An Evaluation of QoE in Cloud Gaming
Based on Subjective Tests
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

I. 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.

Recently, a new type of cloud service has been introduced, which may have the most stringent demands on network QoS to date: cloud gaming. 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. [1], their findings cannot be applied here.

Cloud Gaming enables new business models for games 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 handheld 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 (c.f. [2]) in the US. The company operates several data centers across the US in order to minimize the propagation delay imposed by the physical constraints (c.f. [3]). 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 survey 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.
To the best of our knowledge there is currently no scientific publication dealing with the topic of user perceived quality in Cloud Gaming. We have therefore opted to integrate work in related fields, i.e. video conferencing, conventional gaming, into the design of our user survey.

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

II. SURVEY PARAMETERS AND DESIGN

In Subsection II-A we select the range of the QoS parameters loss and delay whose influence on QoE is tested in our user tests. In Subsection II-B we discuss the attributes of the test group our survey is based on. Our testbed is explained in subsection II-C. Finally, we characterize the actual survey process in Subsection II-D.

A. 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 [4] 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 look at [5] 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 [6] 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.

Three classes of games with different behaviour towards QoE are identified by Claypool et al. in [7]. 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.

Table I 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 noticeable 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).

B. Demographics of the Test User Group

A study performed for Electronic Arts [8] 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.

C. Emulation of Cloud Gaming

Figure 1 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 HaiVision’s 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.

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 II-A. 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.

D. Survey Process

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.

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.
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.

III. IDENTIFICATION OF KEY INFLUENCE FACTORS FOR CLOUD GAMING QoE

![Figure 2. MOS Ratings per Scenario/Game](image)

Figure 2 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.

A. Impact of Symmetric Delay and Loss on QoE

![Figure 3. Delay Scenarios](image)

Figure 3 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 result for each game are plotted as one graph. 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 4. Loss Scenarios](image)

Figure 4 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, each game corresponds to one graph. 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 subconsciously repeated. These facts seem to confirm our assumption.

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

In Figure 5 we used a two-dimensional surface-plot to identify a user tendency on what is perceived worse for each
C. 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 value of scenario A1 was more than twice as high as the MOS value given to scenario A2. Figure 6 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 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 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.

Fig. 5. Mixed Scenarios

The 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 5 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...
with MOS ≤ 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.

D. Towards a Key Quality Indicator

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 [10] to derive the importance of the parameters in our survey. Table II 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 a significant weight, while it appears that the influence of game type, player skill, etc. is negligible. Additionally we used the WEKA [11] implementation of the REPTree algorithm in RapidMiner to construct a decision tree. As the most important parameter the downstream packet loss is at the root of this tree. If there is packet loss on the downstream the downstream delay becomes the next significant influence factor, if not the upstream delay is more important. This suggests a difference in perception for up- and downstream impact factors as seen in the previous section.

IV. CONCLUSION AND OUTLOOK

In this paper 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.

ACKNOWLEDGMENT

The authors would like to thank Prof. Tran-Gia for the fruitful discussion and support in this work.

REFERENCES