ON USER PERCEPTION OF AUTHENTICATION IN NETWORKS

Charlott Lorentzen

Blekinge Institute of Technology
Doctoral Dissertation Series No. 2014:05
Department of Communication Systems
On User Perception of Authentication in Networks

Charlott Lorentzen
On User Perception of Authentication in Networks

Charlott Lorentzen

Doctoral Dissertation in Telecommunication Systems

Department of Communication Systems
Blekinge Institute of Technology
SWEDEN
For my Family
and my dear Friends
Abstract

Authentication solutions are designed to stop unauthorized users from getting access to a secured system. However, each time an authentication process occurs an authorized user needs to wait in expectation of approved access. This effort can be perceived as either a positive or negative experience. If the effort is perceived as a security measure, the effort is usually perceived as a positive experience. On the other hand, if the effort is perceived as a waiting time, the effort is usually perceived as a negative experience. The trade-off between security, user-friendliness and simplicity plays an important role in the domain of user acceptability. From the users' point of view, security is both necessary and disturbing at the same time.

The overall focus in this thesis is on user perception of authentication in communication networks. An authentication procedure, or login, normally includes several steps and messages between a client and a server. In addition, the connection could suffer from low Quality of Service, i.e., each step in the authentication process will add to a longer response time. The longer response times will then infer lower Quality of Experience, i.e., a worse user perception.

The thesis first presents a concept of investigating user perception. A framework is developed in which different criteria and evaluation methods for authentication schemes are presented. This framework is then used to investigate user perception of the response times of a web authentication procedure. The derived result, which is an exponential function, is compared to models for user perception of web performance. The comparison indicates that users perceive logins similarly, but not identically, to how they perceive standard web page loading.
The user perception, with regards to excessive authentication times, is further studied by determining the weak point of the Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAPSIM) with the OpenID service. The response times are controllably increased by emulating bad network performance for EAP-SIM and other EAP methods in live setups. The obtained results show that one task of the EAP-SIM authentication deviates from the other tasks, and contributes more to the total response time. This deviation points out the direction for future optimization.

Finally, this thesis investigates how users of social networks perceive security, and to which extent they contribute to it. One way of contributing to security by creating and using strong authentication credentials, e.g. passwords. Web pages might enforce a password length which is insufficient to provide a strong password. This might then cause problems by giving users a false perception of what constitutes a strong password. The origin of the password problem, namely the construction of passwords, and the user perception of password security is studied. A survey is conducted and the results indicate that the passwords of the respondents are not as strong as the respondents perceive them to be.
Acknowledgements

I would like to extend my greatest thank you to Markus Fiedler for his generous and solid mentorship and supervision throughout my seemingly endless five years of doctoral studies.

I also want to thank my assistant supervisors, Henric Johnson and Patrik Arlos, for their help throughout the years and especially for the help with finishing my thesis.

An extra thanks to Markus and Henric for their inspiring lecturing during my years as a master student. Their inspiration is the reason I decided to pursue doctoral studies and got interested in network security.

Special thanks to my colleagues Niklas and David for constructively criti-cising, supporting, and acknowledging me in my work. And also to Niklas for his guidance in finding my own path in research.

I would also like to thank my colleagues and fellow doctoral students at DIKO and DIDD for fruitful discussions and sharing of ideas.

To my family, and especially Peter, Dad, and Mom, I send my sincerest thanks for always believing I can achieve everything I set out to achieve, and for supporting me endlessly, no matter what.

I would like to thank my friends for still being my friends and for putting up with me, despite the lack of mental presence during the last years.

A big thank you to Inga Johansson for shaking the foundations of my life, just enough to make me see where I went wrong.

For last, a big cuddle to my smallest, but most loyal supporter, Aquila.

Charlott Lorentzen
Karlskrona, March 2014
# CONTENTS

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Publications</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Aim and Scope</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Outline</td>
<td>3</td>
</tr>
<tr>
<td>2 Main concepts</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Authentication</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Quality of Experience and User Perception</td>
<td>9</td>
</tr>
<tr>
<td>3 Research approach</td>
<td>13</td>
</tr>
<tr>
<td>3.1 Problem Statements</td>
<td>13</td>
</tr>
<tr>
<td>3.2 Research methodology</td>
<td>14</td>
</tr>
<tr>
<td>4 Paper Summary</td>
<td>21</td>
</tr>
<tr>
<td>4.1 Paper I</td>
<td>21</td>
</tr>
<tr>
<td>4.2 Paper II</td>
<td>21</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>4.3</td>
<td>Paper III</td>
</tr>
<tr>
<td>4.4</td>
<td>Paper IV</td>
</tr>
<tr>
<td>4.5</td>
<td>Paper V</td>
</tr>
<tr>
<td>5</td>
<td>Discussion</td>
</tr>
<tr>
<td>5.1</td>
<td>User Experiments for Quality of Experience</td>
</tr>
<tr>
<td>5.2</td>
<td>Perception of Security</td>
</tr>
<tr>
<td>6</td>
<td>Contributions</td>
</tr>
<tr>
<td>7</td>
<td>Conclusions &amp; Outlook</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
</tr>
<tr>
<td>I</td>
<td>A criteria-based evaluation framework for authentication schemes in IMS</td>
</tr>
<tr>
<td>II</td>
<td>On User Perception of Web Login – A Study on QoE in the Context of Security</td>
</tr>
<tr>
<td>III</td>
<td>Decisive Factors for Quality of Experience of OpenID Authentication Using EAP-SIM</td>
</tr>
<tr>
<td>IV</td>
<td>On User Perception of Safety in Online Social Networks</td>
</tr>
<tr>
<td>V</td>
<td>Actual and Perceived Password Security in Online Social Networks</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure Page
1.1 Outline of the thesis. 4
2.1 EAP-SIM authentication procedure [15]. 8
3.1 Communication between client and server (A and C), and internal processing in the server (B). 15
3.2 Two ways of presenting QoE results. 17

Paper I

1 Framework of evaluation criteria. 43
2 Methodology of authentication scheme evaluation. 49

Paper II

1 Average RTs mapped to the corresponding delay added in shaper. 60
2 Setup for confirm and login experiments. 61
3 Methodology for the study of authentication schemes. 63
4 OS for corresponding RT for confirm test with different trendlines. 67
5 OS for corresponding RT for login test with different trendlines. 68
6 User perception profile: Frequency of OSs from RTs for different BWs. 72

ix
Paper III

1 Setup of the Ubisafe OpenID authentication, with BTH client side. ... 83
2 SC for EAP-SIM authentication with OpenID. .................... 84
3 RT components versus bi-directional delay added in the shaper... 88
4 Frequency of $T_{N1}^*$, for a bi-directional delay $d=1$ s, bin width 2 ms. 90
5 QoE in terms of MOS versus added delays, using Equation 8. ... 92
6 Setup of the EAP authentication experiments. .................. 93
7 RT components for EAP-MD5 versus bi-directional delay added in shaper. .................... 94
8 RT components for EAP-TTLS-PAP/EAP-TTLS-CHAP versus bi-directional delay added in shaper. .................... 94
9 RT components for EAP-TTLS-MD5/EAP-TTLS-MSCHAPv2 versus bi-directional delay added in shaper. .................... 96
10 RT components for PEAP-MSCHAPv2 versus bi-directional delay added in shaper. .................... 96
11 The decisive factors for each authentication method. ......... 98
12 Security versus complexity for several authentication methods, with 100 ms added delay. .................... 99

Paper IV

1 Mean ratings of user perception of safety with 95% confidence interval, stratified according to behaviour, for categories High and Low. .................... 114
2 Mean ratings of user perception of safety with 95% confidence interval, stratified according to behaviour, for categories Same and Misc. .................... 114
3 Password complexity for teenagers, divided in study programme groups. .................... 119
4 Password complexity for teenagers, divided in male and female. ... 120

x
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Paper</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Interpretation of MOS grades</td>
<td>Paper I</td>
<td>17</td>
</tr>
<tr>
<td>1 Interpretation of MOS grades</td>
<td>Paper II</td>
<td>63</td>
</tr>
<tr>
<td>2 Trendline equations with respective coefficient of determination ($R^2$)</td>
<td>Paper II</td>
<td>70</td>
</tr>
<tr>
<td>1 Decisive factors and Security level for each method</td>
<td>Paper III</td>
<td>99</td>
</tr>
<tr>
<td>1 User perception of safety with regard to response time, for Facebook login, shown in trial order</td>
<td>Paper IV</td>
<td>116</td>
</tr>
<tr>
<td>2 Demographics of teenage participants in the password complexity and identical password surveys</td>
<td>Paper IV</td>
<td>117</td>
</tr>
<tr>
<td>3 Demographics of staff member participants</td>
<td>Paper IV</td>
<td>117</td>
</tr>
<tr>
<td>4 Mean password complexity credits (1–5) for the Facebook account passwords used by the teenage participants</td>
<td>Paper IV</td>
<td>121</td>
</tr>
<tr>
<td>5 Mean password complexity credits (1–5) for the Facebook account passwords used by the staff member participants</td>
<td>Paper IV</td>
<td>122</td>
</tr>
</tbody>
</table>
6 Percentages of teenagers that use identical passwords for their private e-mail and Facebook account. . . . . . . . . . . . . . . . . . 123

7 Percentages of staff members that use identical passwords for their private e-mail and Facebook account. . . . . . . . . . . . . . . . . 123

Paper V

1 Password entropy. The table shows the mean password entropy and standard deviation. Each segment represents a password structure type. Each row represents a study programme and each column represents a gender. . . . . . . . . . . . . . . . . . 144

2 Password entropy. The table shows the mean password entropy and standard deviation. Each segment represents a password structure type. Each row represents a study programme and each column represents a gender. . . . . . . . . . . . . . . . . . 146
LIST OF PUBLICATIONS

The papers are ordered chronologically in each section. The name of the author was Charlott Eliasson until May 2010, when it was changed to Charlott Lorentzen.

INCLUDED PAPERS


Further papers


Chapter 1
Introduction

People spend a lot of time connected to social networks, communities, and online services. Therefore, passwords and login procedures are an essential part of their online security measures. Passwords are a type of credential which users give as tokens to other parties to identify themselves. The process of identifying oneself is called authentication and is performed during a login procedure. Several authentication methods using other credentials than username and password are finding their way into technology. Despite these methods, username and password continue to be the most used authentication credential.

With more activities, information, and services found online, the need for authenticating users and securing user data is getting larger. Therefore, the importance and attention of security is increasing. The increasing attention that security gets, or rather that absence of security gets, is partly because of the increasing threat that users are exposed to. For example, during recent years, two large companies have experienced hackers compromising user data, including usernames and passwords [14, 15]. These events both resulted in compromised accounts for their users, not only for the accounts associated with the mentioned companies, but potentially for the users’ accounts elsewhere with the same username and password.
Many users have the same username and password for different accounts. This redundancy is added to by the possibility to use for example a Facebook or Gmail account as user credential for authentication to other accounts. The combined user data will then be secure as long as their Facebook or Gmail account is not compromised. If their Facebook or Gmail account is compromised, all associated accounts will also be compromised. Despite this risk, users seem to perceive the concept of associating accounts as secure or they simply have not thought it through.

In general, users perceive different perspectives of online security. As they perceive the procedure of an authentication, they get a perception of login security. When they have constructed a password, they get a perception of password security. Despite the usualness of the two concepts and the implications of compromised user accounts, the use of the concepts are rarely regarded by users as something worth spending time and effort to perfect, or even to moderately refine.

Most web pages that require login have username and password as tokens for users to authenticate themselves with. Though several studies have been conducted on finding the actual security for passwords, there is a lack of non-intrusive studies. There is also a lack of studies investigating the user perceived security of passwords.

An authentication may include several steps with requests and responses, back and forth between client and server. If the connection suffers from low Quality of Service (QoS), each step will take longer time and add to the total response time. Furthermore, the addition in total response time will affect the quality of the users perception, or the Quality of Experience (QoE).

When looking at QoE for performance of web pages, users are less happy with longer response times. Since users perceive both performance and security when logging on to a web page, the perception of security also needs to be investigated.
1.1 AIM AND SCOPE

In this thesis, a new authentication method using SIM-card as credential provider will be investigated. User perception of this authentication method will be studied and analysed. Furthermore, regular passwords and how they are perceived, constructed, and used will be investigated.

The studies that comprise this thesis are all aimed at understanding users' perception of authentication. The first paper aims at finding ways of evaluating user perception and performance of authentication. The second paper aims at finding the user perception for login procedures on web pages. The third paper aims at finding factors in performance that affects the user perception which was evaluated in the second paper. The fourth paper continues the work done in the second paper, but focuses solely on the user perception of the security in the login procedure. The focus in the fourth paper is also on teenage users of Online Social Networks. The fifth paper follows the forth with focus on teenage users of Online Social Networks, but aims at evaluating how users perceive their own passwords, with regards to security. The fifth paper also aims at finding weather there is a difference in what users perceive and the actual security of their password.

1.2 OUTLINE

The outline of the thesis is as follows. There are two parts in this thesis, see Figure 1.1. Chapter 1 to Chapter 7 constitute the first part and present an introductory summary for the thesis. Paper I through V constitute the second part and present all studies included in this thesis.
**CHAPTER 1. INTRODUCTION**

<table>
<thead>
<tr>
<th>First part</th>
<th>Introductory Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1: Introduction</td>
<td></td>
</tr>
<tr>
<td>Chapter 2: Main concepts</td>
<td></td>
</tr>
<tr>
<td>Chapter 3: Research approach</td>
<td></td>
</tr>
<tr>
<td>Chapter 4: Paper summary</td>
<td></td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td></td>
</tr>
<tr>
<td>Chapter 6: Contributions</td>
<td></td>
</tr>
<tr>
<td>Chapter 7: Conclusions &amp; Outlook</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second part</th>
<th>Publications &amp; Submitted articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper 1: A criteria-based evaluation framework for authentication schemes in IMS</td>
<td></td>
</tr>
<tr>
<td>Paper 2: On user perception of web login – A study on QoE in the context of security</td>
<td></td>
</tr>
<tr>
<td>Paper 3: Decisive factors for Quality of Experience of OpenID authentication using EAP</td>
<td></td>
</tr>
<tr>
<td>Paper 4: On user perception of safety in Online Social Networks</td>
<td></td>
</tr>
<tr>
<td>Paper 5: Actual and perceived password security in Online Social Networks</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.1: Outline of the thesis.

The outline of the first part of the thesis is as follows. The main concepts are presented in Chapter 2. Overviews of problem statements and used methodologies are presented in Chapter 3. The studies which comprise this thesis are summarized individually in Chapter 4. The studies are discussed, individually and summarized in Chapter 5. The main contributions of this thesis are presented in Chapter 6. The studies are concluded and an outlook is presented in Chapter 7.

Paper I through V constitute the studies included in this thesis, presented in their entirety in the second part of this thesis. Paper I through IV are published studies and Paper V is submitted. Each study is presented as published or submitted, except for the layouts which have been altered to fit the layout of this thesis.
Chapter 2

Main Concepts

2.1 Authentication

Authentication is used by a system to prove the identity of a user: to be able grant authorized users and deny unauthorized users access. Thus, authentication is part of access control. When reaching an access controlled system, the user provides a token that proves the user’s identity, i.e., a credential. If the credential is a valid proof of the user’s identity, the user gains access.

The credential for authenticating a user varies between something you know, something you have, something you are and something you do [16]. These credential types can be used individually or combined. 'Something you know' is typically a password. 'Something you have' is typically a key, e.g., in a smartcard. 'Something you are' is based on biometrics and can be, e.g., a fingerprint or retina pattern. 'Something you do' is a pattern of something you do similarly all the time, e.g., the frequency with which you type the characters in a specific word.

In Section 2.1.1 and 2.1.2 two authentication methods used in the studies of this thesis are presented.
2.1.1 OpenID Architecture

OpenID is an open-source authentication protocol which provides a solution for umbrella login. Users log on to an OpenID server and can then visit any web page that has OpenID enabled and pre-selected, without manually providing credentials for authentication for each web page. Users are commonly identified at web pages 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. The OpenID service should be provided by a trusted part, such as an Internet service provider or mobile operator.

The OpenID architecture supports authentication with SIM-card keys as credential, either using a mobile phone or a SIM-dongle via a computer interface. This support makes OpenID using SIM-card ideal to use as a service from mobile operators towards customers. The keys used for authentication are distributed with the SIM-cards, which are issued by the operator.

If the authentication is done with a SIM-card, as in the studies of this thesis, users needs to know their OpenID username and nothing else. A user is identified via the SIM-card and keys in the SIM-card are used to authenticate the user. As long as the user has the SIM-card, there is no need to remember passwords to be able to login to OpenID enabled pages on the Internet.

2.1.2 EAP-SIM Authentication

Users tend to choose simple credentials (passwords) which they can relate to [16], e.g., the name of a child or a pet. Therefore, user-chosen credentials can cause security issues. When using a smartcard or a SIM-card the credentials for authentication are keys, which are not user chosen. Hence, the issue of simple credentials is not present when authenticating with keys from a smartcard or SIM-card.

The Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAP-SIM) [17] 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.

The messages sent between client, or supplicant, and authenticator are visualised in Figure 2.1 and are briefly described below [17].

**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.

**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_MAC values are equal, i.e. correct, a challenge response is sent to the authenticator.

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

There are mechanisms within this message exchange which assure the validity of the messages for both the authenticator and the supplicant. Protection of the version negotiation is done by including the list of supported version when calculating the keying material. The list of supported versions is attained from the authenticator. The authenticator can use the same credentials to make sure that the selected version from the client is the accurate one and that its integrity is intact.

Verification of correctness is also done when the AT_MAC value calculated by the supplicant is compared to the one that is sent by the authenticator. If two AT_MAC values 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
Supplicant Authenticator

EAP-Request/Identity

EAP-Response/Identity

EAP-Request/SIM/Start

EAP-Response/SIM/Start

EAP-Request/SIM/Challenge

EAP-Response/SIM/Challenge

EAP-Success

Figure 2.1: EAP-SIM authentication procedure [15].

message. Since this random number is used to calculate the AT_MAC value, the peer can ensure that EAP-SIM messages sent by the authenticator are fresh and not replayed.

Improved Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA’) [18] is a newer version of the EAP-SIM. This newer version was not available when the work of this thesis started. Given that 3G network infrastructures and 3G identity modules are also available, this version is an alternative. 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.
2.2 QUALITY OF EXPERIENCE AND USER PERCEPTION

QoS denotes the quality with which a service is delivered via a network. The concept of QoS has amongst others been defined as “something a flow of data seeks to attain” [19]. Another definition of QoS is found in the ITU-T E.800 recommendation [20]:

"Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service".

If network conditions, such as jitter, delay, loss, bandwidth, are bad then the performance of the network will be low and the QoS will be low. The result of QoS parameters is the response time, i.e., the time interval from a request to complete delivery of response data.

In these studies, the QoE denotes the users' perception of the response time or the displaying of results based on response time. To summarise the concepts: QoS affects response times, which effect QoE. The user perception and experience of a service can also be summarised as the QoE. A definition of QoE is found in the ITU-T P:10/G.100 recommendation [21]:

"The overall acceptability of an application or service, as perceived subjectively by the end-user".

The network of excellence Qualinet (European Network on Quality of Experience in Multimedia Systems and Services) states a working definition of QoE. A “working” definition means that the possibility of further evolution and refinements of the definition with the advances in QoE-research are considered.
CHAPTER 2. MAIN CONCEPTS

The current working definition is given in a white paper [22] as:

“Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state.”

High response times when loading a web page or a service, gives a bad QoE [23, 24, 25, 26, 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 seconds users start becoming impatient. At around 10 s there is a highly increased risk that the user will leave the web page or service.

2.2.1 Measuring Quality of Experience

Users' perception of services can be important for service providers to know. Users with a bad perception of a service are more likely to leave the service. A measure of QoE can be used to show user perception. To measure QoE, the user’s perception has to be quantified. Quantifying QoE means translating user perception into numerical and interpretable values. Once quantified, the QoE measure can be compared to QoS parameters or response times. Such a mapping of QoS and QoE will tell a service provider about critical thresholds for QoS.

Users consider their experience based on an almost indefinite number of unconscious parameters. Examples of such parameters are habits, origin, mood, and expectations. These parameters are difficult to measure without extensive user studies or advanced technology, such as eye tracking or heart monitoring.

User models in this thesis are 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.
In this thesis a user group is defined as a group of users participating in an experiment. The user group must be of sufficient size for the resulting answers and ratings to be representative and valid. If an experiment aims to model the representative behaviour of a certain type of users, the characteristic requirements must be fulfilled by all participating users.

In some cases, measuring QoE means posing questions to users about their perception of a service in mind. The questions must then be designed to avoid manipulating or leading the user. Furthermore, the questions should capture the essence of what the user is experiencing. The questions in the user experiment also need to be specific, in order to avoid misunderstandings of what possible ratings concerns. Failing to pose proper questions may cause misleading or invalid answers and results.

2.2.2 Awareness

In Paper I of this thesis, awareness is used to describe a user’s consciousness, good and bad, of a service or application. Another aspect of awareness is the accuracy with which a user perceives a service or component as compared to reality. In the same spirit as in Paper I, the accuracy with which a user perceives reality can be labeled “accurate”, “false negative”, or “false positive”.

“Accurate” is the case where the user perception of a service or component coincides with reality. “False negative” is the case where the user perception of a service or component is worse than reality. “False positive” is the case where the user perception of a service or component is better than reality.

Both false positive and false negative shows a gap between the actual and the perceived service or component. Generally, the false positive cases are the worst cases when considering security: a service or component which is believed to be secure is unsecure in reality.
CHAPTER 3
RESEARCH APPROACH

3.1 Problem Statements

Problem statements are presented as per paper of this thesis.

Problem statements for Paper I, “A criteria-based evaluation framework for authentication schemes in IMS” are:
1.1. How should we evaluate security solutions?
1.2. What methods exist for evaluating authentication?
1.3. What criteria are important when evaluating authentication?

Problem statements for Paper II, “On User Perception of Web Login - A Study on QoE in the Context of Security” are:
2.1. How do users perceive the authentication mechanism which is being tested?
2.2. To which extent are users satisfied with the performance of the authentication mechanism?

Problem statements for Paper III, “Decisive Factors for Quality of Experience of OpenID Authentication Using EAP-SIM” are:
3.1. Are there some parts of the tested authentication mechanism which affects QoE more than the other parts?
3.2. Are these potential parts scalable with regards to network issues?
CHAPTER 3. RESEARCH APPROACH

Problem statements for Paper IV, “On User Perception of Safety in Online Social Networks” are:

4.1. How do users perceive security in OSNs?
4.2. How much do users contribute to their online security?
4.3. How complex are the users’ passwords?

Problem statements for Paper V, “Actual and Perceived Password Security in Online Social Networks” are:

5.1. Previous results indicate that some users fail to contribute to their online security via their password. To what extent do users contribute to their online security, in terms of password security?
5.2. Are users who fail to contribute to their password security aware of this?
5.3. Is there a difference between users’ actual and perceived password security?

3.2 RESEARCH METHODOLOGY

3.2.1 RESPONSE TIME EXPERIMENTS

A response time is defined in this thesis as the end-to-end time for a request to leave the client, until the result of the request is fully visualized at the client. Response times, in general, can be measured or calculated. The time synchronisation of entities in an experiment setup is of utter importance [29]. If the response time is defined as starting in one entity and ending in another entity, so will the measuring of the response time. This means that the entity in which the measurement starts and the entity in which it stops have to be time-synchronised.

The time-synchronisation between entities can be avoided in some cases. In these cases each response time have to be calculated using timestamps from the same entity. This will ensure that all timestamps used in one calculation are time-synchronised. A mixture of timestamps from different entities in the same calculation gives invalid results if time-synchronisation is absent.
3.2. RESEARCH METHODOLOGY

In the experiments of this thesis the response time is calculated using timestamps from the same entity. The timestamp for a request at the client side is compared to the timestamp for an arriving, corresponding response at the client side. Thus, comparing timestamps from different entities, such as client and server, were avoided. The avoidance of comparing timestamps from different entities were due to inaccuracies and time differences that exist between entities [29].

In some cases the network communication times were calculated. The network communication time is defined as the end-to-end time, as perceived by a user, minus the time spent by the server to process the request, B, see Figure 3.1. In those cases the internal server processing time, B, was calculated with timestamps for arrival of request, A, and departure of response, C, at the server. This gave the time that the server spent processing the request, B. This internal server time was then subtracted from the end-to-end response time at the client side. The calculations then left the time for the total communication between client and server.

3.2.2 QoE USER EXPERIMENTS

A user experiment aims to find how users perceive a service or product. If the aim is to evaluate how users perceive a service, then the users are subjected to the service and their perception is logged. The experimental setup can either feature the real service or something that the user would see as the real service,
CHAPTER 3. RESEARCH APPROACH

an emulation. In this thesis, user experiments are done to find how users perceive authentication.

The main advantage of using a real service is that there is less doubt about the users perception of the service. The main advantage of using emulation is that parameters can be manipulated to trigger a broader user response. The interface of the experiments in this thesis features a real service, whereas the setup, which is not visible to the user, allows for manipulation.

For user experiments, Mean Opinion Score (MOS) have been used as a basis for questioning subjects on their perception. MOS is the mean of all opinion scores given by users. Each opinion score is a user rating based on the MOS scale. The ratings for MOS are found in the ITU-T Recommendation P.800 [30]. The ratings found in this recommendation are used for evaluating user experience of both telephone transmission quality and more generally, e.g., for web browsing [31].

The ratings on the MOS scale are both numerical and ordinal, as shown in Table 3.1. The ordinal ratings go from “Bad” to “Excellent”. The numerical ratings go from 1 to 5. Ratings given according to the ordinal scale can be directly translated to the numerical scale [30]. Thus, quantification from ratings to opinion scores is attained. The opinion scores are then used for calculations and statistics.

For the user experiments of QoE, the users were introduced to the experiment and were then exposed to a series of login trials which had different network settings. For each trial, the user gave an opinion score, which was mapped to the corresponding network setting or response time. For example, the network settings could issue different delays in the network. The differences in delays were exposed to the user as differences in response time from one login trial to the next.

When the mapping is done, the result can be visualised in graphs where each point corresponds to one opinion score. An example of this is given in Figure 3.2(a) where each opinion score is shown. A regression curve which is based on all the opinion scores is also shown. Alternatively, the MOS can be calculated separately for different experiment settings, which is shown in
Table 3.1: Interpretation of MOS grades [30].

<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>

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

(b) MOS mapped to additional delay.

Figure 3.2: Two ways of presenting QoE results.

Figure 3.2(b). Then each point corresponds to the QoE for that particular QoS. In this case a point corresponds to the MOS for a particular introduced delay. A third option is to present opinion scores to average response time.

**Issues in Measuring QoE**

The environment is usually important when measuring user experience. 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. Most users do not continuously give ratings to their experiences in a real environment. A user might dislike an experience in an everyday situation with, e.g., an internet connection. However, a user rarely or never put a five scale grading on it.

An issue with measuring QoE is potential diversity in user characteristics and past experience amongst a user group. Users can have different backgrounds
and different moods on the occasion of the experiment. Past experience affects user’ ratings in the form of expectations. If the user is used to waiting times of 20 s, then 10 s might be quick. If the user is used to some 100 ms of waiting time, then 10 s might be unacceptably slow to that user.

A potential issue with measuring QoE is the memory effect [32]. A user is presented with low quality before medium quality and rates the medium quality with rating \( x \). The concept of the memory effect then says that if the user is presented with high quality before medium quality, the user rates the medium quality with a lower rating than \( x \). Therefore, when a user is asked to rate several scenarios directly after each other, the order of the scenarios will affect the resulting ratings.

Experiment scenarios can be mixed and presented in different orders to different users. This mixing will then, with regard to the memory effect, give a greater diversity in the resulting ratings per scenario, than non-mixed scenarios would. However, this diversity may be considered as a “worst case” with regard to spread of values.

Based on the above, the major challenge with user experiments is the setup of a realistic environment. The environment should be realistic enough to provide ratings which are close enough to a real scenario experience. In the experiments to be reported in this work, the setup of the environment was designed such that any underlying changes in network settings or similar were transparent to the user.

QoE FOR MEASURING PERCEIVED SECURITY

The challenges when measuring QoE for security mechanisms or security based, like authentication services, are different in some aspects. 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 they do of their own.
3.2. RESEARCH METHODOLOGY

If the user does not care about a security service, the aspect of security might be lost or compromised. To diminish this potential issue, an experiment can be done on an authentication procedure that is well-known and can be replicated. The potential issue can also be diminished by doing experiments with the actual well-known authentication procedure. An example of such a well-known procedure is a login procedure of a community web page.

3.2.3 PASSWORD SECURITY SURVEY

For the final study of this thesis, a survey was used to evaluate user perception of password security. The survey enquired subjects on their passwords and their perception of their passwords. The questions about the subjects' passwords were strictly quantitative and all evaluation or interpretation was left to the researcher.

The survey did pose questions where the subject had to evaluate their perception of the security of their password. However, these perception questions were stripped of quantitative grading and simply gave the subjects response alternatives in a yes/no manner. Using questions without quantitative grading minimizes the risk of interpretation bias in a subject's answers.

The overall methodology for the password security studies are derived from methods by Creswell [33]. Concurrent embedded design was used for the topics of Paper V [5]. For one topic in Paper V, Concurrent triangulation design was used.

Concurrent embedded design [33] can be used by embedding a qualitative part in a quantitative design, such as a user experiment. One purpose of concurrent embedded design is to strengthen a traditional quantitative design by introducing a qualitative part. All data is collected in the same database and interpreted concurrently.

Concurrent triangulation design [33] can be used when the focus is to cross-validate findings from different parts in the same study. One purpose of concurrent triangulation design is, by collect concurrent qualitative and quantitative data, to offset weaknesses of one data type with strengths of the other. The
two data types are compared after collection and interpretation. Weight can be
given to any or none of the collected data types during comparison. Typically,
the parts are given equal weight.

Avoiding bias and common mistakes are important when designing user sur-
veys and interviews. The validity and credibility threats covered by [34] was
used to identify risks of such threats when designing the password security ex-
periment.
Chapter 4

Paper Summary

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

Authentication can be evaluated in different ways, based on different criteria. A framework for evaluating security with evaluation criteria and existing evaluation methods was presented in Paper I, and later reused throughout the studies of this thesis.

The evaluation framework spans from qualitative to quantitative and from theoretical to practical methods for evaluating security. The framework ends up in a summarized 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 perception of a web login procedure is investigated in this study. The response time for a login is logged and the user’s perception of the login is collected in the form of opinion scores. The procedure is repeated for a set of
response times. The response time is altered by changing the added delay in the network.

For each response time the users' opinion scores are compiled into MOS. The MOS for each response time is then presented in a graph together with trendlines. Corresponding coefficient of determination is also presented, as indicator for goodness of fit.

The results indicate that users' perceptions of a login to a community web page are best represented by an exponential curve. Users feel less secure with longer response times and more secure with shorter response times. The results are also compared to similar studies of web pages without security or authentication. Some similarities found in the comparison are discussed.

4.3 Paper III – Decisive Factors for Quality of Experience of OpenID Authentication Using EAP

The performance of OpenID with EAP-SIM is investigated in this study. The response time for each task of the EAP-SIM method was isolated for each trial. Furthermore, the network delay was increased in a shaper, from 0ms to 1000ms, in steps of 250ms. The performance of each task was evaluated separately, using the user model from Paper II, in order to find the decisive factor for QoE.

The results show that the tasks that were most sensitive to network performance degradation give the greatest impact on QoE. The results further show that additive changes of response time give multiplicative changes of QoE. The order of magnitude of QoE decrease was shown for certain network performance degradations.

Further, the overall performances of other EAP authentication methods are investigated using total response time. The results indicate differences in performance based on the security level. Higher security indicates worse performance.
4.4 PAPER IV – ON USER PERCEPTION OF SAFETY IN ONLINE SOCIAL NETWORKS

User perception of the login procedure for Facebook is investigated in this study. The user perception is evaluated by users giving opinion scores for a set of logins to Facebook. The response times for each login in a set is logged and later mapped to the corresponding calculated MOS.

Further, the users' passwords are investigated by using a survey. Respondents of the survey are enquired about the length and types of characters included in the password. The scores are then mapped to the users' characteristics, such as age, gender and study programme.

The results from the user perception evaluation indicate that the response time is irrelevant for some users to feel safe. These users either do not care or feel safe all the time. The results on user passwords indicate that there might be a difference in password complexity for teenagers of different study programs with different levels of technical education.

4.5 PAPER V – ACTUAL AND PERCEIVED PASSWORD SECURITY IN ONLINE SOCIAL NETWORKS

The construction of passwords for authentication to Facebook and the user awareness of password security are investigated in this study. The awareness of password security is evaluated by comparing actual and perceived password security.

Users' perception of their passwords and the corresponding characteristics of passwords are analysed using a survey. The characteristics of passwords that are enquired of the user are length and types of characters included.

The results of the study indicate that most users have weak passwords and that there is a gap between actual and perceived password security. About half of the respondents consider their password as strong, and about 75 % of those respondents have a weak password.
CHAPTER 4. PAPER SUMMARY
Chapter 5

Discussion

5.1 User Experiments for Quality of Experience

Doing quantitative research using qualitative methods introduces some issues. Demographics and how the subjects are chosen in a study have to be considered. A group must be selected based on some prerequisites. The subjects that chose not to participate should be accounted for when presenting the results. Failing to perform any of these tasks may imply improper results.

When performing a user experiment where the user should provide an opinion on an event, the users’ situation is important to consider. Important aspects of the subjects are for example background, expectations on the event, and previous experience of the event. Such aspects have to be taken into account.

There is a known issue with user experiments on network applications or services, namely the memory effect [32]. This issue needs to be properly mapped so that the implications can be accounted for, either when designing the experiment or when interpreting the results.

Grading perception using a scale means that the scale is interpreted by the user. When designing a user experiment the interpretation of the scale have to
be taken into consideration. Measures must be taken to assure that all users interpret the scale equally, regardless of the scale being text-based or number-based. For example, ranking “1” can be equivalent to the service being so bad that the user stops using the service. Alternatively, ranking “1” can be equivalent to the service performing so badly the user becomes extremely annoyed.

5.2 PERCEPTION OF SECURITY

In Paper IV and Paper V, actual and perceived authentication security based on response time, and perceived password security based on actual password security were investigated. This section will differentiate between investigating perception of login security and perception of password security.

There are four major differences between the password security survey and the user experiments on login security.

1. The selection of the user group and the resulting demographics.
2. The user input: grading versus yes/no, and interpretation possibilities for the user/subject.
3. The direct or indirect connection of the user input to security.

The selection of user groups was done differently in the different studies of this thesis. For Paper II, test subjects were recruited amongst students at Blekinge Institute of Technology. The demographics of this user group are random in many aspects. All test subjects who chose to participate rendered results for the experiment. For Paper IV and Paper V, test subjects were pre-defined by the event in which the experiment was included. The demographics of this user group are homogeneous in many aspects. The test subject who chose to participate in the experiment rendered results to the experiment. The test subject who chose not to participate could be accounted for in the evaluation.

User ratings can be different for different test subjects, based on the interpretation of the scale. Interpretation is done for both numbers and words, e.g., the rating “2” can be interpreted as “Poor” by one test subject and as “Bad” by a second subject. The rating “2” can be defined as “Poor” to the test subjects, but
then there may be an interpretation bias on the word “Poor” between different subjects. A less elaborate scale could be yes and no, or secure and insecure. By using a less elaborate scale the interpretation space is narrowed. If the difference between secure and insecure is defined, then the interpretation space is further narrowed.

The user perception of login security is investigated by evaluating the user perception of response times and calculating the response time. The different test subjects' expectations on how security is connected to response time are unclear. The user perception of password security is investigated by evaluating the user perception of their password's strength and by estimating the actual strength of the password. This gives a direct connection between the evaluations.

5.2.1 PERCEPTION OF LOGIN SECURITY

When establishing a user model for QoE of QoS, there are many parameters that can be taken into consideration. Such parameters are, beside response time, for example parameters that effect user expectations of the login or service, or different kinds of user background parameters. If the user model should be used as a user model of perception of login, for users in general and for login in general, all parameters must be taken into consideration.

If such parameters are not taken into consideration, a user model can still be used, but only for the specific setup that was used when establishing the user model and in some cases only for the enquired user group. If the demographics of the enquired user group correctly represent the demographics of a larger group of people, then the corresponding user model can be used for the entire larger group.

The user model that is established in Paper II, is used in the two succeeding papers, using the same setup. An important factor to point out is that the users targeted in these papers are not general users, but users with similar demographics as the users who were enquired when establishing the user model.
5.2.2 Perception of Password security

A general and valid user model is hard to establish, even for one single service. Therefore, a different viewpoint was elaborated and evaluated. The actual and perceived security of users was investigated as two different parameters in Paper V. The compared results indicated a gap between actual and perceived password security for some test subjects.

The evaluation was performed with teenagers from several Swedish sixth form colleges, which is the last three years of mandatory school in Sweden. Teenagers in sixth form college are generally between 16 and 19 years old. The teenagers’ actual password security and their perceived password security were investigated using a survey. There are several issues when using surveys as a method for evaluating user perception. Several issues are presented and accounted for in Paper V, section 6.3 Validity threats.

Even though the results of the password security experiment can be more broadly used, there are limitations. The Swedish primary and sixth form schools follow the same educational planning, in which is does not say anything about teaching online security. Outside of Sweden, there can be different guidelines for teaching online security. Thus, an investigation of educational planning is needed before considering applications of the results globally. An investigation of similarities and differences in demographics is also needed before considering applications of the results globally.
This chapter summarises and presents the main contributions from the studies in this thesis.

- A framework for authentication scheme evaluation.
- A user model for authentication towards a community using OpenID.
- Identification of a gap between actual and perceived password security amongst Swedish teenagers in sixth form college.

A framework for evaluating authentication schemes was developed and presented in Paper I [1]. The framework covers targeted solutions, detailed and mapped criteria for evaluation and methodology alternatives. There is also a possible usage area for the framework, or parts of the framework, in other authentication and security related user interfaces and applications.

A user model was founded in Paper II [2] and used in follow-up Paper III [3]. The user model covers login to a community using OpenID and SIM-card for our user group. Conclusions based on the model indicate that the users are more patient when waiting for a login to finish than users are when waiting for a regular web page or other IT services to finish loading [31]. Studies with similar user groups can use the user model for benchmarking their system or for comparison.
A lack of awareness of password security amongst teenagers in sixth form college was first suspected through the study presented in Paper IV [4]. A gap between actual and perceived password security was then confirmed in the study presented in Paper V [5]. In the study, about 75% of the teenagers who perceived their passwords as strong actually had a weak password.
Chapter 7

Conclusions & Outlook

The studies in this thesis are aimed at understanding users’ perception of authentication. Response time experiments, user experiments and surveys were chosen as methods of evaluation by first summarising how to evaluate authentication schemes. A framework for evaluation of authentication schemes stands as the methodological basis for the rest of the studies.

The results presented in this thesis give a broad picture of how users perceive authentication: from users’ perception of the performance of authentication on web pages to users’ perception of their own contribution to password security. Some of the results need further study in order to be generally applicable, while other results are generally applicable already.

In this thesis, the areas of user perception and authentication are brought together. The combination of user perception and authentication is not entirely novel regarding password security. However, the combination of methods used is novel. Investigating QoE of authentication from both a performance and security point of view with user experiments was also novel when conducted. Furthermore, investigating user passwords by asking for the characteristics of passwords that are currently in use, instead of asking for the entire password, is novel.
CHAPTER 7. CONCLUSIONS & OUTLOOK

Some of the later results of this thesis indicate that teenage users in Sweden are not prone to enhance their personal security online. The teenagers show a lack in awareness of their personal security online. Furthermore, there are indications that the education of online security in sixth form college is inadequate.

There are several open issues related to the studies of this thesis. Some issues are directly related to the studies of this thesis. Some of the issues are loose ends which were discovered in studies of this thesis, but were impossible to undertake within the scope or time limitations of this thesis. Thus, some suggestions for future work are suggested below.

- When logging in to a web page, some initial test subjects for the studies in Paper II [2] have indicated that insignificant response times are perceived as unsecure, while perceived as well-performing. Implications and consequences of this difference would be interesting to investigate.
- The documented memory effect [32] can be accounted for when designing user experiments. To evaluate and publish mapping and designing to use as a basis when planning user experiments for QoE would be valuable.
- For EAP-SIM some parts of the authentication procedure are more sensitive to bad network performance. This sensitivity makes the protocol scale bad with network problems. It would be interesting to investigate if this scalability issue could be resolved by altering the protocol to send fewer or smaller messages back and forth, without altering the security of the protocol.
- Are the problems of passwords disappearing with passwords or are the passwords not disappearing as soon as one would hope? New methods for authentication are developed, tested, and used. Despite the new and sometimes more secure methods, passwords persist. Even though research is needed for new methods, a solution for the problems of passwords and password usage would be interesting to find.
- The gap between actual and perceived password security will probably have to be solved. Whether the gap in knowledge is regarding passwords, risks, security or different authentication methods, people will get in trouble for not knowing enough. Investigating how to reduce the gap would be interesting.
BIBLIOGRAPHY


35
BIBLIOGRAPHY


PAPER I

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 and Ivar Jørstad*

Blekinge Institute of Technology, Karlskrona, Sweden, (charlott.eliasson, markus.fiedler)@bth.se

*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 self-evident 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
For security regarding authentication the following subcriteria 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 itself, 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.
3 EVALUATION CRITERIA

- 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


http://www.dreamhack.se.


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

Charlott Lorentzen, Markus Fiedler, Henric Johnson, Junaid Shaikh and Ivar Jørstad*

School of Computing, Blekinge Institute of Technology, Karlskrona, Sweden. Email: <firstname.surname>@bth.se

*Ubisafe AS, Oslo, Norway
Email: ivar@ubisafe.no

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/\(<\text{username}\)> , 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

![Figure 1: Average RTs mapped to the corresponding delay added in shaper.](image-url)
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 included in the methodology (see Fig. 3) of the larger study, to quantify user experience. 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
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.

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>

Figure 3: Methodology for the study of authentication schemes [4].

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.

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.
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.

Figure 4: OS for corresponding RT for confirm test with different trendlines.
5 RESULTS

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

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 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 $-0.638$).

### 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.”
- “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
### 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>

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 expectations, like network connection and different kinds of user background.

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.

**Figure 6**: User perception profile: Frequency of OSs from RTs for different BWs.
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
REFERENCES

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

Published in Journal

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

Charlott Lorentzen, Markus Fiedler and Peter Lorentzen

School of Computing, Blekinge Institute of Technology, Karlskrona, Sweden. Email: <firstname.surname>@bth.se

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 increases of the latter, we emulate bad network performance by introducing bi-directional delay between the supplicant (client) and the authentication server. The same procedure is also applied to several other EAP methods. 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 EAP-SIM authentication contributes significantly more to the total response times than the other tasks, 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 give the greatest contribution to RT, when authenticating via OpenID. Furthermore, the study will find the decisive factors of the following authentication methods: EAP Message-Digest algorithm 5 Challenge (EAP-MD5), EAP Tunteled Transport Layer Security (EAP-TTLS) with MD5, EAP-TTLS with Password Authentication Protocol (PAP), EAP-TTLS with Challenge Handshake Authentication
Protocol (CHAP), EAP-TTLS with Microsoft CHAP version 2 (MSCHAPv2), and Protected EAP (PEAP) with MSCHAPv2. The decisive factors for each authentication method will then be compared.

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 knowledge, the combination of EAP-SIM authentication method with OpenID for web authentication has not been investigated yet, nor have there been studies done on decisive factors of different EAP methods.

The organization of this paper is as follows: Section 2 gives an overview of the authentication method and services used for the OpenID EAP-SIM authentication. Section 3 discusses the impact of different network parameters and describes the methodology of the study, and Section 4 describes the OpenID EAP-SIM authentication experiments with corresponding setup, procedure and RT measurements. The results of the latter study 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. Section 7 describes an additional study of decisive factors for several authentication protocols, and presents and discusses the results. Finally, Section 8 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

81
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).
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 $T_{ik}$ denote internal durations within the supplicant, while $T_{nk}$ 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 authentication is a valid method for both supplicant and server (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.
response 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.

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
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-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

![Figure 3: RT components versus bi-directional delay added in the shaper.](image)
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 for EAP-SIM

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:

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

(5)

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

\[ \sum_{k=1}^{3} T_{Ik} \approx 1.8 \text{ s}, \]  

(6)
Thus, equation 5 can be rewritten as

\[ QoE \approx 4.7e^{-0.18}e^{-0.1T_N/s} \approx 3.9e^{-0.1T_N/s} \] (7)

which yields

\[ QoE \approx 3.9e^{-2.2d/s} \] (8)

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

\[ QoE \approx 3.9e^{-1.6d/s}(e^{-0.2d/s})^3, \] (9)

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 \( T_{N2}, T_{N3} \text{ and } T_{N4} \) on QoE. It can be seen that the first part has a significantly higher impact:

\[ \gamma = \frac{e^{-1.6d/s}}{e^{-0.6d/s}} = e^{-d/s} < 1 \] (10)

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, \( T_{N2}, T_{N3} \text{ and } T_{N4} \), 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 and secure tunnel between the supplicant and the authenticator in OpenID with EAP-SIM that does not scale nicely with an increasing delay.

7 Performance of other EAP methods

In this study the authentication methods EAP-MD5, PEAP-MSCHAPv2, EAP-TTLS-PAP, EAP-TTLS-MD5, EAP-TTLS-CHAP, and EAP-TTLS-MSCHAPv2 were studied in a laboratory environment.

The experiments were done in a similar, but laboratory, environment, which excluded the Internet but included all other parties as in the previous study (see Fig 6). Though, the authenticator and the authentication server are two
Figure 6: Setup of the EAP authentication experiments.

separate devices in this setup. The traffic shaper is placed between the authenticator and server, in this experiment setup. The new setup has the network traffic shaper on the server side of the authenticator, since the delay is introduced on the network layer (IP layer), and the traffic between supplicant and authenticator is not sent using IP addresses. This setup gives the same result in addition of network delay as in the previous study for all tasks, except for the first task including the two initial messages, which are only sent to and from the authenticator.

Time stamps were logged during all experiments; both at the supplicant side and the server side, and RTs were calculated for each run in each experiment. The experiments were run 40 times for each delay setting. The network delays added in the network shaper in this experiment were 0 ms, 20 ms, 40 ms, 60 ms, 80 ms, and 100 ms, in both directions.

The EAP methods were analyzed to find the decisive factors. The RTs in these additional experiments are considerably lower, but the linear behavior in RT increase with increased delay shows the same “per packet” proportionality as in the above study.

7.1 Authentication method overview

EAP is the underlying protocol for authentication procedure and TTLS is used to setup a tunnel for the authentication. The tunnel is established between the supplicant and the authenticator for secure data, and between supplicant and
Figure 7: RT components for EAP-MD5 versus bi-directional delay added in the shaper.

Figure 8: RT components for EAP-TTLS-PAP/EAP-TTLS-CHAP versus bi-directional delay added in the shaper.
PAP, MD5, CHAP, and MSCHAPv2 are the authentication algorithms used in the authentication procedure. The security level is a bit different for these algorithms. In PAP the user name and password are sent unencrypted. Without a tunnel PAP would, in other words, be insecure and therefore it is not supported by EAP itself, but by TTLS. In TTLS the tunnel hides the password, so sending it unencrypted is not a security issue. Though, the password is stored as a hash at the authentication server side and can therefore not be stolen.

CHAP was standardised for PPP from the beginning and is only supported by TTLS, since it is not an EAP method. The authentication server challenges the supplicant, which responds to the challenge by proving the possession of a shared secret, i.e. a password. The password is sent as a hash over the network, but must be available in plain text at both the supplicant side (by typing it) and the authentication server side (in a stored file, e.g. /etc/passwd) in order to confirm that the hash in the challenge response is valid.

EAP-MD5 was standardized with EAP from the beginning. MD5 is supported by EAP and can be used with EAP, PEAP or TTLS. Though, the password needs to be available on both supplicant side and authentication server side, as in CHAP. When MD5 is used as EAP-MD5, the ID is sent from the beginning, but when it is used in EAP-TTLS-MD5, the supplicant is anonymous in the beginning and the ID is instead sent in the authentication phase, which is reflected in the response times for those phases, as there is one more roundtrip from supplicant to server for this task.

MSCHAPv2, like MD5 and CHAP, sends a hash of the password, but in this case it does not have to be stored in plain text on both sides. A particular one-way hash of the password, which will then serve as the password, is stored at the authentication server side. The supplicant, who knows the hash algorithm, will be able to produce a matching “password” from the original password, which can be used in the response of a challenge from the authentication server. MSCHAPv2 also provides mutual authentication. MSCHAPv2 can be used with EAP, PEAP and TTLS.
Figure 9: RT components for EAP-TTLS-MD5/EAP-TTLS-MSCHAPv2 versus bi-directional delay added in the shaper.

Figure 10: RT components for PEAP-MSCHAPv2 versus bi-directional delay added in the shaper.
7.2 Results and Analysis

For each method the messages have been separated into four tasks, and what is included in the RT for each task is are presented in a list below. The RTs are all have the supplicant as base. All methods include all the four tasks, except for EAP-MD5 which does not setup a tunnel for the communication, and thus lacks the response time increase from $T_{n3}$.

- $T_{n1}$: Initiation of the connection with the authenticator.
- $T_{n2}$: Negotiation of authentication protocol with the authentication server.
- $T_{n3}$: Setup of a secure tunnel to the authenticator and authentication server.
- $T_{n4}$: Authentication with challenge(s) and response(s).

One of the authentication methods, namely EAP-MD5, does not have a setup of a tunnel. In this method the authentication challenge has the longest RT, $T_{n4}$ (see Fig. 7). The increase in RT of the decisive task, $T_{n4}$, is four times the added delay, and since the delay is added in both directions four times means two round trips between the supplicant and the authentication server.

For EAP-TTLS-PAP and EAP-TTLS-CHAP the task with the RT of largest impact is the setup of the TTLS tunnel (see Fig. 8). The RT of this task, $T_{n3}$, increases with six times the added delay. The increase in RT for $T_{n3}$ is just two times the added delay larger than for the task with the next largest RT, $T_{n2}$, which is the time for the negotiation of authentication protocol. The increase in response time for the authentication challenge is only two times the added delay.

EAP-TTLS-MD5 and EAP-TTLS-MSCHAPv2 also have the largest impact on the total RT from the setup of the TTLS tunnel, with the same increase in RT. The RT of this task, $T_{n3}$, increases with six times the added delay (see Fig. 9), whereas the tasks that has the RTs with the next largest impact on the total RT, namely $T_{n2}$ and $T_{n4}$, are four times the added delay.
The EAP method that has the task which distinguishes itself the most, in terms of largest impact on the total RT, is PEAP-MSCHAPv2. Though the next largest RT, $T_{n3}$, increases with six times the added delay, the RT of $T_{n4}$ increases with ten times the added delay (see Fig. 10).

Fig. 11 shows the RTs with the largest impact for each of the EAP methods in this additional study. As seen in the figure, PEAP-MSCHAPv2 is the method that has the highest RT with $T_{n4}$. EAP-MD5 has the lowest RT with $T_{n2}$, which in fact is lower than the next largest RT for PEAP-MSCHAPv2, and equal to the next largest RT for EAP-TTLS-MD5, and EAP-TTLS-MSCHAPv2.

The decisive factors for each of these methods are derived from the RT of the task with the largest impact on the total RT. The $T_n$ for each method is translated into a dependence on delay based in Equation 5. The decisive factors for each method are presented in Table 1.
7.3 Simplicity versus Security

In a previous study [16] the compromising relationship between simplicity and security is discussed. If one party is increased the other party will sooner or later suffer from degradation. In this section we look at the relationship between

<table>
<thead>
<tr>
<th>Method</th>
<th>Security level</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAP-MD5</td>
<td>$e^{-0.4d/s}$</td>
</tr>
<tr>
<td>EAP-TTLS-CHAP</td>
<td>$e^{-0.6d/s}$</td>
</tr>
<tr>
<td>EAP-TTLS-PAP</td>
<td>$e^{-0.6d/s}$</td>
</tr>
<tr>
<td>EAP-TTLS-MD5</td>
<td>$e^{-0.6d/s}$</td>
</tr>
<tr>
<td>EAP-TTLS-MSCHAPv2</td>
<td>$e^{-0.6d/s}$</td>
</tr>
<tr>
<td>PEAP-MSCHAPv2</td>
<td>$e^{-1.0d/s}$</td>
</tr>
<tr>
<td>EAP-TTLS-CHAP</td>
<td>$e^{-1.6d/s}$</td>
</tr>
</tbody>
</table>

Figure 12: Security vs. complexity for several authentication methods, with 100 ms added delay.
security and simplicity with regard to the six EAP methods, described and analysed above.

The EAP methods have been arranged from highest security level (6) to lowest security level (1) in a simple manner (see Table 1). For example, a plain text password is, within this comparison example, considered to provide a lower security level than a hashed password, and a tunnel is considered to add security. Response time will, in this example, be compliant with complexity, which is the inverse of simplicity. The higher the RT, the lower the simplicity is considered to be. The RT order of magnitude is due to the number of messages sent back and forth for each task.

In Fig. 12 it can be seen that the simplicity is increasing with the decrease in security. The EAP method that is considered to have the lowest security level, EAP-MD5, is in fact the simplest method, and the PEAP-MSCHAPv2, which is considered to have the highest security together with EAP-TTLS-MD5 and EAP-TTLS-MSCHAPv2 has the least simplicity. This is of course a generalisation and the model is not exactly compliant with every existing authentication method. Though, it gives a rough picture of the options that exist when choosing an authentication method. If the situation or solution requires a high level of security, then it can be justified to add complexity to the system, even though it might give higher RTs.

8 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 for OpenID authentication with EAP-SIM, it
could be clearly seen that one of the network times is growing faster than the others, namely the one that refers to the initiation and setup of a secure connection between the supplicant and the authenticator, followed by an ID request from the authenticator. The fact that this initial task has the largest contribution to the total RT and therefore also had the greatest impact on QoE was shown from several angles. The initial is also the task with most packets sent back and forth.

The results for OpenID using EAP-SIM were then connected to the QoE user model that was developed in a previous study of the same system. From this mapping it was shown that the increase in RT for the initial secure connection setup task would result in a degradation of QoE to the extent that it reaches a rating of 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 other EAP based authentication methods were found to have the lower decisive factors than the OpenID solution with EAP-SIM. The largest contribution to the total RT for PEAP-MSCHAPv2 is given by $T_{n4}$, which is the authentication phase including the challenge. The method that had the decisive factor with the lowest impact was EAP-MD5. In EAP-MD5 the initiation of a connection between the supplicant and the server gave the largest impact on RT, with $T_{n2}$. The rest of the authentication methods had almost equal decisive factors, with the largest impact in RT from $T_{n3}$, i.e. the setup of the secure tunnel.

After comparing the simplicity and security levels of the EAP based authentication methods, the compromise between security and simplicity was shown in a quantitative manner. Adding security level will compromise the simplicity in most cases, depending on how the increase in security level is defined.

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, a suitable traffic shaper has to be found before it can be further evaluated. Variable delay will perhaps result in a bit lower RT than constant delay, but that needs 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 for OpenID using EAP-SIM needs to be examined 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.

Acknowledgements

The authors would like to thank Dr. Ivar Jørstad at Ubisafe AS in Olso, Norway, for providing and supporting the authentication server setup, and Johan Lindh for his assistance in gathering parts of the underlying data.

References


proceedings of Australasian Telecommunication Networks and Applications Conference (ATNAC 2010), Auckland, New Zealand, November 2010.


PAPER IV

On User Perception of Safety in Online Social Networks

Published in journal
International Journal of Communication Networks and Distributed Systems
Special Issue on: “Modelling of Emerging Internet Services: Social Networks and Crowdsourcing”
2013
On user perception of safety in online social networks

Charlott Lorentzen, Markus Fiedler and Henric Johnson

School of Computing, Blekinge Institute of Technology, Karlskrona, Sweden. Email: <firstname.surname>@bth.se

Abstract – Today, many people share personal information through online social networks. While these networks offer tremendous facilities to share one’s life with family, friends and the outer world, users are also exposed to the risk to get their accounts compromised, with consequences for their professional and private lives. A strong password can help to ensure privacy, while in contrast, a weak password may lead to a compromised account. On the other hand, it is observed that users tend to shortcut laborious and time-consuming security measures, which alleges that their perception of their safety on the social network might deviate from the real safety situation.

In this article, we investigate how users of social networks perceive safety, and to which extent they contribute to it. First, we examine the question in which way and to which extent a user’s perception of safety is affected by increased response times. An investigation was conducted for teenage users of the online social network Facebook. Its results indicate that the users’ perception of safety differs significantly in face of increased response times that usually imply worse Quality of Experience. Second, in order to find out to which extent users are aware of the risks associated with weak passwords as well as what measures have been taken to keep themselves secure and to ensure privacy, a survey about password complexity in online social networks was conducted with Swedish users. Our results of the survey indicate differences in password complexity between teenagers with different levels of technical education.
1 Introduction

Online social networks (OSNs) represent a platform for connecting, sharing and interacting with other people. During the last decade, the interest in OSNs has grown rapidly. From comprising mostly young users, OSNs now encompass a large variety of users. It is obvious that OSNs have a great impact on society. Not only do young people live parts of their lives through OSNs, but OSN activities may also, for example, affect the chances of getting employed, as employers nowadays use the web and social networks as reliable sources of information about potential employees.

Anecdotal evidence suggests that there is a discrepancy between the perceived and actual impact of OSNs on peoples’ lives. Users may perceive OSN data as something that needs to be secured, but studies suggest that there is a discrepancy between perceived and actual safety [11], and users may choose passwords that are weak while overestimating the password protection mechanism [8, 9]. During the Swedish computer festival Dreamhack Winter 2008, teenagers were asked to participate in an experiment to login to Facebook, and the whole session was recorded for analysis purposes. Merely two persons did not want to participate because of the recording of their password. While participating in the experiment, many of the test subjects said that they care neither about their passwords, nor about the security in OSNs.

This study aims at improving our knowledge with regards to these issues. In particular, we investigate (a) how users of social networks perceive safety, and (b) to which extent they contribute to it. It is seminal work, expected to open up for future and deepened work on the topic of user perception of safety, in particular in the context of OSNs.

Regarding question (a), we observe that login procedures are usually associated with response times for the users, in particular if the entered password needs to be sent via the network and processed remotely. Earlier studies have shown that rising login response times (i.e. the time that elapses between the start of the authentication process by the user, and the completion of the process in the user interface) affect the user-perceived Quality of Experience (QoE).
in a negative way. We will investigate in which sense and to which extend the user perception of safety depends on the login response time. Here, a growing response time can imply that a user feels more or less safe.

Regarding question (b), the agnostic attitude of OSN users with respect to security in general and to their passwords in particular led us to conduct a survey of Swedish Facebook users and their passwords. The study focused on teenagers during a theme day called Researchers’ Night [4]. A similar, but more extensive survey was conducted at an institute of technology in Sweden. Both surveys aimed at providing insight into how Facebook users constructed their passwords, and to what extent they use identical passwords for their private e-mail and Facebook accounts. A previous study [1,2] shows that 75% of Facebook users use identical passwords for their e-mail and Facebook account. Our survey of password complexity amongst teenagers suggests that the level of awareness and knowledge of security might be related to password complexity.

The remainder of the article is organized as follows. In Section 2, we clarify a set of terms used in this study. Section 3 presents the research methodology. We present the results in Section 4, and provide a discussion in Section 5. Finally we draw conclusions and present future work in Section 6.

2 Terminology

The notion password complexity represents a quantitative measure of how strong a password is. A password is constructed by using one type, or combinations of several types, of characters. These types could be upper- and lower-case literals, integers and/or special characters.

Quality of Experience (QoE) [12] refers to how the user perceives or experiences the quality of a particular application, service, or task, in our case based on the end-to-end response time. The conditions of the latter can be changed due to settings or problems with different Quality of Service (QoS) parameters, such as delay, loss, throughput, and so on.

Changes in QoS parameters can give rise to higher response times. A re-
sponse time it defined in this study as the time from entering a user request to receiving the full response to that request, as seen from the user point of view. This means that a response time of a login starts when the user clicks to login, and it ends when the page, to which the user is logging in, is loaded.

In this study, the notion of security represents the technical part of a login, namely the mechanisms, the algorithms and the procedures themselves.

User perceived safety represents the perception for a user when experiencing a procedure including security, or an application, service or procedure that is accessed via a security procedure or mechanism.

3 Research methodology

The surveys in this study are carried out by sending out questionnaires to users. The questionnaires are responded to, collected and analyzed.

3.1 User perception of safety survey

In our earlier work [14], user models have been explored for QoE of login sessions. These have been derived from results of user experiments. Users have been asked to rank their perception of different response times for logging in to an OSN for sharing pictures, and on a server page which was using OpenID [15]. OpenID was used due to its authentication service, with a trusted server (the OpenID server) that authenticates the users towards the web pages that they want to access. Only one authentication is needed from a user, and that is towards the OpenID server when starting a session.

The QoE of a login session, or the user perception of the response time of a login session, were previously modeled as the changes of the QoE with regard to the changes in RT. The QoE is represented by the Mean Opinion Score (MOS) [13]. The MOS is a mean value of the opinion scores given by users and stretches from 1 to 5, where 5 represents the highest grade. The QoE of a login session in form of user ratings of login response times was studied as follows: The
response time perceived by the users was controlled through configurable extra delays introduced into the path between client and server. For each authentication, the users ranked their experience of the corresponding response time on the numerical MOS scale. The value pairs (response time, rating) were plotted, and regressions were applied in order to identify the trends the ratings (representing QoE) follow as function of the response time (representing QoS). The user models of QoE are expressed as \( QoE = f(QoS) \). Reference [14] contains a set of QoE models obtained for login procedures. Amongst the linear, logarithmic, exponential and power matchings, the exponential relationship [7, 10] is taking a prominent position. Besides of being the only basic relationship yielding very similar formulae in two independent experiments, summarized by

\[
QoE \approx 4.7e^{-0.10T_R/s}, \tag{1}
\]

it has the appealing property that additive delays are turned into factors affecting the QoE. Indeed, according to (1), each extra second of response time costs approximately 10% in terms of user ratings. The latter can be compared to the web user model [19] expressed by

\[
QoE \approx 4.8e^{-0.15T_R/s} . \tag{2}
\]

Here, each extra second of response time decreases the user ratings by approximately 15% by trend. Although users are obviously more patient when it comes to waiting for logins as compared to waiting for web pages, an increasing login response time has a negative impact on user perception.

In the user perception part of this study, the users from the teenager group that participated in the password complexity survey were asked how safe they felt when logging into Facebook. A questionnaire was used in order to enable quantification of user perception of the login procedure. A delay was added on the client computer, in order to yield higher response times. The login was performed five times, and for each trial the delay was changed and the participants were asked to give a rating on a continuous scale, going from “completely insecure” to “completely secure”. A continuous rating scale with minimum and
maximum given was used, in order to give the users more flexibility, i.e., not to lock them on a scale with only a few numerical or verbal options. The sequence of the delay was 0 s, 2 s, 6 s, 0 s and finally 4 s, with one delay representing one trial. A sequence of continuously increasing delay, from 0 to 6 s, was rejected due to a previously noticed anticipation, among test users, of the next trial to have a higher response time. The latter could, for some test users, result in a lack of participation at the end of the survey, yielding non-interpretable answers on the questionnaire.

After collecting the data, the continuous scale was quantified between 0 and 100. After analyzing the results for each user, the ranking behaviour was categorized into one of four categories, as follows:

- **High**: Higher safety ratings for longer response times;
- **Low**: Lower safety ratings for longer response times;
- **Same**: Same safety rating all through, with a variation of less than $(\pm)4$;
- **Misc**: Miscellaneous, cases not covered by the previous categories.

These survey results are not claimed to be statistically significant results. Though, they can be used as an indication as to whether or not this issue should be further investigated.

### 3.2 Password complexity survey

**Teenagers** participating at the Researchers’ Night were given a questionnaire that included questions about age, gender, study programme and time spent for surfing per week. Respondents were also asked if they had a Facebook account, what kind of characters their password contained, and whether they used identical passwords for their private e-mail and Facebook accounts.

**Staff** from a Swedish institute of technology composed a second user group in the survey. In a second questionnaire, they were asked about age, gender, department, employment form, subject, education, and year of graduation. Further they were asked if they had an account on Facebook or another OSN. If
that was the case, they got to answer which type of characters the password contains and if they use identical passwords for their private e-mail and Facebook accounts. The two questionnaires contained questions that addressed the same type of information, namely information on password complexity and personal information. The data were collected and stratified in groups of study programmes and gender.

The password complexity was realized by assigning one credit for each type of character that the password contained. The recognized types of characters are upper case letters, lower case letters, numbers and special characters. The latter character type includes all characters that are not included in the other three categories. One credit was also given for not using identical passwords for Facebook and e-mail account. A password will receive a minimum of one (1) credit and a maximum of five (5) credits. For instance, the password “AZaz19#!” will yield maximum credits (five credits) if a user uses it for Facebook but not for an e-mail account.

4 Results

This section describes the user groups and the results of the two surveys of this study. The results of the user perception of safety survey are presented in Section 4.1 and provide a complementary view on earlier results on QoE for login sessions. In Section 4.2 the results of the password complexity survey are presented.

4.1 User perception of safety

The teenagers on Research Friday were asked to rate their perception of safety of different response times for login on Facebook. From the 142 teenagers that answered this survey, 89 responses were accurate and used as result. 53 out of the survey answers were answered incorrectly or not answered at all, due to misunderstandings or lack of interest in the survey.
In Figure 1 and 2 it can be seen that different groups of users rate differently for an equal addition in delay. In the first category “High”, higher response times
are interpreted as more secure communication, despite of potentially lower QoE as indicated by equation 1. In other words, we observe a trade-off between user perceived safety and user perceived QoE. 10% of the user ratings belong to this category.

In the second category “Low”, higher response times are interpreted as less secure communication, which is the same type of trend as for the QoE. For this user group a short/fast authentication procedure is preferable, in terms of both user perceived safety and QoE. The ratings of user perceived safety in category Low are complementary to those of category High (see Figure 1). 36% of the user ratings belong to this category.

In the third category “Same”, the mean ratings are hardly dependent on changes in response time and correspond to the best ratings in category “Low”, yielded for vanishing response times. The users in this category appear to be agnostic to response times. 22.5% of the user ratings belong to this category.

The user results that were categorized as “Misc” did not seem to follow any identifiable pattern. The average ratings are found below the corresponding ratings in category Same with a rather small dependency on the response time. 31.5% of the user ratings belong to this category.

The standard deviation (see Table 1) varies from 14.2 to 25.5, and for the case with same ratings in all trials (Same) it varies from 14.2 to 14.8. One reason for the standard deviation being lower for the group Same is that many respondents in this group gave high ratings (e.g. 100) in all trials. The latter can also be seen in the mean rating for the group Same, which is higher in all trials than for all the other groups. Obviously, this category contains ratings from users that might not care at all about the experiment.

4.2 Password complexity survey

The first user group of teenagers from Researchers’ Night for the password complexity survey included a total of 142 teenagers from secondary school, in the ages of 17 to 19 years. 137 out of the 142 collected questionnaires were answered according to specifications and used for analysis of password complexity.
Table 1: User perception of safety with regard to response time, for Facebook login, shown in trial order.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>43.1 (16.7)</td>
<td>53.1 (24.9)</td>
<td>65.0 (22.7)</td>
<td>47.5 (22.0)</td>
<td>63.1 (24.1)</td>
</tr>
<tr>
<td>Low</td>
<td>77.6 (17.4)</td>
<td>66.0 (22.3)</td>
<td>45.0 (18.4)</td>
<td>77.0 (17.4)</td>
<td>51.7 (25.5)</td>
</tr>
<tr>
<td>Same</td>
<td>82.9 (14.7)</td>
<td>82.8 (14.3)</td>
<td>83.4 (14.2)</td>
<td>82.5 (14.8)</td>
<td>82.8 (14.3)</td>
</tr>
<tr>
<td>Misc.</td>
<td>69.8 (21.8)</td>
<td>70.0 (18.5)</td>
<td>67.1 (22.4)</td>
<td>71.9 (19.0)</td>
<td>65.4 (25.5)</td>
</tr>
</tbody>
</table>

Each column is a trial with a certain delay in seconds, trial number @ delay.

and identical passwords, while five were answered incorrectly or not answered at all.

There were 67 male and 70 female teenagers (see Table 2 for group details), out of the 137 valid responses, and most of them attended the study programmes: Commerce, Hotel and Restaurant, Natural Science, and Technology. Two students attended the Social Science study programme. The questionnaires were stratified in groups according to the level of technology that is taught in the study programme. Commerce, Hotel and restaurant, and Social Science were called group H; Natural Sciences is one group called N; and the last group is called T and includes Technology.

In the second user group of staff at a Swedish institute of technology for the password complexity survey, 106 questionnaires were received. Out of the 106 questionnaires, 68 were answered according to specifications and used for the analysis of password complexity, whereas 38 were answered incorrectly or not answered at all. For the analysis of identical passwords, 79 questionnaires were answered according to specifications and used, but 27 were not. For the analysis of identical passwords, one requirement was that the staff member had an account for Facebook. This was considered when checking for validity of the answers.
**Table 2:** Demographics of teenage participants in the password complexity and identical password surveys.

<table>
<thead>
<tr>
<th>Teenager survey</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group H</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
</tr>
<tr>
<td>Male</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
</tr>
</tbody>
</table>

1 H = Commerce, Hotel & Restaurant, Social Science; N = Natural Science; T = Technology

**Table 3:** Demographics of staff member participants.

<table>
<thead>
<tr>
<th>Staff member survey</th>
<th>Number of participants (Identical passwords)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Number of participants (Password complexity)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
</tr>
</tbody>
</table>

2 A = administration; B = healthcare and sociology; C = planning and cultural aspects of technology; D = technology
Out of the 68 answers used in the analysis of password complexity, there were 30 male and 38 female staff members (see Table 3 for group details). For the identical password analysis, there were 50 female and 29 male staff members out of the 79 used answers (see Table 3 for group details). They are, in both cases, affiliated with one of the departments, in this survey called group A1, A2, B1, B2, C1, D1, and D2. The departments denoted Ai are associated with administration, Bi are associated with healthcare, psychology and sociology, C1 is associated with the cultural aspects of technology and with planning, and Di are associated with technology. Notable for group B1 is that there are no male staff members participating in this survey, which is perspicuous as almost 80% of the staff in this department are females.

4.2.1 Password complexity credits

Figure 3 reveals differences between the teenagers of the three groups. It can be seen that the mode of the credits, yielded by passwords in group T, is at four (4) credits, whereas for group H and N the mode is at three (3) credits. When looking at Table 4, group T has a higher mean password complexity than group H and N, which was suggested by the modes in Figure 3.

Furthermore, the teenage participants were stratified according to gender. The trend for male versus female teenagers in password complexity is not obvious from Figure 4, where the numbers are different for each password complexity level. Though, statistical analysis (Table 4) indicates that male teenagers have a slightly higher mean password complexity in group H and N, but also when combining all three groups. However, in group T, the male and female teenagers have approximately the same mean password complexity. Standard deviations are found around 1 for all groups but for group H, for which they are a bit smaller, indicating more homogeneity.

For both male and female teenagers, group H has a lower mean password complexity as compared to group N, which in turn has a lower complexity than T (see Table 4). The trend of group T having higher password complexity than the other two groups can also be seen in general, for both genders together.
For the staff, the results in Table 5 indicate that the A and D groups have the highest mean password complexity when comparing both male and female staff members together. The highest mean password complexity for the female staff members was isolated to the D groups and then the A groups. When looking at the male staff members the results are a bit hard to interpret due to the small number of staff members in three of the groups. Though, the A groups have the highest mean password complexity for male staff members.

When comparing female and male password complexity within the groups, in Table 5, the D groups differ from the others in that the female staff members have a higher mean password complexity than the male staff members. The mean password complexity is higher for male staff members in all the other groups, except for group B1, which does not have any male staff members participating in this survey. For all groups together, the male staff members have
Figure 4: Password complexity for teenagers, divided in male and female. 
NOTE: Credits are given per type of characters included in the password 
and for not using identical passwords for e-mail and Facebook account. 
The total number of credits per password is used, and a larger number of 
credits indicate a more complex password.

4.2.2 Identical passwords for Facebook and e-mail account

In Table 6 it can be seen that teenagers from group T, to a slightly higher 
degree than group N, use identical passwords for Facebook and e-mail account, 
whereas group H has the lowest degree of identical passwords usage. For female 
teenagers in group T, the use of identical passwords is significantly higher (83 %), 
as compared to group N (42 %) and group H (36 %). For male teenagers, group 
H and T have approximately the same degree of using identical passwords, and 
group N has the highest degree.

Comparing male and female in general, including all three groups, female 
teenagers have a higher tendency of using identical passwords for Facebook and 
e-mail account. Though, for group T the difference between female and male
Table 4: Mean password complexity credits (1–5) for the Facebook account passwords used by the teenage participants.

<table>
<thead>
<tr>
<th>Teenage survey</th>
<th>Password complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Group H</td>
<td>2.29 (0.61)</td>
</tr>
<tr>
<td>Group N</td>
<td>2.82 (1.17)</td>
</tr>
<tr>
<td>Group T</td>
<td>3.33 (1.21)</td>
</tr>
<tr>
<td>All groups</td>
<td>2.76 (1.11)</td>
</tr>
</tbody>
</table>

3 H = Commerce, Hotel & Restaurant, Social Science; N = Natural Science; T = Technology

Credits are given per type of characters included in the password and for not using identical passwords for e-mail and Facebook account. The total number of credits per password is used. A larger number of credits indicate a more complex password.

teenagers of 52%, is significantly larger than the general difference of 8%. The differences between male and female are lower for groups H (6%) and N (3%).

It can be seen in Table 7 that 18% of the staff use identical passwords for their private e-mail and Facebook account. Group A1 is the department with the lowest degree of identical passwords usage, namely none, and group A2 has the highest degree of identical password usage. For female staff members, group C1 has the highest degree of identical password usage, at 50%, and for male staff members the corresponding figure is 33%, which is found in group B2.

Comparing male and female staff members within the different groups, it can be seen that group B2 is the only group in which males to a higher degree than females use identical passwords. In general, the female staff members use identical passwords to a higher degree than the male staff members. This is also reflected in the ratio for all groups together, where the female staff members have a usage of identical passwords of 22%, whereas for male staff members the corresponding ratio is 10%.
Table 5: Mean password complexity credits (1–5) for the Facebook account passwords used by the staff member participants.

<table>
<thead>
<tr>
<th>Staff survey</th>
<th>Password complexity</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Both</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Group A1</td>
<td>3.20 (1.10)</td>
<td>4.00 (*)</td>
<td>3.33 (1.03)</td>
<td></td>
</tr>
<tr>
<td>Group A2</td>
<td>3.62 (1.33)</td>
<td>4.17 (0.98)</td>
<td>3.79 (1.23)</td>
<td></td>
</tr>
<tr>
<td>Group B1</td>
<td>2.67 (1.00)</td>
<td>**</td>
<td>2.67 (1.00)</td>
<td></td>
</tr>
<tr>
<td>Group B2</td>
<td>2.50 (0.71)</td>
<td>3.50 (0.71)</td>
<td>3.00 (0.82)</td>
<td></td>
</tr>
<tr>
<td>Group C1</td>
<td>2.50 (0.71)</td>
<td>3.00 (*)</td>
<td>2.67 (0.58)</td>
<td></td>
</tr>
<tr>
<td>Group D1</td>
<td>4.00 (0.00)</td>
<td>3.25 (1.26)</td>
<td>3.57 (0.92)</td>
<td></td>
</tr>
<tr>
<td>Group D2</td>
<td>4.25 (2.22)</td>
<td>3.60 (0.74)</td>
<td>3.74 (1.15)</td>
<td></td>
</tr>
<tr>
<td>All groups</td>
<td>3.32 (1.22)</td>
<td>3.68 (0.91)</td>
<td>3.48 (1.12)</td>
<td></td>
</tr>
</tbody>
</table>

4 A = administration; B = healthcare and sociology; C = planning and cultural aspects of technology; D = technology

* Only one user in this category. ** No user in this category.

Credits are given per type of characters included in the password and for not using identical passwords for e-mail and Facebook account. The total number of credits per password is used. A larger number of credits indicate a more complex password.
4.2 Password complexity survey

Table 6: Percentages of teenagers that use identical passwords for their private e-mail and Facebook account.

<table>
<thead>
<tr>
<th>Teenager survey</th>
<th>Identical passwords</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
</tr>
<tr>
<td>Group H</td>
<td>36 %</td>
</tr>
<tr>
<td>Group N</td>
<td>42 %</td>
</tr>
<tr>
<td>Group T</td>
<td>83 %</td>
</tr>
<tr>
<td>All groups</td>
<td>44 %</td>
</tr>
</tbody>
</table>

*H = Commerce, Hotel & Restaurant, Social Science; N = Natural Science; T = Technology

Table 7: Percentages of staff members that use identical passwords for their private e-mail and Facebook account.

<table>
<thead>
<tr>
<th>Staff survey</th>
<th>Identical passwords</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
</tr>
<tr>
<td>Group A1</td>
<td>0 %</td>
</tr>
<tr>
<td>Group A2</td>
<td>38 %</td>
</tr>
<tr>
<td>Group B1</td>
<td>18 %</td>
</tr>
<tr>
<td>Group B2</td>
<td>0 %</td>
</tr>
<tr>
<td>Group C1</td>
<td>50 %</td>
</tr>
<tr>
<td>Group D1</td>
<td>20 %</td>
</tr>
<tr>
<td>Group D2</td>
<td>20 %</td>
</tr>
<tr>
<td>All groups</td>
<td>22 %</td>
</tr>
</tbody>
</table>

*A = administration; B = healthcare and sociology; C = planning and cultural aspects of technology; D = technology

*No user in this category.*
5 Discussion

The results from the survey of user perception of safety are somewhat inconclusive, and contradictory trends are observed. Since this study was performed only on groups of teenagers at a particular event, a more thorough study is needed to distinguish groups for different safety perceptions, or in any case trends or models for safety perceptions. A model for perception of safety would be an important tool in making OSNs more secure and to increase usability and user-friendliness.

Low password complexity appears to be one of the major security issues in OSNs today. From the results of this work for the user group of teenagers, there seems to be a relation with the study programme, and thus with the amount of technical education therein. When stratifying according to study programme and not gender, the respondents with the highest mean password complexity attend the study programme that has the highest amount of technology education. The programmes with the lowest amount of technology education have the respondents with the lowest mean password complexity. This could indicate that the more users are aware of technology, security, and risks, the more competent they are in constructing a strong password. Therefore, problems with compromised passwords could be minimized already in school, by extending today’s teaching to the online world.

There have been suggestions that about 75% of Facebook users use identical passwords for their Facebook and their private e-mail account [1]. The results of this work regarding the use of identical passwords indicate that the ratio is different for our user groups. For staff members, only 18% use identical passwords, and for teenagers the corresponding ratio is 40%. Although the first figure might be biased by the fact that all of the 250 000 accounts had publically available usernames and were in fact hacked, we hopefully see a positive trend. Still, the order of magnitudes of all figures may suggest that many users do not prioritize their personal security on Facebook, and that more awareness would be needed to improve this situation.
6 Conclusion and Future work

This study aimed at revealing password usage for teenagers of secondary school and staff at an institute of technology. Previously, studies were also performed for user experience of login sessions. The results of the latter are extended in this study to get a perspective of user perceived safety in OSNs.

The study was conducted with two surveys: one for teenagers about user perceived safety; and one for teenagers and staff members about password complexity. In a previous study, we described user QoE as a negative exponential function of response times. The corresponding user model also suggested that users are slightly more patient with login session than with regular web pages.

The “user perception of safety” survey indicates that some users may not care about security and/or feel safe all the time, independent of the amount of time it takes to login to Facebook. Amongst the remaining users we observe both falling and rising safety ratings for increased response times. These results indicate that the picture of user perceived safety is so far diverse, and that a deeper study of this topic is needed. In particular, we need to clarify how users perceive the safety of login sessions in correlation with their QoE. Furthermore, a clustering algorithm could be used to determine possible categories. Additional user experiments and interviews may shed new light onto the issue, and finally lead to user models for user perceived safety.

The survey results also indicate a difference in password composition for teenagers of different study programmes. This might suggest that the level of awareness and knowledge is somewhat related to password complexity. It also indicates that bad password strategies may be “taught away”, or that there would be a point in having good password strategies and online safety taught in primary or secondary school to increase security awareness. In any case, the participants in this study do not use identical passwords for Facebook and e-mail to the same degree as in a previous study of 250 000 Facebook accounts [1], which is somehow encouraging. Indeed, further research on this subject might show if similar behaviour is present in different user groups throughout the world.
References


PAPER V

Actual and Perceived Password Security in Online Social Networks

Submitted to journal
2014
Actual and Perceived Password Security in Online Social Networks

Charlott Lorentzen and Niklas Lavesson

School of computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden

Abstract – Today, it is common to associate application and program accounts to e-mail and Online Social Network accounts. A successful attack that results in a compromised e-mail or Online Social Network account could then propagate to compromise all other connected accounts and applications. Furthermore, people who are careless, or inadequately informed about online risks and threats, are easily deceived by malicious users or programs. Websites sometimes enforce a password length. However, this length might not be sufficient to provide a strong password. For inexperienced or uninformed users, a length requirement might give the user a false indication that the minimum length yields a strong password. In such situations there is a gap between the user’s actual and perceived password security. This study investigates the origin of the password problem, namely the construction of passwords and the user awareness of password security. Furthermore, this study investigates the potential gap between actual and perceived password security. A survey is conducted with teenagers from sixth form colleges in Blekinge County, Sweden. The results of the survey indicate that the passwords of the respondents are not as strong as the respondents perceive them to be. Based on the results, suggestions for best practices on password security are discussed.
1 Introduction

Facebook enforces a password length of six characters, together with a meter which tells when the password is of enough quality. The minimum requirement for password construction on Facebook may be insufficient for creating strong passwords. Meters which tell the quality of the password will make users create longer passwords according to Ur et al. (2012). However, if users are unaware of what constitutes a strong password, the length requirement may give a false indication that a password with the minimum length is strong enough to withstand an attack. In reality, however, a password of length six with a simple pattern can be compromised. In such a situation there exists a gap between the user's actual and perceived password security.

People who are careless, or inadequately informed about online risks and threats, are easily deceived by malicious actors. One compromised password can lead to multiple hijacked accounts due to the propagation of accounts. For example an e-mail account can function as backup for a forgotten password on an Online Social Network (OSN) account. Then, if the e-mail account is hijacked, the OSN account can be hijacked. Today, it is common to associate accounts by connecting applications and program accounts to e-mail and OSN accounts. A successful attack that results in a compromised e-mail or OSN account would then propagate to compromise all other connected accounts and applications.

Considering the possible consequences of using a weak password, there are incentives to strengthen users’ password security. During the last three to four decades, researchers have investigated and analyzed the issue of weak user passwords from different perspectives and have suggested possible solutions. Despite the suggested solutions the problems of weak passwords persist. This study investigates the construction of passwords and the user awareness of password security. The focus is put on teenagers since they are common users of password protected OSNs. The study also presents best practice suggestions for password security based on the found results.

This empirical study primarily investigates the potential gap between actual
and perceived password security. A survey is conducted with teenagers from sixth form colleges† in Blekinge County, Sweden, participating in Researchers’ Night‡ 2012. A quantitative analysis of the data from the survey is performed to explore differences of password entropy, password security awareness, and password construction.

This study deviates from most similar studies by analyzing the characteristics of the password, instead of the actual password, and by investigating the gap between actual and perceived password security. The results of this study can be used as a basis for creating stronger Internet services and for development towards better security education.

The outline of this paper is as follows. Related work is described in Section 2. Special phrases, variables, and words used in this study are defined and explained in Section 3. The methods for data collection and analysis are described in Section 4. The results are presented in Section 5, and then discussed and analyzed in Section 6. The paper is then concluded in Section 7 and an outline of future work is presented in Section 8.

2 Related work

Password construction, reuse, and misuse have been widely discussed and preventive methods have been developed for decades, but the issues of weak passwords persist. Several authors agree that training is the key to overcome the problems Adams and Sasse (1999); Gehringer (2002); Leach (2003); Yan (2002); Gaw and Felten (2006); Florencio and Herley (2007); Albrechtsen (2007); Kuhn and Garrison (2009). However, training have different forms, like advices, guidelines, or policies. Others argue that enforcing similar policies directly on sites will have a similar effect, for example with proactive password checking Klein (1990); Yan (2002); Dell’Amico et al. (2010); Wanli et al. (2010); Ur et al. (2012).

†The sixth form collage is the final mandatory school in Sweden, for teenagers of ages 16–19 years old.
‡Researchers’ Night, which is held in about 300 cities throughout Europe, is supported by the European Commission in an effort to show what researchers really do for society.
Such enforcements would then be technical solutions to the same basic problem; that users are without sufficient password security education, and often unaware about the nature of security attacks. Related work for both of these concepts are presented below.

Klein (1990) identified password security problems, discussed crackability, and suggested to solve the problems of weak passwords with a proactive password checker for checking password contents at the point of construction. Yan (2002) measured password strength to evolve the proposedly inadequate methods of proactive password checking, established by Klein (1990).

Referring to Klein (1990) and Yan (2002), two recent studies by Wanli et al. (2010) and Dell’Amico et al. (2010) aim at improving proactive password checking. Wanli et al. (2010) aim for improvements with a new indicator for password quality. Dell’Amico et al. (2010) aim for improvements with a method of performing a set of known attacks on a set of known passwords to know which are possible to compromise. These studies also indicate that most passwords are still weak and that most password strategies are inadequate for users to create strong passwords. Similar indications were already made by Adams and Sasse (1999).

Adams and Sasse (1999) issued a survey on password construction, use, and memorization, and established that policies were needed to help users with their password strategies. Gehringer (2002) elaborated, on the same topic, that policies were needed for proper password strategies.

Leach (2003) argues that employees’ common sense of security and decision-making skills influence how they use information security. Leach concludes that a company can benefit from strengthening the employee’s common sense of security as well as from training their security decision-making skills.

Continuing the work of Adams and Sasse (1999); Yan et al. (2004), Gaw and Felten (2006), and Florencio and Herley (2007) conducted separate surveys on password security. While Yan et al. (2004) analyzed how users constructed and used passwords within the study, Gaw and Felten (2006) and Florencio and Herley (2007) performed studies on actual password use on the web. Florencio and Herley (2007) performed a large-scale opt-in study using Windows Live
All three studies, as well as the literature study by Cisar and Markovic Cisar (2007), concluded that users need help in the form of policies and guidelines for constructing and managing passwords.

Albrechtsen (2007) interviewed employees regarding their experience of information security. Albrechtsen concludes that there is a gap between the understood role of the employees in the total information security work and the actual behaviour of employees. Furthermore, the employees perceived information security workshops as one of the most effective ways to influence individual security awareness.

The continuing problems of creating sufficiently strong passwords, and confirming to the user that the strength is sufficient, were also analyzed and discussed in three recent papers. Kuhn and Garrison (2009) analyzed password requirements on websites in 2007 and 2009. Kuhn and Garrison concluded that even though websites required stronger passwords in 2009 than in 2007, the requirements were generally still insufficient to provide a recipe for strong passwords. Helkala (2011) also indicate that users often fail to create strong passwords, even if password creation strategies are divided and specialized with regard to construction and management. Ur et al. (2012) found that meters for telling the quality of a password during creation led users to create longer passwords.

Though the work of this study lies between the two concepts above, it is closest to the idea that training is the key to overcome the problems of password security. The main difference from the above studies is methodological, and more specifically with regards to avoiding bias. We were unable to find discussions about potential biases in the reviewed studies. Nevertheless, potential biases will be discussed below.

When performing user surveys, there might be risks of bias in the answers from the users. Windows Live Toolbar users had to opt-in to be part of the study by Florencio and Herley (2007), and grant a tool access to log their user data, which include password data. An approach where users have to grant access to their passwords might bias the studied user group by including users who are less security aware than the average user. There is a risk that such a
study design might reveal unrealistically weak passwords.

In the study by Yan et al. (2004), users constructed passwords within the study. Such an approach might lead to passwords with different levels of security than for the average password that will be used in real life. If users construct a password in a study or in other ways have to reveal their own password, there is a risk of biased results regarding the received user data.

According to Fahl et al. (2013), the two major categories of recent studies of password security and usability suffer either from ethical issues or from lack of ecological validity. Studies in the first category are based on real world password from leaked or stolen passwords lists. These studies have an ethical issue in obtaining passwords lists through criminal activity. Studies in the second category are based on user studies. These studies suffer from the concern that user may not behave in the study as they do in real life. Consequently, the relevance of the results and whether they are transferable to the real world are hard to tell.

In this study, passwords are neither enquired for nor obtained in any other way. Just and Aspinall (2009) enquired test subjects about characteristics of the users’ secret. Using the same method, test subjects of this study were enquired about their passwords’ characteristics, such as length and types of characters included. The aim of avoiding to obtain the actual passwords from test subjects is both to avoid ethical issues and to maintain ecological validity.

3 Terminology

In this study, passwords are characterized by classes of characters, length, and structure. Each character class represents a group of characters: lower case letters, capital letters, digits, and special characters.

Since the respondents in this study are Swedish teenagers, the lower case and capital letters are based on the Swedish alphabet, which comprises 29 characters. The last three characters in the Swedish alphabet are missing in the English alphabet. Though, a dictionary attack can be successful on a password that
includes letters formed as a word. Swedish words can contain these last three 
letters of the Swedish alphabet. Thusly, these additional Swedish letters are 
also classified as letters.

Individual characters from 0 to 9 are classified as digits. The special char-
acters group represents characters which are excluded from the other classes. 
Based on a Swedish keyboard layout there are 36 special characters, 10 digits, 
and 29 each of lower case and capital letters. In this study, 36, 29, 29, and 10 
are used as the sums of the total number of characters for each respective class 
of characters.

In this study, the password structures are categorized as pattern-based, ran-
don, or mixed. These types are referred to as pattern-based passwords, random 
passwords, and passwords with mixed structure. A pattern is defined as a non-
random sequence of characters. A pattern can for example be a birthday, name, 
registration number, or word.

Simple examples of passwords of each structure type are given below.

- Pattern-based passwords: Name, 12345, pa55word
- Random passwords: z4Sk53, 8y5#kS2, %3o!s9j
- Passwords with mixed structure: name*2j3!, user7w&l, u4!0word

Pattern-based passwords contain patterns, and no random characters. Random 
passwords contain only random characters. Passwords with mixed struc-
ture contain both patterns and random characters. Pattern-based passwords 
can be easy to guess and easy to brute-force. Random passwords cannot be 
guessed based on dictionaries, rainbow tables or something connected to the 
user. Passwords with mixed structure can partly be guessed. Thus, passwords 
with mixed structure are harder to guess than pattern-based passwords, but 
easier to guess than random passwords.

The length of a password is commonly defined as the total number of char-
acters, of any class, that are included in the password. The length and the total 
number of possible characters form the password entropy (Florencio and Herley, 
2007; Cisar and Maravic Cisar, 2007; Burr et al., 2011). The number of possible 
characters is the sum of the total number of characters for each included class of
characters. For example, if capital letters and digits are used, the total number of possible characters is 39 ($29 + 10$). In this study, the password entropy is used as a measure of password strength, which is also adopted by Florencio and Herley (2007) and Cisar and Maravic Cisar (2007).

The length of a password will determine the average time for a brute force attack to attain the password (Scarfone and Souppaya, 2009). The presence of patterns in a password enables attacks on the password, such as brute-force attacks, dictionary attacks or social engineering (Scarfone and Souppaya, 2009). The attack success or failure is determined by the absolute strength of a password. The absolute strength is commonly determined by the length, classes of characters included, and the structure.

Vu et al. (2011) states that brute-force attacks are mostly used on passwords of 6-12 characters. We therefore define passwords as weak in this study if they contain 12 or less characters. Entropy gives a value for the distribution, not for the single event a password is. Therefore, our definition of weak is based on length instead of entropy. The current trend is that passwords need to be stronger in order to assure that they are uncompromised. Consequently, the passwords we define as weak will be regarded as weak as long as the trend is consistent.

There are also, independent of the absolute strength, factors that affect the strength of a password. Such factors are for example change frequency, use of a single password on multiple sites, writing passwords on physical notes, orally stating your password to anyone, etcetera. Such factors are not discussed in this study.

4 Methodology

This study focuses on password construction, password entropy, and awareness of password entropy, and password security education. Regarding password construction it is interesting to find out the contents and length of a password. From these characteristics it is then possible to derive the password entropy. Further-
more, it is interesting to know whether teenagers are aware of the strength of their password and whether teenagers are taught about password security at school.

In this study, characteristics of a password will be found without retrieving the password itself. Avoiding to retrieve the password is done to prevent introduction of possible biases, which might have been introduced in the large-scale studies by Florencio and Herley (2007) and Yan et al. (2004).

The method for evaluating password strength in this study will be the same as Florencio and Herley (2007) and Cisar and Maravic Cisar (2007) used; password entropy as a measure of password strength. The new method for measuring password strength by Wanli et al. (2010), and methods used by Helkala (2011) are inapplicable in this study, since both methods require the actual password instead of the characteristics of the password.

Concurrent embedded design (Creswell, 2009) is used for all topics in this study. Concurrent collection of quantitative and qualitative data and all data integrated into one database are chosen for this study from the concurrent embedded design choices. Furthermore, the main approach is quantitative, and the qualitative approach has a supporting role. The qualitative approach is represented by a questionnaire.

Concurrent triangulation design (Creswell, 2009) is used for the survey part on education of password security. Concurrent collection of quantitative and qualitative data is chosen for this study from the concurrent triangulation design choices, and results are compared in the discussion and analysis in Section 6.

4.1 Survey

A survey was constructed with a questionnaire and an interview part. The questionnaire included user characteristics, social network memberships, password usage, and password characteristics. User characteristics included gender, year of birth, study programme, and study year. Social network associations included memberships and if the subject considered its information and account in need of securing. Password usage included change frequency and multipe usage.
Password characteristics included length, character types, perceived strength, and whether the password included any pattern.

The last questions of the questionnaire concerned to which extent the teenagers were educated at school about password construction, password use, and risks with weak passwords. The questionnaire was tested on two colleagues and one teenager to check if the questions were comprehensible. Language issues that arose during the tests were fixed.

The questionnaire was made with both quantitative and qualitative questions. The qualitative questions had answers that could be quantified in an ordinal or nominal scale. The questionnaire had a fixed set of questions. One of the questions was exposed to the respondents based on their previous answers.

During the survey, teenagers from the sixth form college programmes of natural sciences and technology answered the survey individually. Before answering the survey, practical instructions were presented in groups of about 20 teenagers at a time. In contrast to e-mail surveys, the response rates are high in group-administered questionnaires. About 15 of the attending teenagers failed to participate in the questionnaire, 131 teenagers participated, and 120 of the answers were used in this study.

Facebook has a high coverage in the studied user group; 126 out of 131 respondents had a Facebook account. Therefore, we chose to use the answers of respondents with a Facebook account. Furthermore, an extra validity check on answers was enabled by exclusively including answers of respondents with a Facebook account. Facebook enforces a password length of six characters. Consequently, all respondents who answered that their password was shorter than six characters were excluded. Five respondents were excluded for lacking a Facebook account. Six respondents were excluded for mismatching answers in password characteristics.

The demographics and variations of the 122 teenagers that participate are considered diverse enough to be used as a subgroup of Swedish teenagers. The respondents of the questionnaire were born between 1993 and 1996, and were in study year two or three. From the technology study programme there were 43 participants, and 77 from natural sciences. The group consists of 54 women
and 64 men.

Based on previous discussions and a pre-study made in 2011 (Lorentzen et al., 2013) with a similar group of teenagers, there were some expectations on the results of this study. The expectations are presented as a premise for the design of the survey. Most teenagers were expected to think that their password is strong, when it had less than three classes of characters and less than 10 characters in length. The most used character classes were expected to be lower case characters and numbers. The part of the questionnaire about password security education is constructed with an underlying belief that people in general value education on password security as important.

Apart from the questionnaire, a group interview was also conducted as a discussion about education of password security with each group of about 20 teenagers. The results of the interview were analyzed together with the corresponding results of the questionnaire, and the results were compared.

The questionnaire was constructed to enable data analysis and comparison.

4.2 Data Analysis

All questionnaire answers were checked for mismatching information, such as a password having one character, but three classes of characters. The answers were sorted and matched based on being qualitative or quantitative. The results for different groups were then compared.

To check if respondents were aware of their actual password strength, the password entropy was compared to the corresponding perceived password strength. Password entropy was calculated using the general formula for entropy, $H$ (Burr et al., 2011):

$$H = \log_2(b^n)$$

(1)

where $b$ is the base, or the number of possible characters in the password based on the used character classes, and $n$ is the number of characters, or the length of the password. For example a password of 12 characters, taken from all classes of characters, has a password entropy of 80. The lowest password
entropy is achieved by having the lowest base, 10, and the shortest length, one. Such a password has a password entropy of just over three. Since the entropy rises with the base and length, there is no upper limit for the password entropy.

The password entropy was used as a measure of the password strength. The usage of password entropy as password strength as well as the general formula for entropy was also adopted by Florencio and Herley (2007) and Cisar and Maravic Cisar (2007). The general formula for entropy could be used without observing the actual password. Other formulas for entropy required the whole password for calculating entropy.

By using the general formula for entropy the different password structures failed to impact the resulting password entropy. To visualize the difference, the passwords were divided according to structures when presented and discussed. The results for each group were also compared with care, taking into account that a random password is stronger than a pattern-based password of the same length.

The answers from the questionnaire about education of password security were compared for different groups. The results from the interview are presented in Section 6, and also compared to the corresponding results of the questionnaire.

5 Results

5.1 Password construction

To be able to analyze password construction and entropy as indicators of password strength, the respondents are enquired about the contents of their password.

More than 93% of the passwords contain 6–15 characters. 20% of the passwords contain 13 or more characters. The shortest passwords contain six characters, due to the password policy of Facebook. The longest password contains 23 characters.
The most frequently used character class is lower case letters. The next most used character class is digits. Lower case letters and digits are the only character classes which are found as the single character class of some passwords. Furthermore, lower case letters and digits combined is the most used combination. Numbers are given below for use of character classes in a percentage of the passwords.

- Lower case letters: 96 %.
- Digits: 73 %.
- Special characters: 17 %.
- Lower case letters as the single class: 26 %.
- Digits as the single class: 3 %.
- Lower case letters and digits combined: 30 %.
- Lower case letters, capital letters, and digits combined: 23 %.
- All four classes of characters combined: 13 %.
- All remaining combinations of character classes: 3 %.
- One class of characters: 29 %.
- Two or less character classes: 61 %.
- Three or less characters classes: 87 %.

5.2 Password entropy

Password entropy increases as the length and the character base increases. One interpretation of high password entropy is that it reflects a strong password. Therefore, it is relevant to compare password entropies. The classic example of a low entropy, pattern-based password is a single word.

Table 1 shows the mean password entropy for different passwords structures, namely: pattern-based passwords, random passwords, and passwords with mixed structure. The highest password entropy is observed for male technology teenagers with random passwords. The lowest password entropy is observed for female technology teenagers with pattern-based passwords.

When comparing male groups, it is observed that the technology males have
Table 1: Password entropy. The table shows the mean password entropy and standard deviation. Each segment represents a password structure type. Each row represents a study programme and each column represents a gender.

<table>
<thead>
<tr>
<th>Password entropy</th>
<th>Mean (Standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td><strong>Pattern-based passwords</strong></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>63.30 (9.13)</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>49.13 (8.06)</td>
</tr>
<tr>
<td>Both study programmes</td>
<td>55.57 (11.04)</td>
</tr>
<tr>
<td><strong>Random passwords</strong></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>84.08 (34.23)</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>54.90 (16.85)</td>
</tr>
<tr>
<td>Both study programmes</td>
<td>71.31 (31.02)</td>
</tr>
<tr>
<td><strong>Mixed passwords</strong></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>59.71 (26.09)</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>54.47 (14.23)</td>
</tr>
<tr>
<td>Both study programmes</td>
<td>57.65 (22.02)</td>
</tr>
<tr>
<td><strong>All structures</strong></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>66.80 (26.55)</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>52.44 (12.68)</td>
</tr>
<tr>
<td>Both study programmes</td>
<td>60.27 (22.44)</td>
</tr>
</tbody>
</table>

A higher mean password entropy for all password structures than the natural sciences males. Though, for passwords with mixed structure the difference in password entropy is minor. The same comparison between female groups shows a higher mean password entropy for natural sciences females and pattern-based passwords. There are no technology females with random passwords. For passwords with a mixed structure the values are similar for technology females and natural sciences females.
When comparing the different password structures it is observed that the random passwords have the highest mean password entropy in five out of the nine columns, and for the passwords with mixed structure in the remaining four. When comparing mean password entropies between male and female groups, the least similar values can be observed between males and females of technology, with pattern-based passwords. The most similar values for males and females can be observed for technology teenagers with passwords with mixed structure. When comparing male and female groups, regardless of study programme, it can be observed that there is a difference in mean password entropy between males and females with random passwords.

5.3 Password entropy awareness

To understand respondents’ awareness of password strength, the respondents were enquired to choose the alternative answers that most correctly characterized their password. Table 2 presents the possible answers and the mean password entropy for each answer.

For the passwords which are perceived as strong, the random passwords have the highest password entropy, and for the passwords which are perceived as weak it is the pattern-based passwords. For the passwords for which the user is unaware of the password’s strength, the passwords with mixed structure have the highest password entropy. For passwords for which the user is indifferent of the password’s strength, the pattern-based passwords have the highest password entropy.

When comparing the mean password entropy for each password structure, it can be observed in Table 2 that the passwords that are perceived as strong have the highest password entropy for pattern-based and random passwords. For the passwords with mixed structure, the passwords that are perceived as strong and the passwords that have an uncertain password strength have the highest password entropy. Overall, the passwords that are perceived as strong have the highest password entropy.

Another interesting aspect to look at for awareness of password entropy in
Table 2: Password entropy awareness. The table shows mean password entropy and standard deviation. Each row represents a password structure. Each column represents a perception of the password strength.

<table>
<thead>
<tr>
<th>User’s perception of own password</th>
<th>Password entropy</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Standard deviation)</td>
<td>Pattern-based</td>
<td>Random</td>
<td>Mixed</td>
<td>All</td>
</tr>
<tr>
<td>Strong password</td>
<td>61.12 (13.68)</td>
<td>67.85 (30.77)</td>
<td>61.50 (21.39)</td>
<td>63.39 (23.23)</td>
<td></td>
</tr>
<tr>
<td>Unaware</td>
<td>49.96 (15.25)</td>
<td>48.90 (17.74)</td>
<td>62.87 (28.44)</td>
<td>55.13 (22.52)</td>
<td></td>
</tr>
<tr>
<td>Weak password</td>
<td>53.51 (12.94)</td>
<td>- (-)</td>
<td>39.64 (3.74)</td>
<td>50.98 (12.92)</td>
<td></td>
</tr>
<tr>
<td>Indifferent</td>
<td>52.49 (9.40)</td>
<td>45.68 (17.62)</td>
<td>47.41 (11.63)</td>
<td>49.60 (10.65)</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Strong password = My password is strong and impossible to compromise,
Unaware = I do not know if my password is strong or weak,
Weak password = My password is weak and probably possible to compromise,
Indifferent = I do not care if my password is strong or weak.

Table 2 is the number of passwords that constitute each value of mean password entropy. This will depict the amount of the respondents that are aware of the effect that a pattern in a password have for the strength of the password.

None of the random passwords are perceived as weak, which can be observed in Table 2. About 82 % of passwords that are perceived as weak are pattern-based passwords. About 31 % of the passwords that are perceived as strong are random passwords, and about 47 % are passwords with mixed structure.

Furthermore, 9 % of the respondents answer that their password is weak, 34 % that they are unaware of the strength of their password, and 48 % that their password is strong. The rest of the respondents are indifferent to the strength of their password.
5.4 Password security education

To get a notion of how much teenagers are educated about password security and its importance in school, the questionnaire enquired about three things on the topic. The teenagers were enquired about to which extent they were educated in school about how to create a strong password, why a strong password is important, and for what personal information can be used if an account is hijacked.

The vast majority of the respondents answered that they are educated to some extent or not at all about how to create a strong password (93 %), about the importance of a strong password (79 %), and about how personal information can be misused (82 %). Between 25–35 % of the respondents think that they are educated about password security at school to some extent. Between 5–12 % of the respondents think that they are educated about password security at school moderately. Between 0–6 % of the respondents think that they are educated about password security at school extensively/completely.

6 Analysis

In this study, the password strength is based on the length, the classes of characters included, and the structure. The password entropy is derived from the length and the classes of characters that are included. The structure of the password can be pattern-based, random or mixed.

Researchers have pointed out that most successful brute-force attacks are conducted on passwords of 6-12 characters, irrespective of the password’s entropy and structure. Therefore, we define passwords with 12 or less characters as weak with regard to length. Consequently, in this survey, at least 80 % of the passwords can be considered weak.

28 % of the respondents’ passwords include only one class of characters. 20 % of the passwords include only one class of characters and are pattern-based. As discussed previously, these passwords can be compromised by dictionary attacks.
or social engineering.

Considering the threat of dictionary attacks and social engineering, a password would require a random component to be secure at a length of 13 characters or more. In this study, less than 13% of the passwords have both random components and more than 12 characters. Including more than one character class in a password, combined with an appropriate length, and a random or mixed structure, apply on less than 11% of the respondents’ passwords.

The survey enquires about the perceived password security. In some cases there are insignificant differences between the entropies of passwords perceived as weak and passwords perceived as strong. About 50% of the respondents perceive their password as strong. Since at least 80% of the passwords in this survey are weak, at least 30% of all the respondents have a false awareness of security concerning their password.

This problem of false awareness can be more appropriately visualized by looking at respondents who perceive their passwords as weak and strong, separately. Considering the respondents who perceive their passwords as weak, when in fact they have strong passwords, there is no issue of vulnerability online. If considering the respondents who perceive their password as strong, about 75% of them have passwords which are weak with regards to length. Thusly, about 75% of the respondents who consider their password to be strong have a false awareness of security. These respondents are the most vulnerable group, since they are not aware of their weakness and may act online as though their password is strong.

6.1 Suggestions for best practice

There are several suggestions for improving password security among teenagers, based on the results of this study.

Password security education may need to be improved to increase the teenagers’ knowledge of password security. The policy for primary school decrees that secure, or rather safe, usage of the Internet and source criticism should be taught. In the context of this study there is one major problem with the
policy. A recent study of IT usage and IT competence in the Swedish primary school and sixth form college indicates that around 50% of the teachers need a higher competence in secure use of IT. It is also important to emphasize that this judgment was made by the teachers themselves.

Teenagers should also be taught about attacks on passwords and accounts as a topic in one of the existing subjects. By increasing this knowledge, teenagers will have a higher understanding of why it is important with strong passwords. This might also increase teenagers’ motivation to construct stronger passwords.

Password security theory and practices in education will need to be designed with care. It is important to relate to facts about password creation and attacks on passwords, to motivate and create an understanding of why strong passwords are needed.

For websites with login procedures, the information on what constitutes a strong password has to be updated. If the website Facebook allows a password of six characters without a warning, users might think that this is good enough. Furthermore, websites like Facebook, that propagate their login to authenticate the user on other accounts, should warn the user of the implications of such propagations. Then, the user can make an informed decision about whether to prioritize the security risks or the benefits of propagating their account to other accounts.

6.2 Validity threats

We have investigated the password security of teenagers using a qualitative and quantitative survey. Feldt and Magazinius (2010) cover validity threats and validity analysis for quantitative and qualitative research methods. Credibility concerns confidence and accuracy of the results according to Feldt and Magazinius (2010). For this study, we have identified six relevant threats to validity and credibility.

When a survey achieves a low response rate there is a risk of bias. This potential bias must be acknowledged. We used a group administered survey with individual responses. This way, we achieved a response rate of about
Sometimes respondents fail to complete specific parts or specific questions of a survey. The results of such parts or questions can be refutable. To avoid refutability we piloted all questions, and all questions that were used as a basis for statistical analysis were made mandatory. Since the survey was done digitally, the mandatory questions could be enforced.

Respondents that provide invalid information increase the respondent bias. This bias was addressed when designing the survey. Some questions require quantitative answers that can be compared with other answers to determine consistency. The answers given by some respondents were contradictory for the mandatory questions. In those cases the entire set of answers for that respondent was discarded. This helped increase the consistency and dependability of the answers.

Another risk with surveys can be described as socially desirable responding. This risk is most apparent when respondents are enquired to disclose threatening or embarrassing information. Since passwords are regarded as sensitive information, they will be subject to similar risks. The survey was designed without personal questions that could single out respondents. Consequently, we made sure to provide total anonymity to minimize the risks. The anonymity was also thoroughly explained to the respondents.

According to Feldt and Magazinius (2010), conformability is typically associated with risks of researcher influence on responses and respondents. Reactivity is a risk related to the researcher presence influencing the setting and respondents behavior. Researcher bias is a risk related to the researchers’ assumptions and personal beliefs affecting the survey. These risks were apparent in this study as the researcher was present during the responding of the survey. The risk of reactivity was handled by introducing an assistant, who was unrelated to the research and the topic, to assist the respondents with question issues on the digital questionnaire. The risk of researcher bias was minimized by omitting to introduce the research prior to the questionnaire.

Regarding the definition of a weak password, there is a risk of bias in the results. Many of the passwords in this study are weak according to the employed
definition. The definition of a weak password could be broadened to employ an even larger set of passwords. However, we are then unable to find any compatible definitions of a weak password.

7 Conclusions

We characterize a password by three variables: the length, the classes of characters included, and the structure. The length and the classes of characters form the password entropy. The character sequence defines the password structure. The sequence can be random, contain patterns, or be mixed. These variables are used as a basis for data collection and analysis of teenagers’ online passwords. Previous work frequently uses variables like password entropy as indicators of password security.

The length of a password is vital when suffering from a brute force attack. Our survey came to the conclusion that 80% of the passwords of the respondents are short enough to be compromised within a reasonable time and with reasonable resources. A random password with all classes of characters included and a minimum appropriate length has a password entropy of over 80. This password entropy is exceeded by less than 15% of the respondents’ passwords. The results on password length imply that, in best case, 20% of the respondents have passwords that can be considered strong. The results on password entropy conclude that, in this study, many of the passwords have low entropy.

The survey enquires about the actual and perceived password security. Almost 50% of the respondents perceive their password as strong. About 75% of the respondents who consider their password to be strong have a false awareness of security, and a gap between actual and perceived security. These respondents are the most vulnerable. Since these respondents are not aware of their password’s weakness and the gap, they may act online as though their password is strong.

Many previous studies enquire participants for actual passwords. In contrast to these previous studies, this survey enquired participants for password
characteristics. This approach enabled us to get answers from users, who would have refrained from giving their actual password. Furthermore, the answers we got enabled data analysis and comparison of results.

Besides investigating the gap between actual and perceived password security, this study also aimed at investigating the education of password security. This was done with a questionnaire and an interview discussion within the survey. The answers of the two survey parts are somewhat inconclusive, due to the design of the questions or the characteristics of the population.

The survey results indicate a lack of password security education. The respondents seem to have acquired any knowledge they might have about password security indirectly through other subjects or elsewhere. The inadequate security education might explain the gap between actual and perceived password security. We therefore hypothesize that the gap can be decreased through increased formal education.

8 Future Work

There are several interesting topics and open issues to investigate based on this initial study.

A large-scale study on the same topic could confirm or deny the gap between actual and perceived password security. Such a study could also explore the source of the gap. Large-scale investigations of the gap and its sources could motivate an introduction of online security education in primary school.

Related to the topic of online security education in primary school, a study could be done on how this education should be introduced. The question of when to introduce Internet security education could also be evaluated. Answers to these questions can be important since the usage of the Internet reaches kids of low ages.
Bibliography


<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3G</td>
<td>3rd Generation</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</td>
</tr>
<tr>
<td>EAP-SIM</td>
<td>Authentication and Key Agreement</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>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>NAI</td>
<td>Network Access Identifier</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute for Standards and Technology</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</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 or Service 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>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>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
</tbody>
</table>
ABSTRACT

Authentication solutions are designed to stop unauthorized users from getting access to a secured system. However, each time an authentication process occurs an authorized user needs to wait in expectation of approved access. This effort can be perceived as either a positive or negative experience. If the effort is perceived as a security measure; the effort is usually perceived as a positive experience. On the other hand, if the effort is perceived as a waiting time; the effort is usually perceived as a negative experience. The trade-off between security, user-friendliness and simplicity plays an important role in the domain of user acceptability. From the users’ point of view, security is both necessary and disturbing at the same time.

The overall focus in this thesis is on user perception of authentication in communication networks. An authentication procedure, or login, normally includes several steps and messages between a client and a server. In addition, the connection could suffer from low Quality of Service, i.e., each step in the authentication process will add to a longer response time. The longer response times will then infer lower Quality of Experience, i.e., a worse user perception.

The thesis first presents a concept of investigating user perception. A framework is developed in which different criteria and evaluation methods for authentication schemes are presented. This framework is then used to investigate user perception of the response times of a web authentication procedure. The derived result, which is an exponential function, is compared to models for user perception of web performance. The comparison indicates that users perceive logins similarly, but not identically, to how they perceive standard web page loading.

The user perception, with regards to excessive authentication times, is further studied by determining the weak point of the Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAPSIM) with the OpenID service. The response times are controllably increased by emulating bad network performance for EAP-SIM and other EAP methods in live setups. The obtained results show that one task of the EAP-SIM authentication deviates from the other tasks, and contributes more to the total response time. This deviation points out the direction for future optimization.

Finally, this thesis investigates how users of social networks perceive security, and to which extent they contribute to it. One way of contributing to security by creating and using strong authentication credentials, e.g. passwords. Websites might enforce a password length which is insufficient to provide a strong password. This might then cause problems by giving users a false perception of what constitutes a strong password. The origin of the password problem, namely the construction of passwords, and the user perception of password security is studied. A survey is conducted and the results indicate that the passwords of the respondents are not as strong as the respondents perceive them to be.