Predicting the perceived quality of a First Person Shooter: the Quake IV G-model

A.F. Wattimena[1,2]

[1]TNO Information and Communication Technology
Brassersplein 2, P.O. Box 5050,
2600 GB Delft, The Netherlands

[2]Vrije Universiteit
Faculty of Sciences
Business Mathematics and Informatics
Amsterdam, The Netherlands

frank.wattimena@tno.nl

R.E. Kooij[1,3], J.M. van Vugt[1], O.K. Ahmed[1]

[3]Delft University of Technology
Dept. of Electrical Engineering, Mathematics and Computer Science
Mekelweg 4,
2628 CD Delft, The Netherlands

{robert.kooij|jeroen.vanvugt|kamal.ahmed}@tno.nl

ABSTRACT
This paper describes the development of an end-to-end quality measurement method that allows us to quantify the perceived quality of Interactive Gaming, with an emphasis on the so-called First Person Shooter (FPS) game Quake IV. We conducted a number of subjective experiments to quantify the impact of network parameters on the perceived quality of this recent FPS game. Making use of a multi-dimensional regression analysis we developed the Quake IV G-model which enables us to predict a gamer's Quake IV quality rating (expressed in a Mean Opinion Score) based on measured ping and jitter values. Our G-model shows a very high correlation ($R = 0.98$) with the subjective data.

Categories and Subject Descriptors
H.4.m [**Information Systems Applications**]: Miscellaneous

General Terms
Measurement, Performance, Experimentation

Keywords
On-line games, perceived QoS, Mean Opinion Score, First Person Shooter, regression analysis, Interactive Gaming

1. INTRODUCTION
Interactive Gaming has taken an enormous flight since the Internet enabled gamers to play over a globally accessible network. A number of factors contribute to this explosive growth of online gaming, of which the ongoing penetration of broadband connections and the expansion of online console platforms are only two [11]. Increasingly, gaming looks to be one of the mainstay applications that help to strengthen retention and loyalty of high-speed Internet consumers. By providing a superior gaming experience that produces reliable performance, broadband operators can likely retain relationships for years with the next generations of customers [14].

Netgames'06, October 30–31, 2006, Singapore.

The main goal of this paper is to provide insight into the quality experience of gamers. A more specific goal of our research is the development of an end-to-end quality measurement method that allows us to quantify the perceived quality of Interactive Gaming. In this paper we will only consider so-called First Person Shooter (FPS) games, in particular Quake IV. We will call the obtained model, which will be designed to capture the users' quality of experience (expressed as a Mean Opinion Score (MOS), see [8]), the Quake IV G-model, analogous to the VoIP speech quality prediction model (E-model) [7].

Many papers investigate the impact of delay, jitter and packet loss on gameplay for FPS games, see for instance [1], [3], [5], [6], [9], [10], [12], [13]. Most of these works have focused on network performance metrics that game players can tolerate. However, in [13] the authors have also conducted subjective experiments, and quantify the relation between network characteristics (i.e. delay and packet loss) and perceived QoS. Our approach is novel in the following respects; at first, we construct a model for the prediction of the experienced gameplay that depends on more than one network parameter simultaneously. Secondly, we introduce the approach of relating measured network impairment to perceived QoS, instead of using the a-priori impairment values of the configured network emulator. And last, we suggest an objective measurement method for determining the network impairment.

Obviously, so far our model is only based upon subjective measurements for Quake IV. We have chosen to use Quake IV [16], [17] for this user experiment, because it is one of the newest First Person Shooters on the market and has (by our knowledge) not yet been used for research purposes. In order to make our model more applicable we intend to do more subjective tests in the near future, for popular FPS games such as Quake III, Unreal Tournament, Counter Strike, Halo etc.

2. THE EXPERIMENT

Based on the findings in the literature we have decided to emulate the network parameters delay (from now on called ping), jitter and packet loss during the subjective experiments. The network emulator Netem [15] was used in order to vary ping, jitter, and loss. Netem is by default enabled in the kernel of the Linux distribution 2.6. The simulated values for these parameters are given in Table 1.

Table 1: Ping, jitter and loss values simulated

Ping [ms]	Jitter [ms]	Packet loss [%]
0,20,40,80,160,320	0,10,20,40,80,160	0,2,5,10,20,40

Netem was configured such that the additional ping time was evenly divided between the client server and server client directions. For the introduction of packet loss a uniform distribution was used to drop random packets. Jitter can be defined as the variance of the ping or delay. Inspired by the different causes of jitter, we have chosen to add variable delay to the network in two different ways. The first variant is implemented with the Pareto option enabled in a Netem script, which is configured in the following way:

```
tc qdisc add dev eth0 root netem delay 80ms
20ms distribution pareto

tc qdisc add dev eth1 root netem delay 80ms
20ms distribution pareto
```

The above commands induce a mean delay of 80 ms in each direction (ping 160 ms), combined with a variable delay (jitter) of 20 ms Pareto distributed in both directions (total of 40 ms jitter). The second variant is implemented with a self-made script which resembles the occurrence of route flapping. During intervals of 10 seconds the ping jumps between a base delay (determined by the current ping factor value) and a constant increased delay (determined by the jitter factor value). Here, a jitter of 80 ms means that the difference between the base delay and the increased delay is 80 ms. For more thoughts about the use of jitter in user-experience trials, we refer to [2].

During our two test sessions, six gamers simultaneously played Quake IV in a Free-For-All setting. All players had prior experience with FPS games. The FFA setting is one of the more popular FPS game modes in which players constantly engage in battles to collect the most frags (kills) given a certain time limit. This mode is particularly suited for our experiment, because it is very combat oriented in comparison to for example team games, which are heavier on strategic aspects. Each Netem scenario lasted five minutes. After these five minutes of gameplay, the gamers were asked to give their opinion of the gaming quality. They could select one of the following five opinion scores, motivated by the ITU-T ACR scale, see [8]: *5: Excellent gaming quality, 4: Good gaming quality, 3: Fair gaming quality, 2: Poor gaming quality, 1: Bad gaming quality*

Each player had to assess a total of 33 scenarios. The 33 scenarios included both network impairment caused by a single factor, as well as impairment consisting of combinations of ping, jitter, and loss. Because a visible ping value itself can influence the player's performance and joy, the gamers were not allowed to use the scoreboard function of Quake IV, which shows the current ping.

During the experiment the gaming server kept track of all kills and deaths of the participating players with the purpose of identifying a relation between these objective performance measures and the chosen impairment factors.

The gaming test setup consisted of 6 client PC's which were all connected to the gaming server through a Gigabit switch. The Netem network emulator was placed between the switch and the server. Both the gaming clients and the dedicated server were running the latest version of Quake IV (v1.1) plus a modification called X-Battle (v0.22) [18]. This modification is used in a large number of high profile Quake IV leagues and competitions. It enables gamers and administrators to customize Quake IV to their specific needs and addresses a lot of bugs which are still present in the original version of the game.

All of the scenarios were played on the map "*q4dm7: Over the edge*". It is a medium sized map which perfectly suits our 6 player Free for All setting. We have chosen to use the same map for the entire experiment, because otherwise our objective performance measures would not be comparable between scenarios.

Both the client PC's and the server computer were equipped with the latest generation of hardware (AMD Athlon 4000+ and 2 Gigabyte PC3200 RAM and in addition NVIDIA Geforce 7800 GTX graphical cards for the client PC's) in order to minimize the probability of performance drops. It is common knowledge that reduced hardware performance can negatively influence the subjective quality experience as well as the objective performance measures.

3. RESULTS

A first look at the experiment data shows us that the individual subjective opinions of all the participating players have a high correlation (between 0.74 and 0.95) with the MOS per scenario. This indicates that all of our test users were able to provide more or less consistent opinion scores.

3.1 Factor analysis

Ping

The influence of ping on MOS is depicted in Figure 1, which contains both MOS values and 95% confidence intervals.

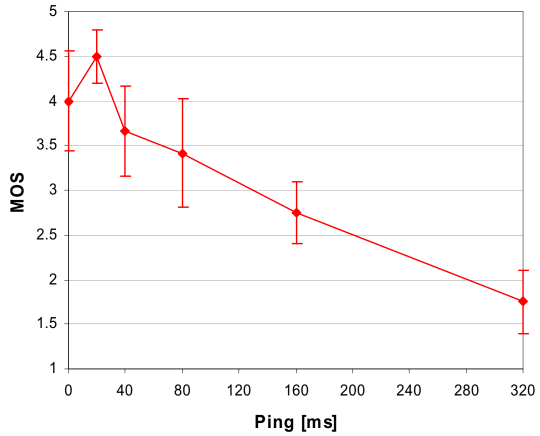

Figure 1: Influence of ping on MOS

The results of our gaming experiments clearly indicate that higher ping times negatively influence the subjective quality experience of gamers (MOS) as well as the objective performance measures (kills). These results were expected and are totally in line with the outcomes of earlier conducted experiments reported in literature.

Jitter

The influence of the impairment factor jitter is globally similar to that of ping, see also Figure 2. Both of our measures (MOS and kills) show a negative trend when plotted against increasing jitter values, for both implemented jitter variants. The bars in Figure 2 again represent 95% confidence intervals.

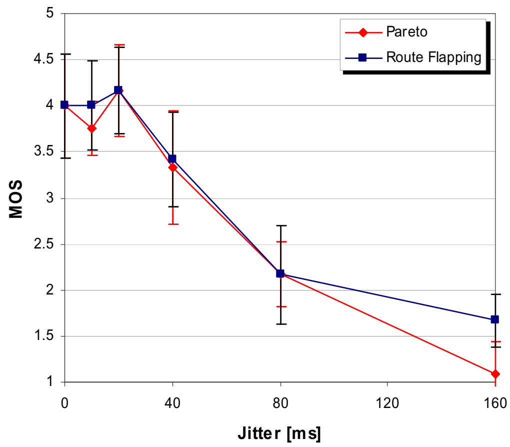

Figure 2: Influence of jitter on MOS

Packet loss

Our experiment data confirm the results of the user study on Quake III [13]. The authors of that paper conclude that packet loss hardly affects the perceived quality of Quake III. Our plots of MOS and total kills (not shown in this paper) against loss illustrate a constant behavior, in contrast to the graphs of ping and jitter which show declining trends. These outcomes show that the Quake IV engine is very robust when it comes to dealing with packet loss. The built-in loss handling algorithms are very effective in the prediction of lost gaming data. Unfortunately, the Quake IV network algorithms have not been documented yet, at the time of this writing.

3.2 Regression Analysis: measured network impairment approach

In the previous section we analyzed the impact of the three separate impairment factors on the multiplayer shooter Quake IV. The logical next step is to examine the combined effect of ping, jitter and loss on our subjective and objective measures. For this purpose we use a multi-dimensional regression analysis. The measure that we use to express how well a model explains subjective data is the linear correlation coefficient R, also known as the Pearson correlation. The correlation coefficient is related to the coefficient of determination R^2, which is a measure for how well the model explains the variation in the subjective data and can be 1 for maximum correlation. A correlation coefficient $R > 0.90$ is considered high while a correlation coefficient $R > 0.95$ is considered very high.

A robust and practical way of relating ping and jitter to opinion scores would measure the realized network conditions, instead of using the a-priori impairment as configured in Netem. In order to measure the realized network impairment we have used a variant of the Chirp method [4].

For each of the 33 Netem network degradation scenario's an objective measurement of the network quality between the clients and the server was determined. This was done in the following way. First 100 standard pings (ICMP Echo Request) were sent from client to server and vice versa. The average value over these 200 pings gives one ping_average value in ms for each scenario. Secondly, a series of 300 UDP packets was sent from client to server and vice versa. The UDP packet size including UDP and IP header was 100 bytes. Each UDP packet was sent 50 ms after the other. At the receiving side the transmission time of each UDP packet from sender to receiver was logged. The transmission time of the fastest UDP packet was shifted to 0 ms, after which all shifted transmission times were averaged. This was done for both directions and the average of these is called jitter_average.

All measurements of the 33 Netem network degradation scenarios were used as a training set to develop a gaming model for the prediction of the number of kills and the MOS, based on the objective ping and jitter measurements.

The correlation R between the number of kills or the MOS based on only one indicator, ping_average or jitter_average, is around 0.8 for each of the 4 combinations. When a model is based on 2 indicators, ping_average and jitter_average, the correlation coefficient is 0.93 for the number of kills and 0.98 for the MOS. The correlation between the model and the subjective data for MOS is depicted in Figure 3.

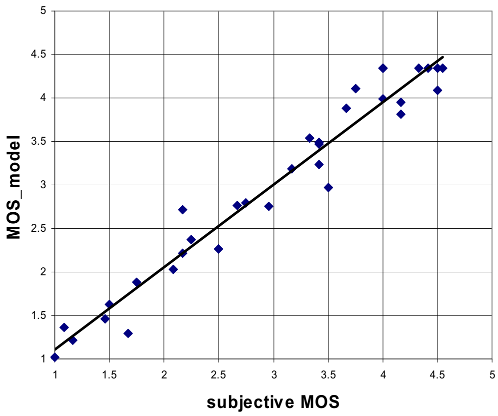

Figure 3: Comparison of subjective MOS with MOS based upon our regression model, R = 0.98

The network impairment X is given by the following equation

$$X = 0.104 * ping\_average + jitter\_average.$$

The mapping for the MOS values is given by:

$MOS\_model = -0.00000587\ X^3 + 0.00139\ X^2 - 0.114\ X + 4.37.$

Note that the above mapping defines our Quake IV G-model.

In order to validate our Quake IV G-model we have run an additional Quake IV experiment. During this short validation test 6 players competed in 8 FFA sessions of 5 minutes. Although the setup of this test was slightly different from our original experiment and consisted of other impairment scenarios, the G-model appears to be proficient in predicting the subjective quality experience. Based on the measured network impairment X (consisting of the weighted combination of ping and jitter measurements), the MOS estimation of the FPS G-model shows a 0.92 correlation with the MOS given by the test users.

4. CONCLUSIONS

In this paper we have examined the (simultaneous) impact of ping, jitter, and packet loss on the game-experience for the FPS Quake IV. Our experiment results demonstrate that ping and jitter have a significant negative effect on both the subjective (MOS) and objective gaming quality (kills), while packet loss goes unnoticed for values up to 40%. Especially the introduction of jitter in the network had a large negative effect on the perceived quality of the Quake IV players.

Making use of a multi-dimensional regression analysis we have come up with the Quake IV G-model which enables us to predict a gamer's Quake IV quality rating (expressed in MOS) based on measured ping and jitter values. The correlation between the subjective data and our model is very high ($R = 0.98$). A follow-up validation experiment showed that the model is very accurate in estimating these MOS values.

In the near future we plan to validate how well our G-model performs for other FPS games. In order to do so we will repeat our experiment approach for popular FPS games such as Quake III, Unreal Tournament, Counter Strike, Halo etc. We also plan to further analyze the occurrence and simulation of jitter as found on the Internet and in home networks, since the outcomes of our experiment stressed the importance of this impairment factor. At the moment we are also analyzing the traffic characteristics of Quake IV based upon logged packet traces. These results could be used to contribute to the development of a general FPS traffic model.

5. ACKNOWLEDGMENT

This work is partially supported by the IST-507295 Multi-Service Access Everywhere (MUSE) project, see www.ist-muse.org.

Finally, we would like to thank all of the participating players for their valuable contribution to this project.

6. REFERENCES

[1] G. Armitage, "Sensitivity of Quake3 Players To Network Latency," Poster session, SIGCOMM Internet Measurement Workshop, San Francisco, November 2001

[2] G. Armitage, L. Stewart, "Thoughts on Emulating Jitter for User Experience Trials," Proceedings of NetGames2004 Workshop, ACM SIGCOMM2004, Portland, Oregon, USA, August 2004

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

[4] J. G. Beerends and J. M. van Vugt, "A data chirp measurement method for assessing end-to- end quality of data links", White contribution COM 12-C11 to ITU-T Study Group 12, Delft, The Netherlands, 2005

[5] 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, U.S.A.

[6] 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

[7] ITU-T Recommendation G.107, "The E-model, a Computational Model for Use in Transmission Planning", 2005

[8] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality", 1996

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

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

[11] N. Shah, C. Haigh, "The video game industry, An Industry Analysis, from a VC perspective", Hanover (NH), U.S.A., 2005

[12] 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

[13] S. Zander, G. Armitage, "Empirically Measuring the QoS Sensitivity of Interactive Online Game Players," Proc. Australian Telecommunications Networks & Applications Conference 2004 (ATNAC 2004), Sydney, Australia, December 2004

[14] Camiant, http://www.camiant.com/perspectives/gaming.htm

[15] Netem, http://linux-net.osdl.org/index.php/Netem (as of May 2006)

[16] id Software, http://www.idsoftware.com/

[17] Ravensoft, http://www.ravensoft.com/

[18] Xbattle, http://www.xbattle.de/ (as of May 2006)