Objective and Subjective Evaluation of the Influence of Small Amounts of Delay and Jitter on a Recent First Person Shooter Game

Peter Quax Patrick Monsieurs
Wim Lamotte
Expertise Center for Digital Media
Limburgs Universitair Centrum
Universitaire Campus
B-3590 Diepenbeek, Belgium

Danny De Vleeschauwer
Natalie Degrande
Alcatel Bell NV
Network Strategy Group
Francis Wellesplein 1
B-2018 Antwerpen, Belgium

ABSTRACT

There have been several studies in the past years that investigate the impact of network delay on multi-user applications. Primary examples of these applications are real-time multiplayer games. These studies have shown that high network delays and jitter may indeed influence the player's perception of the quality of the game. However, the proposed test values, which are often high, are not always representative for a large percentile of on-line game players. We have therefore investigated the influence of delay and jitter with numbers that are more representative for typical access networks. This in effect allows us to simulate a setup with multiplayer game servers that are located at ISP level and players connected through that ISP's access network. To obtain further true-to-life results, we opted to carry out the test using a recent first person shooter (FPS) game, Unreal Tournament 2003. It can, after all, be expected that this new generation of games has built-in features to diminish the effect of small delay values, given the popularity of playing these games over the Internet. In this paper, we have investigated both subjective perceived quality and objective measurements and will show that both are indeed influenced by even these small delay and jitter values.

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous

General Terms

Measurement, Human factors

Keywords

On-line games, Network influences, Analysis

SIGCOMM'04 Workshops, Aug.30+Sept.3, 2004, Portland, Oregon, USA.
Copyright 2004 ACM 1-58113-942-X/04/0008 ...$5.00.

1. INTRODUCTION

Multi-user applications that are deployed on the Internet are inevitably influenced by the delay that is present on the global network. The influence of these network anomalies is especially apparent when considering highly interactive real-time applications, with networked computer games as prime examples. There have been a number of studies that were carried out to measure this influence in an objective manner, with many of these having focussed on multiplayer games as representative applications [6, 7, 10, 11].

Ardent game players often cite network delay (referred to as "ping time" or "ping") as the main cause for degradation in their performance and/or scores. This factor is often seen as culprit because of built-in ping features in modern multiplayer games, with that ping number in itself possibly influencing the players performance. It has been shown that high amounts of delay may indeed influence player's performance to a certain degree. Results however are rarely consistent, especially when considering different types of games [10, 11]. A secondary problem with some of these results is that they are often based on non-realistic numbers, i.e. they consist of test cases with delays that run up to the order of seconds. In many cases however, Internet gamers are connected to a dedicated game server that is placed directly in their ISP's data center, which leads to a significant reduction in delay. This network delay is in most cases composed of a number of contributing factors, such as access network delay [8], backbone delay and server-level delay. In the setup with dedicated servers at ISP level, the backbone delay is ruled out, leaving only the access network delay and (minimal) server-processing delay. Most of the so-called 'hard-core' gamers often simply choose not to connect to game servers that show a ping higher than a few 100 ms [1, 2]. This remark has an influence on a number of results that are shown in previous work on this subject and is the reason why, in this work, we chose to concentrate on low delay values.

We had three major questions in mind when starting work on this paper. Can a player effectively determine whether his/her connection is influenced by lag without consulting diagnostic tools, i.e., based on his/her perceived game quality and/or performance ? Is there a bound below which the influence of delay and jitter on the players performance is minimal or even non-detectable ? Will small amounts of delay and jitter influence the score on modern FPS games that were developed for use over the Internet? The latter question is directly based on the observations described in the previous paragraph. We were however also interested to know to what degree a user is able to determine the amount of impairment he/she is facing, e.g. in comparison to the other players on a server. It is interesting to know whether players will rate the network as sub-optimal if they are effectively impaired by means of lag. Another proposed explanation for this degradation in perceived quality could, for example, be the influencing by other player's comments.

2. RELATED WORK

A number of papers describe the effect of delay and jitter on multiplayer games. In [11], these effects are described on a real-time strategy game, Warcraft III, while [10] presents similar work based on a racing game. In [11], the authors conclude that the influence of delay on user performance is minimal, even when facing large delays. This is attributed to the nature of these strategy games. The authors of [10] conclude that for racing games, a delay of up to 50ms is not regarded as critical, while all delay values over 100ms do show significant impact on the realism of the game.

There have been relatively few efforts to map the perceived subjective quality of multiplayer games under varying network conditions. Most of the existing work on this subject is focussed on relating the player-server connection time to the measured network delay. This way, it is proposed that it is possible to derive whether a specific amount of lag is acceptable to a game player [1, 2]. When a player only connects for a short amount of time with a given server (typically less than 10 seconds), this is expected to indicate that network conditions are perceived as unfavorable. In this setup however, it is difficult to get accurate measurements. It is also hard to derive delay and jitter bounds as there is no means to get consistent feedback from the players on their reasons for leaving the game server or their perceived quality of experience.

There is also the issue of QoS techniques that can be employed specifically for game use. In [7], the authors question whether there is need for QoS support for networked games, based on observations of when players join and/or leave specific servers, due to changing ping times. They conclude that players do not seem to react instantaneously to adverse network conditions, and seem to tolerate surprisingly high delay values when playing on popular servers. Moreover, compensation techniques are often employed in many of today's networked multiplayer games. Some of these are described in [3].

3. MEASUREMENT SETUP

As described in the introduction, the main objective of the experiment is to determine the influence of small amounts of delay and jitter in a network on the quality of game play of a highly interactive game. It can be expected that recent games, which are developed with global deployment on the Internet in mind, may have better performance than older games when faced with adverse network conditions. Because of this, Unreal Tournament 2003 [5] was selected for our experiment. Unofficial sources report that UT2003 has built-in jitter compensation, although no detailed technical information is available.

Our original experiment was set up for 14 players to participate, each using an identical PC. A dedicated Unreal Tournament server was installed on an additional machine. To simulate delay and jitter on the network connection, a router was placed between the switch and the dedicated server machine. On this router, the software NISTNet [9] was used to introduce delay and jitter on specific network streams. Settings for NISTNet were left at their default values (standard distribution and correlation factors for jitter.) The dedicated server was configured using the default settings, the most important being the server tick rate, which was set at 25, and the InstaGib mutator. The server tick rate represents the rate at which the dedicated server recalculates the state of the entire world (i.e. player positions, object properties,...) and distributes this state to the connected clients. In case of a tick rate of 25, the server updates the internal state 25 times per second, which is consistent with a standard frame rate for video playback. This way, a delay below 40ms should have little influence. Using the InstaGib modification of the original UT2003 game, a single hit suffices to get killed. This modification also facilitated recording each players activity, as only kills (or frags) can be logged server-side. The type of game was so-called 'deathmatch', which means that players engage each other in one-to-one combat, instead of playing in teams.

To measure the effect of delay and jitter, it is necessary to simulate various different network conditions. In this case, these conditions range from a negligible round-trip delay and jitter of 20 ms +/- 5 ms to a maximum of 100 ms +/- 95 ms. The minimal round-trip delay has been selected from results described in [8] (typical access network delay). A total of 20 different settings of delay and jitter were selected, each called a 'scenario'. NISTNet was configured to split the delay and jitter equally over upstream and downstream traffic towards a particular (impaired) user. Every scenario was tested in a session that lasted for 7 minutes. During every scenario, about half of the participants experienced the introduced lag and jitter. In the following text, these players will be designated as being 'impaired'. The other players were unaffected, i.e. their settings were set to the minimum amounts (20ms +/-5 ms). We opted to subject the non-impaired group of players to these minimal amounts in order to simulate the minimal delay that is always present on a typical access network. The set of affected players changed after every configuration. Originally, 14 players were scheduled to participate in the test. Unfortunately, 2 players were unable to attend, and as a result, PCs 2 and 14 were not used.

In order to obtain answers to the initial questions that are stated in Section 1, the influences of delay and jitter were measured using two approaches. First of all, the number of times a player effectively killed another player ('kills'), and the times he was killed himself ('killed') were logged at the server. This provides us with an objective measurement of the quality of the game play, and enables comparison between different sessions. Secondly, after every session, the participants had to fill out a short questionnaire regarding the network quality they experienced. The following questions had to be answered after every session: Rate the quality of the network (0(=worst)-1-2-...-8-9-10(=best)). How much did the quality of the network influence your gameplay ? (0(=not at all)0-1-2-...-8-9-10(=very much)). Do you think the quality of the network influenced your score? (Yes/No). Remarks on this scenario. For questions 1 and 2,

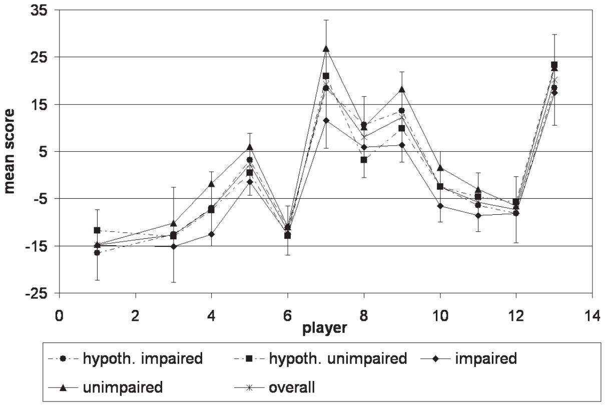

Figure 1: Mean score for all players : over all 20 scenarios, only over the impaired, only over the unimpaired scenarios, over the hypothetical impaired and over the hypothetical unimpaired scenarios, respectively.

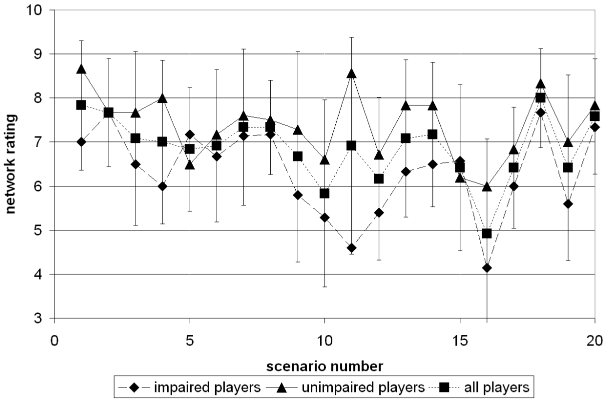

Figure 2: Mean rating for all unimpaired, all impaired and all 12 players, respectively, as a function of scenario number.

the participants were not allowed to select the values 0 and 10 (as described in ITU-T Recommendation P.800.1).

4. MEASUREMENT ANALYSIS

4.1 Objective Observations

The objective measurements consist of statistics about the number of times each player has been killed and/or has killed other players during each of the 20 scenarios. In order to analyze these objective results, the difference between the number of kills and the number of times being killed for each player p and in each scenario c, is used as a score : $S(p,c) = \#\text{kills}(p,c) - \#\text{killed}(p,c) \quad \forall p \in P, \forall c \in C$ with P the set of 12 players and C the set of 20 scenarios. A positive (negative) score indicates that a player has killed more (less) other players than that he was killed himself. As such, the score as defined above is an objective measure of the performance of each player in each scenario.

Figure 1 shows the mean score for each player over all scenarios where he is not affected by impairment (triangles connected with solid lines) and the mean score over all scenarios where he is subjected to impairment (diamonds connected with solid lines). The mean score over all 20 scenarios is also shown (stars connected with solid lines). The vertical error bars indicate the standard deviation of the latter. This figure shows the trend that the mean score of the impaired games is consistently lower than the mean score of the unimpaired games. This is an indication that network impairment, as defined in section 3, does have a negative influence on the affected players' performance.

The question can be asked whether players that are not directly subjected to impairment, are hampered when other players in the game session are subjected to high impairment conditions. In other words : can, at high impairment levels, all players be considered as being hampered, no matter if they are directly subjected to the impairment or not. In order to get a notion of that, all 20 scenarios are divided in 2 categories : hypothetical impaired and hypothetical unimpaired scenarios. Where the hypothetical impaired scenarios are the 13 scenarios where the delay and jitter values are at least 60 ms and 50 ms, respectively. The hypothetical unimpaired scenarios are the other 7 scenarios. Now for each player, his mean score over the 13 hypothetical impaired scenarios, irrespective of really being exposed to the impairment or not, is calculated, as well as his mean score over the 7 hypothetical unimpaired scenarios. If in general these 2 mean scores do not show significant differences compared to the differences between real impaired and real unimpaired scenarios, one can not state that all players in hypothetical impaired scenarios (irrespective of directly being exposed to the impairment) can be considered as being impaired. In Figure 1, the circles and squares, connected with dashed lines represent the mean scores for all players for the hypothetical impaired and unimpaired scenarios, respectively. This figure shows that in general the difference between the hypothetical impaired and unimpaired scenarios is inferior to the difference between the real impaired and unimpaired scenarios. As such, players that are not directly subjected to impairment, are not hampered when other players in the game session are subjected to high impairment conditions.

4.2 User Perspective

4.2.1 Overall Rating

The answers given to the first question 'Rate the quality of the network' are used to see whether in general players experience impairment as degrading network quality. A distinction is made between players that are exposed to impairment and players that are not subjected to impairment. For each scenario, the mean rating of the network quality is calculated for the non-impaired and for the impaired players, respectively. This is shown in Figure 2. The mean rating of all players is also given, as well as the standard deviation. The figure indicates that overall, players that are subjected to increased delay and jitter rate the network quality poorer than the other players. In other words : increased delay and jitter are experienced as degrading network quality.

Besides this general tendency, the above figure provides more interesting information when looking at scenarios 1, 5, 11 and 15. For that purpose, the players are divided in two categories : the 'optimists' and the 'complainers'. The selec-

		Players													
		1	3	4	5	6	7	8	9	10	11	12	13	Delay	Jitter
Scenario	1	X	X	X	X	X							X	40	20
	5						X	X	X	X	X	X		100	80
	11	X	X	X	X	X								100	65
	15						X	X	X	X	X	X	X	100	95
	Optimism	18	6	6	3	1	8	15	14	17	19	15	17		
			Complainers												

Table 1: Summary of impairment settings for scenarios 1, 5, 11 and 15; the degree of optimism is also given.

tion of both is done based on their network rating behavior and is explained in the following. For all 20 scenarios, the mean rate given by all 12 players is determined, irrespective of being impaired or not. Next, for all players it is checked in each scenario whether they rate this scenario better or worse than this mean rate. The more scenarios a player rates better, the higher his degree of 'optimism'. Finally, the mean degree of 'optimism' over all 12 players is determined. Players with a degree of optimism, lower than this overall mean degree, are labeled 'complainers', the players with a degree higher than this overall mean degree, are labeled 'optimists'. This method classifies players 3 to 7 as 'complainers' and all other 7 players as 'optimists'. Note that applying the above methodology to only the impaired scenarios, leads to the same classification of players. Table 1 summarizes for scenarios 1, 5, 11 and 15 which players are subjected to impairment, indicated with an X on a colored background, as well as the corresponding impairment settings. Their degree of optimism is also given.

As can be seen in Figure 2, for scenarios 5 and 15, respectively, the players that are exposed to impairment rate the network quality slightly better than the players that are not exposed to impairment. Examining both scenarios more in detail explains what happens here. On the one hand both scenarios are quasi-identical : severe impairment conditions and identical sets of impaired and unimpaired players (except for player 13). On the other hand the top 4 of the complainers are concentrated in the group of unimpaired players for these two scenarios (see Table 1). This forces the mean rating of the unimpaired players for these 2 scenarios downwards compared to the mean rating of the impaired players.

Considering Table 1 one sees that, regarding the configuration of the complainers' impairments, scenarios 1 and 11 are exactly the opposite from scenarios 5 and 15 : the top 4 of the complainers now belongs to the group of impaired players. Following the above reasoning, the difference between the mean rating of the unimpaired and the impaired players, respectively, should blow up. For scenario 11 this is obvious in Figure 2. For scenario 1, this effect is less pronounced (but still present). Taking into account the very mild impairment conditions in scenario 1 and the severe impairment conditions in scenario 11, one can conclude from this that players are able to distinguish to some level between different degrees of impairment.

4.2.2 Worst and best rated scenarios

Since the rating scales of different players might be different and there is no real means to hallmark them, only the best rated and the worst rated settings of all players

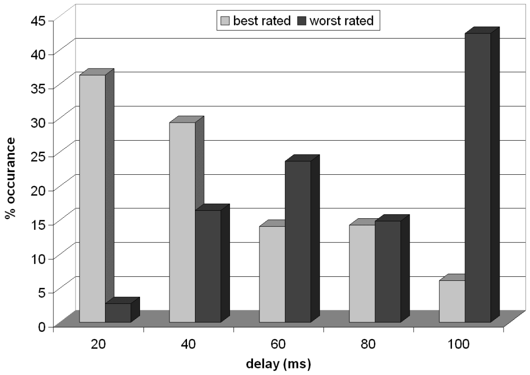

Figure 3: Occurrence of delay values and their ratings.

will be used in the following analysis. The main idea is : the more a certain impairment scenario has been given the best (worst) rating, the lesser (more) a negative impact on the network quality is experienced. Note that many players have given their best/worst rating to more than 1 scenario, as such 1 player can cause different scenarios to be present in the analysis. First we will only consider network delay, in a second step the jitter will be examined.

For practical reasons, it was impossible to cover all possible scenarios in the experiment. As such, not all players have been exposed the same number of times to all different delay values. This implies that, even if the players had to choose randomly the ratings of the different scenarios, not all scenarios would have equal chance to get e.g. a best rating. This discrepancy has been taken into account. The final corrected figures, normalized to 100, for the number of times (a scenario with) a certain delay values has been rated as best or worst are shown in Figure 3.

The trend that can be observed from this figure is that it is less likely that a higher imposed delay will receive a best quality rating, and on the other hand that it is more likely that the network will receive a worst quality rating.

An analogue analysis for the imposed jitter values did not yield useful information. Moreover, the statistical relevancy for this analysis was very low since almost no jitter values were present in more than two best/worst rated scenarios.

4.2.3 Indication for delay and jitter bound

Based on the cases where the players give the 'best' or 'worst' rating to the network, we concluded that the players are able to predict whether or not they are hampered. Next we try to estimate from which level of delay and jitter they are able to predict that they are hampered. In contrast to the previous paragraph we consider all experiments in this analysis (not just the ones where the players give either their worst or best score) and consider all three questions on the questionnaire.

We want to test the hypothesis, if the players are able to predict that they feel congestion for a (delay,jitter) pair (d,j) imposed in scenario c. If they feel hampered for this pair (d,j), then the players that experience this kind of impairment, should give a low value to question 1, a high value to question 2 and answer "yes" to question 3.

In order to translate the subjective rating given by the

		Question 1							Question 2							Question 3						
Delay	20 ms	0							0							0						
	40 ms	0	0	0					1	0	0					1	0	0				
	60 ms	1	0	0	0				1	0	0	0				1	0	0	0			
	80 ms	1	0	1	1	1			1	1	1	1	0			1	0	1	1	1		
	100 ms	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
		5	20	35	50	65	80	95	5	20	35	50	65	80	95	5	20	35	50	65	80	95
		Jitter							Jitter							Jitter						

Table 2: Results of the hypothesis tests.

players to question 1 in binary values (0 meaning that the player indicates that he does not feel hampered, 1 meaning that he esteems he does), we translate the ratings given by a player into 0 if the rating given is larger than the average value given by this particular player and 1 otherwise. For question 2, we set the value to 0 if the rating given is smaller than the average given by the player, and to 1 otherwise. So based on the three questions we have a table (one for each question) with an indication whether or not player p feels hampered in scenario c : $Q_i(p, c)$ indicates (is 1) if player p feels hampered in scenario c, according to his answer to question i.

To perform the hypothesis test, we predict whether or not player p is impaired in scenario c. Player p is impaired in scenario c, if his network path was affected and if for the delay and jitter value imposed in scenario c, the player is able to feel impairment, the latter of which is exactly the hypothesis we want to test. So, we define $R(p, c; H(c)) = \text{AND}(T(p, c), H(c))$, where $T(p, c)$ is 1 if the player was subjected to increased delay and/or jitter and 0 otherwise. $H(c)$ is the hypothesis that the (delay,jitter) pair (d,j) used in scenario c is felt by the user as hampering his performance and AND(.,.) is the Boolean and-function.

If $Q_i(p, c) = R(p, c; 0)$ this supports the fact that the players do not experience impairment in scenario c (more precisely, for the (delay,jitter) pair (d,j) used in scenario c) and similarly $Q_i(p, c) = R(p, c; 1)$ endorses the hypothesis that the players do feel impairment in scenario c. So, the hypothesis test consists of determining for each scenario c, which value of $H(c)$ maximizes $\sum_{p \in P} \text{EQ}(Q_i(p, c), R(p, c; H(c)))$ where EQ(.,.) is the Boolean equality-function. Table 2 gives the results of this hypothesis test for the three questions in the questionnaire.

This table indicates that a round-trip delay below 60ms is not felt as an impairment and that the jitter does not play a prominent role in whether or not the player feels hampered for the particular FPS game considered in this paper. These findings are supported by Figure 3 where a large gap from 40 to 60 ms delay is present for best rated delay and consistent with the racing game conclusions in [10].

5. CONCLUSIONS

In this paper, the objective and subjective influence of delay and jitter as present in typical access networks on the quality of game play of a highly interactive game is investigated. This is done by means of an experimental setup consisting of a LAN in which 12 players were present. They fought each other in the FPS game Unreal Tournament 2003. A router in the network simulated delay and jitter on the network connection for part of the players. The data recorded during the experiment is analyzed afterwards.

The main conclusions are that network impairment, as defined in this paper, does have a negative influence on the affected players' perceived game quality and performance. Players that are not directly subjected to impairment however, are not hampered by possible impairment present for other players in the gaming session. From a user perspective, it was seen that the players' perception of the quality of the game depends on the size of the delay the network introduces. Finally, there are indications that from 60 ms round-trip delay on, the player experiences the impairment as disturbing.

6. ACKNOWLEDGEMENTS

Part of this research is funded by IWT project number 020659 and the Flemish Government.

7. REFERENCES

[1] G. Armitage. Sensitivity of quake3 players to network latency. In *ACM SIGCOMM Internet Measurement Workshop 2001*, Berkeley, CA, USA, Nov. 2001.

[2] G. Armitage. An experimental estimation of latency sensitivity in multiplayer quake 3. In *11th IEEE Int. Conf. on Networks (ICON 2003)*, Sydney, Australia, 2003.

[3] Y. W. Bernier. Latency compensation methods in client/server in-game protocol design and optimization. In *Proc. of Game Developers Conference'01*, 2001.

[4] M. Borella. Source models of network game traffic. In *Computer Communications, vol. 23, no. 4*, pages 403–410, 2000.

[5] Epic Games, Atari. *Unreal Tournament 2003*.

[6] T. Henderson. Latency and user behaviour on a multiplayer game server. In *Proc. of the Third Int. COST264 Workshop on Networked Group Communication*, pages 1–13. Springer-Verlag, 2001.

[7] T. Henderson and S. Bhatti. Networked games: a qos-sensitive application for qos-insensitive users? In *Proc. of the ACM SIGCOMM workshop on Revisiting IP QoS*, pages 141–147. ACM Press, 2003.

[8] T. Jehaes et al. Access network delay in networked games. In *Proc. of the 2nd workshop on Network and system support for games*, pages 63–71, 2003.

[9] National Institute of Standards and Technology. *NISTNet.* World Wide Web, `http://snad.ncsl.nist.gov/itg/nistnet`.

[10] L. Pantel and L. C. Wolf. On the impact of delay on real-time multiplayer games. In *Proc. of the 12th int. workshop on network and operating systems support for digital audio and video*, pages 23–29. ACM Press, 2002.

[11] N. Sheldon et al. The effect of latency on user performance in warcraft 3. In *Proc. of the 2nd workshop on Network and system support for games*, pages 3–14. ACM Press, 2003.