Quantifying the Influence of Network Conditions on the Service Quality Experienced by a Thin Client User

Barbara Staehle[1], Andreas Binzenhöfer[1], Daniel Schlosser[1] and Björn Boder[2]

[1] University of Würzburg, Institute of Computer Science, Würzburg, Germany
{bstaehle, binzenhoefer, schlosser}@informatik.uni-wuerzburg.de
[2] DATEV eG, Nürnberg, Germany
bjoern.boder@datev.de

Abstract. Quality of Experience (QoE) is a measure for the satisfaction of a user with a given service. Despite the popularity of the term, there is a lack of understanding of how to capture or quantify QoE. In this paper we therefore take a step forward in this direction using the example of thin client architectures, where the client is only used as a dumb terminal while the applications run on a central server. The satisfaction of a thin client user is thus heavily influenced by the quality of the network connection between the client and the server. We study this influence by varying different network parameters like packet loss, jitter, and delay in a controlled testbed environment. The service quality is assessed using objective measurements on application level (oQoE) and verified by means of a survey among test persons capturing subjective user satisfactions (sQoE).

1 Introduction

Most software products and especially office applications are run locally on the client computer or on network file systems. In thin client architectures, however, the application as well as all data processing is performed by a central server. The clients themselves are very cheap computers, often without a hard-drive, and are merely used as interactive displays. This design paradigm offers many advantages like lower hardware and administration cost, lower energy consumption, and a more efficient use of resources. Especially small and mid-size organizations can thus achieve a much higher degree of security and reliability if all their applications are hosted, managed and maintained by an Application Service Provider (ASP) who can spread the support and maintenance costs among a large number of customers [1]. A thin client architecture also enables the simple integration of home office workers into the corporate environment. The downside, however, is that the response of the application to the input of the user is no longer direct but has to be transmitted over the network.

While most thin client architectures had initially been designed for local area networks, they are increasingly being used in wide area networks or over leased lines. In such environments typical network characteristics like packet loss, jitter, or delay play a decisive role. Naturally, the *Quality of Service* (QoS) offered by the network directly influences the *Quality of Experience* (QoE) which expresses the satisfaction of the user with the service: In the presence of network disturbances remote applications will no longer react correctly to user input like key strokes or mouse movements. However, it is non-trivial to translate measured QoS parameters into an indicator for the QoE perceived by the end user.

In this paper we therefore study and quantify the influence of typical network parameters on the QoE achieved in a Citrix MetaFrame environment. In particular, we vary different network characteristics in a controlled testbed and analyze how such changes in the QoS affect the service quality on application layer. To be able to quantify the QoE we automate the typical behavior of office application users and record the time required for specific tasks in dependence of the network conditions. To evaluate to what extend this *objective* QoE (oQoE) metric translates into real user satisfaction, we compare these findings to a survey with test persons which asses the personally experienced *subjective* QoE (sQoE). Our results are a first step toward deriving a measure for the satisfaction of a thin client user based on measurements on network layer. They can furthermore be regarded as a means to help in establishing Service Level Agreements as well as in deciding whether a given connection quality suffices to support Citrix-based applications.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of related work. The measurement setup as well as the measurement methodology are explained in detail in Section 3. We show different ways to analyze the measurements and discuss the results and their implications in Section 4, while Section 5 concludes this work.

2 Related Work

The concept of Quality of Experience is used to describe to what amount a user is pleased with the results of a service he used. In contrast to Quality of Service, QoE is a subjective measure which is not only based on the performance of the service itself but also on the individual expectations and perceptions of the user. Measuring the performance of a well defined process is a typical task in technical sciences and understanding the satisfaction of a human being is widely studied in psychology. Bringing those two areas together and quantifying QoE, however, is a rather unexplored topic. White papers introducing the idea of measuring QoE based on the characteristics of a service, are e.g. [2] or [3], but it has to be kept in mind, that since long, works dedicated to quantifying the user experience without using the term "QoE" exist. An example are the efforts of the group around Rubino, aiming at quantitatively evaluating the user perceived quality of a video or audio stream. Their Pseudo-Subjective Quality Assesment (PSQA) technique [4] is based on Random Neural Networks and allows to map network QoS conditions to user satisfactions.

Other examples for mapping QoS to QoE, are the *Mean Opinion Score* (MOS) [5] and the *E-model* [6] which allow to express user experienced quality for multi-media transmissions like VoIP and IPTV. To generate the MOS, a number of test users have to listen to a set of standardized test sentences read aloud over the communication medium under examination. Afterwards, they have to rate the audio quality using numbers between 1 (bad) and 5 (excellent). The E-model allows to compute a scalar "R-factor" based on connection delays and equipment impairment factors. It ranges between 100 (excellent) and 0 (poor). Several researchers have worked on the problem of deriving the QoE of a Skype user from the current network conditions. Results of an exemplary study, which examines the influence of several codecs on the user perceived quality are presented in [7]. For the case of edge-based multimedia services, [8] hypothesizes an exponential relationship between the network QoS and the end user QoE. Exemplarily, MOS is used as a QoE measure for Skype, but the relationship is assumed to be adoptable to other cases, too.

In contrast, research on thin client computing has focused on typical traffic characteristics and neglected the user perceived service quality. In [9], the traffic caused by thin client based office applications is characterized and in [10] the differences between several thin client architectures are analyzed. While the end user QoE has not been of major interest, there are some papers approaching this topic. One example is [11], where the usability of Microsoft PowerPoint via WAN has been evaluated. For this purpose, the completion time of several PowerPoint tasks which were shown to the user via a Citrix thin client architecture have been analyzed. Test users report "on machine" experience for high bandwidth links like Fast Ethernet in contrast to slower links which cause the execution times to be up to twice as long. In [12] the response time of text editing, presentation creating, and image processing applications accessed via VNC was used as a measure for the interactive thin client user experience. For this purpose, all response times were classified to be either "crisp", if smaller than 150 msec, "unusable", if over 5 sec and "noticeable to annoying", "annoying" or "unacceptable" if the response times were in between. Based on this simple QoE measure, it could be shown that the performance of highly interactive applications is more sensitive to higher delays in the network than that of simpler applications.

However, none of these studies considers the implications of a decreased network QoS for the QoE of an office user. In this paper, we therefore focus on the question in how far objective performance measures, like task completion times, can be used to assess the satisfaction of a real user in the same situation.

3 Measurement Setup and Methodology

To be able to capture the QoE of a thin client user we set up a realistic testbed environment and define a simple metric which quantifies user experienced service quality by measuring the time it takes to complete a specific task under preset network conditions. To obtain stochastically significant results it is necessary to automate the measurement process for obtaining multiple measurement results under the same network conditions. The measurement infrastructure which we used for this purpose is described in the following.

3.1 Testbed Environment

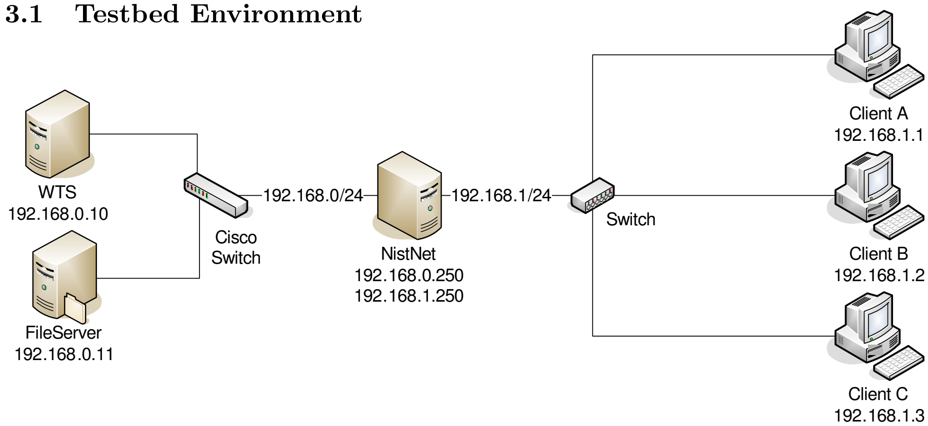


Fig. 1. Overview of the measurement setup

In order to emulate a typical thin client architecture used in productive environments, we set up a testbed as depicted in Figure 1. For the server side we use two 3.4 GHz Intel Xeon servers with 3.5 GB RAM each, running on Windows 2003 Server standard edition with Service Pack 1. The Windows Terminal Server (WTS) in Figure 1 hosts the entire Microsoft Office 2003 product family and runs Citrix Presentation Server 4.0 to make these applications available for remote users. The second server is set up as a file server and stores the user data. The clients use version 9.237 of the ICA client to access the applications hosted on the server. ICA is short for "Independent Computing Architecture", it uses TCP/IP, and is the proprietary protocol used by Citrix products. On the ICA client, data compression and session reliability were enabled and the color depth was set to the default value of 16 bit.

To control delay, jitter, and packet loss of the end-to-end connection we put a Dual Pentium III 500MHz computer with 512 MB RAM running OpenSuSE 10.0 and the network emulator NIST Net 2.1012.c [13] in the middle of the communication channel. On the client side we use Pentium IV 2.6 GHz machines with 1 GB RAM running Windows XP with Service Pack 2. All hosts are connected using 100 Mbit links. Note that we dimensioned the hosts and the network in such a way, that none of these components is a bottleneck and the performance of the applications is only affected by the emulated network conditions.

3.2 Organization of the Measurements

The service quality a user experiences under probabilistic network conditions is not deterministic but also varies with the quality of the communication channel. In order to repeat measurements multiple times under the same conditions, we implemented a control software in Perl which automates the entire measurement process. It consists of a client which controls the measurements and several server components installed on all participating machines which receive simple instructions from the client. An automated emulation run is organized by the client according to the following pattern:

- Configure network parameters on the NIST Net machine
- Start user application, e.g. Word, on the thin client
- Start recording packets
- Execute test
- Stop recording packets
- Rename and collect results

Each emulation run is performed for the duration of an hour. Within this hour the network emulation settings on the NIST Net machine remain unchanged. The client starts the corresponding application and a test consisting of three typical office tasks is performed several times (cf. Section 3.3). During the entire time we record all network traffic using WinDump [14]. Once the measurement is over, we collect all data and reset the testbed. The tests that were done in the first five minutes and the last five minutes of the hour are discarded in oder to prevent incorrect data caused by a transient time which might occur at the network emulator.

3.3 Automation of User Tasks

Inherently, Quality of Experience (QoE) is a subjective performance measure and is therefore difficult to quantify. We approach this problem by defining an objective QoE (oQoE) metric as the time required to complete typical office tasks on application layer. To make changes in the duration of a test measurable, we need to design short and well-defined tasks. We chose three popular MS Office products and created three typical subtasks for each of them. Word is mainly used for editing text, the main purpose of Excel is manipulating and representing data and most PowerPoint users create multimedia presentations. We sketch the user tasks mimicking this behavior in Table 1.

Table 1. Outline of the automated user tasks

Word	Excel	PowerPoint
Typing: Enter text, misspell, delete and retype some words	**Typing:** Enter some values into cells of the spreadsheet	**Slide Design:** Click slide design button to choose a new slide design from list
Scrolling: Scroll several pages up and down using the scrollbar on the right	**Selecting:** Select cells moving mouse from top-left to bottom-right	**Insert Picture:** Insert picture using the file option menu
Menu: Browse menu entries using the mouse pointer	**Diagram:** Create simple bar chart using diagram assistant	**Zoom:** Change zoom level of the current slide twice

In order to obtain repeatable and comparable measurement results, we need to generate the exact same user behavior in each emulation run. Therefore the entire user input is done automatically using the open-source tool AutoHotkey [15], which is able to carry out keystrokes as well as mouse movements and clicks defined in a simple macro language. The tool additionally enables us to measure the duration of each task and to verify that the intermediate steps of each task are successfully completed. To do so, AutoHotkey simply waits until some predefined pixels turn to a specific color. Under optimal network conditions, i.e. without any delay or packet loss, the duration of a test is deterministic and will therefore be used to normalize the oQoE values in the following.

4 Measurement Analysis and Exemplary Results

4.1 Identifying Critical Scenarios

There is a vast number of network parameters like packet loss, packet delay, packet reordering, or jitter, as well as asymmetric network conditions which all might have a different influence on the QoE of different applications. Evaluating the performance of all tasks described in Section 3.3 under all imaginable network conditions, would result in an infeasible amount of experiments to conduct and data to analyze. As a first step, we therefore identify the most critical QoS conditions to concentrate on. Initial experiments showed that packet loss (PL) and round trip time (RTT) are the two network QoS parameters which mainly influence the QoE. In the following, we apply a design-of-experiments based approach to identify which of the examined subtasks are particularly vulnerable to such degraded network conditions.

In Fig. 2 we illustrate the change of the duration of the different considered Word tasks, when both parameters increase from a *low* value (-) to a *high* value (+) which could be experienced by a user remotely accessing an application via Citrix. In particular, we choose PL(-) = 0 % and PL(+) = 2 %, and RTT(-) = 0 msec and RTT(+) = 200 msec as typical low and high values, respectively. For all four possible combinations of RTT and PL, we evaluated the performance for the chosen applications by repeating the corresponding tasks represented in Table 1 for the duration of an hour and measuring their length. Each point shown in Fig. 2 for a low or high value represents the mean over the task durations that where observed, when the parameter given on the x-axis was set to the indicated level and the other parameter was either at its low or high level. To make the results comparable, we show the normalized change in the test duration which is obtained by dividing the measured test lenghts by the time it takes the task to complete under perfect network conditions. The error bars show the corresponding 95 percent confidence intervals.

Fig. 2 shows that under the network QoS conditions we considered in this preliminary experiment, the round trip time has a more significant influence on the duration of the individual Word tasks than the packet loss. The Menu test, e.g., takes on average 30 % longer to finish if the RTT increases, but only 12 % longer if packet loss is increased. The Typing test is least influenced by bad

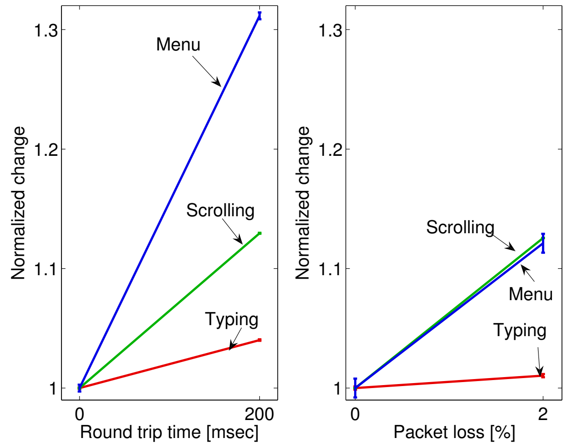


Fig. 2. Influence of RTT and PL on Word

network conditions: The average test duration increases only slightly while again packet loss has a smaller influence. The Scrolling test, in contrast is influenced to a similar extent by both QoS parameters.

The examination of the Excel tasks lead to very similar results (figure omitted): The Typing test is only slightly influenced by a degrading network quality, while the Diagram test and the Selecting test suffer more. We observe again, that both tests take longer to finish if the network delay is increased than if more packets are lost.

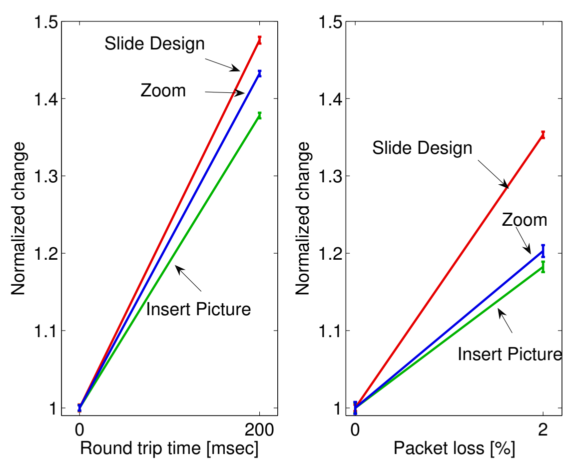


Fig. 3. Influence of RTT and PL on PowerPoint

Fig. 3 illustrates the normalized changes for the PowerPoint tasks under increasing RTT and PL. A comparison with Fig. 2 shows that PowerPoint seems to be more sensitive to a decreased network quality than Word. Recall, however, that we defined no simple text editing, but rather complex and interactive tasks for PowerPoint (cf. Table 1). Text editing under PowerPoint would be affected likewise than under Word. Increasing the round trip time has a similar effect on the PowerPoint Zoom and Slide Design Tests as on the Menu test under Word. The tasks of inserting a picture into PowerPoint under a heavily delayed connection is even more severely influenced than the Word Scrolling test. In analogy to the Word results, packet loss influences the Zoom and Insert Picture tests less significantly than does the RTT. This is, however, not the case for the Slide Design test, which needs longer to complete if PL is increased than if the RTT is increased.

According to the performance of the tasks we investigated, Word is the most robust of the three tested office applications in terms of decreased network conditions. Recall, however, that we tested more complex task under PowerPoint and that not all tasks of a specific application are affected to the same extent. To obtain a first overview, we compared mean test durations together with confidence intervals. The latter already illustrated, that the single test durations are not deterministic and often vary strongly, even if the network conditions do not change. Mean values are thus suitable for a first comparison, but do not give any insight into the exact distribution of the measured values. In the next section we will therefore have a more detailed look on the different test durations obtained under varying network conditions.

4.2 Mapping Network Parameters to QoE Values

In order to study the influence of different network conditions on the QoE, we repeat the Word, Excel and Powerpoint tasks described in Table 1 for one combination of PL and RTT during one hour. In this way, we obtain up to 100 oQoE measurements for each considered network QoS and each examined task.

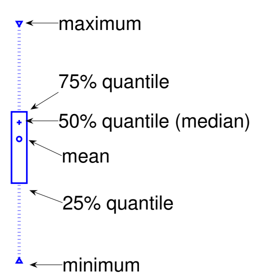
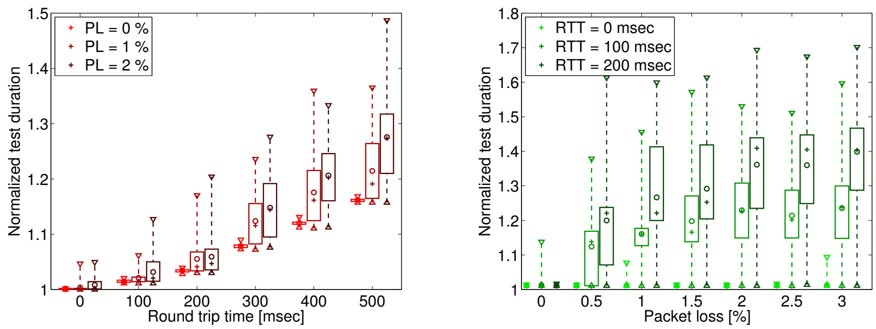


Fig. 4. Illustrating a distribution

To gain an overview of the statistical characteristics of the test lenghts measured under the same network conditions, we use a modified box plot as shown exemplarily in Fig. 4. All measured values lie within the interval given by the minimum and the maximum, i.e. by the lower and upper triangle, respectively. The mass of the distribution, i.e. 50 % of all observed values, lies within the box given by the 25 % and 75 % quantile. The length of this box, the inter quartile range (IQR), is thus a measure for the statistical dispersion of the observed test values. The positions of the mean and the median in the box furthermore characterize the skewness of the distribution: If both are roughly on the same position in the middle of the box, the mass of the distribution lying above the mean value is equal to the mass lying below the mean. If, in contrast the median is not equal to the mean as in Fig. 4, this is an indicator for statistical outliers, i.e. a small number of extreme test durations which significantly influence the mean value. In the example, the median is furthermore very close to the 75 % quantile, indicating, that many users, i.e. approximatley 25 %, experienced very similar test durations.

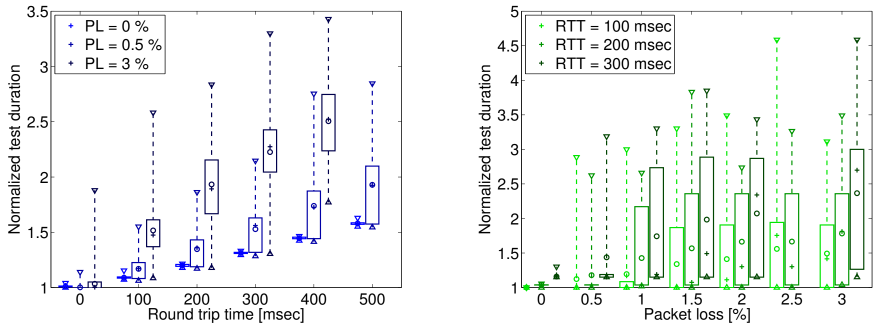


(a) Typing test vs. RTT (b) Scrolling test vs. PL

Fig. 5. Detailed Analysis of Word

As before, we analyze *normalized test durations*, i.e. the observed test duration is divided by the time it takes the same test under optimal conditions to complete. This way we are able to compare the tasks of the different applications to each other. The analysis of the normalized Typing test duration depicted in Fig. 5(a) shows that typing in Word documents suffers from an increasing RTT, but only to a small extend, which confirms the results of Fig. 2. The normalized durations are plotted against an increasing RTT on the x-axis. For each RTT value, we show the results of measurements done for three different PL values. Within such an RTT group, the PL is increasing from left to right and is visualized by different colors. Observe, that 0 % packet loss leads to almost deterministic results, while for 2 % packet loss the sizes of the IQR boxes become relatively large indicating heavily varying test durations. The minimal and maximal test durations show furthermore, that a lossy connection may lead to different user experiences: When there is only 1 % packet loss, the maximum test duration already increases significantly, whereas in most cases the tests completed much more quickly. Furthermore, note the near to linear increase of the fastest observed test, which only depends on the RTT and not on the PL.

In Fig. 5(b) we examine the behavior of the Scrolling test under increasing packet loss. Note that scrolling is affected in a very special manner by lossy connections: while under all network conditions, some perfect test runs can be observed, the maximum test duration increases strongly as soon as packets are lost. This corresponds to some users encountering no problems while scrolling, whereas others may notice severe disturbances. This problem is further illustrated by small IQR boxes identical to the minimum going together with very high maxima, which represents a nearly homogeneous user experience except for a very limited number of users. Such situations, which make the test behavior and thus the user satisfaction hard to predict were, e.g. in this test observed for PL = 0 % or 1 % and RTT = 100 msec.



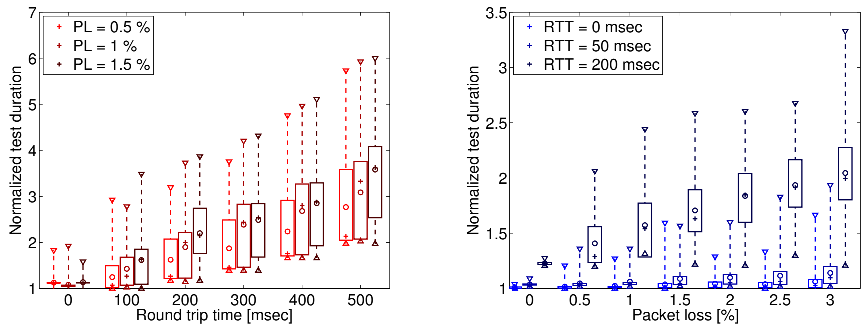
(a) Diagram test vs. RTT (b) Selecting test vs. PL

Fig. 6. Detailed Analysis of Excel

Typing values into an Excel spreadsheet is only very slightly influenced by increasing packet loss or round trip time. In Fig. 6 we therefore concentrate on the tasks of creating a diagram and selecting cells. The results of the Diagram test under increasing RTT are shown in Fig. 6(a) and illustrate that the increase of the RTT in combination with only a small packet loss probability of 0.5 % already leads to a significantly higher and also more varying duration of the Diagram test as compared to the optimal conditions. The experiment furthermore reveals, that PL = 3 % results in highly varying and significantly increased test durations. In the scenario with (RTT = 500 msec, PL = 3 %) we encountered too many Citrix connection failures to obtain enough values for statistically significant statements. The overall tendency of the test behavior again illustrates that for moderate network QoS conditions the minimum test duration is linearly increasing with the RTT, while the individual test outcomes are very dispersed, as soon as packets may be lost.

Selecting cells in Excel is also influenced more heavily by packet loss than the examined Word tasks. Fig. 6(b) illustrates that the measured test durations were up to 4.5 times longer in situations with very bad network conditions. Also note, that we were not able to obtain enough test values for the (RTT = 300 msec, PL = 2.5 %) scenario due to failures of the Citrix connection. The figure furthermore reveals, that while the mean and the worst case durations of the Selecting test are both increasing with the packet loss, the minimum duration always stays very close to the optimum case. Again, a small number of users will encounter no problems when selecting cells, while others will find this task very annoying or even impossible to complete.

We already saw earlier, that the examined PowerPoint tasks are more sensitive to increased packet loss and round trip times than the examined Word tasks. This fact can also be observed in Fig. 7(a), where the mean test durations increase with the round trip time (from left to right over the entire figure) and packet loss (within the three bars representing the measurement data for the



(a) Slide Design test vs. RTT (b) Zoom test vs. PL

Fig. 7. Detailed Analysis of PowerPoint

same RTT). Observe the strongly varying measured test durations, which indicate the sensitivity of this task to the network conditions. For degraded network conditions, the general user satisfaction with this task will thus be very hard to predict. Also note, that even without any delay (RTT = 0 msec), a small packet loss can sometimes have a significant impact and cause the task of choosing a slide layout to take twice as long as under optimal conditions.

The Zoom task represented in Fig. 7(b) again reveals, that the influence of PL on the test duration is not dramatical as long as the RTT is small, i.e. either 0 or 50 msec. As soon as the round trip time increases to 200 msec, the test duration is affected heavily. A closer look at the measurement results with RTT fixed to 200 msec, unveils that the increase of the mean test duration (depicted by the circle) is mainly due to a small number of very long test durations. In the (RTT = 200 msec, PL = 0.5 %) scenario, e.g., the cross representing the median is very close to the lower edge of the box indicating the 25 % quantile. Thus, the majority of tests need significantly less time to complete than the mean test duration, while a few tests require significantly more time.

The general insight which can be gained from the experiments shown in this section is that a delayed network connection always leads to an increase of the test duration. A lossy connection however, first of all leads to more varying test durations: while in some cases, the tests can still be completed in optimal time, the test duration increases dramatically in some other cases. Moreover, to which extent the absolute test durations increase differs strongly among the different applications and tasks. It is now of interest, in how far the normalized test durations which we used to express QoE in this section actually correspond to the satisfaction of real users. In the next section, we therefore validate our results by considering the opinion of real users.

4.3 Subjective vs. Objective QoE

In order to put the objective QoE (oQoE) in relation to the subjective QoE (sQoE) of real users, we conducted a survey with 150 graduate, undergraduate and high school students. All test persons were given the same exercise which required them to type a text, use the scroll bar, and search menu entries under

Microsoft Word for a period of approximately ten minutes. Each test person was working on one of the client computers in our measurement setup and did not know the actual network conditions. We acquired the satisfaction of the test persons with a questionnaire. In particular, the overall application quality as well as the quality of typing, scrolling and menu access tasks had to be classified using the following values: 1 (perfect), 2 (some disturbances), 3 (annoying), and 4 (unacceptable).

In Fig. 8, the dots illustrate the (RTT,PL) combinations which were used during our experiments with test persons. In particular, we chose RTT values varying in four steps between 0 and 180 msec, as a typical range which could be experience by an end user working in a Citrix environment. To investigate in how far realistic packet loss probabilities influence the service quality experienced by the end-user, we set PL to 0, 0.25, 0.5, 1 and 2 % respectively. This results in a total of 20 different scenarios to investigate. To obtain at least 10 sQoE ratings per measurement point (except for the best and worst conditions), we reduced the number of investigated scenarios as indicated in Fig. 8.

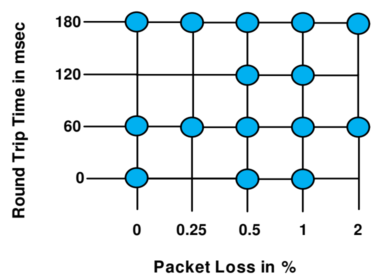
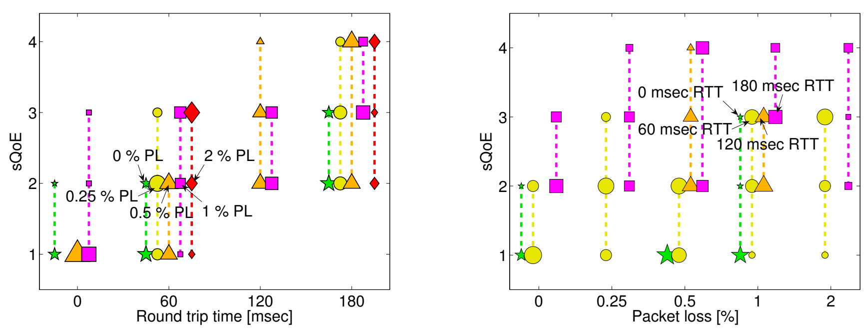


Fig. 8. Considered scenarios



(a) sQoE versus RTT

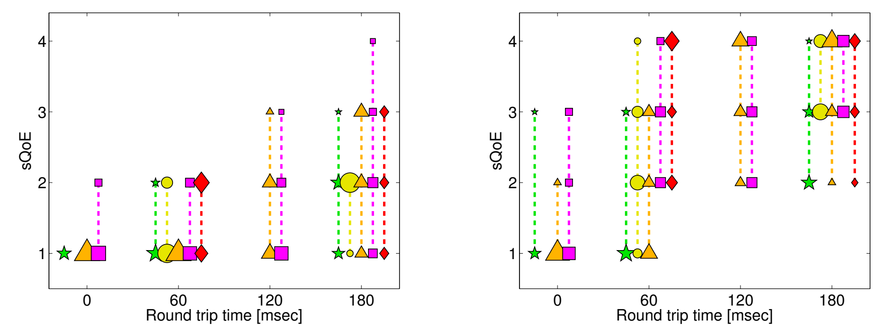
(b) sQoE versus PL

Fig. 9. Overall user satisfaction

In Fig. 9 we depict the results of the question regarding the overall user satisfaction. We try to give a graphical overview of the sQoE ratings both versus RTT and PL. In Fig. 9(a), the RTT is shown on the x-axis, while the different colors and markers within one RTT group indicate the different PL values, increasing from left to right. The sizes of the markers represent the number of the users which observed the corresponding sQoE level. Note that only 5 users were faced with the optimal scenario of no delay and no packet loss. The two green stars representing the sQoE of those five test persons are thus smaller than the markers of the other QoS conditions. In this case, four test persons indicated that they were totally satisfied (sQoE = 1) with the situation, where one test person claimed to have experienced small disturbances (2). Likewise for the (RTT = 120 msec, PL = 0.5 %) scenario, the three orange triangles correspond to one user finding the situation unacceptable (4) and four users which either picked (3) or (2), respectively.

By using a larger number of test persons, one could visualize the survey results using the same box plots introduced in the previous section or compute the correlation between the two metrics to obtain a more detailed mapping of oQoE to sQoE. Despite its simplicity, our approach still allows to roughly validate the results of the preceding sections. Fig. 9(a), e.g., qualitatively confirms the main statements of the oQoE plots: The user satisfaction is degrading with an increasing RTT as illustrated by the concentration of the mass to the upper right corner. Considering the markers representing the sQoE values for a fixed RTT but varying packet loss (i.e. the markers within one RTT group) more closely, one can also observe, that an increasing packet loss leads to an on average slightly degraded and more varying user satisfaction.

The same conclusions can be drawn from Fig. 9(b) which represents the overall user satisfaction in dependence of the packet loss. Observe that the markers are less clustered than in Fig. 9(a), i.e. the user opinions are less uniform and that a few users always report a perfect application quality. It can nicely be seen, that the stars at the left of each group, standing for the lowest investigated RTT value, always correspond to more satisfied users than the other markers representing slower connections for the same PL value.



(a) Typing sQoE versus RTT (b) Scrolling sQoE versus RTT

Fig. 10. User satisfaction with specific tasks

In accordance with our oQoE result, users gave different sQoE ratings for different tasks under the same network conditions. Take the satisfaction of the users with typing and scrolling in dependence of an increasing RTT, shown in Fig. 10 as an example. Note that on average typing (cf. Fig. 10(a)) is rated far better than the overall application quality shown in Fig. 9(a). The latter is in turn rated very similar to the Scrolling sQoE depicted in Fig. 10(b). Note that only one test person classified an RTT of 180 msec as unacceptable for the

Typing test, whereas most people indicated, that they were not satisfied with the overall quality of Word and Scrolling in particular under these circumstances.

Interestingly, none of our test persons reported a perfect experience for the Scrolling test, as soon as the RTT increased to moderate 120 msec. The same is true for using the menu, whereof the majority was already annoyed under mediocre network conditions. Note that under these conditions, no test user reported a “perfect” overall application quality, too. Psychology teaches that bad experiences do influence the peoples opinion much more heavily than positive ones, which might be an explanation for the relatively bad overall sQoE ratings going together with good Typing sQoE scores.

5 Conclusion

As end users are no longer willing to pay more and more money for even more bandwidth and connection speed, Internet Service Providers have to investigate new markets and concepts they can charge for. Charging for Quality of Experience instead of charging for Quality of Service is seen as one new possibility. During the last years, the attention of service providers has thus shifted from the network core to its edges and thus, the user satisfaction and the network performance have become equally important. In contrast to QoS, QoE, however, is much more difficult to assess as it is influenced by a large variety of parameters. Besides network parameters like packet loss, delay, or jitter, the QoE of an application is also affected by the expectations of the individual users, their subjective perception, as well as the type of service offered. This is probably the reason why so far most researchers avoided the problems of quantifying QoE and rather concentrated on analyzing QoS.

In this paper we therefore made a first step toward concretizing the concept of QoE. We discussed two different possibilities for quantifying the user experience, namely objective and subjective QoE. As a simple measure for the oQoE, we use the normalized time it takes to complete automated short tasks which are typical for the considered application. We used the example of MS Office applications in a thin client environment and showed in how far the different tasks are influenced by different network parameters. Our studies illustrate the importance of considering the entire distribution of the measurement results instead of only concentrating on mean values, as the statistical characteristics of the distribution provide valuable insights as well. For this purpose, we introduced a modified box plot which shows the most important distribution characteristics at a glance. Finally, we conducted a survey among test persons to examine the influence of the network conditions on the sQoE. The results of the survey matched our oQoE measurements to a satisfying degree. The time it takes to complete a specific task is therefore a simple and quick way to quantify the service quality experienced by an end user, when working with a thin client based office application.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable remarks. Special thanks go to Markus Weiß for his help during the experiments.

References

1. Walsh, K.R.: Analyzing the Application ASP Concept: Technologies, Economies, and Strategies. Communications of the ACM **46** (2003)
2. Nokia: Quality of experience (QoE) of mobile services: Can it be measured and improved? White paper (2006)
3. Spirent Communications: Delivering Optimal Quality of Experience (QoE) for IPTV Success. White paper (2006)
4. Rubino, G.: Quantifying the Quality of Audio and Video Transmissions over the Internet: the PSQA Approach. In: Design and Operations of Communication Networks: A Review of Wired and Wireless Modelling and Management Challenges. Imperial College Press (2005)
5. ITU-T Recommendation P.800.1: Mean Opinion Score (MOS) Terminology (2002)
6. ITU-T Recommendation G107: The E-model, A Computational Model for Use in Transmission Planning (1998)
7. Markopoulou, A., Tobagi, F., Karam, M.: Assessing the Quality of Voice Communications Over Internet Backbones. IEEE/ACM Transactions on Networking **11** (2003)
8. Hoßfeld, T., Tran-Gia, P., Fiedler, M.: Quantification of Quality of Experience for Edge-Based Applications. In: ITC20, Ottawa, Canada (2007)
9. Emmert, B., Binzenhöfer, A., Schlosser, D., Weiß, M.: Source Traffic Characterization for Thin Client Based Office Applications. In: 13th EUNICE, University of Twente, the Netherlands (2007)
10. Schlosser, D., Binzenhöfer, A., Staehle, B.: Performance Comparison of Windows-based Thin-Client Architectures. In: ATNAC'07, Christchurch, New Zealand (2007)
11. The Tolly Group: MetaFrame XP Presentation Server Windows-based Application Access Performance and Functionality (2003)
12. Tolia, N., Andersen, D.G., Satyanarayanan, M.: Quantifying Interactive User Experience on Thin Clients. IEEE Computer **39** (2006)
13. Carson, M., Santay, D.: NIST Net: A Linux-based Network Emulation Tool. SIGCOMM Computer Communications Review **33** (2003)
14. WinPcap. http://www.winpcap.org/ (last accessed: January 2008)
15. AutoHotkey. http://www.autohotkey.com/ (last accessed: January 2008)