There is no doubt that security mechanisms, such as authentication, are required in Information and Communication Technology, but they come at a price: Users need to spend additional time and effort to authenticate themselves. With this in mind, user perception of authentication is an important factor for successful use of authentication solutions. If users perceive an authentication procedure as time-consuming and difficult, they might ignore or try to bypass it. Therefore, user-perceived Quality of Experience (QoE) should be investigated. QoE is a challenging area as it, in this case, covers network performance and security as well as Human Computer Interaction and user experience.

Throughout this work, authentication performance is investigated, starting with a framework for evaluating security architectures and authentication solutions in general. Criteria for user-friendliness, security and simplicity are described and the evaluation methods span from theoretical to practical, and qualitative to quantitative methods. The latter two aspects are investigated by a study of user experience of web authentication with OpenID using the EAP-SIM authentication method. The user experiments resulted in several user models of QoE. One particular user model for QoE, the exponential relationship between QoE and network level performance, was then used in further experiments on performance evaluation of OpenID authentication using EAP-SIM. The latter was done to determine the decisive factors for QoE of the authentication method in use. The results from these experiments show that the combination of OpenID and EAP-SIM for authentication over a secure tunnel is not appropriate to use over networks with high delays. The latter implies the need for improvements of the authentication procedure of OpenID using EAP-SIM, which should be addressed in the future. The user model of QoE obtained in this study will even help to quantify the performance aspects of future authentication procedures.
USER PERCEPTION AND PERFORMANCE OF AUTHENTICATION PROCEDURES

CHARLOTT LORENTZEN

APRIL 2011
SCHOOL OF COMPUTING,
BLEKINGE INSTITUTE OF TECHNOLOGY
For my Family
and my dear Friends
Abstract

There is no doubt that security mechanisms, such as authentication, are required in Information and Communication Technology, but they come at a price: Users need to spend additional time and effort to authenticate themselves. With this in mind, user perception of authentication is an important factor for successful use of authentication solutions. If users perceive an authentication procedure as time-consuming and difficult, they might ignore or try to bypass it. Therefore, user-perceived Quality of Experience (QoE) should be investigated. QoE is a challenging area as it, in this case, covers network performance and security as well as Human Computer Interaction and user experience.

Throughout this work, authentication performance is investigated, starting with a framework for evaluating security architectures and authentication solutions in general. Criteria for user-friendliness, security and simplicity are described and the evaluation methods span from theoretical to practical, and qualitative to quantitative methods. The latter two aspects are investigated by a study of user experience of web authentication with OpenID using the EAP-SIM authentication method. The user experiments resulted in several user models of QoE. One particular user model for QoE, the exponential relationship between QoE and network level performance, was then used in further experiments on performance evaluation of OpenID authentication using EAP-SIM. The latter was done to determine the decisive factors for QoE of the authentication method in use. The results from these experiments show that the combination of OpenID and EAP-SIM for authentication over a secure tunnel is not appropriate to use over networks with high delays. The latter implies the need for improvements of the authentication procedure of OpenID using EAP-SIM, which should be
addressed in the future. The user model of QoE obtained in this study will even help to quantify the performance aspects of future authentication procedures.
Acknowledgements

• It is with deep gratitude that I would like to thank Markus Fiedler for his mentorship and supervision during my first years of PhD studies, and for his patience, support and inspiration while working together.

• Further, I would like to thank Henric Johnson och Patrik Arlos for their supervision and support during my first years of PhD studies.

• Thanks also to David Erman for his concrete and valuable criticism and massive support, to the Security Group for fruitful discussions on security, and to my colleagues and fellow PhD students for discussions and ideas.

• Special thanks to Lotta Runesson for always supporting me at work, and helping me with anything that ever needed the slightest attention.

• My family, Mom, Dad and my husband, Peter, thank you for always believing I can achieve anything and for supporting me endlessly.

• Finally, I would like to thank my friends for putting up with me, despite the lack of mental presence during the whole period of writing this thesis.

Charlott Lorentzen
Karlskrona, April 2011
## Contents

<table>
<thead>
<tr>
<th>1 INTRODUCTION</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Motivation</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Background</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Contributions</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Thesis Outline</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 AUTHENTICATION METHODS</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 OpenID architecture</td>
<td>7</td>
</tr>
<tr>
<td>2.2 EAP-SIM authentication</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Alternative authentication solution</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 QUALITY OF EXPERIENCE</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 From Quality of Service to Quality of Experience</td>
<td>11</td>
</tr>
<tr>
<td>3.2 On Quantifying QoE</td>
<td>12</td>
</tr>
<tr>
<td>3.3 Measuring Quality of Experience</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 PAPER SUMMARY</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Paper I</td>
<td>17</td>
</tr>
<tr>
<td>4.2 Paper II</td>
<td>17</td>
</tr>
</tbody>
</table>
I A criteria-based evaluation framework for authentication schemes in IMS 21

II On User Perception of Web Login – A Study on QoE in the Context of Security 39

III Decisive Factors for Quality of Experience of OpenID Authentication Using EAP-SIM 61

5 Conclusions & Outlook 81

Bibliography 85
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Paper</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Methodology of authentication scheme evaluation.</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>EAP-SIM authentication procedure.</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>Two ways of presenting QoE results.</td>
<td>15</td>
</tr>
</tbody>
</table>

**Paper I**

1. Framework of evaluation criteria. 27
2. Methodology of authentication scheme evaluation. 33

**Paper II**

1. Average RTs mapped to the corresponding delay added in shaper. 44
2. Setup for confirm and login experiments. 47
3. Methodology for the study of authentication schemes. 49
4. OS for corresponding RT for confirm test with different trendlines. 51
5. OS for corresponding RT for login test with different trendlines. 53
6. User perception profile: Frequency of OSs from RTs for different BWs. 56
Paper III

1  Setup of the Ubisafe OpenID authentication, with BTH client side. 67
2  SC for EAP-SIM authentication with OpenID. . . . . . . . . . . . 68
3  RT components versus bi-directional delay added in the shaper. . 72
4  Frequency of $T_{N1}$, for a bi-directional delay $d = 1$ s, bin width 2 ms. 74
5  QoE in terms of MOS versus added delays, using Equation 8. . . . 76
# List of Tables

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Interpretation of MOS grades.</td>
<td>14</td>
</tr>
</tbody>
</table>

**Paper II**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Interpretation of MOS grades.</td>
</tr>
<tr>
<td>2</td>
<td>Trendline equations with respective coefficient of determination ($R^2$).</td>
</tr>
</tbody>
</table>
List of Publications

The papers are ordered chronologically in each section. The name of the author was, until May 2010, Charlott Eliasson, then she changed her family name to Lorentzen.

Included papers


Further papers


Chapter 1

Introduction

Users of the Internet are exposed to an increasing number of security threats, whether the users are surfing the web, using a service, such as instant messaging or VoIP, or sending e-mail. Solutions to protect users are developed, but users are rarely asked what they want. Of course users want to be safe, but since many security solutions do not make sense to the user, are too intruding, or take too much time, users tend to bypass them if possible. It is not until an intrusion occurs that the user suddenly considers it to be a disaster, if she discovers the intrusion at all. So, users need to be protected, but preferably with subtlety. Though, there is a lack of research in the area of user perception of security services and mechanisms.

The areas of Quality of Experience (QoE) and Security are not by themselves new, but combined they form a rather unexplored area. Since security is an area that keeps growing in importance, and since authentication is increasingly used, e.g., on the web just to mention one domain, this area is quite challenging. If increasingly many people are using a service, it should be known how they perceive it. If it is known how a service is perceived, the provider of the service can, at least, try to keep within the limits of what the user thinks is good performance, but this can be a challenge [11]. It is known [12] that users tend to leave services that continuously perform badly, without even complaining
to the corresponding support first. It is also known [12] that; in average, an unsatisfied user tells 13 people about the bad service she has experienced; for every user that calls customer service with a problem there are 29 more users with problems who will never call; and 90% of users will not complain before leaving a service. This implies that gaining, or even keeping, customers is a lot more difficult if there are many unsatisfied users, or many users that have left the service all together.

Performance of security is not a new area either, and there have been much work done on security solutions using simulations, but there are not much work done based on measurements, and even less that is based on measurements done in real environments. Though, work based on measurements done in laboratories can serve as points of reference.

There are studies done regarding what is called Sense of Security [13–16], which is similar to QoE of security. Amongst others, a study by Murayama et al [17] shows a model for evaluating Sense of Security, or as they also call it, Anshin (Japanese for ease of mind). In the study, the cognitive factors “information on providers, services and systems”, “security technology” and “quality of user-interface” are taken as arguments for the functions “trust”, “knowledge” and “intuitive”, respectively. These functions are then weighted with experience information and a parameter that says which of the three should have the greatest impact. Finally the three functions are added together to get the emotion value $A$ (Anshin) for the Sense of Security, spanning from 1 to 10.

The area of this work, including the area of QoE, and security, poses several questions.

a) How do users actually perceive security?

b) Which are the limits for patience, regarding security mechanisms?

c) Can a determined relation between QoE and security mechanisms be useful?
1.1 **MOTIVATION**

How users perceive authentication would have to be studied with user experiments as basis. These would then be compared to existing user experiments done for user experience, e.g., of visiting web pages. The results of the user experiments of authentication would then show that users are; more patient; less patient; or equally patient; with authentication as compared to a regular web page.

A determined relation between QoE and Security, for example user experience of an authentication, is expected to result in a user model, which can be used for performance testing of the authentication in question. As long as the scenario and the environment are the same for the user, the user model can be used even if several parameters are changed behind the user interface. Performance evaluations can then be performed using different Quality of Service (QoS) parameters, and also different authentication mechanisms.

1.2 **BACKGROUND**

The work presented in this licentiate thesis started within the EUREKA!-funded project Mobicome, and more specifically the project package of Authentication and Identity management, where it focused on finding a way of evaluating authentication mechanisms. Such an evaluation framework (see paper [1]) was developed for the project, together with Ubisafe AS, Oslo, Norway. Evaluation frameworks for security solutions were hard to find when that part of the work started. The framework developed in the course of this thesis work consists of different existing evaluation methods and evaluation criteria, spanning from theoretical and qualitative methods, such as TOWS analysis or SWOT [18] as it was initially called, to practical and quantitative methods, such as user experiments and measurements [19] in real environments, see Figure 1.1. In this figure the evaluation methods span from qualitative on the right hand side to quantitative on the left hand side. From the top down, the evaluation framework progresses from choosing evaluation method, to performing the actual evalua-
tion, and finally visualising the results of the evaluation. The two right-most evaluation methods are used in the second [2] and third [3] paper, respectively.

After the framework for evaluating authentication mechanisms was developed, a suitable mechanism was to be tested together with UbiSafe AS for the IMS platform of the project. The mechanism and the whole solution were then tested from user experience point of view. Regarding QoE and user perception, there are several studies done on other aspects, such as web browsing [20] and download times [21]. Many of these studies result in an exponential or similar relation between user ratings and decreasing quality of the delivered service, i.e. increasing response times. This part of the thesis work gave a major contribution to similar work in the form of an exponential user model for QoE of web authentication with OpenID, which is adaptable to similar scenarios, i.e., most general web authentication (see paper [2]).

The exponential user model for QoE of web authentication could then be used for the results of a performance evaluation of OpenID authentication using the Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAP-SIM), see paper [3]. The study visualises response times of the different parts of the OpenID authentication using EAP-SIM, and points out the decisive factors for QoE. The user model is used together with response times of the decisive factors in order to see the order of magnitude of impact
1.3. CONTRIBUTIONS

from the response time on the QoE.

1.3 CONTRIBUTIONS

The main contributions in this thesis are summarised below.

- A framework for evaluating security architectures, see paper [1].
- QoE user model for web authentication with OpenID, which is adaptable to similar scenarios, i.e. most general web authentication, see paper [2].
- Performance evaluation of OpenID authentication using EAP-SIM, which reveals limits for network delay that should be kept in order to keep a sufficiently high QoE, see paper [3].

1.4 THESIS OUTLINE

The structure of the thesis is as follows. First, the authentication mechanisms used throughout the thesis are presented in Chapter 2. An introduction to QoE is given in Chapter 3, including a discussion of how to measure QoE and of problems related to the latter. When the technical background is covered, Chapter 4 provides a summary of the included and related papers. The thesis is concluded in Chapter 5 and future work is outlined. Finally, the three papers that constitute this thesis are presented.
Chapter 2

Authentication Methods

Authentication methods and architectures used throughout the thesis work are described in this chapter. In the last part of this chapter (Section 2.3) an alternative authentication method is briefly presented.

2.1 OpenID Architecture

OpenID [31] is an architecture for simplifying password management. Users log on to an OpenID server and can then visit any page that has OpenID enabled, without providing new credentials for authentication.

Users are commonly identified at other servers with a public OpenID username, consisting of the OpenID server URL and the username for the OpenID server, e.g., openid.ubisafe.no/dh, where dh is the internal username of the user in the OpenID server.

Though, if the authentication is done with a SIM-card, as is the case in this thesis, the user needs to know neither her username nor her password (PIN-code not used in this scenario). The user is identified via the SIM-card and the keys in the SIM-card are used to authenticate the user. As long as the user has her SIM-card she does not have to remember usernames and passwords to be able
CHAPTER 2. AUTHENTICATION METHODS

to authenticate oneself to any OpenID enabled pages on the Internet.

2.2 EAP-SIM AUTHENTICATION

The Extensible Authentication Protocol Method for GSM Subscriber Identity Modules, EAP-SIM, uses the SIM credentials of an actual SIM-card for authentication. If the SIM-card should be used for both EAP-SIM and for GSM/GPRS, the EAP-SIM security level will be the same as for GSM security mechanisms. The latter scenario could be desirable when an operator provides a security or authentication solution using EAP-SIM. What could compromise the security of EAP-SIM is compromised secrecy of the SIM triplets.

2.2.1 EAP-SIM messaging

The messages sent between client, or supplicant, and authenticator are visualised in Figure 2.1 [32] and are briefly described below.

**EAP-Request/Identity**: Authenticator requesting the identity of the supplicant.

**EAP-Response/Identity**: Supplicant responding with the identity, i.e., the International Mobile Subscriber Identity (IMSI).

**EAP-Request/SIM/Start**: Contains a list of supported versions of EAP-SIM.

**EAP-Response/SIM/Start**: Supplicant selects version and responds.

**EAP-Request/SIM/Challenge**: Authenticator sends a challenge and a message authentication code attribute (AT_MAC).

**EAP-Response/SIM/Challenge**: GSM authentication algorithm is run, and a copy of the AT_MAC is calculated. If the AT_MACs are equal, i.e., correct, a challenge response is sent to the authenticator.
2.2. EAP-SIM AUTHENTICATION

**Figure 2.1: EAP-SIM authentication procedure.**

**EAP-Success:** If the challenge response is correct, the supplicant is authenticated and the authenticator sends a success message to the peer.

Within this message exchange, there are mechanisms that help both the authenticator and the supplicant to assure that certain messages are not compromised. Protection of the version negotiation is done by including the list of supported version, which is attained from the authenticator, when calculating the keying material. Then the authenticator uses the same credentials to make sure that the selected version from the client is the accurate one and that it has not been compromised on the way to the authenticator. Verification of correctness is also done when the AT_MAC calculated by the supplicant is compared to the one that is sent by the authenticator. If they do not match, the supplicant sends a message called EAP-Response/SIM/Client-Error, which terminates the authentication attempt. The peer also chooses and sends a random number in the EAP-Response/SIM/Start message. Since this random number is used to calculate the AT_MAC, the peer can ensure that EAP-SIM messages sent by the authenticator are fresh, i.e. not replayed.
CHAPTER 2. AUTHENTICATION METHODS

2.3 ALTERNATIVE AUTHENTICATION SOLUTION

When using a smartcard or a SIM-card the user does not choose her own secret for authentication towards the server. It is a security issue when users choose their own secrets, since they tend to choose simple secrets that they can relate to, e.g., the name of a child or a pet. Hence, the latter security issue is not present when authenticating with a smartcard or SIM-card.

A newer version of the EAP-SIM, namely Improved Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA’) [33], may be used, given that 3G network infrastructures and 3G identity modules are also available. The Identity is then the user’s Network Access Identifier (NAI), instead of the IMSI, and the message exchange does not include the EAP-SIM specific SIM/Start messages.
Chapter 3

Quality of Experience

The term QoE is slightly differently interpreted, depending on the area of research \[11, 34\]. In this work QoE refers to user experience, the quality of a particular service as an end-user experiences it, based on the end-to-end response time. In this chapter we will take the detour via QoS to QoE, and clarify the meaning of QoE and the relationship between QoE and QoS.

3.1 From Quality of Service to Quality of Experience

QoS is a known concept within the networking and telecommunication area. QoS has amongst other been defined as “something a flow of data seeks to attain” \[22\]. QoS parameters such as delay, loss and jitter can have great impact on QoS for services such as VoIP, streaming video, etc. Furthermore, loading web pages can be affected by such QoS parameters, when the response times become large.

The latter leads us into QoE, which is an upcoming concept that increasingly many telecommunication researchers are considering in their research. High re-
CHAPTER 3. QUALITY OF EXPERIENCE

response times, e.g., when loading a web page, gives a bad QoE [23–27]. Early studies [27, 28] showed that for response times of around 1 s users start noticing that there is a delay, at around 4 s users start becoming impatient and at around 10 s there is a highly increased risk that the user will leave the web page/service/etc., this time or even for good.

When comparing QoE and QoS, one can say that QoS is the quality of the traffic delivery by the network, while QoE is the quality of the experience that the user has. One parameter from QoS is response time, which is the total time for a service to deliver all data starting from the request of a user.

QoS parameters such as jitter can be high, i.e., bad, without the QoE being affected. The other way around; QoS can be good, and the QoE can still be affected, e.g., by problems on the client machine. There are QoS technologies that can help improve the traffic throughput. There are also techniques to achieve high QoE despite of QoS issues such as jitter, e.g., by having a buffer when showing a video.

3.2 ON QUANTIFYING QOE

Quantifying QoE means translating user perception into numerical and interpretable values. To quantify users’ experience of a service or a network is important for providers of such, in order to know when the QoS is too low. The latter can be achieved by mapping QoE to QoS parameters, which is possible once QoE has been quantified.

Users consider their experience based on an almost indefinite number of unconscious parameters, such as habits, origin, mood, expectations, etc. These parameters are difficult to measure without extensive studies and/or advanced technology, such as eye tracking, heart monitoring, etc. User models in this work are therefore derived from user ratings, which are answers on questions about perceived quality. The aim of a user model is to visualise the main part of the users’ experiences. The user model is then a representation of a mean user, based on all users’ experiences.
Numerical values for QoE, or user models, make it easier to compare QoE to the parameters of QoS. E.g., a particular high delay in a certain network results in a specific, and probably bad, value of QoE for using the network.

3.3 MEASURING QUALITY OF EXPERIENCE

There must be a user group of sufficient size in an experiment in order for the resulting answers and ratings to be representative, and thereby valid. In this work a user group is defined as a group of users participating in an experiment. The size of the user group is defined by the variations of the results in each experiment scenario. If the experiment aims to model the behaviour of a certain type of users, the characteristic requirements must be fulfilled by all participating users.

When measuring QoE one poses questions to users about the service in mind. The questions must not be manipulative in any way and the questions should capture the essence of what the user is experiencing. If users are rating their experience the questions in the user experiment also need to be specific, in order to avoid misunderstandings of what the rating concerns. If the latter is not taken into consideration the results might be misleading or unusable.

3.3.1 MEAN OPINION SCORE

The ratings for the Mean Opinion Score (MOS) are found in the ITU-T Recommendation P.800 [29]. The ratings found in this recommendation are primarily used for evaluating user experience of telephone transmission quality, but have recently also been used for evaluating user experience in general, e.g., for web browsing [20].

The ratings for MOS goes from “Excellent” to “Bad”, as shown in Table 3.1. After all ratings have been given, the ratings are then translated to numbers (1 through 5) that can be used for calculations and statistics.
CHAPTER 3. QUALITY OF EXPERIENCE

Table 3.1: Interpretation of MOS grades [29].

<table>
<thead>
<tr>
<th>MOS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretation</td>
<td>Bad</td>
<td>Poor</td>
<td>Fair</td>
<td>Good</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

3.3.2 Problems in Measuring QoE

When measuring user experience, the environment is usually important. The user should be in a realistic environment and be faced with a realistic scenario. Even then, the user might not give the same rating as in a real situation. In a real environment, most users do not continuously give ratings, like the ones for MOS, to their experience. Users might dislike an experience with for example a connection speed, but they rarely put a five scale grading on it in an everyday situation.

Another problem with measuring QoE is diversity in user characteristics and past experience. All users have different backgrounds in some way, and all users have different moods on the occasion of the experiment. Past experience affects the ratings. If the user is used to waiting times of 20 s, then 10 s might be quick, but if the user is used to some 100 ms of waiting time, then 10 s might be unacceptably slow to that user.

The same user can give different ratings in exactly the same scenario at different occasions. If a user is presented with bad quality before medium quality, the medium quality gets a better rating than if medium quality is preceded by high quality, i.e., a memory effect occurs [30]. So, when a user is asked to rate several scenarios directly after each other, the order of the scenarios will affect the resulting ratings. If the scenarios are presented in different orders to different users, there will be a greater diversity in the resulting ratings, but then it might be considered as a “worst case” with regard to spread of values.

Based on the above, the major challenges with user experiments, where QoE should be measured, is first of all the setup of an environment that is realistic enough to provide ratings that are close enough to a real scenario experience.
3.3. MEASURING QUALITY OF EXPERIENCE

![Graphs showing QoE and QoS mapping](image)

(a) OS mapped to response time, with exponential regression.

(b) MOS mapped to additional delay.

**Figure 3.1:** *Two ways of presenting QoE results.*

In the experiments to be reported in this work the setup of the environment was done to be as realistic as possible.

### 3.3.3 MAPPING QoE AND QoS

After finding the ratings of users, each opinion score (OS) should be mapped to the corresponding response time, or the underlying delay. In the mapping the OS is the QoE rating, and the response time is the parameter impacted by various other QoS parameters, such as delay or loss. So, the relation here is that QoS problems may cause increased response time, which in turn may cause decreased QoE, as the users rate what they perceive and not single packet losses or delays, which may go unnoticed.

When a mapping is done, the result can easily be visualised in graphs. An example of this are given in Figure 3.1(a) where each OS is shown, as well as a regression curve, which is based on all the OSs. The result can also be visualised as QoE mapped directly to QoS as MOS mapped to delay, which is shown in Figure 3.1(b), or as OS mapped to average response time.
CHAPTER 3. QUALITY OF EXPERIENCE

3.3.4 Quality of 'Security Experience' aspect

When QoE or user perception is desired to obtain for security mechanisms or security based services, like authentication services, the challenges are a bit different. The challenge of getting the users in a realistic environment still exists, but there is now also the additional challenge of presenting a security service that the users care about like it was their own. The latter is a challenge because if the user does not care about the security service, the aspect of security might be lost or compromised. To bypass this potential issue, an experiment can be done on an authentication procedure that is well-known and can be replicated, or by doing experiments with the actual well-known authentication procedure, e.g., a login procedure of a community, which is done in this work.
This chapter gives summaries of the papers included in this thesis [1–3], and short summaries of the published papers related to the thesis work.

4.1 Paper I – A criteria-based evaluation framework for authentication schemes in IMS

A framework for evaluating security with existing evaluation methods was presented in the paper [1], spanning from qualitative and theoretical to quantitative and practical methods, and ending up in a summarised evaluation based on several presented criteria and finally visualisation of the results.

4.2 Paper II – On User Perception of Web Login - A Study on QoE in the Context of Security

User experiments on the authentication architecture OpenID and the EAP-SIM authentication method were performed in this study [2]. A client was situated in Karlskrona, Sweden, and the authentication server in Olso, Norway. Users logged in on a web page, the network delay was varied for each trial, and the
users gave ratings for QoE of the log in for each new delay setting. The ratings were then mapped to the corresponding response time and regression curve were found for the relation of QoE and network performance. The relations were presented as user models for web authentication, based on the regression lines.

4.3 **Paper III — Decisive Factors for Quality of Experience of OpenID Authentication Using EAP-SIM**

In this study [3] the performance of the authentication architecture OpenID together with the authentication method EAP-SIM was tested. The network delay was increased in a shaper, from 0 ms to 1000 ms, in steps of 250 ms. Each task of the EAP-SIM method was isolated and the performance of each task was evaluated, in order to find the decisive factor for QoE. The tasks that were most sensitive to network performance degradation give the greatest impact on QoE, and additive changes of response time gives multiplicative changes of QoE. The order of magnitude of QoE decrease was shown for certain network performance degradations.

4.4 **Related papers**

This section contains short summaries of papers that are related to the thesis, but not included.

4.4.1 **Quality of Experience and Quality of Service in a Service Supply Chain**

A study with user experiments of map service quality was presented in the related paper [4]. This paper covers the network part of the study. The service
4.4. RELATED PAPERS

supply chain for the map service is presented and the impact of the latter is discussed.

4.4.2 Mapping service quality - comparing quality of experience and quality of service for Internet-based map services

This paper [5] is a product of the same study as the previous paper. In this paper it was seen that previous experience of a user highly impacted the user patience with the used map service.

4.4.3 Evaluation of Authentication Schemes Based on Security, User-Friendliness and Complexity

This paper [7] presents the pre-study of the first included paper [1]. The primary evaluation criteria security, user-friendliness and complexity are analysed and discussed, and the criteria that are found between them are presented.

4.4.4 Releasing the potential of OpenID & SIM

The use of OpenID and SIM in web services is discussed and scenarios of such usage are presented in this related paper [9]. The scenario of authenticating and taking ones authentication to other web sites is the basis for the user experiments in the second paper [2].

4.4.5 Security and Privacy Issues for the Network of the Future

This related paper [10] presents and summarises the security and privacy issues in the Internet of today and networks of the future. Problems that will be brought from today’s networks to networks of the future, as well as new
problems, are discussed. The work of this thesis constitutes one section of the paper.
A CRITERIA-BASED EVALUATION FRAMEWORK FOR AUTHENTICATION SCHEMES IN IMS

IN PROCEEDINGS OF
INTERNATIONAL CONFERENCE ON AVAILABILITY, RELIABILITY & SECURITY
INTERNATIONAL WORKSHOP ON SECURITY AND USABILITY
FUKUOKA, JAPAN, MARCH 2009
A criteria-based evaluation framework for authentication schemes in IMS

Charlott Eliasson, Markus Fiedler
Blekinge Institute of Technology, Karlskrona, Sweden
(charlott.eliasson, markus.fiedler) @bth.se
Ivar Jørstad
Ubisafe AS, Oslo, Norway, ivar@ubisafe.no

Abstract – The IP Multimedia Subsystem (IMS) is regarded as one of the most prominent enablers for successful service provisioning across different access network technologies and devices. While new paradigms, e.g. seamless communication, enter the IMS, existing solutions, e.g. for authentication, need to be redefined, which is one of the major activities within the EUREKA!-funded Mobicome project, involving operators, manufacturers and academia. As there exist several candidate solutions for providing seamless authentication, there is a need for a set of criteria that helps to select the candidate that fulfils those criteria in a bestpossible way.

Given this background, this position paper proposes a framework of criteria for the evaluation of authentication schemes in IMS. The primary criteria are security, user-friendliness and simplicity. Inbetween these criteria, the secondary criteria can be found. These are awareness, usability and algorithms. Each criterion, both primary and secondary, is then also divided into one or several substantiating sub-criteria. The discussion of the criteria is followed by a description of the evaluation methodology, which comprises both qualitative and quantitative evaluations such as SWOT analysis, use of NIST and ISO guidelines, user rankings, and measurements of authentication times. The paper is concluded with an outlook on future work, including studies and experiments.
1 Introduction

In the EUREKA!-funded Mobicome project [1] eleven partners are participating in order to specify, implement and evaluate prototypes of an improved version of the next generation network IP Multimedia Subsystem (IMS). The major idea is to extend IMS with better functionality for supporting seamlessness, both regarding network and device switching. Seamlessly working security solutions are a selfevident must in this context. In order to satisfy the customer while meeting technical boundary conditions, solutions must work in a satisfactory manner in many respects. As part of the Mobicome project we are going to study and evaluate several SIM-based and non SIM-based authentication schemes for use in an IMS platform.

The first steps towards improving the IMS authentication procedure will be done by experimenting on and evaluating a web-based solution. This will be a server-client based system, where several components are working together in order to emulate an IMS environment. It was pre-evaluated at the DreamExpo, during Dreamhack Winter 2008 [2] in Sweden. Dreamhack Winter is the largest computer festival in the world and has been so since 2004. The input from this event gave the impression that users do not have a different attitude towards waiting for a webpage containing authentication (login) than for a webpage that does not. The major problems seem to be related to scalability, meaning that users have so many accounts, for example on different communities, that it is not feasible for them to care about security anymore.

Consequently, many users’ attitude towards security in digital information and communication services can be considered jaded, eventually reaching a level of ignorance, but security is still needed. A user that for example is hacked will start care there and then. This would conclude that a security solution, ultimately, does not make life difficult for the user, but still protects him/her.

In the first stages of this evaluation, we will therefore investigate whether the above observations, made at [2], gave an accurate impression or not. Though, the largest and most important part of this evaluation targets the optimal solution for authenticating users in an IMS.
On this background, this paper is a position paper, stating started and planned work on the evaluation of candidate security solutions for seamless IMS-based communications. First, the primary criteria of evaluation, namely security, userfriendliness and simplicity, are discussed. Then, the secondary criteria, which include awareness, usability and algorithms, are described. Along with the criteria, both sub-criteria and corresponding parameters are outlined. In case there are any well-known best-practices, these are also mentioned. After this, the methodology for the evaluation is described. Finally, an outlook is presented at the end of the paper.

2 Target solutions

Within the Mobicome project, the evaluation [3] is focused on a set of well-known authentication schemes. Furthermore, the IMS does not place any special requirement on the authentication scheme per se, but most likely on the environment and the handling of authentications.

2.1 SIM-based solutions

The current SIM-based authentication scheme candidates for the evaluation are listed below:

- Early IMS might be used in IMS implementations that do not yet support all “Full IMS” requirements.

- ISIM: An IP Multimedia Services Identity Module (ISIM) is an application running on a Universal Integrated Circuit Card (UICC) smart card in a 3G mobile telephone in IMS.

- USIM: A Universal Subscriber Identity Module (USIM) is an application for UMTS mobile telephony, also running on a UICC smart card.
2.2 Non-SIM-based solutions

The current non-SIM-based authentication scheme candidates for the evaluation are as follows:

- Username/password is used in the HTTP 1.0 Basic Authentication, where a client/web browser provides credentials through the HTTP request.

- Smartcard: The smart card consists of integrated circuits (ICCs) embedded in a pocket-sized card. A smartcard may include a secure crypto processor, a secure file system and the capacity to provide security services like confidentiality of information in the memory.

- SMS-based authentication: A SMS-based authentication scheme involves providing the user with a secret key in an SMS, carried through the mobile network.

- MAC address authentication uses the quasi-uniqueness in the link-layer address

3 Evaluation criteria

We started out with the criteria that are now called primary criteria, i.e. security, user-friendliness and simplicity. After further investigation and discussion the interesting so-called secondary criteria were found at the intersections between the primary criteria. The framework of these criteria is illustrated in Fig. 1.

3.1 Primary criteria

3.1.1 Security

This document defines security in the context of IMS authentication as the level of security that is obtained for the user and the system when using a certain
authentication scheme. For security regarding authentication the following sub-criteria will be considered.

- Authentication level is a criterion taken from the National Institute for Standards and Technology (NIST) [4] Electronic Authentication Guideline [5]. This criterion states that the highest level will be achieved by three-factor authentication, combining “something you know”, “something you have” and “something you are”. Though, this is not an optimal solution for Mobicome and a two-factor authentication that combines “something you have” with “something you know” is regarded as the highest level of authentication in this evaluation.

- (Automatic) Trust with possible timeout and reauthentication: When the user equipment (UE) has been authenticated, the UE will remain authenticated for some time until a timeout occurs or until the UE disconnects or the session is broken. If a UE remains authenticated until the session is broken, there is a greater risk for malicious use than if a UE needs to continuously authenticate itself. The more often it reauthenticates it-
self, the higher the security. There are of course disadvantages of a high frequency of reauthentications, such as lowered performance or higher response times, but from the point of view of security it is an advantage.

- Known attacks: If an attack is possible, then it is just as bad whether it is known or not. Though, if it is known it is a good choice not to use that particular authentication scheme, or to find a solution to make the attack impossible or useless for the attacker. If an authentication scheme has a known possible attack and no solution of how to shield itself, then the level of security is compromised.

### 3.1.2 User-friendliness

We define the user-friendliness as how probable it is that a typical user is able to authenticate without extra help or guidance. Furthermore, the user-perceived quality [6] or Quality of Experience (QoE) [7] for a user during the (TISPAN [8]) authentication is also a measure of userfriendliness. The user-friendliness might be low if the authentication takes too long or if the user does not fully understand what (s)he is doing or how to do it. The latter is closely related to usability. The following sub-criteria will be considered.

- End-user experience: If a tool has been built without considering so-called dumb users, then it will probably not be user-friendly. In the case of an authentication scheme it is of utter relevance that a user can use the user interface (UI) on the UE. Otherwise the user will not receive the desired service. The end-user experience is highly affected by the response time [6, 7] and is typically quantified by the Mean Opinion Score (MOS) [9]. While first used in the context of telephony, MOS is increasingly determined for subjective experiments for any type of application or service involving data communications.

- Authentication time: If a scheme takes too much time to authenticate a UE, then the user will probably think that there is something wrong with the service and/or the UE or that someone has tampered with the UE or
the service account. The authentication time is perceived by the user as a part of the response time. Already after one second the user’s flow of thoughts are broken, and after four seconds (s)he might get impatient and develop negative thoughts about the service [7]. If it takes even longer, e.g. ten seconds, the user might lose patience and leave the service [6], this time or even forever.

- Password difficulty: Preferably, the user should not have to worry about remembering a long password or transferring a password from SMS (or similar) to another source to be able to be “secure enough”. The authentication scheme should be able to provide security with user-friendly means.

- Functionality: The UI of the authentication scheme should have all the functionality that is needed, but still be easy to use. A very broad selection of functionality can even be counter-productive in terms of usage if a user for example does not find a desired function in the jungle of functions. For a UI of an authentication scheme this is very obvious. It does not have to have a lot of functionality, and if it does the user will probably get confused.

3.1.3 Simplicity

In the context of what the authentication scheme adds to the system, the authentication solution should be as simple as possible and still be sufficient as an authentication scheme. If the latter adds complexity to the system, the level of simplicity decreases. Simplicity is also closely related to scalability, in terms of effort and overhead. The following sub-criteria are considered:

- Execution time/speed: The time it takes the scheme to authenticate the user in the system. If the authentication, for example, involves several steps it will probably take longer time to authenticate, as in each step there is a risk of additional delay. The latter is called a service supply chain [10], and ultimately the time of each step is measured and compared in order to find the most critical link [10, 11].
• Performance impact on system: A simple and straightforward authentication scheme with a sufficiently long key will most probably not affect the performance of the system. The longer and the more complex the scheme, the greater the impact on performance will be. Such impacts can come from delay, loss, etc. initiated by the size of the data sent by the authentication scheme, by the process time of the scheme, etc.

• Performance impact on UE: If a process is computationally heavy or if the authentication runs several parallel processes, it might affect the performance of the UE.

3.2 Secondary criteria

These criteria can be found in the intersections of the primary criteria, as outlined in Fig. 1.

3.2.1 Awareness

Awareness of the security service can be found as a criterion in the intersection between user-friendliness and security (see Fig. 1). The user can be aware of a service or an application both in a good and a bad way. The good way can in the case of authentication be that the user feels secure because of something that made him or her aware of being correctly authenticated. This can of course be a false feeling, if the user feels secure beyond the actual authentication level, or if the feedback says “authenticated” when the user is actually not authenticated. The latter is called false positive feedback and is a most critical error. There can also be false negative feedback, which is not as critical, as this will not harm the user or his integrity. However, it will hinder the user from performing the desired task. Bad awareness can be caused by permanent, annoying so called positive feedback. If a user gets annoyed or irritated by the provided feedback, then (s)he might not want to use the application or service, which will of course be a failure for the provider. The user might instead turn off the feedback, which in turn might lead to misuse or denial of service for the user. Based on
this, the following sub-criteria will be considered.

- **Positive awareness:** The user feels secure and does not have any doubts about the authentication scheme working properly. If nothing is shown for example during login, how will the user then know that the login was performed properly or that the UE is not tampered with? Compare with SIM “login” where nothing is actually shown.

- **Negative awareness:** On the other hand, if the information towards the user is too heavy or the user is always aware of the authentication in a very obvious and attention-stealing way, then (s)he will be bothered by it and probably not want to have it there.

- **Understanding:** If the user understands what (s)he is doing (is aware of what (s)he is doing), then it is easier to feel secure. Ergo, if the authentication process is easy to understand, the user might feel more secure. Erroneous usage may lead to confused users and even malfunctioning and compromised security.

- **Feedback:** This is a widely discussed area. Should the user really get any positive feedback, or only negative? Regardless of that, if there should be feedback, there is still the question of in what form and on what level the feedback should be.

### 3.2.2 Usability

User-friendliness and simplicity together form the usability of a service (see Fig. 1). Usability is a concept that tells how well a user actually can use a service or application. It should be easy enough to be used by a typical user, but it should also have sufficient functionality. We define usability within this study as the ability for a typical user to use the scheme, based on how the scheme acts. This is closely connected to userfriendliness, but also to simplicity. If the scheme is complex and adds a huge execution time to the authentication process, then the user might not be able or willing to use it as intended. It also
increases the risk of authentication failure. The following sub-criteria, found in [12], will be considered.

- Effectiveness: If a user can successfully perform a desired task of a scheme, then the execution of the task is effective. It involves the outcome of the desired task.

- Efficiency: To reach a desired outcome of a task, there can be more or less effort and resource put into it. Depending on how much effort and resource that have to be used, the scheme is more or less efficient.

- Satisfaction: This criterion is comparable with QoE and more or less connected to user-friendliness.

3.2.3 Algorithms

The simplicity and the security of a service are both based on the algorithm (see Fig. 1). This criterion is related to how the algorithm handles the task and how well both comply with each other. Both algorithms and tasks are by themselves related to both security and simplicity. A complex algorithm should provide a higher level of security and a less complex algorithm (that might be needed in some cases) should provide a level of security that is as high as possible, given the level of complexity.

- Security-complexity ratio, which should be as high as possible, serves as the sole sub-criterion. An algorithm that is widely complex and still does not provide a higher level of security than some other less complex algorithm, will most likely not be used as much. A complex algorithm should provide a higher level of security and a less complex algorithm (that might be needed in some cases) should provide a level of security that is as high as possible, given that level of complexity. On the other hand, if an algorithm is very simple and still provides a high level of security, it will most likely be used wherever possible.
4 Methodology

In this section, the methodology of the evaluation is discussed. An outline of the methodology can be seen in Fig. 2, and each part will be explained in this chapter. To the left in Fig. 2, the rather qualitative methods can be found, and to the right, the rather quantitative methods are located. We will apply the criteria discussed in previous chapters.

4.1 SWOT

Regarding the evaluation, there will first be a theoretical and conceptual investigation of the previously mentioned authentication schemes using SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis. The SWOT analysis has mostly been used in the context of societal studies [13]. All known and documented aspects will be considered and the outcome will be a first judgement of each authentication scheme.
4.2 NIST & ISO guidelines

The authentication level will be evaluated conceptually, since it is based on what mechanisms are used in the scheme. The NIST [4] Electronic Guideline on Information Security [5] will be used as reference material, but the Common Criteria (CC) [14] will have the most significant role. A Protection Profile (PP) that suits IMS authentication and the functionality will have to be developed. It will also most probably have to be modified to fit these functionalities. The CC does not have any standards for evaluating cryptography. CC is commonly used as part of a certifying process of products.

4.3 User rankings

The aspects of user-friendliness, awareness and usability can be evaluated “on top of” the IMS environment with real users, where their reactions are considered as subjective results. The user reactions will be documented through observations and also by users giving their subjective judgements of performed tasks. The user judgements can be given according to the MOS-scale, from 1/worst to 5/best, which can then be compared to response times for the same tasks. Moreover, well-recognised results such as the ones reported in [7] can be taken into account.

For some criteria the users can also be thoroughly interviewed to get the required subjective result. The interviews will be done in a wider context in order to not tamper the results by only asking one kind of question connected to one criterion. Users will be asked to give their judgment of tasks with deliberately imposed variations of system behavior. The users can also be interviewed about their experience with regards to awareness and feedback. Furthermore, the users will be asked questions about what they thought of the scheme and how they felt about using it. These questions will be closely connected to functionality. Reference material can be the System Usability Scale (SUS) [12].
4.4 Measurements in a real IMS environment

The aspects of security, simplicity and algorithms can be evaluated in a real IMS environment. The latter can be complemented with instruments in order to obtain objective results through measurements. Wiretaps and measurement points can be installed in the IMS environment in order to perform and control the traffic measurements [11].

Security as well as scalability has several sub-criteria that can take their data from measurements in the IMS environment. The response times of the system can be measured before and after adding the authentication scheme. The scheme can then perform its tasks and the packet flow will be recorded via wiretaps and then analyzed to make differences visible. These times are expected to be different for different tasks, but also for different algorithms.

5 Conclusion and Future work

In this position paper we described the work and results of finding a framework of criteria for evaluating authentication schemes for an IMS environment. This work was done within the EUREKA!-sponsored Mobicome project. Within this framework, we distinguished between primary and secondary criteria, the latter of which are found in the intersection between the corresponding primary criteria. The evaluation methodology that is supposed to use these criteria was also outlined and discussed.

The next step that will be taken is to evaluate the authentication schemes is a web-based solution, whose authentication part will work more or less like an IMS environment. As outlined above, this system will have several wiretaps deployed in strategic locations. These will allow for observing and analyzing authentication messages and their occurrence in time, which will tell us how different parts of the system will respond to the different schemes.

As target authentication solutions will have been implemented, the above-described evaluation methodology and framework of criteria will be applied in
order to determine the best-suited solution. In particular, due to its generic set-up, the wiretap-based test-bed described above will be used for evaluating user-perceived performance related to the different solutions.

As work proceeds, we also anticipate the need for a refinement of the above criteria, with the addition and use of Common Criteria [14] and a suitable Protection Profile.

References


PAPER II

On User Perception of Web Login – A Study on QoE in the Context of Security

In proceedings of
Australasian Telecommunication Networks and Applications Conference
Auckland, New Zealand, November 2010
Abstract – As there is a lack of such studies, this paper investigates user perception (Quality of Experience, QoE) of the response times (Quality of Service, QoS) of a web authentication procedure, in particular a login to a community web page. Comparing the results to well-known user perception of web performance, we show that the users perceive logins in a similar way as standard web pages, which means that similar limits on user patience apply. The derived QoE-QoS relationship, an exponential function, serves then as the basis for assessing the performance of authentication algorithms in the domain of user acceptability.
1 Introduction

From the user point of view, security has shown to be ambivalent; on one hand it is felt necessary, but on the other hand it is felt disturbing. Authentication solutions, for instance, are designed to keep undesired and unauthorized users out; however, allowed users need to spend some effort and waiting time when logging into a system. The question remains to which extent this effort is perceived as positive as security is increased, or negative through the waiting time spent in the process. Excessive waiting times imply the risk of users leaving the service.

There are a few new authentication services coming up, of which OpenID [1] is one. This authentication is a solution that makes password management easy for the users. Only one username is required to log into several web pages once the users are authenticated to their OpenID provider. More of these security solutions are expected in the future, where ease of use is prioritized.

The demand of secure login schemes increases, and users’ willingness to accept a security service depends on the ability to meet the users’ expectations. For that reason, defining User Experience (UE) or Quality of Experience (QoE) [2] in combination with a security system is essential, both quantitative and qualitative. However, defining and measuring QoE is a challenge and involve interdisciplinary research. While user perception of regular web pages is pretty well known, perception of security enabled web pages, e.g. web pages requiring authentication, is hardly addressed in literature.

In this paper, which is part of a larger study on evaluating authentication methods for an IP Multimedia Subsystem platform [3,4] within the EUREKA!-funded Mobicome project [5], we evaluate QoE in the context of authentication. How do users perceive the Quality of Delivery (QoD) in the form of response time (RT) for security, and in particular for authentication? Responses to and rankings of perceived security are searched for. A comparison is then performed to previous studies in QoE of web pages. Is perception of security services, i.e. authentication, different from perception of other information technologies, i.e. web browsing? And if so, to which extent?
From the answers to these questions and from the performed experiments, we aim to find a model for user perception of security in web pages. This model will be presented in this paper, together with usage possibilities and constraints. The user model will, in the proceeding of the work presented in this paper, be used to further evaluate similar authentication methods for Internet services.

The organization of this paper is as follows: Section 2 clarifies technical terms that are used throughout the paper. Section 3 describes the methodology of the study and Section 4 describes the experiments with corresponding setup, procedure and RT measurements. The results (qualitative and quantitative) are presented in Section 5, followed by a discussion of the results in Section 6. Finally, Section 7 provides a conclusion of the paper and points out future work.

2 Technical background

2.1 OpenID

An OpenID identity is a unique URL which contains the trusted provider and the username. The provider is the host of the URL, in our case Ubisafe AS, in Norway, and the username/URL will be openid.ubisafe.no/<username>, but the provider can also be for example Yahoo, or any other site that provides OpenID as authentication service. With OpenID one can use the same password or authentication credential and username for every site that one has to authenticate oneself to.

If users are going to use OpenID, e.g. for www.facebook.com, then they first have to enable OpenID at that web page. Then, in the future, when the users log into openid.ubisafe.no (or the provider in question) it is possible to visit www.facebook.com and provide the OpenID username/URL, and the login to www.facebook.com will be completed without an additional password. This applies for all OpenID-enabled pages during one web browsing session.

OpenID was chosen according to the requirements of seamlessness and of the IMS platform of the Mobicome project [6].
2.2 Quality of Experience

The notion of QoE and UE have been widely discussed in [7], where the concept of QoE refers to the totality of end user experience of the delivered service [8].

QoE of a service is usually evaluated based on the time it takes for the service to perform a given task, i.e. the RT, which is a measure of the QoD. Quality of Service (QoS) problems (such as loss, delay and other variations) lead to increased RTs, i.e. QoD problems, which then translate into low QoE. Thus, the QoE measure is the user’s perception and judgement of the QoD, and the QoD is affected by QoS parameters.

User perception of web pages and perception of download times have also been evaluated and discussed as QoE in several studies such as [9,10], and [11].
Another definition of QoE is found in ITU-T P:10/G.100 recommendation [2]: “The overall acceptability of an application or service, as perceived subjectively by the end-user”. As described in the ITU-T P:862 recommendation [12], the QoE can be measured by user tests and is then typically expressed with a Mean Opinion Score (MOS). There exist other methods to estimate QoE such as the E-model [13], instrumental metrics (PESQ) [12], and even a neural network approach [14], but these are mostly used for voice and video.

We denote, in this paper, the user perception and user experience by the term QoE. For evaluating the latter we use MOS, with the preset grades in the ITU-T P:800 recommendation [15], seen in Table 1.

Studies have shown that users get less interested with the increasing RT of a service. A user browsing a web page notices a couple of 100 ms of delay, gets bored after about 4 seconds, and after 10 seconds the risk of the user leaving the page is high [7].

### 2.3 Supply chains

The total RT for an authentication often consist of times from several parts of the communication. There are several messages going back and forth in an authentication, e.g. challenges, each depending on the previous message being received and processed. Furthermore the packets will most likely be sent to a server, which in turn has to call another server or database to retrieve information to complete the request from the client. These kind of chains of messages, and/or chains of servers and databases are called Supply Chains (SC) [16]. These can be very large and also quite complex.

In [16] it can be concluded that SCs are susceptible to growth in delays, due to its nature. The longer and the more complex a SC is, the more susceptible it is to growth in delay. The effect of the SC can be seen in Fig. 1, showing the average RT for the login procedure in this study. When adding a delay of 1 s, the average RT is about 7 s, and when adding a delay of 2 s the RT is about 13 s. This SC mostly consists of a large chain of messages going to and from the SIM-dongle during the authentication procedure, while showing
only one RT to the user since it is all contained in the same task. Though, as
can be seen in Fig. 2, some messages in the SCs for this task can also have a
quite complex route, with authentication requests going all the way to the HLR
(Home Location Registry) at the network operator that issued the SIM card.

3 Methodology

This study uses measurements of RTs and user rankings of those, which is in-
cluded in the methodology (see Fig. 3) of the larger study, to quantify user expe-
rience. The experiments aimed to give quantitative results on visiting security-
enabled web pages of the considered system, including the login procedure.

A web login application will be used as a test case to find out the user
satisfaction, the QoE, with regards to the RT. For a web login, the QoE is mainly
affected by the delay of reply messages after a request has been submitted.

3.1 Quantitative

For each experiment the time for starting and completing a task was logged.
From these, a total time for completing each task could be calculated, which
is what we call the RT of the system for that particular task. The RT for a
particular task can be different each time the task is performed, since the total
RT depends on, not only the authentication method, but on the conditions of
the whole SC for that task.

MOS is a mean value of the Opinion Scores (OS) given by users. It has
been used in several studies before, amongst others to evaluate RT in a map
service [10], and to evaluate RTs of web pages [11] and download times [9].
Delay is put on the procedure in order to increase the RT. The users will then
rank that experience according to the MOS-scale (see Table 1), and the ranking
will be mapped to the RT of that particular trial.
Figure 2: Setup for confirm and login experiments.

Table 1: Interpretation of MOS grades [15].

<table>
<thead>
<tr>
<th>MOS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretation</td>
<td>Bad</td>
<td>Poor</td>
<td>Fair</td>
<td>Good</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

3.2 Qualitative

After the ranking, users were asked questions of more qualitative character of how they perceived the login procedure and the service, with regards to security. The questions were asked in order to find out the user perception of the service itself and also to underline the result of the user rankings.
4 Experiments

Two sets of experiments were performed in the study. The performance of the system is shown in Fig. 1, where it can be seen that the actual RTs for the respective delays were larger than the added delay. Sometimes single RTs were as high as 16 seconds. The first experiments considered a full OpenID login including the visited community web page before and after the users log into the OpenID server. The second experiment considered only the login procedure on the OpenID server web page, but the total procedure from the first experiment was still shown and explained to the users, to create an understanding of what they were testing.

4.1 Setup of confirm experiment

The first experiment setup consisted of a client computer with a SIM dongle, a traffic shaper for adding delay on the network interface of the client, and a server situated in Oslo, Norway, as shown in Fig. 2. All computers in the setup had a fixed line network connection.

The experiment setup in this study closely resembles a real-life setup, where server and client/user are situated on different geographical sites, as opposed to a setup in a regular laboratory environment where client and server are situated in the same room or in some cases even on the same computer. The only “artificial” part in this experiment setup is the traffic shaper, which is needed to provoke longer RTs. The delays that were added in the shaper were 0 ms, 100 ms, 200 ms, 300 ms, 500 ms, 1 s and 2 s.

In total there were 35 participants in the experiment, all students at Blekinge Institute of Technology, with varying experience of web services, and of ages between 22 and 30 years.

After telling the users about the procedure and OpenID, the users browsed to MyGeolog.com. There they chose OpenID as login method and entered the given user name. When clicking login, the users were then redirected to the web page of the OpenID server. Here “SIM-dongle” was chosen as authentication
method and when clicking “login”, the server and the SIM dongle exchanged information in order to authorize the users via the SIM card, and gave access to the original website. Then the users pressed “confirm” and were redirected to the original website.

This total procedure is done only once. Then, when visiting the community again, the users only need to choose OpenID as login procedure at the community and then confirm their identity.

After performing each test (i.e. one login procedure) the users were asked to rank their experiences. The tasks that were ranked are the confirm task together with the display of the MyGeolog.com page when the users are logged in. In other words, the last part of the authentication and the redirection, which is perceived as one task by the users, are ranked together. Even though the authentication has already been made at this stage, this is not apparent to the users since there is still the task of confirming one’s identity before the page of MyGeolog.com is shown. I.e. the latter is perceived by the users as the last part of the authentication.

A modified version of Fasterfox [17] was used in Firefox to get the RTs of the web pages. RTs were recorded and, according to the modification, logged to a file on the client. Each RT was then mapped to the opinion score that the users provided for that particular test.
4.2 Setup of login experiment

The second experiment setup consisted of a client with a SIM dongle, a traffic shaper for setting the bandwidth (BW) for the client, thus inducing delay, and a server, still situated in Oslo. The BWs used were maximal BW (which changed with the used type of network connection), 2000, 1000, 750, 500, 350, and 250 B/s. Bandwidths below 250 B/s gave a timeout and were not usable in the experiment. In this experiment we also used different non-fixed line network connections, such as wireless and mobile, together with the traffic shaper in order to force a greater variety of RTs. However, the RTs did not go much below 4 s because of the wireless or mobile network connection.

The tests were performed on the login procedure on the OpenID server web page. However, the users were shown how it should work when accessing, or using, it from another web page, such as MyGeolog.com, like in the confirm experiment. This was done to ensure that the users knew what they were doing and what the purpose of the task was. This time all the components of the authentication procedure were ranked together, as it is perceived as one RT by users. In this experiment the total response time consists of the authentication towards the OpenID server and the loading of the next page following instantaneously.

The RTs were recorded in Firefox in a similar fashion as in the confirm test setup. In this experiment the add-on Firebug Net Panel History Overlay [18] for the add-on Firebug [19] (the older version 1.2.0b6, which supports the history add-on) was used for recording the RT for web pages. Firebug gives detailed information on all times of a web page loading. Though, the authentication was done via a Java applet on the OpenID server web page before the actual web page started loading, and the times for java applet communication are neither recorded by Fasterfox nor by Firebug. This was then solved in the supplicant (see Fig. 2) with code that logged detailed timestamps at the client side to a log file, for all communication steps of the authentication.

Both the timestamps in Firebug and the timestamps in the log file were taken from the client computer, so the timestamps could then be compared to
make sure that they would give the correct total RT. The total RT for the login procedure, as the user perceived it, was then the time for the authentication added to the time for loading the web page.

5 Results

The user experiments gave some noteworthy results, both qualitative and quantitative, that are presented in this section. For the quantitative results of the login experiment we provide an equation for predicting a MOS from RTs in future evaluations of authentication solutions. This equation comes from the trendline for the MOS from the user experiments performed in this study. Out of the candidates for the trendline, an exponential function had the bests fit and
was therefore chosen according to the results in [11], and [9]. The trendlines were matched to the (OS, RT) pairs according to the least square principle, and is shown together with the coefficient of determination \( R^2 \) [20] as indicator: the closer \( | R^2 | \) gets to one, the better the trend is captured.

### 5.1 Quantitative results

The experiments gave OSs from all users that were then mapped to the corresponding RTs. These are shown, for the confirm experiment in Fig. 4, and for the login experiment in Fig. 5 together with matchings, as trendlines, of the values. The same trendlines as in [9] are used, derived from regressions of \( y \) vs. \( x \) (linear), \( y \) vs. \( \ln(x) \) (logarithmic), \( \ln(y) \) vs. \( x \) (exponential), and \( \ln(y) \) vs. \( \ln(x) \) (power).

In the web test [9], where users were waiting for pictures to download completely, the exponential and logarithmic trendlines are of similar quality and outperform both the linear and the power trendline.

In the confirm test, the logarithmic trendline matches best, followed by power, exponential and linear trendlines. Fig. 4 reveals that both logarithmic and power trendlines better approximate the behaviour for small RTs. Indeed, all non-linear trendlines take into account the convex trend of the measurements, implying a slower decay as RT values increase (i.e. \( y'(x) < 0 \), but \( y''(x) > 0 \)).

In the login test, the exponential trendline is superior to power, logarithmic and linear trendline. In Fig. 5 it is shown that the non-linear trendlines are found pretty close to each other. In particular, the comparison of the exponential trendlines for confirm and login test reveals similar sets of coefficients in Table 2, i.e 4.702 versus 4.836 as factor, and \(-0.097\) versus \(-0.105\) in the argument of the exponential function. This motivates us to focus on the exponential trendline in the sequel.

Comparing the confirm and login tests with regular web usage [9], a certain difference in the argument of the exponential function can be seen. In the web case, the curve is decaying more quickly (\(-0.15\) instead of \(-0.1\)). As seen in Fig. 4, some users do not really think that the service is bad despite the high
5.2 Qualitative results

Users in the login tests were asked what they thought of the service of OpenID, using SIM. The results were a bit different between the users, as expected, but surprisingly enough many of the users were united in some points.

- “I would not use it in the beginning.”
- “Is it really safer than passwords? I can lose the SIM-card.”

RT, and still at up to 12 or 13 s they think it is fair (3) or even good (4). This points at slightly increased user patience in the case of logins as compared to web usage. A similar evidence is seen from the comparison of the power trendlines between web test and confirm test in Table 2, where the latter displays a slower decay (power $-0.488$ instead of $-0.638$).

**Figure 5:** OS for corresponding RT for login test with different trendlines.
5 RESULTS

Table 2: Trendline equations with respective coefficient of determination ($R^2$).

<table>
<thead>
<tr>
<th>Type</th>
<th>$R^2$</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>0.990</td>
<td>$y = 4.836 e^{-0.150x}$</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>0.988</td>
<td>$y = -1.426 \ln(x) + 4.469$</td>
</tr>
<tr>
<td>Power</td>
<td>0.912</td>
<td>$y = 5.339x^{-0.638}$</td>
</tr>
<tr>
<td>Linear</td>
<td>0.966</td>
<td>$y = -0.318x + 4.158$</td>
</tr>
<tr>
<td>Confirm test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>0.618</td>
<td>$y = 4.702 e^{-0.097x}$</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>0.691</td>
<td>$y = -1.31 \ln(x) + 4.895$</td>
</tr>
<tr>
<td>Power</td>
<td>0.643</td>
<td>$y = 5.407x^{-0.488}$</td>
</tr>
<tr>
<td>Linear</td>
<td>0.966</td>
<td>$y = -0.2482x + 4.462$</td>
</tr>
<tr>
<td>Log in test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>0.807</td>
<td>$y = 4.836 e^{-0.107x}$</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>0.720</td>
<td>$y = -1.687 \ln(x) + 5.576$</td>
</tr>
<tr>
<td>Power</td>
<td>0.791</td>
<td>$y = 11.065x^{-0.860}$</td>
</tr>
<tr>
<td>Linear</td>
<td>0.705</td>
<td>$y = -0.206x + 3.921$</td>
</tr>
</tbody>
</table>

- “Good solution... IF it works!”
- “If I forget my phone at home I won’t be able to log in anywhere!”
- “If I don’t pay for the service I can wait an extra second.”
- “When it is too fast, or too slow, it feels like something is wrong.”

Uncertainty of what the service does and what it means for the users, how it really works, could be solved with detailed information. The risk of losing the SIM-card could be compared to that of losing a cellular phone with a SIM-card in. The last point is perhaps the most interesting. If a user chooses to pay for
a service, the expectations are probably high, but if the service is free the user might be more patient when experiencing higher RTs, i.e. worse QoD.

5.3 User perception profile

The approximation formula $OS = f(RT)$ is used in order to construct a direct link between performance of the authentication solution under study and the (predicted) user perception. The exponential approximation is used as it has shown to be the most generic of all the trendlines, in particular with respect to the obtained parameters.

Fig. 6 shows, for all different test cases from 250 B/s to 2 MB/s of bandwidth, what the predicted OS would be with our user perception profile, or user model, as tool. The system in question, which is the same system as in the previous experiments, was tested with different bandwidth constraints. The RTs that was logged for the different bandwidths were then transformed into OSs according to the formula

$$OS = 4.7e^{-0.1RT/s}$$

as shown in Fig. 6. It can be seen that the case of 250 B/s would not be accepted by the users and that bandwidths from 700 B/s and higher would be well accepted by users. So, with the user model it is possible to benchmark the system, within the QoE domain, before it is put in use.

6 Discussion

When comparing the exponential trendlines in the cases of the web experiment, the login experiment and the confirm experiment, we saw that the equations were similar, but a bit higher for both the login and confirm experiments (see Table 2). The two last-mentioned include a login procedure. This would imply that users doing a login are a bit more patient than users of regular web pages, which also applies for other waiting times in the context of IT services [21].

The result of this study can also be compared to results of previous similar
studies, such as [10] and [11]. The results from the login and confirm experiment will then, again, indicate that using a login gives a bit more user patience.

The results imply a direct relationship between the performance of the authentication method and the corresponding user ranking. This enables the expression of performance directly in expected user ranking of login. I.e. QoS and QoD can be directly matched to QoE. This kind of models for expressing relationships between QoD and QoE are valuable for companies, such as different kinds of service providers, on the purpose of QoE benchmark testing systems (as shown in Fig. 6) before delivery or release.

However, models for producing QoE from QoS or QoD are also hard to produce because of the large number of parameters that must be taken into consideration. Parameters that should be considered in such a model are not only RTs and user rankings, but for example a set of parameters that affect the user

![Figure 6: User perception profile: Frequency of OSs from RTs for different BWs.](image)
Another very interesting aspect of Fig. 4 and 5 are the points when the ratings that are lower than 3 start appearing. A company providing a service would not want to get rankings of 2 and definitely not rankings of 1. From both graphs (Fig. 4 and 5) we can clearly see that around 4 seconds users already start ranking the QoE as 2. In the graph from the confirm test (Fig. 4) we can also see that there are rankings of 1 at about 7 seconds, and rankings of 2 already at about 4 seconds. The latter implies that service providers should be careful, already when RTs start growing larger than 4 seconds.

7 Conclusion and Future work

This paper shows a study of user experience of web pages including the security feature of authentication. After several experiments, both with web experience with and without login, we came to the conclusion that there are differences on what users think about delay in web experience with security as compared to web experience without security. The results that we got from the experiments follow the line of the previous studies regarding QoE of web browsing, such as [9] and also other similar studies, such as [10] and [11], however with different time constants. The results indicate slightly higher user patience when security is involved.

The best fitting trendline of the QoE, for the login and confirm experiments, follow an exponential shape. The latter equation was also presented as a user model for further QoE evaluation without user tests. The question of whether a user will be satisfied with a certain QoD can be answered without extensive user experiments, which will save a lot of time, thus money.

Within a more extensive study, that will follow the work presented in this article, the user experience model will be used to further evaluate different authentication algorithms based on the QoD. A user will not know what authentication algorithm is used and the experience of the user will be based on the QoD. As long as the user interface and the environment are the same as
in the experiments in this study, the model that was produced are intended to be used. The user interface used in these experiments was a login on a web page, which leaves us with a large area of use for the user model. The latter, in turn, will be refined if necessary, as new pairs of OS and RT measurements will become available.

The next step is the evaluation of different authentication algorithms in the method together with OpenID. The user model presented in this paper will be used on the performance parameters of the evaluation to receive a QoE evaluation result. Though, e.g. for the One-time password (OTP) authentication method, new user experiments still need to be conducted, since that would mean a totally new experience for the user in the context of OpenID. A OTP authentication will include e.g. a cell phone, on which the OTP will be received.

8 Acknowledgment

The authors would like to thank Syed Usman Ali for his contribution and hard work on carrying out the confirm experiments during his Master Thesis work.

References


[16] M. Fiedler, C. Eliasson, S. Chevul, and S. Eriksén: Quality of Experience and Quality of Service in a Service Supply Chain. EuroFGI IA.7.6 Workshop on Socio-Economic Issues of Future Generation Internet, Santander, Spain, June 2007.


PAPER III

Decisive Factors for Quality of Experience of OpenID Authentication Using EAP-SIM

In proceedings of
European Teletraffic Seminar
Poznan, Poland, February 2011

Selected for publication in
Journal of Advances in Electronics and Telecommunications
Special Issue on “Recent Advances in Teletraffic”
July 2011
Decisive Factors for Quality of Experience of OpenID Authentication Using EAP-SIM

Charlott Lorentzen, Markus Fiedler, Henric Johnson
School of Computing, Blekinge Institute of Technology, Karlskrona, Sweden. Email: <firstname.surname>@bth.se
Ivar Jørstad
Ubisafe AS, Oslo, Norway Email: ivar@ubisafe.no

Abstract – When using the web, large response times are bones of contention for users, i.e. they damp the Quality of Experience (QoE). Though, if one knew the cause of a large response time, one can examine what can be done about this obstacle. In this paper, we determine the weak point of the Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAP-SIM) with the OpenID service with regards to excessive authentication times, which determine the response times. In order to provoke controlled raises of latter, we emulate bad network performance by introducing bi-directional delay between the supplicant (client) and the authenticator (server). Based on a recent, exponential relationship between QoE and response time, we then identify, quantify and compare the decisive factors for QoE reduction as functions of the components of the authentication times. The results we obtain clearly show that one task of the authentication contributes significantly more to the total response times than the other task, which points out the direction for future optimisation of user perception of authentication times.
1 Introduction

OpenID [1] is a system that allows for automatic confirmation of a user’s identity when visiting other authentication-enabled sites or communities supporting OpenID. In other words, a user is always authenticated, but only the first authentication will have to be initiated by the user. OpenID is promising to be used in a seamless environment, where the goal is to remain seamlessly authenticated while switching network, device or even application.

An authentication procedure typically produces a chain of messages before completing. And the more messages the chain consists of, the greater the risk of unacceptable response times (RT) becomes. These kinds of chains of messages, and/or chains of requests to different servers and databases form service chains (SC) [2] that can be quite large and complex.

We previously discovered significant RTs for the authentication to the OpenID server when using networks with low bandwidth, such as mobile networks. Such waiting times challenge user patience [3] and increase the risk of users trying to bypass or turn off security features. Once the Quality of Experience (QoE) [4] is really bad, users might even abandon the service [5]. The concept of QoE refers to the totality of end user experience of the delivered service [6], and in this case of the RT of the service. Typically different parts, or steps, of a service contribute with a certain RT and there might be a specific part that contributes more than others to the total RT.

Given the background above, this paper will identify and quantify the decisive factors for QoE of Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAP-SIM) with the OpenID authentication service as functions of network impairments in form of additional delay. The study will show what parts of the EAP-SIM authentication gives the greatest contribution to RT, when authenticating via OpenID.

There are several studies dealing with the evaluation and optimisation of different EAP authentication methods in different network environments and scenarios, such as studies on performance evaluation of EAP for roaming [7] and handover [8] in WLAN environments. However, to the best of our knowl-
edge, the combination of EAP-SIM authentication method with OpenID for web authentication has not been investigated yet.

The organization of this paper is as follows: Section 2 gives an overview of the authentication methods and services used. Section 3 discusses the impact of different network parameters and describes the methodology of the study, and Section 4 describes the experiments with corresponding setup, procedure and RT measurements. The results and the following analysis of the results are presented in Section 5, followed by a discussion of the results in aspects of QoE in Section 6. Finally, Section 7 provides a conclusion of the paper and points out future work.

2 Technical background

2.1 OpenID

OpenID is a service that handles one’s authentications. If users are logged in to their OpenID server, they are automatically logged in at visited web pages that have previously been enabled with OpenID for their particular user account.

An OpenID identity is a unique URL which contains the trusted provider and the username. The provider is the host of the URL, in our case Ubisafe AS, in Norway, and the username/URL will be openid.ubisafe.no/⟨username⟩, but the provider can also be for example Yahoo, or any other site that provides OpenID as authentication service. With OpenID users need one password or authentication credential and username to be able to authenticate oneself to all enabled sites, and the password or authentication credential only needs to be used when logging in at the OpenID server.

When a user account have OpenID enabled, for example on a community, the user log in at the OpenID server and then it is possible to visit the community and provide the OpenID username/URL, and the authentication to the community will be completed without an additional password. This applies for all OpenID-enabled pages during one web browsing session.
OpenID was chosen according to the requirements for seamless network and service access, to be provided by the IMS platform of the Mobicome project [9, 10].

2.2 EAP-SIM Authentication

In the EAP-SIM authentication method the actors are the SIM, the supplicant, the authenticator and the authentication server. The authenticator and the authentication server can be the same actor or situated in the same physical device (see Fig. 1). The supplicant is the user client, and the SIM-card is connected to the supplicant. The user just plugs in the SIM-card or makes sure the supplicant has access to it, and the supplicant is then the entity that communicates with the SIM-card. The authentication procedure is as follows.

- A connection (physical and virtual) is established between the supplicant and authenticator.
- The authenticator requests the identity (ID) from the supplicant (EAP-Request/Identity).
- The supplicant produces a response and sends its ID to the authenticator (EAP-Response/Identity).
- The authenticator challenges the supplicant in one or several steps to verify the ID of the supplicant (EAP-Request/SIM/Start and EAP-Request/SIM/Challenge).
- The supplicant handles and responds to the challenge(s) from the authenticator (EAP-Response/SIM/Start and EAP-Response/SIM/Challenge).
- If everything is in order with the challenge response the supplicant is authenticated and a success notification is sent from the authenticator (EAP-Success).

EAP does not have a retransmission feature. So, for certain packets a loss will result in a failed authentication procedure. When authentication is done
via a secure connection, lost packets might be handled by TCP (depending on the connection type) and the effect would be a higher RT instead of a failed authentication.

2.3 Authentication service chain

When a user requests to login to the UbiSafe OpenID server on the web page [11], a chain of messages to supply the service, i.e. a SC, is started. The SC for this authentication method is visualised in Fig. 2. In the sequel, let $T_{I_k}$ denote internal durations within the supplicant, while $T_{N_k}$ refer to durations involving network communications. The user request enters the supplicant which starts the setup of a connection to the authenticator after making sure the SIM-card is fully functional (duration $T_{I1}$). Once the connection is established, the authenticator sends back a request (end of time $T_{N1}$) for the ID of the supplicant, as described in the list in the previous section (Sect. 2.2). Please, note that $T_{N1}$ includes the time for setting up a connection between the authenticator and the supplicant, before sending the ID request to the supplicant. The setup of a connection is made visible with the dotted line in Fig. 2 and the vertical dots below it, since these messages are not EAP-SIM messages.

When the supplicant receives requests from the authenticator, the SIM-card is needed to produce a response to the requests (time $T_{I2}$) since the SIM-card has the ID, and the keys, before sending the response to the authenticator. The SIM-card is also needed to produce a response (time $T_{I3}$) for the SIM-challenge.

![Diagram showing the setup of the UbiSafe OpenID authentication service chain.]
Two further durations shown in Fig. 2, namely $\epsilon$ (between user click and initiation of the authentication) and $\delta$ (between completion of the authentication and displaying the results to the user) have shown to be of minor importance in the context of this study. We will therefore assume that the RT is well approximated by the authentication time, i.e. the sum of its internal and network...
components.

3 Network impact

In the course of our work, we seek to determine the RTs for the different part of the authentication method in question. The user perception is of course based on a whole RT, but if the greatest contribution to the RT was found, then it might also be minimised or made more scalable for large network delays with the goal to preserve a good QoE.

Timestamps were recorded for each task within the authentication, and RTs were calculated from the start timestamp and end timestamp for each task.

The objective of calculating RTs is to see whether there are any parts of the authentication that contribute most to the total RT or whether some parts are particularly sensitive to degradations in network performance. For this reason, we measure and compare the parts of the RT, in particular to see differences between RT for different parts of the authentication process, as well as their impact.

On network level, in most cases, adding delay or imposing bandwidth constraints gives a similar effect, namely higher RT values. In this study, bad network performance is emulated with a traffic shaper situated in the supplicant that shapes bi-directionally on the network interface of the latter.

Bandwidth can in some shapers be difficult to use for provoking the results we are looking for. For a single packet message the bandwidth constraint might never give any effect as the shaper might use a previous packet arrival to determine the next one. The latter was visible in the trials with bandwidth that were done early on in this study, where the RT for some parts did not change when decreasing the bandwidth and finally a timeout was received without any change in the RT for those parts.

Loss will result in higher RTs because of necessary retransmissions, but if one packet is lost in each part of the authentication, it will have the same impact.
on the RT for each part. To compare the effect of loss for each part would be quite difficult because of the encrypted traffic for the authentication. Therefore, loss has not yet been used as network performance degradation parameter in this study.

*Delay* is added to every packet that passes the traffic shaper, and the delay can be constant or variable. Variable delay has been tested in a previous project for map services [2]. Even though a constant delay might not be the most realistic case for emulating delay, it has shown a crucial enabler for the quantitative results presented in this study.

4 Experiments

The experiment setup consisted of a client computer with a SIM dongle, a traffic shaper for adding delay on the network interface of the client, and a server situated in Oslo, Norway, as shown in Fig. 1. All trials on the client computer were carried out on campus, during the same period of the day to withhold consistency, namely during evenings when most personnel were not at work. The delays that were added by the shaper were 0 ms, 250 ms, 500 ms, 750 ms, and 1 s in *both* directions. The timestamps were recorded via a JavaScript. Even though JavaScript logging of timestamps have proven to be only fairly accurate [12], the accuracy is sufficient for this experiment. Although the shaper adds constant delays on network level, the corresponding RT values are varying slightly, as there are many random impacts affecting the way between user and authentication service. Nevertheless, the chosen delays allowed to change the order of magnitudes of the RT such that trends regarding QoE could be clearly seen [3].

The experiment considered the login procedure on the Ubisafe OpenID server web page. On the web page “USB-SIM Dongle” was chosen in the Java applet handling the login. After clicking “Login”, the Java applet logged timestamps for starting and ending all parts of the EAP-SIM authentication (see Fig. 2). When the login was completed and the new page was loaded, a logout was done and then the procedure was repeated.
For each delay the experiment was done 45 times, and the log file was saved for later analysis. Although caching was disabled and cookies were not saved, the first five trials for each delay were discarded in order to avoid any potential bias of the measurements. The results were averaged, and 95% confidence intervals were calculated; the latter have however shown to be too small to be visible in the plots of Figure 3.

5 Results and Analysis

The authentication procedure consists of 16 steps, from initiation to success, including both internal processing time ($T_I$) and communication time spent in the network ($T_N$). Though, it can be abstracted down to seven steps, of which three (indexed by I1 to I3) are internal durations and four (indexed by N1 to N4) are external communication, i.e. network communication outside the supplicant (cf. Fig. 2). These steps are formalised as

$$T_R - \delta - \epsilon = \sum_{k=1}^{4} T_{Nk} + \sum_{k=1}^{3} T_{Ik} = T_N + T_I$$

where $T_R$ is the total RT, $T_N$ is the total time for the network communication steps, and $T_I$ is the total time for the internal processing steps, including communication between the supplicant and the SIM-card. As indicated before, we assume $\delta \to 0$ and $\epsilon \to 0$.

When comparing $T_N$ and $T_I$, the main contributions to the RT change with the increase of the RT. In case of no or low delays, RT is dominated by the processing time, $T_I$. For high delays, the RT is instead dominated by the network communication time, $T_N$. For a delay of 1 s the RT of about 24 s consist of more than 90% of network communication time, while for a transparent shaper, the relation is almost the opposite, as the processing time takes up about 70% of the total RT.

The steps including network communication are affected by the delay, whilst the internal communication, e.g. communication between supplicant and SIM-
Figure 3: RT components versus bi-directional delay added in the shaper.

dongle, is not affected. The parts of the RT that are interesting in the results are therefore $T_{N1}$, $T_{N2}$, $T_{N3}$ and $T_{N4}$, whereas the rest of the tasks or steps include only internal processing times that will not change with regard to increasing delay.

$T_{N1}$ provides the largest contribution to the total RT when there is no delay added, as can be seen in Fig. 3. It can also be seen that $T_{N1}$ is most sensitive to additional delay of the four network communication steps. When the additional delay is 1 s, both ways, $T_{N1}$ is already about 16 s long. For the remaining three tasks the RT grows equally fast, but substantially slower than for task N1, and they do start at a lower RT values in the case of no additional delay introduced by the shaper. Comparing the four tasks behind $T_N$, it is also for $T_{N1}$ that the most roundtrips in communication can be seen, due to the setup of a connection.

For $T_{N1}$, the linear growth in $T_R$ with respect to changes of the network...
delay added in the shaper \( d \) is eight times as large as compared to \( T_{N2}, T_{N3} \) and \( T_{N4} \):

\[
T_{N1} \approx 16d = 8 \times 2d
\]  \hspace{1cm} (2)

while

\[
T_{N2,N3,N4} \approx 2d
\]  \hspace{1cm} (3)

Thus, for the total network time, we get the approximation

\[
T_N \approx 16d + (3 \times 2d) = 22d.
\]  \hspace{1cm} (4)

The fact that the tasks with one round trip get a RT of double the delay (cf. Equation 3), as the delay is added in both directions, might indicate a relation between the number of packets sent back and forth and the factor of growth. If two messages, counting both ways, gets two times the delay, then 16 times the delay should indicate 16 messages when counting both ways, and thus eight round trips. Eight is also the factor between \( T_{N1} \) (Equation 2) and the other components (Equation 3).

When looking at the distribution of the RT values, they are a bit different from trial to trial, and from delay to delay. Most of the distributions are similar to a normal distribution, but in some cases with a (rather short) tail on the right-hand side. In some cases there are a few values that are bigger than the average, the median and the 90\% percentile. Such values belong to a so-called tail and can be quite bad when it comes to (perceived) network performance. However, for a RT in the order of magnitude of around 16 s, a parts of a second is not that large of a difference. In Fig. 4 the tail value is about 70 ms from the center of the distribution and the RTs are all larger than 16 s. Short tails, in this order of magnitude, do not have to be considered.

6 QoE aspects

Considering the user perception of the RT of the authentication, or QoE, and the changes of the latter with regard to the changes in RT, a previously researched user model [3] is used. Equation 5 was developed in a previous study for the
same system and in the same environment and represents basically a Mean Opinion Score (MOS) [13], which enables us to use the equation in this study in a straightforward manner:

\[ QoE \approx 4.7e^{-0.1RT/s} \]  

Obviously, each additional second factor of the network part of the \( RT \) yields a relative damping of the QoE by factor 0.9. From the measurements, it has been observed that processing time is approximately constant, which can be formulated as

\[ \sum_{k=1}^{3} T_{Ik} \approx 1.8 \text{ s}, \]
Thus, equation 5 can be rewritten as

\[ QoE \approx 4.7e^{-0.18}e^{-0.1T_{N}/s} \]
\[ \approx 3.9e^{-0.1T_{N}/s} \]  

which yields

\[ QoE \approx 3.9e^{-2.2d/s} \]  

Obviously, for the fixed line connection used in this experiment, it can be seen in Equation 8 that the QoE cannot exceed 3.9.

In Figure 5 it can be seen that, because of the exponential slope, already at 150~200 ms of delay, the MOS has gone below the rating “Fair”, and at 250 ms of delay the MOS is closing in on the rating “Poor”. The QoE reaches the MOS value 1, or “Bad”, at about 650 ms of added delay. Since the MOS scale goes from 1 to 5 when users are rating, values below 1 can be transformed to 1 as shown in Equation 2 in [15].

Dividing Equation 8 into the parts that grow equally gives

\[ QoE \approx 3.9e^{-1.6d/s}(e^{-0.2d/s})^{3} \]  

where the first e-term shows the impact of $T_{N1}$ and the second e-term shows the joint impact of the remaining times, namely $T_{N2}$, $T_{N3}$ and $T_{N4}$ on QoE. It can be seen that the first part has a significantly higher impact:

\[ \gamma = \frac{e^{-1.6d/s}}{e^{-0.6d/s}} = e^{-d/s} < 1 \]  

Equation 10 describes the QoE damping factor $\gamma$ between the impact of the connection setup time, $T_{N1}$, and the impact of the remaining network times, $T_{N2}$, $T_{N3}$ and $T_{N4}$, as function of the one-way delay $d$ introduced by the shaper. It can be seen that, as $d$ is growing, the damping impact of the connection setup supersedes the one of all the remaining network communication times. Even for low delays $d$, the connection setup time has a greater impact than all the remaining times, though in a lower order of magnitude.
Assume that one could reduce the number of messages during the connection setup by 50%, and thereby also reduce the connection setup time $T_{N1}$, one would observe a much less critical impact of the delay on the QoE, namely a factor of $e^{-0.2d/s}$.

As far as we can tell from this study, it is the setup of a connection between the supplicant and the authenticator that does not scale nicely with an increasing delay. The setup of a secure tunnel could, in theory, be skipped when authenticating with the SIM-card. No username and password are needed in this authentication, and as far as security without a tunnel is concerned, the provider of the SIM-card (i.e. the mobile operator) still has information from the keys for the SIM-card. The latter means that a secure communication is possible without an initial negotiation that will take a lot of time and transmissions. However, this requires the authenticator to be the operator or card-issuer, or at least the authenticator has to cooperate with the operator.

Figure 5: QoE in terms of MOS versus added delays, using Equation 8.
7 Conclusion and Future work

This paper has described the study of finding the most vulnerable part of EAP-SIM authentication method using the OpenID authentication service. After initial tests of the methods and the network connection, the experiments were performed with constant delay added in a shaper on the network interface of the supplicant, or client machine. A constant delay, which was increased for each trial, was added to provoke a change in RT for the different parts of the authentication.

When analysing the results, it could be clearly seen that one of the network times is growing faster than the others, namely $T_{N1}$, which refers to the initiation and setup of a connection between the supplicant and the authenticator, followed by an ID request from the authenticator. The fact that $T_{N1}$ was the largest contribution to the total RT and therefore also had the greatest impact on QoE was shown from several angles. N1 is also the task with most packets sent back and forth.

These results were then connected to the QoE user model that was developed in a previous study. From this mapping it was shown that the increase in $T_{N1}$ would result in a degradation of QoE that reaches 1 (lowest grade, meaning “bad”) at about 650 ms of added delay, in both directions. To achieve a factor of the impact on the RT that is more tolerable and scalable with large delays, one would need to reduce the number of messages during the connection setup. The reduction of the latter by factor two would entail a significant gain in scalability, seen from a smaller damping factor.

The accurate impact of variable delay, and loss also, could be evaluated in the future. Since bandwidth already has been tried without giving a realistic result, it will not be further evaluated unless a suitable traffic shaper is found. Variable delay will perhaps result in a bit lower RT than constant delay, but it need to be proved, or counter-proved. Also, loss would be interesting to evaluate as network performance parameter.

The setup of a connection between the supplicant and the authenticator needs to be examined a bit closer, to see if there is a possibility to optimize it.
Since the supplicant and the authenticator shares information from the SIM-card, there might be other possibilities to setup a connection. Then it might also be a good option if the OpenID authentication service is provided by the operator issuing the SIM-card. These possibilities will be closer examined with regards to trustworthiness and functionality in future work.

References


This work has examined users’ experience of an authentication procedure, combining QoE with security. The latter has been found to be an quite unexplored area. The work started with the construction of an evaluation framework for security mechanisms, which is presented in the first paper [1]. The evaluation framework aims at making evaluation of security mechanisms easier by taking existing evaluation methods and criteria and combining them, spanning from theoretical and qualitative methods to practical and quantitative methods.

Within the Mobicome project it was decided to use the authentication architecture OpenID together with the EAP-SIM authentication method. This solution was then evaluated, in a real setup with the server in Oslo, Norway, and the client in Karlskrona, Sweden, using user experiments. The study [2] resulted in user ratings for response times for the solution, and user models for web authentication.

Furthermore, the performance of the solution with OpenID and EAP-SIM was tested in the same environment setup. The authentication procedure was divided into tasks, and the tasks that were the most sensitive to network performance were identified. The previously presented user model [2] was used to quantify the impact of the network performance degradations on QoE, and the most sensitive tasks were pointed out as decisive factors for QoE. In this case
the setup of a tunnel for the authentication procedure is the decisive factor for QoE, and the impact factor that it has is significantly larger than the other impact factors.

**OUTLOOK**

Based on the results in [3], a deeper exploration of the security method of OpenID using EAP-SIM and the “encapsulation” is needed to see if an improvement of the method is possible. The investigation would then also include other methods that use tunnels and encapsulation, e.g., EAP-TTLS and PEAP, to compare the performance and to determine an optimal solution to the problem.

Another issue that was observed during this work relates to very short authentication time. The question arouse whether there exists a lower threshold for authentication time, below which users do not feel secure anymore. Instant login for example might not feel secure because a feeling of fake security might appear. This is an assumption based on user comments in the study from the second paper [2], and it could be investigated by performing quantitative and/or qualitative user experiments.

There is also an issue with the memory effect in user experiments [30]. The essence of the memory effect says that users rate lower when the response time is increasing than they do when the response time is decreasing, and higher for the opposite conditions. With regard to the latter, there might be a possibility to show the extent of the memory effect by finding the highest and lowest user model for a user group. This could be done by performing user experiments with continuously increasing response time and then continuously decreasing response time. This could then show the two extreme-case user models for the memory effect.

Another way to test the QoE of security can be to do user experiments with authentication from a fixed user interface, but with varying or no authentication method in the background. The user experiment could then be performed on
two different user groups, with the same characteristics of the test subjects, where one group would know about the changing of authentication methods and one would not. The users would then rate the QoE of security based on response time, but with different expectations and background knowledge. The interesting part would then be to compare the results of these experiments and see if there are any differences and what the differences are.


[26] Zona Research Inc. The economic impacts of unacceptaable web-site download speeds, April 1999.


## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3G</td>
<td>3rd Generation</td>
</tr>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>AT-MAC</td>
<td>Message Authentication Code Attribute</td>
</tr>
<tr>
<td>BTH</td>
<td>Blekinge Tekniska Högskola (Blekinge Institute of Technology)</td>
</tr>
<tr>
<td>BW</td>
<td>BandWidth</td>
</tr>
<tr>
<td>CC</td>
<td>Common Criteria</td>
</tr>
<tr>
<td>EAP</td>
<td>Extensible Authentication Protocol</td>
</tr>
<tr>
<td>EAP-AKA’</td>
<td>Improved Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement</td>
</tr>
<tr>
<td>EAP-SIM</td>
<td>Extensible Authentication Protocol Method for GSM Subscriber Identity Modules</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>GPRS</td>
<td>General packet radio service</td>
</tr>
<tr>
<td>HLR</td>
<td>Home Location Registry</td>
</tr>
<tr>
<td>HTTP</td>
<td>HyperText Transfer Protocol</td>
</tr>
<tr>
<td>ID</td>
<td>IDentity</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia Subsystem</td>
</tr>
<tr>
<td>IMSI</td>
<td>International Mobile Subscriber Identity</td>
</tr>
<tr>
<td>ISIM</td>
<td>Multimedia Services Identity Module</td>
</tr>
<tr>
<td>ISP</td>
<td>Internet Service Provider</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>ITU-T</td>
<td>Telecommunication Standardization Sector of ITU</td>
</tr>
<tr>
<td>MAC</td>
<td>Message Authentication Code</td>
</tr>
<tr>
<td>MOS</td>
<td>Mean Opinion Score</td>
</tr>
<tr>
<td>Acronyms</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>NAI</td>
<td>Network Access Identifier</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute for Standards and Technology</td>
</tr>
<tr>
<td>OpenID</td>
<td>Open IDentity</td>
</tr>
<tr>
<td>OS</td>
<td>Opinion Score</td>
</tr>
<tr>
<td>OTP</td>
<td>One-Time Password</td>
</tr>
<tr>
<td>PESQ</td>
<td>Perceptual Evaluation of Speech Quality</td>
</tr>
<tr>
<td>PP</td>
<td>Protection Profile</td>
</tr>
<tr>
<td>QoE</td>
<td>Quality of Experience</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RT</td>
<td>Response Time</td>
</tr>
<tr>
<td>SC</td>
<td>Supply Chain</td>
</tr>
<tr>
<td>SIM</td>
<td>Subscriber Identity Module</td>
</tr>
<tr>
<td>SMS</td>
<td>Short Message Service</td>
</tr>
<tr>
<td>SUS</td>
<td>System Usability Scale</td>
</tr>
<tr>
<td>SWOT</td>
<td>Strengths, Weaknesses, Opportunities, and Threats</td>
</tr>
<tr>
<td>TISPAN</td>
<td>Telecoms &amp; Internet converged Services &amp; Protocols for Advanced Networks</td>
</tr>
<tr>
<td>TOWS</td>
<td>Threats, Opportunities, Weaknesses, and Strengths</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UI</td>
<td>User Interface</td>
</tr>
<tr>
<td>UICC</td>
<td>Universal Integrated Circuit Card</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>USB</td>
<td>Universal Serial Bus</td>
</tr>
<tr>
<td>USIM</td>
<td>Universal Subscriber Identity Module</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
</tbody>
</table>
ABSTRACT

There is no doubt that security mechanisms, such as authentication, are required in Information and Communication Technology, but they come at a price: Users need to spend additional time and effort to authenticate themselves. With this in mind, user perception of authentication is an important factor for successful use of authentication solutions. If users perceive an authentication procedure as time-consuming and difficult, they might ignore or try to bypass it. Therefore, user-perceived Quality of Experience (QoE) should be investigated. QoE is a challenging area as it, in this case, covers network performance and security as well as Human Computer Interaction and user experience.

Throughout this work, authentication performance is investigated, starting with a framework for evaluating security architectures and authentication solutions in general. Criteria for user-friendliness, security and simplicity are described and the evaluation methods span from theoretical to practical, and qualitative to quantitative methods. The latter two aspects are investigated by a study of user experience of web authentication with OpenID using the EAP-SIM authentication method. The user experiments resulted in several user models of QoE. One particular user model for QoE, the exponential relationship between QoE and network level performance, was then used in further experiments on performance evaluation of OpenID authentication using EAP-SIM. The latter was done to determine the decisive factors for QoE of the authentication method in use. The results from these experiments show that the combination of OpenID and EAP-SIM for authentication over a secure tunnel is not appropriate to use over networks with high delays. The latter implies the need for improvements of the authentication procedure of OpenID using EAP-SIM, which should be addressed in the future. The user model of QoE obtained in this study will even help to quantify the performance aspects of future authentication procedures.