Subjective-Based Quality Assessment for Online Games

Dorel Picovici
Institute of Technology Carlow
Kilkenny Road, Ireland
+ 353 59 9175438
Dorel.Picovici@itcarlow.ie

David Denieffe
Institute of Technology Carlow
Kilkenny Road, Ireland
+ 353 59 9175410
David.Denieffe@itcarlow.ie

Zenun Kastrati
University of Prishtina
Mother Teresa Avenue 9, Kosovo
+381-38-244 183
Zenun.Kastrati@fiek.uni-pr.edu

ABSTRACT

This paper describes a new game assessment metric for the online gamer. The metric is in line with a mathematical model currently used for network planning assessment. In addition to the traditional network-based parameters such as delay, jitter and packet loss, new parameters based on subjective assessment are introduced. The metric aims to estimate game quality as perceived by an online game player. In order to validate and calibrate the proposed metric a subjective game quality assessment is also proposed. Two 5-point scales are introduced: a game-quality scale and a game playing-effort scale. The mean average of each scales termed, as Mean Opinion Score (MOS), will indicate the game quality (MOS_{GQE}) and the playing-effort required (MOS_{GPE}). Reported evaluation results indicate a high level of correlation when compared with other algorithms. Comparative results have been carried out for three online games.

Categories and Subject Descriptors

C.2.3 [**Network operations**]: Network monitoring, D2.8 **[Metrics]:** Performance Measures, H.3.4 **[Systems and software]:** Performance evaluation-*efficiency and effectiveness.*

General Terms

Algorithms, Measurement, Performance, Experimentation, Verification.

Keywords

Online game quality assessment, objective/subjective game quality assessment, end-user opinion estimation.

1. INTRODUCTION

This paper introduces a new assessment metric capable of estimating end-user/players perception and performance for online games. The metric is in line with an ITU-T recommended network planning model [1] and the estimation of the end-user/player perception is termed as the Mean Opinion Score Game Quality Estimation (MOS_{GQE}). The assessment can be achieved in real or non-real time for various categories of games such as: games of skill, games of chance, games of strategy, and simulation games.

DISIO 2010 March 15, Torremolinos, Malaga, Spain.

Traditionally the end-user perception for online games has been measured using subjective game quality assessment only. This can be seen in the work of Armitage [2] and Armitage and Stewart [3].

The proposed model, which represents and includes an alternative, to the models based on network impairments only, could have several applications:

a) *Game service providers*: the impact caused by network impairments that may arise during a multiplayer game session will be quantified by the MOS_{GQE}. Using the MOS_{GQE}, the game service providers could enhance the online games quality and the overall Quality of Service (QoS).

b) *Game players*: the MOS_{GQE} can be used to inform the game player about the game quality at a point of time before starting and for the duration of playing the game. If the players' choice is to start/join an online game, the MOS_{GQE} will provide the game player with a continuous feedback about the network state and game quality.

c) *Game Developers*: game developers could take into account the MOS_{GQE} in developing and or adapting an aspect of the online game which may be affected by network impairments.

d) *Network Service Providers*: those providers who aspire to the provision of large-scale distributed interactive applications as part of their network service could use the MOS_{GQE} to assist the policing of traffic and control congestion particularly at the access points or in a wireless environment.

The research motivation is based on the continuous growth in the popularity of online games. The availability of Internet has provided an extensive infrastructure with global connectivity for the games industry to develop and deploy online games. Over the last decade, existing research reports revealed that online games have more severe requirements that are not fulfilled by the Internet's best effort model when compared with bi-dimensional Internet interfaces such as World Wide Web (WWW). Monitoring and measuring the end-user requirements is thus of great importance for both game service providers and game players. This is particularly relevant with the launch of 'network-ready' consoles such as the Playstation 3 and the provision of network based services to consoles such as Xbox LIVE.

Research carried out by US Entertainment Software Association (ESA) revealed that the percent of online game players is as high as 44% from total game players [4]. The statistics indicate that the

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8721
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8721

average game player has been playing games for 12 years and the average game player is 33 years old. Between 1999 and 2005 there was as increase from 9 to 25 percent of game players with the average age over 50 years old. These statistics demonstrate a high level of online game players and in average each player provides revenue for game service industry for almost 30 years.

Researchers from University College London also published results of a questionnaire targeted at the online game community [5]. The aim of this research was to provide an insight into "What Online Gamers Really Think of the Internet". The main results show the followings:

a) For 75 % of the questionnaire respondents the average weekly time expenditure with online gaming was 5-10 hours.

b) 73% of the respondents consider themselves advanced players and 68% of them determine the purchase of their game equipment according the game requirements.

c) The network condition has a huge impact for online game players. 86% consider network problems as "annoying", 73% consider these problems as the main attribute the majority of game disruptions and 60% of players will abandon the game.

d) The network delay has been identified as the major component of network problems. 69% of respondents are using simple metrics such as "ping" to identify the network delay. A high proportion (60%) consider individual network delay to be a disadvantage and a majority of 85% would prefer game servers where such networks effects are equalized for all players.

e) A significant percentage of players (26%) are not able to adjust the game play in the presence of network problems and a majority of 85% wants to break the user isolation from the network and consider beneficial any form of feedback about the network state.

In order to maintain a large number of online players it is becoming essential for the game service providers to estimate game player perception and performance for their games. On the other hand, advanced information about game playing conditions would allow game players to select different online games, different networks and ultimately different tariffs.

Following this introduction, this paper is broken into 5 further sections. Section 2 outlines the current subjective and objective models and their weaknesses. Section 3 proposes a new model based on the concepts of a "Game Rating Factor" and leading to the MOS_{GQE}. A Subjective Game Assessment mechanism for comparison to the MOS_{GQE} is outlined in Section 4. Section 5 details the model calibration and evaluation, while results and conclusions are presented in Section 6.

2. EXISTING MODELS

Much work has been done at examining, at many different layers, features and metrics to quantify and compensate at in an effort to improve player reaction in the online environment. The focus on distributed interactive applications and the adaptation of the application at both client and server has led to changes in the way games are designed and developed for online purposes. The work of Mauve [6] and the proposal for a Real-Time Transport Layer Protocol with interactive focus shows the work carried out at the Application and Transport Layer level.

Cognizance must be given to the Network Layer and the impairments that occur or the improvements, which can be made. With the move to wireless media and the growth of high-speed access networks, the changing nature of the provision of online interactive applications such as games has meant that distributed bandwidth-intensive features now require network awareness. In terms of network awareness, user perception in terms of player tolerance is crucial. There are, as outlined by Armitage, two distinct approaches for discovering player tolerance to network disruptions. The first is to build a controlled lab environment in which to test small groups of players under selected conditions or secondly to monitor player behavior on public servers over thousands of games. In conducting any experiment requiring assessment of psycho-perceptual factors, there are possible problems with the parameters defined and examined and any assumptions made.

Ubicom's [7] Quake-3 G-model defines R as an impairment factor given by the equation:

$$R = (W_L \times L + W_J \times J)(1+E) \quad (1)$$

where:

W_L is the Latency Weighting Factor set to 1 for this case;

W_J is the Jitter Weighting factor;

E is the packet loss;

L is the latency or one-way delay in ms;

J is the Jitter as defined in RFC 1889 in ms.

A successive approach proposed by Wattimena [3] uses a similar model named the Quake IV G-Model to predict the perceived quality of a First-Person Shooter. The perception is given by the following mapping function:

$$MOS\_model = -0.00000587 X^3 + 0.00139 X^2 - 0.114 X + 4.37 \quad (2)$$

where:

X – the network impairment -is given by the following equation:

$$X = 0.104 * ping\_average + jitter\_average. \quad (3)$$

The approaches described above have at least one of the following weaknesses:

- Although they claim to estimate a user perception, parameters used are based on network components which do not fully predict the user satisfaction.
- Both models are only defined for one specific game.
- Using simple network measures (Latency, ping, jitter) which assume the network is symmetrical.
- Implementing limited subjective testing for a short period of time with a small user pool.
- While a high level of correlation between the subjective and the proposed models is shown, this could be justified only for one defined game in restricted testing conditions.

The proposed model overcomes these weaknesses by:

- Extending the network based model to include all parameters-packet loss must be considered in the bandwidth-intensive application environment

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8721
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8721

- Including new parameters not proposed or used previously such as the user-based experience/knowledge factors and distortions introduced by user equipment.
- Moving from a game-specific model to a wide range of existing online games (including console based games)
- Using a number of users for subjective testing in line with accepted test models used in telecommunications based scenarios recommended by ITU-T
- Expecting a level of correlation in excess of 95%.

3. PROPOSED MODEL

Most of existent game quality assessment techniques take into consideration only network impairments therefore the measured games quality is only correlated with the network impairments (delay, jitter and to a limited extent packet loss) [8]. To estimate the player's overall perception of games quality the proposed model for game quality assessment extends the traditional objective game quality methods by introducing the end-user experience/knowledge. As shown in Figure 1 the proposed objective game quality assessment takes into consideration the following parameters:

a) end-user experience

b) distortions introduced by game client equipment (memory, graphic card) and I/O devices (screen, keyboard, and joystick)

c) distortions introduce by the network (end-to-end delay, jitter, packet loss)

d) distortions introduced by game server (number of users, game type, game capability to adapt to network distortions)

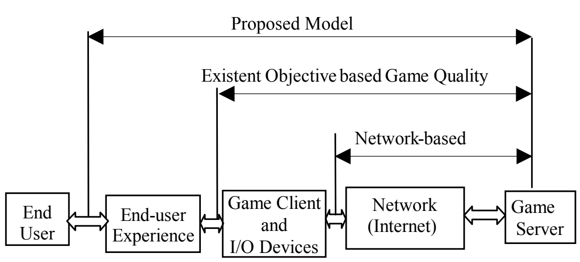

Figure 1: Game Quality Assessment.

Using the above-mentioned parameters a "Game Rating Factor" (*GRF*) is proposed. The *GRF* is inspired from an International Telecommunication Union; Telecommunication Standardization Sector (ITU-T) recommended computational model (E-model) [1]. The model is used to assess the combined effects of variation in several parameters that may affect end-user perception of speech quality. The computation of the *GRF* can be described as follows: a maximum value that reflects the highest level of game quality will be reduced in proportion with the distortions caused by various impairment parameters. Mathematically, *GRF* can be calculated using the following equation:

$$GRF = GRFMAX - IGCD - IN - IGS + A \qquad (4)$$

GRFMAX is the maximum Game Rating Factor (90)
IGCD: impairment factor representing all impairments due to Game Client and I/O device
IN: impairment factor representing all impairments due to network connection between the game server and game client
IGS: impairment factor representing all impairments due to Game Server
A: represents the end-user experience with online games (max 10)

3.1 Impairments due to Game Client and I/O device (*IGCD*)

To evaluate this parameter, a subjective test was carried out. Data related to game client's computer in terms of memory size, processor speed, graphic card and display was collected and combined into a as follows:

$$IGCD = 0.061 * (IGCDm + IGCDp + IGCDgc + IGCDd) \qquad (5)$$

Where:

IGCDm: represent the impairment due to the memory

IGCDp: represent the impairment due to the processor speed

IGCDgc: represent the impairment due to the graphic card

IGCDd: represent the impairment due to the display

For each of the above-introduced parameters, a score of a maximum value 5 is given. An example of such scoring is:

- if 3 GB memory is used, than a score of 5 is given
- if the system is using only 256 MB, than a score of 1 is given

The total sum impairment is 1.46, which is calculated based on the calculation made in [1], The decrease of this number depends on the quality of the above-mentioned devices.

3.2 Impairments due to Network Connection - *IN*

This factor represents all impairments due to network connection between game server and game client. This is comprised by two components *ID* and *IPL*, where the factor *ID* represents the impairment caused by delay *D*, and *IPL* represents the packet-loss dependent Effective Equipment Impairment Factor. The factor *ID* is 0 when the delay is less or equal to 20 ms. A small 40-byte IP Packet (48 bytes including PPP overhead) takes 11.4 ms on a V.90 upstream and 7.5 ms on a V.90 downstream; contributing 19 ms to any RTT measured using 40 byte packets [9]

Mathematically, *IN* can be implemented as follows:

$$IN = ID + IPL_{EIF} \qquad (6)$$

For $D \leq 20$ ms:

$$ID = 0$$

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8721
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8721

For $D > 20$ ms:

$$ID = 25\left\{\left(1+X^6\right)^{\frac{1}{6}} - 3\left(1+\left[\frac{X}{3}\right]^6\right)^{\frac{1}{6}} + 2\right\} \quad (7)$$

with:

$$X = \frac{\log\left(\frac{D}{20}\right)}{\log 2} \quad (8)$$

IPL_{EEF} , which represent the difference between impairment due to the packet loss and the effective equipment impairment factor (IPL_{EIF}=IPL-EIF), is calculated as:

$$IPL_{EIF} = IPL + (95 - IPL)\cdot\frac{Ppl}{\frac{Ppl}{BurstR} + Bpl} \quad (9)$$

:

Where

Bpl is packet lost Robustness factor and for low loss ratios which are not greater than 2%.

Ppl is the probability of packet loss, and

BurstR is so-called Burst Ratio, which is defined as:

$$BurstR = \frac{\text{Average length of observed bursts in an arrival sequence}}{\text{Average length of bursts expected for the network under "random" loss}}$$

(10)

3.3 Impairments due to Game Server - IGS

The factor IGS is the sum of all impairments which may occur more or less simultaneously with the game server. Mathematically is calculated as follow:

$$IGS = 20\left[\left\{1+\left(\frac{X_{IGS}}{8}\right)^8\right\}^{\frac{1}{8}} - \frac{X_{IGS}}{8}\right] \quad (11)$$

Where:

$$X_{IGS} = \frac{64}{NCS} + \frac{4}{SL} + 0.2*(64 - \frac{4}{SL}) \quad (12)$$

And

NCS represent the Number of Clients in the Server

SL represents the Location of Server

3.4 Advantage Factor A

The advantage factor A, relates to the game player's experience. Table 1 illustrates wording and values that can be used to assess how familiarized is the player with the game.

Table 1. User Experience/Skills

Player's Experience/Skills	Advantage (A)
First Time	0
Intermediate	5
Advanced	10

4. MEAN OPINION SCORE MOS_{GQE}

The proposed model uses a 5-point scale to measure the quality of an online game. The mean average of the scale, known as Mean Opinion Score (MOS), will indicate the game Quality (MOS_{GQE}), and mathematically is calculated using the following function:

For $GRF<0$ $MOS_{GQE}=1$

For $0<GRF<100$

$MOS_{GQE} = 1+0.035*GRF + GRF(GRF - 60)*(100-GRF)*7*10^{-6}$

For $GRF>0$ $MOS_{GQE}=5$

(13)

The quantity evaluated from the score is represented by the symbol MOS_{GQE} and represents the overall game quality, as perceived by a player.

Table 2. Game Quality Scale

Wording	Scale
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

The game quality scale is a scale from 1 to 5. Details of wording and score of this scale are given in Table 2.

5. PERFORMANCE EVALUATION

5.1. Test bed Implementation

The test bed used in this research work was built using two, Dell PowerEdge 2950 Servers (2.66 GHz Quad Core Processors, 32 GB of Memory and 1.6 TB HDD each).

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8721
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8721

The fist server was operating under Fedora core 9 with kernel 2.6.25 operating system[10]. A Fedora based network simulator (NCTUNS-5.0 [11]) was also used to simulate relevant game traffic. The network traffic generated by the simulator was monitored and analyzed using Wireshark network analyzer [12]. The second server was built as a dedicated online game server using Microsoft Windows Server 2003 as an operating system.

5.2. Model Calibration

To make the proposed model applicable and usable a calibration process was carried out. The reference model selected for calibration process is detailed in [1]. Linear dependence between the proposed and reference model was investigated using Pearson product-moment correlation coefficient. An additional fixed delay of 50, 150 and 500ms for different game such as: Counter Strike, Unreal Tournament 2004 and Need for Speed Underground 2 were introduced and their contribution to MOS_{GQE} analyzed.

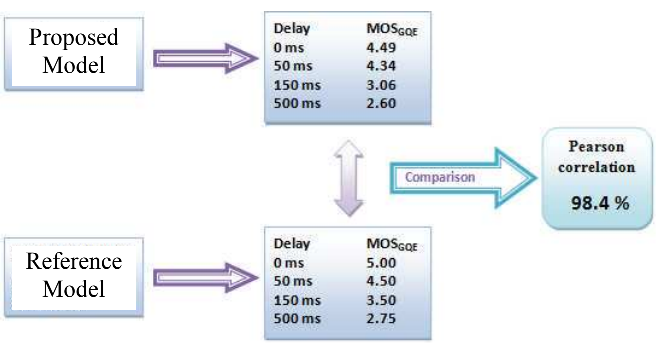

Figure 2. Correlation between proposed and reference model.

As is depicted in Fig. 2, a correlation coefficient of 98.4% for was indicated. This high value of correlation was deemed to be satisfactory for the proposed model.

5.3. Performance Evaluation

To evaluate the performance of the proposed model, several test cases conditions were investigated. Details of each test case are presented in Table 3.

Table 3. Performance evaluation

Test Case	Condition
1	Delay=20 ms
2	Delay=20ms, packet loss= 2%
3	Delay range [20-500]ms, loss rate 0% and 2%
4	Loss rates range [0-6]%

To simulate the Game Traffic Network, a Network Topology Layout based on [13, 13] was created. As detailed in Fig. 3 the network contains one Server dedicated as game Server and four hosts which are connected to the Server through the Router. To monitor and capture the network traffic a network analyzer (Wireshark) is used. Figure 4 illustrates some basic capturing features.

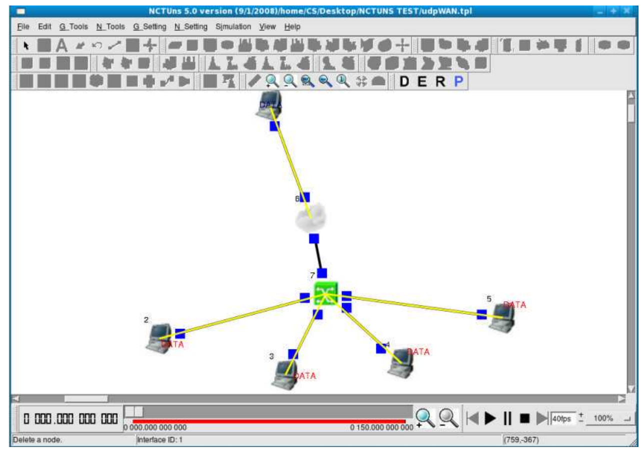

Figure 3. Network simulator configuration.

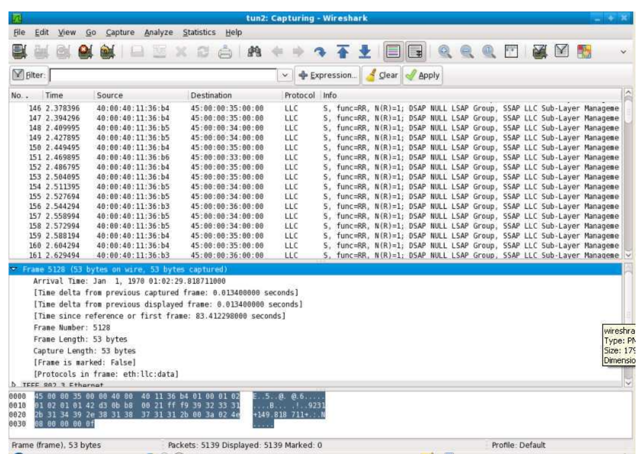

Figure 4. Basic network analyser features.

6. RESULTS AND CONCLUSIONS

As described in Section 5.3, the proposed model was tested under several test cases. The layout of the network was simulated using a network simulator (NCTUNS-5.0) and the generated network traffic monitored using a network analyzer (Wireshark). Table 4 details the obtained results in terms of Pearson Correlation with the reference model. As detailed below, this comparison was achieved for three different online games. The reported result clearly indicates a high level of correlation

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8721
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8721

Table 4. Pearson correlation results

Game	Pearson Correlation
Counter Strike	96.7
Unreal Tournament 2004	98.2
Need for Speed Underground 2	98.4

Test results indicated that the proposed model is generally effective in predicting the corresponding online game quality as perceived by an end-user. Further study is well underway to investigate more testing cases. Based on new games and new test cases scenarios, further adjustments for the proposed parameters used within the proposed model may be required.

7. AKNOWLEGMENTS

The authors would like to acknowledge the EU Tempus Programme Project No. CD_JEP-19090-2004 for financial support during this research work.

8. REFERENCES

[1] ITU-T Recommendation G. 107. The E-model, a Computational Model for use in Transmission Planning. Geneva, (March 2003).

[2] G. Armitage, Sensitivity of Quake 3 Players to Network Latency, Poster Session, SIGCOMM Internet Measurement Workshop, San Francisco, (November 2001).

[3] G. Armitage, L. Stewart, Thoughts on Emulating Jitter for User Experience Trials, Proceedings of NetGames 2004 Workshop, ACM SIGCOMM2004, Portland, Oregon, USA, (August 2004).

[3] 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. Proceedings of NetGames 2006, Singapore. (October 2006).

[4] 2006 Essential Facts about the Computer and Video Game Industry, Entertainment Software Association, E[3], (May 2006).

http://www.theesa.com/archives/files/Essential%20Facts%202006.pdf

[5] M. Oliveira, T. Henderson, What Online Gamers Really Think of the Internet, Proceedings of NetGames 2003, Redwood City (California), USA, (May 2003).

[6] Martin Mauve. How to Keep a Dead Man from Shooting. Proc. of the 7th International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services (IDMS) 2000, Enschede, The Netherlands, (October 2000), 199-204.

[7] Ubicom White Paper, OPScore or Online Playability Score: A Metric for Playability of Online Games with Network Impairments, (2005). http://www.ubicom.com/news/whitepapers/whitepapers.html

[8] M. Dick, O. Wellnitz, and L. Wolf Analysis of Factors Affecting Players' Performance and Perception in Multiplayer Games, 4th Workshop on Network & System Support for Games (NetGAmes'05), 1-7.

[9] G.Armitage, M.Claypool, P.Branch, "Networking and Online Games - Understanding and Engineering Multiplayer Internet Games," John Wiley & Sons, UK, (April 2006) (ISBN: 0470018577).

[10] Fedora operating system. http://www.fedoraproject.org

[11] NCTUns Network Simulator and Emulator, http://nsl10.csie.nctu.edu.tw/

[12] Wireshark network protocol analyzer http://www. wireshark. org.

[13] T. Lang, G.J. Armitage, P. Branch, H. Choo, "A Synthetic Traffic Model for Half-Life," Australian Telecommunications Networks & Applications Conference 2003, Melbourne (ATNAC 2003), Australia, (December 2003).

[14] T. Lang, P. Branch, and G. Armitage , A Synthetic Traffic Model for Quake 3, Proc. of Int. Conference on Advances in computer entertainment technology, (2004), 233-238.

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8721
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8721