Gaming in the Clouds: QoE and the Users' Perspective

Michael Jarschel, Daniel Schlosser, Sven Scheuring, Tobias Hoßfeld

University of Würzburg, Institute of Computer Science, Germany
Email: {michael.jarschel,schlosser,sven.scheuring,hossfeld}@informatik.uni-wuerzburg.de

Abstract

Cloud Gaming is a new kind of service, which combines the successful concepts of Cloud Computing and Online Gaming. It provides the entire game experience to the users remotely from a data center. The player is no longer dependent on a specific type or quality of gaming hardware, but is able to use common devices. The end device only needs a broadband internet connection and the ability to display High Definition (HD) video. While this may reduce hardware costs for users and increase the revenue for developers by leaving out the retail chain, it also raises new challenges for service quality in terms of bandwidth and latency for the underlying network. In this paper we present the results of a subjective user study we conducted into the user-perceived quality of experience (QoE) in Cloud Gaming. We design a measurement environment, that emulates this new type of service, define tests for users to assess the QoE, derive Key Influence Factors (KFI) and influences of content and perception from our results.

Keywords: Cloud Gaming, QoE, Survey

1. Introduction

With the introduction of fast and reliable core networks and wide-spread availability of broadband internet access, a trend towards moving more and more services away from the end devices to remote data centers has established itself. This is widely referred to as Cloud Computing. While initially only services with few requirements towards the delivery network like email were established in the cloud, these days a wide range of applications and services is available to users remotely. This results in greatly increased requirements on network quality of service (QoS) as users expect higher standards to be met. In a recent study cf. Figure 1 Cisco projects that, while the

traffic volume for all types of services will increase over the next few years, multimedia and interactive services like internet video and online gaming will experience a particularly high growth rate.
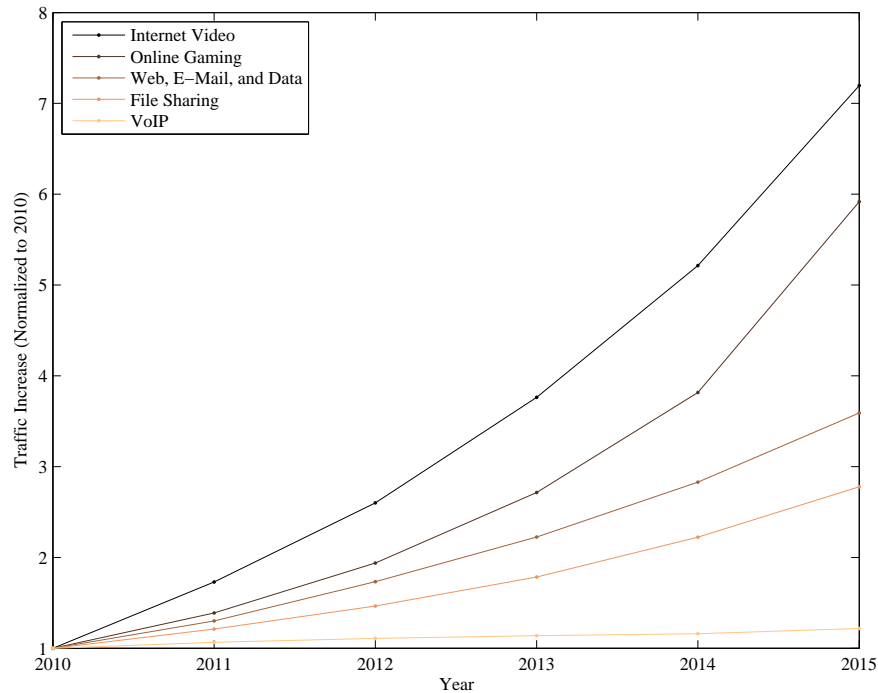

Figure 1: Cisco Traffic Projection

Recently, a new type of cloud service has been introduced, which combines internet video and online gaming and may have the most stringent demands on network QoS to date: cloud gaming. This new paradigm has been subject of a case study by Ojala et al. [1] underlining its potential from a business point of view. Yet, business is not the only field from which cloud gaming has received attention. As early as 2009, Ross [2] identified Gaming as the "Killer-App" for Cloud Computing and Chang [3] even believes that "gaming will save us all".

The service essentially moves the processing power required to render a game away from the user into a data center and streams the entire game experience to the user as a high definition video. Traditionally only multiplayer games use the network. Several clients are connected to one server, which

controls the game environment, receives input commands and sends out status updates. The amount of data exchanged is usually quite small. However, in cloud gaming the entire user experience has to be delivered through the network. This is where Cloud Gaming is significantly different from conventional online gaming in terms of network QoE. While in conventional online gaming the user experience is generated at the client and therefore the network does not have any influence on the presentation, it may greatly affect the quality in Cloud Gaming. In addition cloud gaming enables remote play of games, which were never designed to be played over a network, which may also change the users' perception. Thus, while there have been many studies on the QoE of Online Gaming, e.g. [4], their findings can not be applied here.

Cloud Gaming enables new business models for game- as well as cloud providers. In addition this solution has advantages for users as well as game developers. On the one hand users no longer need to purchase powerful hardware to run new games and can play on virtually any device, which is able to display HD video. On the other hand developers no longer have to fear software piracy as the software never leaves the cloud and also can reduce development costs by focussing on one specific platform. This allows developers to spend more time on improving the quality of the product instead of worrying about compatibility. Even small non-HD end devices can be supported e.g. by reducing the resolution. A smaller resolution requires less bandwidth and the video can be decoded at the end device with less energy consumption enabling the use of hand-held mobile devices as clients.

However, from a network point of view there are several challenges to overcome to operate such a service in the quality expected by the users. Unlike conventional video streaming or web applications Cloud Gaming does not require either a relatively high constant downlink bandwidth or low latency, but both. The only provider currently offering such a service to mainstream users is OnLive (cf. [5]) in the US. The company operates several data centers across the US in order to minimize the propagation delay imposed by the physical constraints (cf. [6]). It is the goal of this study to investigate these parameters based on actual user perceptions to identify KFI for QoE in cloud gaming. To achieve this goal, subjective user surveys are required. Therefore, we have designed a local testbed at the University of Würzburg that emulates a cloud gaming service. This testbed is used to provide a test person with a game experience similar to that of a cloud service. We have developed a series of tests to gauge a user's reactions to varying settings of propagation delay and packet loss. Based on this setup we performed a sur-

vey with test persons and derive general conclusions on the impact of certain QoS parameters on QoE and identify influences of content and perception from the results.

This article represents an extension of our work in [7] published via IEEE Xplore at FINGNet 2011 in Seoul, Korea. In addition to the results published there, we discuss new related work in the field, take rater reliability into account, and mention recent developments.

The remainder of the paper is structured as follows. First we discuss related work in Section 2. We then examine the demographics and reliability of the participants this study relies on in Section 4. In Section 3 we derive the definition of the test from related work and describe the testbed used to emulate cloud gaming. Section 5 discusses the influence of the identified QoS parameters on the QoE. Finally, in Section 6 we give a conclusion and provide an outlook on future work.

2. Related Work

Even though cloud gaming is relatively new, there is a number of papers on the topic. Nave et al. [8] describe an architecture for cloud gaming as developed in the European FP6 Integrated Project Games@Large. Pigora et al. [9] discuss the benefits of applying a cloud gaming approach to training and education by introducing their solution called 'Nexus Web' as an example. They describe implementation challenges and give a rough estimate of the QoS. QoE, however, is not mentioned. We overcame the implementation challenges in this paper by procuring special purpose hardware. Chan [10] simulates the impact of a wireless environment on cloud gaming using Opnet and draws conclusions regarding the QoS and its scalability. Additionally, Chan found that a moving user will experience a significant drop in the QoS. However, he also does not discuss user-based QoE. Chang et al. [11] propose a methodology for quantifying the performance of several VDI solutions in a gaming scenario. To this end they use a classic 2d game and capture the graphics output at server and client for a comparison of quality. However, they do not incorporate actual user feedback and current 3d video games.

Szigeti et al. [12] recommended guidelines to setting QoS parameters for interactive video or video conferencing traffic. As interactive video is related to cloud gaming, we have taken their recommendations into account when designing our tests. However, cloud gaming has different QoE characteristics from interactive video and therefore these values do not apply exactly.

4

Three classes of games with different behavior towards QoE are identified by Claypool et al. [13]. These are "Omnipresent" (e.g. real-time strategy games), "Third-Person Avatar" (e.g. role-play games), and "First Person Avatar" (e.g. First Person Shooters). We adopt these classes and chose one game from each for the purpose of our tests in order to account for the effects of varying content. Additionally, Claypool et al. also give a latency range in which each type of game performs well. We have based the choice of latency values for our tests based on these results.

3. Survey Parameters and Design

In Subsection 3.1 we select the range of the QoS parameters loss and delay whose influence on QoE is tested in our user tests. In Subsection 4.1 we discuss the attributes of the test group our survey is based on. Our testbed is explained in subsection 3.2. Finally, we characterize the actual survey process in Subsection 3.3.

3.1. QoS Parameters of the Survey

An IP network connection maybe influenced by numerous factors: delay, jitter, packet loss, packet re-ordering or packet duplication to mention only a few. However, to Cloud Gaming in its current form only two parameters are relevant for the QoE - packet delay and loss. Delay affects the time a user's action is executed and the results are perceived. In Cloud Gaming this would be the time from the pressing of a controller button to the intended action. All other influence factors result in the application not being able to display a video frame or execute an input command in time. These effects are handled by the network encoding or treated by the application identical to packet loss. To meet the real-time constraint the software cannot wait for one packet to be delivered for an arbitrary amount of time or in an arbitrary order. As a consequence the program will have no choice, but to drop the data resulting in loss. From the user's point of view, lost or late packets lead to the same quality degradation independent from the underlying cause e.g. network congestions or jitter. Therefore, all of these effects can be investigated by just examining the influence of packet loss. Pantel et al. propose in [14] that a delay greater than 100 ms should be avoided based on study of two racing games. We take this value as a starting point for designing our own initial subjective tests. The next QoS parameter we consider in our tests is loss. Since there is no reference value for loss in Cloud Gaming, we take a

5

look at [15] by Szigeti et al., which gives guidelines for the related field of video conferencing. It states that loss should be no more than 1 percent, one-way latency should be no more than 150 ms, and jitter should be no more than 30 ms. In [16] by Henderson et al. the authors describe the effect that degraged QoS can dissuade players from joining a networked game, but those already connected to a server are more tolerant towards bad QoS. We consider this effect in relation to Cloud Gaming, but it affects only the usage of the service, i.e. users might quit the service or not subscribe to it. In this paper we focus on influences occuring while using the service.

The three classes of games we investigate as determined by Claypool et al. [13] again are "Omnipresent" (e.g. real-time strategy games), "Third-Person Avatar" (e.g. role-play games), and "First Person Avatar" (e.g. First Person Shooters).

Table 1 gives an overview of the specific scenarios we define. The table gives the scenario id as well as the specific settings for delay and loss. Finally, it also gives the direction - client to server or server to client - to which the parameters are applied. The first scenario (B) we introduce is the baseline, which is essentially a setting in which all parameters are set to zero. We do so in order to check for the placebo effect, i.e. some of the test subjects could imagine a distortion where there is actually none, simply because they find themselves in a test situation. Additionally, we define three delay-only scenarios (D1-3). These are our subjective perception threshold for delay at 160 ms round-trip time (RTT), a noticable disturbance of play at 400 ms RTT and 600 ms RTT where players should no longer be able to play. Here the delay is identical on up- and downlink. This results in the input commands being received late by the game service and the feedback video being delayed also. Having considered delay, we then introduce two scenarios with symmetric packet loss of 0.3 and 1 percent per link (scenarios L1,L2) being the only source of disturbance. The effect of packet loss on the downlink are a notable fragmentation of the video as well as lost keystrokes on the uplink. After looking into delay and packet loss individually we are interested in the question, which parameter is dominant and has a larger influence on the QoE. To determine this, we create two mixed scenarios combining delay and packet loss (scenarios M1,M2). Finally, we introduce two scenarios with asymmetric settings to investigate whether applying the same parameters on either the up or the downlink changes the outcome of the QoE perception (scenarios A1,A2).

Table 1: Test Scenarios (taken from [7])

Scenario ID	Delay	Packet Loss	Direction
B	0 ms	0.0%	both
D1	80 ms	0.0%	both
D2	200 ms	0.0%	both
D3	300 ms	0.0%	both
L1	0 ms	0.3%	both
L2	0 ms	1.0%	both
M1	40 ms	1.5%	both
M2	180 ms	0.3%	both
A1	120 ms	1.0%	client to server
A2	120 ms	1.0%	server to client

3.2. Emulation of Cloud Gaming

Figure 2 depicts our testbed setup from a logical point of view. The idea of this setup is to replicate the basic infrastructure of OnLive and its competitors intend to use to deliver the game experience to their customers. Hence, three individual components have to be reproduced. The hardware shown on the right hand side of the Figure replaces the data centers. To replace the servers which would usually render the game we use a conventional PlayStation 3 gaming console. This device is optimized for gaming and the games running on it are optimized for its hardware. Therefore, the risk of false results caused by erratic behavior of the rendering hardware is minimal. The images created by the Playstation are then streamed to the client via a special purpose hardware, called Spawn Box. The Spawn HD-720 is capable of streaming the output produced by many modern consoles over an IP network to its client software (Spawn Player). This software is a modified version of the well-known VLC media player. It displays the video and transmits the client input to the Spawn box, which in turn relays it to the game console. The Spawn Player is configured for smooth replay at the best possible quality i.e. a video resolution at three quarters of 720p and a video codec bitrate at 3 MBit/s. The box uses HaiVisions MAKO-HD hardware, which was originally designed for the purpose of high definition video conferencing and hence uses progressive H.264 video encoding. Both video and user input are transmitted through the network via a RTP/UDP connection.

7

In the center of the Figure the component emulating an IP WAN, e.g. the Internet, is represented by a cloud. In fact this is a computer running the Linux-based network emulator NetEM on Debian Lenny. The NetEM software is capable of producing a variety of effects a wide area network could have on a packet stream. However, we only use it to introduce fixed delay as well as random loss as explained in 3.1. A client is represented by an Intel Pentium IV personal computer in our experiment running the Spawn Player software on Windows XP as seen on the left of the Figure.

For the purposes of conducting the survey, we introduce a fourth component. A web-server that controls the simulation by remotely configuring the WAN-simulator and displaying the frontend of the QoE poll as well as storing its results.
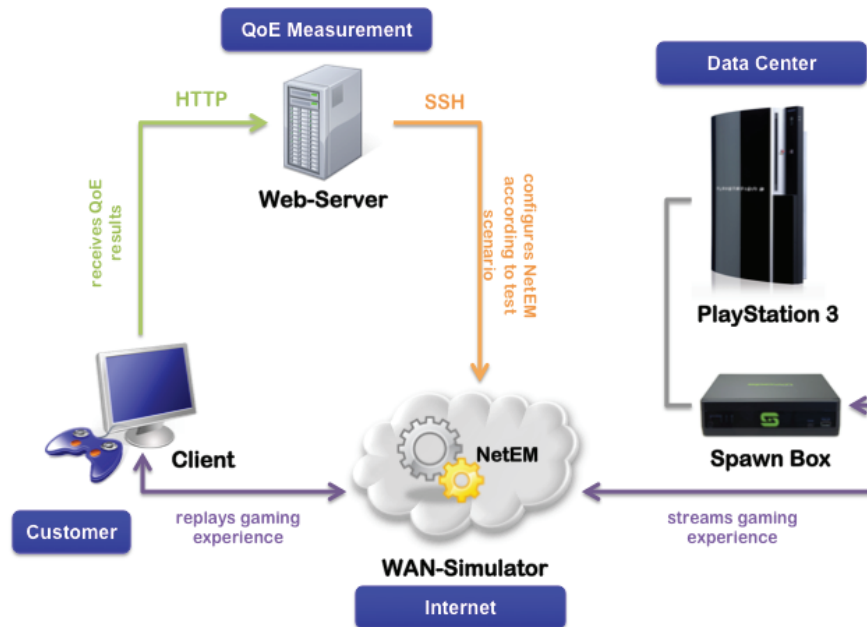

Figure 2: Testbed Setup - Logical View (taken from [7])

3.3. Survey Process

The test participant is asked to sit down at the client pc. The client pc is equipped with two monitors, that serve two different purposes. While on the first monitor the researcher conducts the opinion poll and could control

8

the test, the subject is to play the game on the second display. First we create a unique identifier for each participant and store his age. Next the player can pick one of our three games according to the three classes defined in [13]. We chose Pro Evolution Soccer for the omnipresent perspective (slow-pace gameplay), Final Fantasy XIII for the 3rd person perspective (medium-paced gameplay) and Gran Turismo HD Concept for the 1st person perspective (fast-paced gameplay) (cf. 3.1). The participants are allowed to repeat the test using another game. Subsequently we interview the test person on whether or not they favor games of that particular genre in order to determine if the test participant is potentially biased by their preference. We then ask the participant to estimate his/her skill in gaming as explained in Subsection 4.1.

Following these initial questions the subject is allowed to explore the game and its controls in 10 minutes of free play time. During this period the game is intentionally not affected by any distortions, so that the player can use this experience as a reference point (perfect experience) to the scenarios introduced in the testing phase. Every test subject is supposed to experience every scenario we introduce exactly once during the test. To avoid biased results caused by a specific sequence of scenarios, we decided this sequence to be randomly generated with the exception of always starting with the baseline. Each scenario lasts for about 1 minute. At the end of a scenario the researcher asks the participant for his current game experience, i.e. the quality of experience perceived by the player. This rating is expressed by the so called Mean Opinion Score (MOS) [17] for perceived quality of experience. Each experience was mapped to a value ranging from 1 to 5 with increasing values implying increasing quality ranging from bad to excellent. We left it to our participants to decide, which aspect of their experience image quality or responsiveness they weighted the most in their rating, since we intend to express the entire game experience by this value. With all ten scenarios being completed we then ask the test participant whether or not they are willing to pay a monthly fee for the overall experience they just made on the understanding that they can play any game they wanted to. We do this in order to get an overall impression of how the tests are perceived. Finally, we informally interview our participants on their general attitude towards the idea of Cloud Gaming and the potential they attribute to the concept.

4. Rater Reliability

In this section we have a look at the demographics of our survey participants 4.1 and subsequently determine the reliability of their ratings 4.2.

4.1. Demographics of the Test User Group

A study performed for Electronic Arts [18] in 2005 polled 3000 people in Germany aged 14 and above for the purposes of in-game advertising. It argues that only five percent of all gamers actually play often and are so called "intense gamers". By contrast the major percentage encompasses two groups: 24 percent are what Electronic Arts calls "casual gamers" and 54 percent of the interviewees are considered to be "leisure gamers". The study implies that most gamers and therefore most potential users of cloud gaming in Germany play on an occasional basis. Hence, the sample we took was aimed at getting a representative share of the target population defined by playing on a regular or occasional rather than an intense basis. Our sample is made up of 58 participants. Participants were often unsure whether they played on a regular or occasional basis. Therefore, we changed the question and asked the participants how they perceived their skill at gaming, which seems to be a less vague indicator. 15.2 percent of the participants consider themselves to be skilled gamers, while 44.6 percent think that their gaming skill is "medium", and 39.2 percent even judge themselves as "low". These percentages can be mapped to the groups of "casual gamers" and "leisure gamers". We can conclude that most of our test subjects do not play on an intense basis and thus our sample should lie within the target audience of Cloud Gaming.

4.2. Rater Reliability and Diversity

In order to determine the reliability of our rater, we use two measures - intra- and inter-rater reliability as described by Hoßfeld et al. [19]. Intra-rater reliability determines the consistency of ratings made by one single individual. We use the Spearman rank correlation coefficient to quantify both measures. This coefficient determines whether the relationship between two variables can be described with a monotone function. Here this would be the ranking given by the person and the value of the network parameter in question. Ideally, the ranking should change proportionally with the value of the network parameter resulting in different results for each setting. No repetition of values in the ranking would result in a Spearman rank correlation coefficient with an absolute value of one. In Figure 3 a CDF for the intra-rater reliability of our users is given. We consider users with a Spearman rank correlation of greater than an absolute value of 0.60 to be reliable and thus consistent in their ratings. The Figure shows that roughly 80 percent of our users fall into this category.

Inter-rater reliability on the other hand describes the degree of agreement between multiple users given the same test. Figure 4 gives the inter-rater reliability of our users in different scenarios. For Scenarios in which only packet loss is applied, we see a high absolute value of the Spearman rank correlation between users between about 0.75 and 0.9, which clearly indicates loss as an important factor. However, in delay-only scenarios the picture is not so clear. Especially for the fast-paced game the ratings are very diverse. Based on the parameter weights 5.4 the overall inter-rater reliability lies just around our cut-off point of 0.6. This indicates a significant difference in perception dependent on the user. This is underlined by the standard deviation of the opinion score (SOS) for reliable and overall users as shown in Figure 5 and a relatively high SOS parameter a of about 0.3-0.35. This shows that the assessment of QoE in cloud computing is not trivial as even with a larger number of test participants this deviation will not be significantly lower.
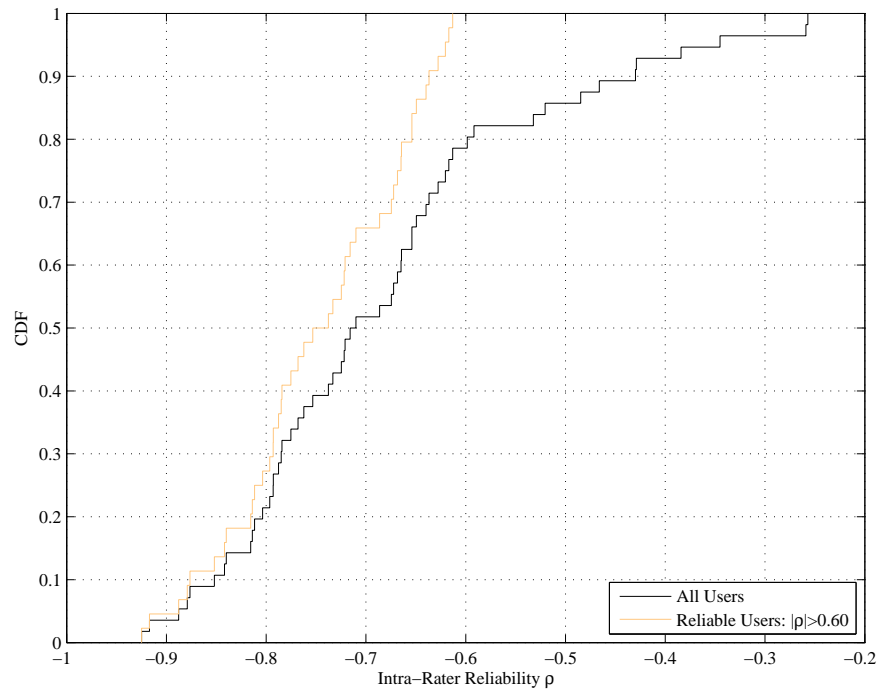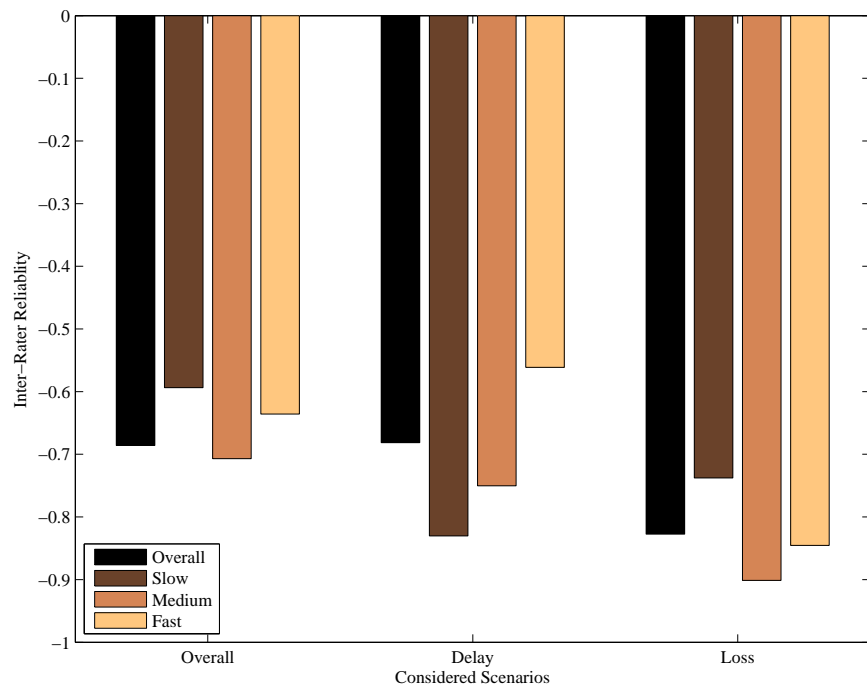
Figure 3: Intra-Rater Reliability
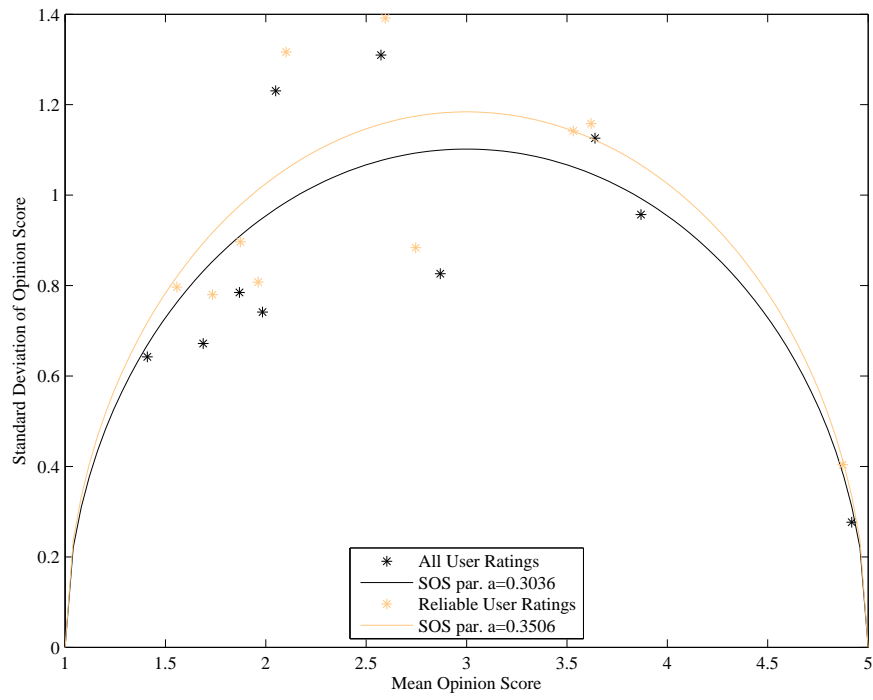
Figure 4: Inter-Rater Reliability

Figure 5: SOS

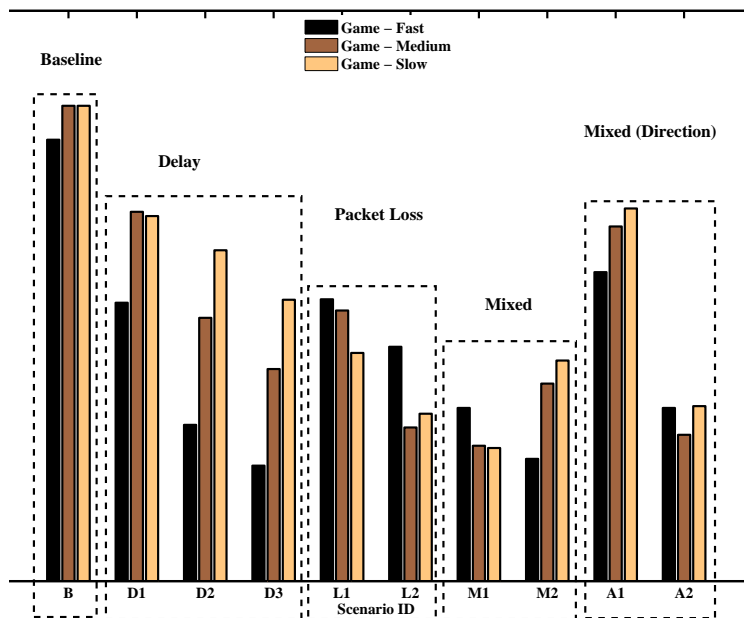# 5. Identification of Key Influence Factors for Cloud Gaming QoE

Figure 6: MOS Ratings per Scenario/Game (taken from [7])

Figure 6 illustrates the surveyed MOS value for each game in each of our scenarios. The plot is based on the data of 79 test runs, respectively 790 user votes. The y-axis indicates the MOS for a particular scenario, denoted by its scenario ID on the x-axis. At first glance it is apparent that the MOS values of each scenario differ from game to game. This variation is most remarkable in the bi-directional delay scenarios (D1-3). It seems the slower the gameplay gets the better the ratings become. For instance, scenario D2 is rated at 1.2143 MOS (bad) in combination with the racing simulation (fast), while it is rated at a value of 2.2308 MOS (poor) using the role play game (medium) and with the soccer simulation (slow) even scores a MOS value of 2.96 (fair). We therefore suspect that faster games are more delay-sensitive than slower ones. This agrees with the classification of Claypool et al.. It is reasonable that the influence of delay on Cloud Gaming is similar to its influence on conventional games.
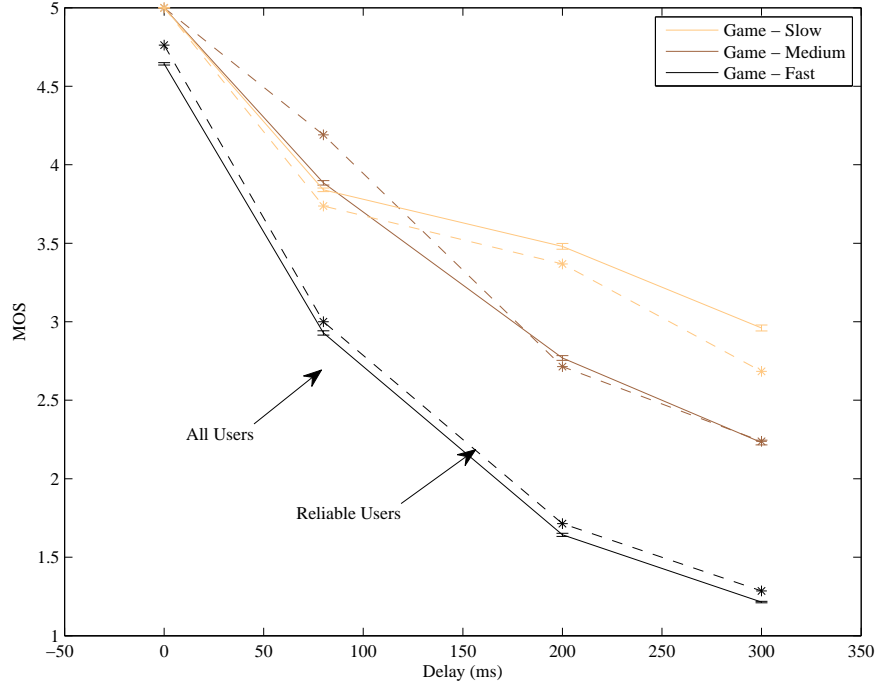
15

Figure 7: Delay Scenarios (taken from [7])

Figure 7 illustrates the measured MOS values for the bi-directional delay scenarios. The delay values are shown on the x-axis and the y-axis gives the corresponding MOS values. The values for the x-ticks are taken from scenarios B and D1-3. The results for each game are plotted as two graphs - one for all raters and one for reliable raters only. Confidence intervals are given for each MOS value. The intervals are small, hence we can conclude, that the MOS values are stable and enough ratings were collected. The difference between all users and the reliable group appears to be marginal. We observe that all graphs decrease with increasing delay. As suspected, there is a decline of MOS values with increasing delay. Furthermore, the plot confirms that the racing simulation appears to be most delay-sensitive for its graph runs below the others. Up to a delay of 80 ms the user experience has the same quality for role play game (medium) and soccer simulation (slow). The delay value of 80 ms was chosen to lie in the area of threshold where

players start to notice the delay. While the delay is recognized in the racing simulation and rated with a MOS value of 3, only some people detected it in the role play game and the soccer game resulting in a MOS value of 4 for both. At a delay of 200 ms, however, the graph of the soccer game is clearly above the role play game graph which allows us to draw the conclusion that indeed the slower the game is, the less delay influences the user rating.
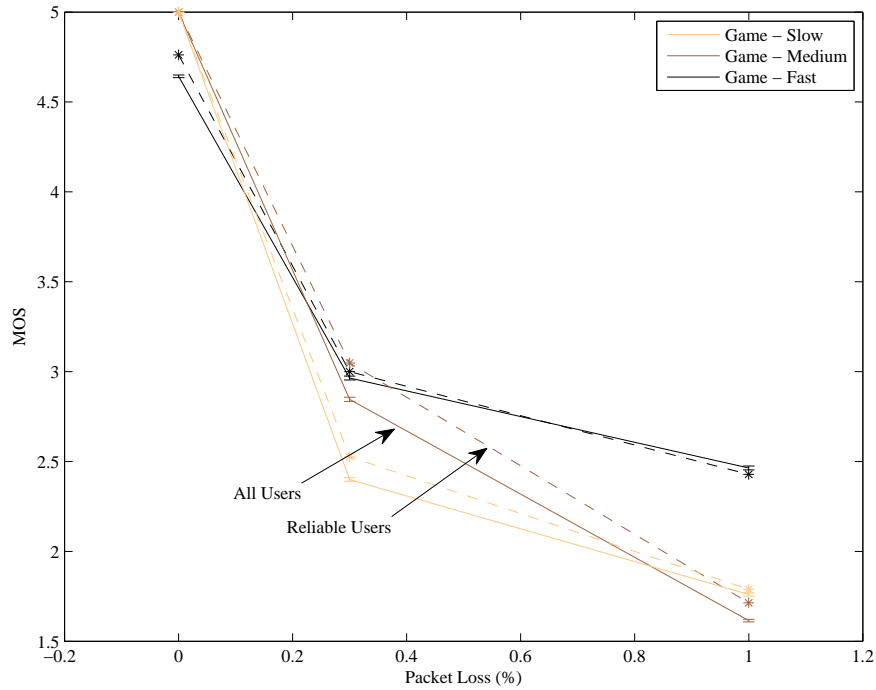
Figure 8: Loss Scenarios (taken from [7])

Figure 8 visualizes the surveyed MOS values for packet loss. MOS values are shown on the y-axis, while the values for packet loss are on the x-axis. We used the packet loss values of scenarios B,L1 and L2 for the x-ticks. Again, the results for each game are plotted as two graphs - one for all raters and one for reliable raters only. Also, the confidence intervals are for each MOS value are again small, indicated a stable value. Same as before, the gap between all users and the reliable group is negligible. It becomes obvious that all graphs drop with increasing packet loss. Consequently, we also conclude that there is a decay in MOS with increasing packet loss. We assume this is due to the fact, that with increasing packet loss the video quality degrades more and more. We note that in essence the racing simulation, the most upper graph, appears to be most resilient towards packet loss. This might be a result of the circumstance that in fast paced games the player never really focusses on his environment as it is changing rapidly and thus degraded video quality becomes less important. Furthermore, fast paced games have a much higher command input rate, than slower games. Here a lost keystroke is often

18

subconsciously repeated. These facts seem to confirm our assumption.

5.2. What is Perceived Worse by Users: Delay or Loss?

In Figure 9 we used a two-dimensional surface-plot to identify a user tendency on what is perceived worse for each game: packet loss or delay. On the x-axis the reader can observe the MOS values of scenario M1, while on the y-axis the MOS values of scenario M2 are denoted. Each point displayed as a square represents the rating for both scenarios. The z-axis, i.e. the color of a square indicates the frequency of a rating combination. The darker a square is, the more participants voted for this combination of MOS scores. For instance, the black square in the upper left plot (game - fast) at the coordinates (2,1) implies that 36 percent of all users that judged the racing simulation rated scenario M1 with a MOS value of 2 and scenario M2 with a MOS value of 1. Additionally we delineated the angle bisector in each plot. Squares that are located left or above this line indicate a preference towards scenario M2, while squares that are located right or below the bisector indicate a favor for scenario M1. Squares that lie exactly on the angle bisector express neutrality, i.e. the MOS value for scenario M1 equals that given to scenario M2. In Figure 9 we observe that about 50 percent of all people that rated the racing simulation considered scenario M1 and M2 equally bad. The remaining 50 percent, however, show a clear tendency towards scenario M1. This further reinforces the assumption made when looking at loss only, that fast games seem to be more tolerant towards loss than others. Furthermore, we see that the delay-intensive test is perceived worse. This fits with our results so far. Delay appears to be the decisive factor in fast paced games. Players of fast games would rather accept higher packet loss rates than they would tolerate high delays, for the gameplay and the players success in the game very strong depend on their ability to react swiftly. The plot for the medium-paced game (rpg) shows quite an opposite trend. Here most of the participants lean towards scenario M2. In the role play game over 50 percent of the users prefer the delay-intensive scenario over the loss-intensive, while about 40 percent remain neutral. This game therefore appears delay-resilient, but loss-intolerant. Players of medium-paced games would prefer high delay over high packet loss rates, since they are more interested in what they see (i.e. video quality) than in responsiveness. The reason for this is two-fold. On the one hand responsiveness is not that decisive for the gameplay and the players success in the game. On the other hand the ability to immerse in the simulated world is far more important in games like this. For
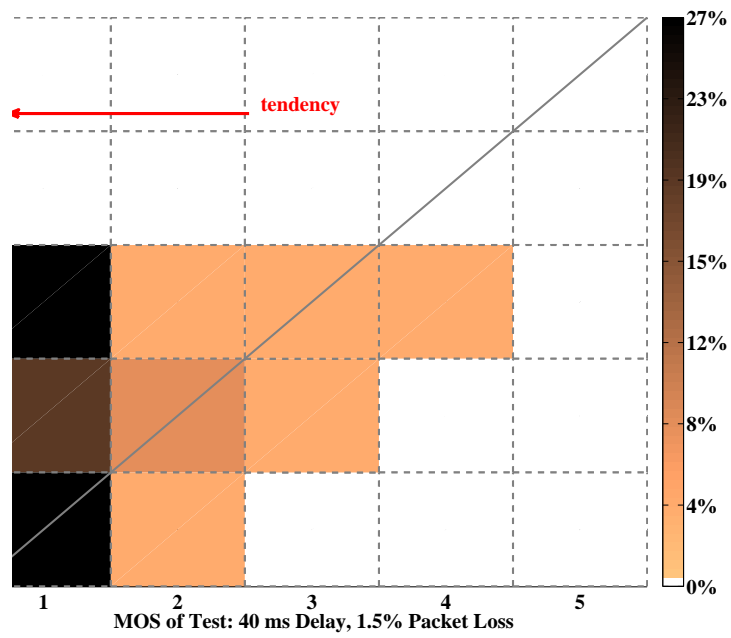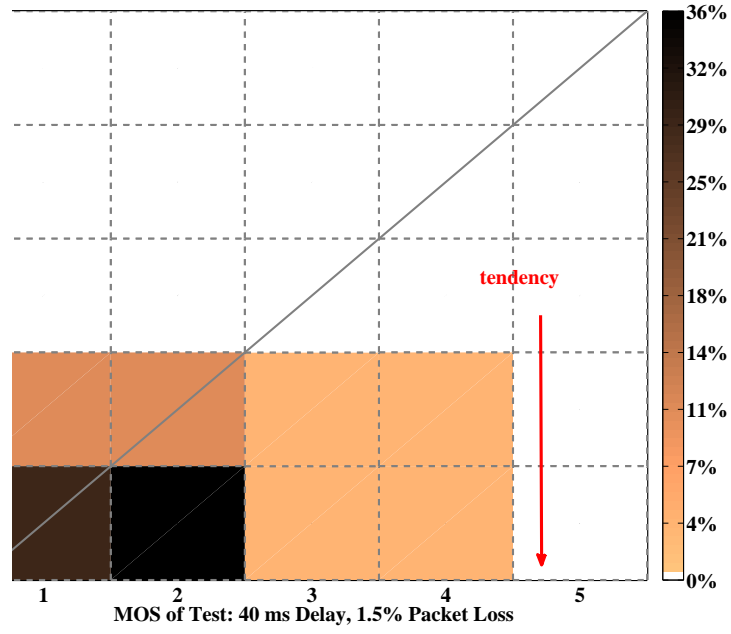
Figure 9: Mixed Scenarios (taken from [7])

the slow-paced game we could not derive any clear tendency. We observe a content-dependency and as we have seen, the question which parameter is perceived worse cannot be answered globally.

5.3. Evaluation of Asymmetric Network Conditions on QoE

Finally we have a look at the results of the asymmetric scenarios A1 (client to server connection disturbed) and A2 (server to client connection disturbed). The results for each of these scenarios contrast each other, although they use the same parameters albeit in different directions. The MOS
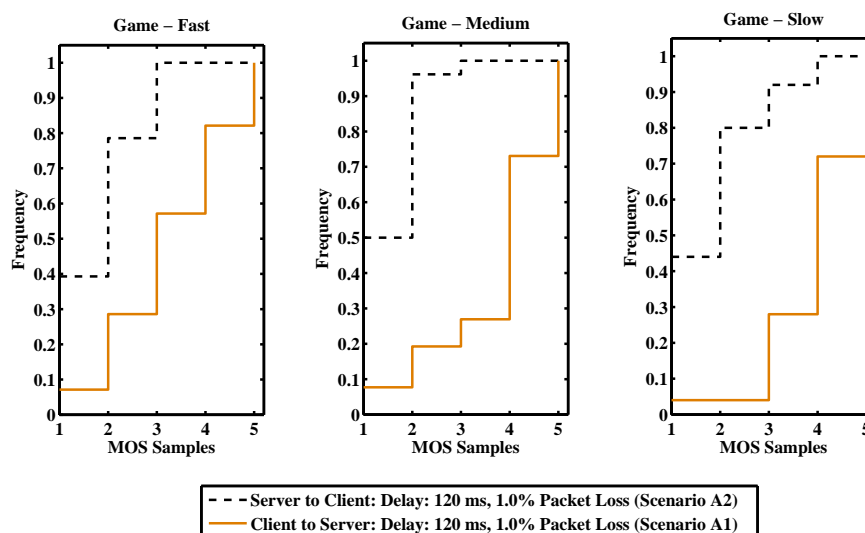

Figure 10: Asymmetric Scenarios (taken from [7])

value of scenario A1 was more than twice as high as the MOS value given to scenario A2. Figure 10 shows the cumulative distribution functions (CDF) using the MOS value as random variable. Since MOS values are discrete, we see a stair-plot. Each plot displays one game. The first observation we make is that in all games the graph of scenario A2 slopes upwards much faster than the graph of scenario A1. This indicates that generally more test participants disliked the distortion of the server to client connection. For instance, while in the fast-paced game less than 30 percent of the test participants rated scenario A1 with a MOS value 2, almost 80 percent rated scenario A2 the same. Hardly anybody rated scenario A2 better than a MOS value of 3,

except for the slow-paced game where less than 10 percent gave a MOS value of 4. The explanation for this tendency is quite obvious: Server-to-client loss of one percent results in massive video distortions, while client-to-server loss of one percent remains virtually invisible. Very few of the test persons ever knowingly complained about control inputs being dropped. However, a packet loss of one percent can very well compromise more than 20 percent of the picture in the video stream. The graph of the role play game increases the fastest. Over 90 percent rated it with $MOS \leq 2$. Again, this is linked to the way people experience the game. Role play gamers want to immerse into the world of game, therefore video distortions of this magnitude can hardly be tolerated as they greatly decrease the visual experience. Comparing the client-to-server graph of each game, we observe that it continuously bottoms out the slower the game gets. This means the less participants rate client-to-server distortions as bad the slower the game becomes. Although not all people consciously recognized the dropping of control inputs, it had an impact on their rating. If a soccer player will not pass the ball immediately, the test subject will simply press the button again as these games often do have an inherent delay to a players action. If a vehicle in a racing game will not turn immediately, however, it might be too late and the player might crash into a wall. We come to understand that server-to-client packet loss due to video distortion is far more critical for many Cloud Gaming applications than client-to-server packet loss, which might not even be knowingly perceived in a great deal of cases. Client-to-server packet loss only becomes grave, if a missed input potentially results in the player using the game. The delay of 120 ms was hardly recognized, no matter in which direction.
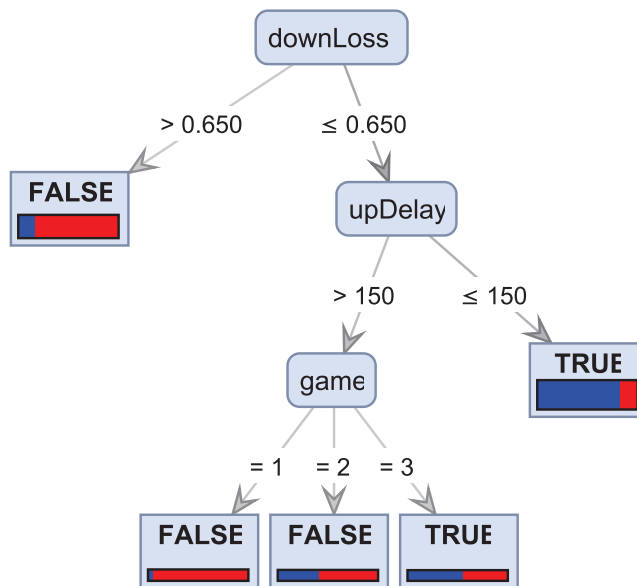
Figure 11: Decision Tree

So far we have derived several qualitative influences that different parameters have on the QoE of a cloud gaming application. However, for a service provider it is also important to know, how significant the influence of a certain parameter is compared to others. This way the service provider can structure the service in such a way as to ensure a minimum level of QoE at all times. We have used the standard data mining and statistics tool Rapidminer [20] to derive the importance of the parameters in our survey. Table 5.4 lists the parameters and their assigned weights based on the information gain calculated by the tool for samples yielding a fair quality, i.e. a MOS value of three and above. It identifies downstream packet loss as the most important parameter for QoE in cloud gaming in our survey with a maximum weight of 1, followed by downstream delay, which is already significantly less important with weight of 0.583. This shows, that the downstream transferring the video has a statistically higher impact on QoE than the upstream with the upstream packet loss and delay at weights of 0.370 and 0.212 respectively. However, both upstream parameters still have

23

Parameter	Weight
Downstream Packet Loss	1.0
Downstream Delay	0.583
Upstream Packet Loss	0.370
Upstream Delay	0.212
Type of Game	0.067
Player Skill	0.006
Player Attitude Towards Game	0.006
Player Age	0.0

Table 2: Weight of Parameters Based on Information Gain (taken from [7])

a significant weight, while it appears that the influence of game type, player skill, etc. is negligible. Additionally we used the WEKA [21] implementation of the REPTree algorithm in RapidMiner to construct a decision tree. This method tries to construct a subset of specific decision rules from a general rule covering the entire data set, i.e. the test results, by recursively splitting it based on information gain. The rule at the root of this tree, i.e. the first split, signifies the most important parameter for the decision. The decision we want to make here is whether the game quality is acceptable, i.e. true, or bad, i.e. false. The resulting tree based on our test results is illustrated in Figure 11. Our tree has the downstream packet loss as the most significant parameter at its root. A loss value of greater than 65 percent will result in a bad experience. If this is not the case, the upstream delay becomes the next significant influence factor. Here a delay of less than 150 ms will result in an at least acceptable experience. However, if the delay is higher, the game type becomes the next decisive factor. Under these circumstances only the slow game can yield acceptable results. This again suggests a difference in perception for up- and downstream impact factors as seen in the previous section.

6. Conclusion and Outlook

In this article we present our findings on the user-perceived QoE in Cloud Gaming. We introduce a test setup to perform a subjective survey on this topic and evaluate the results. We determine that the perceived game experience is not only dependent on the QoS parameters of delay and loss, but

has to be put into context with the content. While this is very similar to QoE in conventional gaming, we also derive a unique effect in Cloud Gaming. In Cloud Gaming it is far more important for players in which direction packet loss occurs than in conventional gaming. Generally, the quality of the video plays an important role. This is especially true for games that rely on impressive visuals. Based on these results we can confirm that Cloud Gaming is indeed a viable option for the future. While in our survey only about 15 percent of the participants were willing to pay a monthly fee, all were generally open to using such a service, if provided in a good quality. This raises two questions. With a considerable amount of computing power and bandwidth required per user, how would such a concept scale and be financially successful? Do technologies that do not rely on the transmission of full pictures but graphics API calls like Microsoft RemoteFX provide an advantage in this field? This questions will be investigated in our future work.

References

[1] A. Ojala, P. Tyrvainen, Developing cloud business models: A case study on cloud gaming, IEEE software 28 (4) (2011) 42–47.

[2] P. Ross, Cloud computing's killer app: Gaming, Spectrum, IEEE 46 (3) (2009) 14–14.

[3] T. Chang, Gaming will save us all, Communications of the ACM 53 (3) (2010) 22–24.

[4] K. Chen, P. Huang, C. Lei, How sensitive are online gamers to network quality?, Communications of the ACM 49 (11) (2006) 34–38.

[5] OnLive.
URL http://www.onlive.com/

[6] OnLive Blog.
URL http://blog.onlive.com/2010/01/21/beta-testing-at-the-speed-of-light/

[7] M. Jarschel, D. Schlosser, S. Scheuring, T. Hoßfeld, An Evaluation of QoE in Cloud Gaming Based on Subjective Tests, in: Workshop on Future Internet and Next Generation Networks (FINGNet-2011), Seoul, Korea, 2011.

[8] I. Nave, H. David, A. Shani, Y. Tzruya, A. Laikari, P. Eisert, P. Fechteler, Games large graphics streaming architecture, in: Consumer Electronics, 2008. ISCE 2008. IEEE International Symposium on, IEEE, pp. 1–4.

[9] M. Pigora, S. Waldron, The 3d world in your browser: A server rendering approach, in: The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Vol. 2010, NTSA, 2010.

[10] D. Chan, On the feasibility of video gaming on demand in wireless lan/wimax.

[11] Y. Chang12, P. Tseng, K. Chen, C. Lei, Understanding the performance of thin-client gaming.

[12] T. Szigeti, C. Hattingh, End-to-end qos network design: Quality of service in lans, wans, and vpns (networking technology) (2004).

[13] M. Claypool, K. Claypool, Latency and player actions in online games, Communications of the ACM 49 (11) (2006) 45.

[14] L. Pantel, L. Wolf, On the impact of delay on real-time multiplayer games, in: Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video, ACM, 2002, p. 29.

[15] T. Szigeti, C. Hattingh, End-to-end qos network design, Cisco Systems, 2005.

[16] T. Henderson, S. Bhatti, Networked games: a QoS-sensitive application for QoS-insensitive users?, in: Proceedings of the ACM SIGCOMM workshop on Revisiting IP QoS: What have we learned, why do we care?, ACM, 2003, p. 147.

[17] P.800 : Methods for subjective determination of transmission quality.

[18] P. Kabel, Spielplatz Deutschland.
URL http://presse.electronic-arts.de/publish/page205796005792144.php3?1=1&aid=162&spieleid&PreSelectCategory=0&Kategorie=&ParentID=&DOWNLOADID=773

[19] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, Quantification of YouTube QoE via Crowdsourcing, in: IEEE International Workshop on Multimedia Quality of Experience - Modeling, Evaluation, and Directions (MQoE 2011), Dana Point, CA, USA, 2011.

[20] RapidMiner.
URL `http://rapidminer.com`

[21] WEKA.
URL `http://sourceforge.net/projects/weka/`