MOS than jitter. As expected, an increased delay has a negative impact on the subjective quality. Regarding the strong correlation we can conclude that packet loss will lead to a strong reduction in the perceptual quality. Note, a lost packet will be retransmitted by the TCP agent, introducing one additional RTT equal twice the actual delay. An increase of delay variation also has a negative impact on the subjective quality. Although the impact is smaller than that seen for FPS games (see also [11]).

4.1 Principal Component Analysis

Furthermore, we used a well known multivariate statistical method the Principal Component Analysis (PCA) [16] for determining the relevance of the two parameters delay and jitter. PCA allows us to propose a compact description of our data set. In our case, two components proved to be sufficient at first for an adequate modelling of the variance of the data (see Figure 4), because the total variability of the first two components is over 95%. Variability describes the percentage of the data sets variance that is covered by the variance of particular component. The horizontal axis represents Principal Component 1 (PC 1) and the vertical axis represents PC 2 in Figure 4. Each of the parameters is represented in the figure by a vector. The direction and length of the vector indicates how each parameter contributes to the two principal components in the graph. Moreover, Figure 4 shows a strong correlation of delay and MOS with component 1 and jitter with component 2. Moreover, the directions of the vectors show mutual independence of delay and jitter. Finally, we draw the conclusion that delay and jitter are KPI parameters for our game class.

4.2 Metric design

The previously reported experience determines our further steps in metric design and allows us to estimate MOS out of these parameters. We propose a single linear metric for the MMOG game class. We designed our metric using two objective parameters.

$$\text{MOS} = 5.17 - 0.012 \cdot \text{Delay} - 0.018 \cdot \text{Jitter}. \qquad (1)$$

To validate the performance of our proposed metric, we used the Pearson (linear) correlation factor. In order to provide a detailed overview we calculated the Pearson correlation factor separately for both resolutions. The performance of the proposed metric is 97.14%, depicted at the scatter plot in Figure 5.

5. CONCLUSION

In this paper we proposed a perceptual quality metrics for WoW which is a member of the MMOG game class and investigated its features. We evaluated the simultaneous impact of delay and jitter (delay variation) on the obtained MOS score. A packet loss was not considered as the transport protocol of WoW is TCP. The TCP agent will take care of lost packets by retransmitting the lost segments. Therefore, a packet loss can be interpreted as an increased delay. We had 10 volunteers which played short sessions of three to five minutes in 15 different settings. Each setting was then benchmarked according to the absolute category rating. The users did not know the actual setting of the network under test. We then applied a PCA analysis to the dataset. The PCA analysis shows that MOS was influenced more by the delay, than the jitter, although, the jitter impact is not negligible. The presented results clearly demonstrate that it is possible to estimate subjective perceived quality of the MMOG game class. Finally, we can conclude that the MOS score for WoW mainly depends on delay, but in fact it is much less sensitive to these network parameters than for example FPS games.

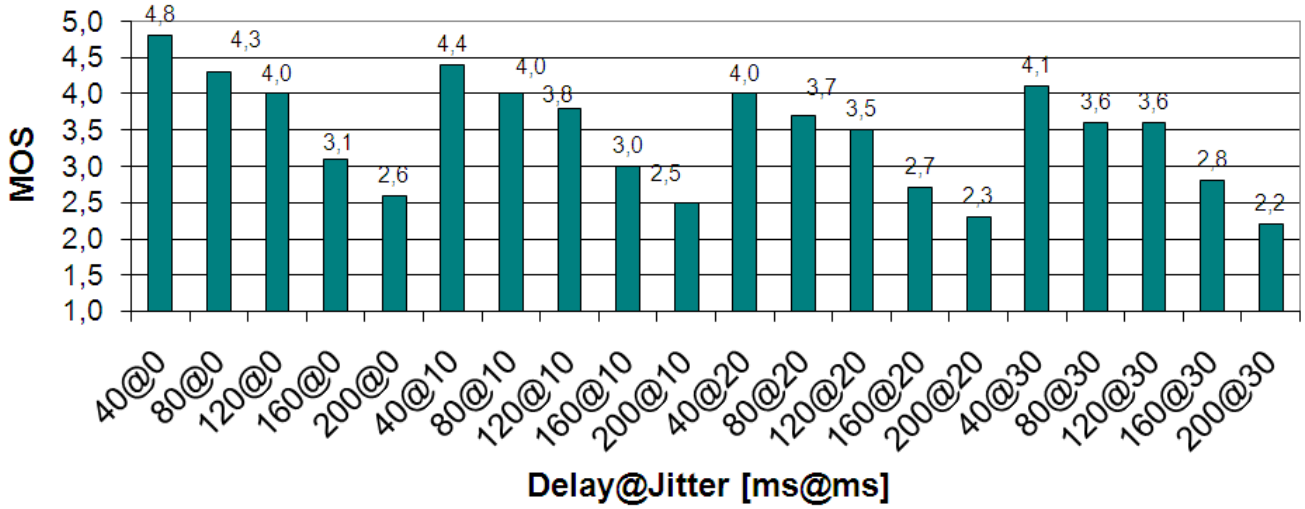

Figure 3: Obtained MOS results.

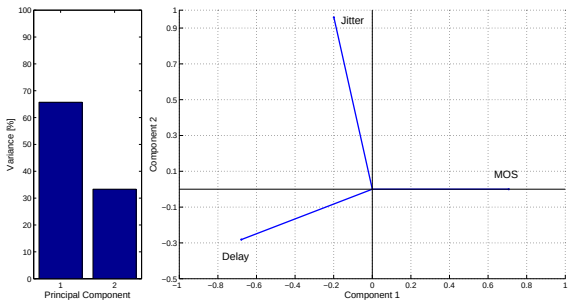

Figure 4: Visualization of PCA results.

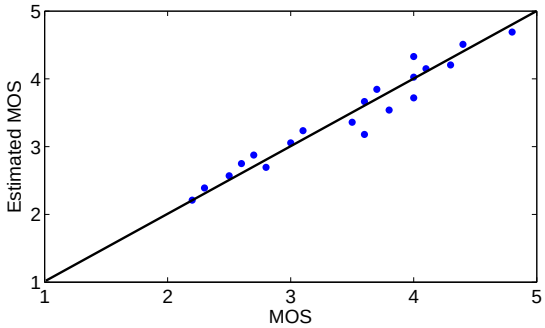

Figure 5: Predicted vs. subjective MOS results.

ACKNOWLEDGMENT

The authors would like to thank mobilkom austria AG for supporting their research. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG.

REFERENCES

[1] P. Svoboda, F. Ricciato, R. Pilz, E. Hasenleithner "Composition of GPRS, UMTS traffic: snapshots from a live network," IPS-MOME 2006 4th International Workshop on Internet Performance, Proc. of IPS-MoMe 2006, Salzburg, Austria, pp. 40 – 51, 27.02.2006 - 28.02.2006.

[2] Blake Snow,"GigaOM Top 10 Most Popular MMOs," Jun. 2007, http://gigaom.com/2007/06/13/top-ten-most-popular-mmos/.

[3] P. Svoboda, W. Karner, M. Rupp, "Traffic Analysis and Modeling for World of Warcraft a MMOG," Proc. of the ICC 2007, Jun. 2007.

[4] M. Dick, O. Wellnitz, L. Wolf, “Analysis of factors affecting Players’ Performance and Perception in Multiplayer Games,” NetGames’05, Hawthorne (NY), U.S.A., 2005.

[5] Ubicom White Paper, “OPScore, or Online Playability Score: A Metric for Playability of Online Games with Network Impairments,” 2005, http://gamer.ubicom.com/pdfs/whitepapers/IP3K-DWP-OPSCORE-10.pdf.

[6] G. Armitage, L. Stewart, "Limitations of using Real-World, Public Servers to Estimate Jitter Tolerance Of First Person Shooter Games," Proc. of ACM SIGCHI ACE2004 Conference, Singapore, Jun. 2004.

[7] T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, M. Claypool, “The Effect of Loss and Latency on User Performance in Unreal Tournament 2003®,” NetGames’04, Oregon, USA, 2004.

[8] T. Henderson, S. Bhatti, "Networked Games – a QoS sensitive application for QoS-insensitive users?", ACM SIGCOMM 2003 Workshops, Aug.25 & 27, Karlsruhe, Germany, 2003.

[9] M. Oliveira, T. Henderson, "What Online Gamers Really Think of the Internet?", NetGames’03, Redwood City (California), U.S.A., 2003.

[10] S. Zander, G. Armitage, "Empirically Measuring the QoS Sensitivity of Interactive Online Game Players," Proc. of ATNAC 2004, Sydney, Australia, Dec. 2004

[11] A. F. Wattimena, R. E. Kooij, J. M. van Vugt, O. K. Ahmed, "Predicting the perceived quality of a first person shooter: the Quake IV G-model," Proc. of 5th ACM SIGCOMM, Singapore, Oct. 2006.

[12] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality," Aug. 1996.

[13] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Sep. 1999.

[14] Olivia Nemethova, Michal Ries, Eduard Siffel, Markus Rupp, ”Quality Assessment for H.264 Coded Low-Rate and Low-Resolution Video Sequences,” Proc. of IASTED Internat. Conf. on Communications, Internet and Inf. Technology (CIIT), Virgin Islands, pp. 136-140, Nov. 22-24, 2004.

[15] Philipp Svoboda, Markus Rupp, "Modelling Online Gaming in Wireless Applications," Proc. of EuroISMA, Internet and Multimedia Systems and Applications Conference 2005, Grindelwald, Switzerland, pp. 417-422, Feb. 21-23, 2005.

[16] W. J. Krzanowski, "Principles of Multivariate Analysis," Clarendon press, Oxford, 1988.