Linköping Studies in Science and Technology

Dissertations. No. 1768

Web Authentication using Third-Parties in Untrusted Environments

by

Anna Vapen

Department of Computer and Information Science
Linköping University
SE-581 83 Linköping, Sweden

Linköping 2016

Copyright © 2016 Anna Vapen

ISBN 978-91-7685-753-3

ISSN 0345-7524

Printed by LiU Tryck 2016

URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-127304>

Abstract

With the increasing personalization of the Web, many websites allow users to create their own personal accounts. This has resulted in Web users often having many accounts on different websites, to which they need to authenticate in order to gain access. Unfortunately, there are several security problems connected to the use and re-use of passwords, the most prevalent authentication method currently in use, including eavesdropping and replay attacks.

Several alternative methods have been proposed to address these shortcomings, including the use of hardware authentication devices. However, these more secure authentication methods are often not adapted for mobile Web users who use different devices in different places and in untrusted environments, such as public Wi-Fi networks, to access their accounts.

We have designed a method for comparing, evaluating and designing authentication solutions suitable for mobile users and untrusted environments. Our method leverages the fact that mobile users often bring their own cell phones, and also takes into account different levels of security adapted for different services on the Web.

Another important trend in the authentication landscape is that an increasing number of websites use third-party authentication. This is a solution where users have an account on a single system, the identity provider, and this one account can then be used with multiple other websites. In addition to requiring fewer passwords, these services can also in some cases implement authentication with higher security than passwords can provide.

How websites select their third-party identity providers has privacy and security implications for end users. To better understand the security and privacy risks with these services, we present a data collection methodology that we have used to identify and capture third-party authentication usage on the Web. We have also characterized the third-party authentication landscape based on our collected data, outlining which types of third-parties are used by which types of sites, and how usage differs across the world. Using a combination of large-scale crawling, longitudinal manual testing, and in-depth login tests, our characterization and analysis has also allowed us to discover interesting structural properties of the landscape, differences in the cross-site relationships, and how the use of third-party authentication is changing over time.

Finally, we have also outlined what information is shared between websites in third-party authentication, defined risk classes based on shared data, and profiled privacy leakage risks associated with websites and their identity providers sharing data with each other. Our findings show how websites can strengthen the privacy of their users based on how these websites select and combine their third-parties and the data they allow to be shared.

Populärvetenskaplig sammanfattning

Allt fler människor tillbringar stora delar av sina liv på Internet för att arbeta, uträtta myndighetsärenden, sköta sin ekonomi och umgås med vänner. Ofta när vi talar om Internet i dagligt tal menar vi *World Wide Web* (WWW), webben, som består av en mängd mer eller mindre sammanlänkade webbsidor där Google och Facebook är några av de mest populära. Från att ha bestått till största delen av statisk information har webben nu utvecklats till att blir mer personlig. Många webbsidor låter sina användare registrera ett personligt konto för att kunna skapa sig en personlig profil med eget innehåll på webbsidan, samt anpassa webbsidan utifrån sina önskemål. Personliga konton är centrala för sociala nätverk som Facebook, men används även av exempelvis nyhetssidor på vilka användaren kan logga in för att kommentera och dela nyhetsartiklar.

Att webben blir mer personlig har också lett till att många användare har ett flertal konton på olika webbsidor att hålla reda på. För att logga in på sitt personliga konto måste användaren autentiseras, det vill säga, bevisa att hon/han är den rättmätiga ägaren av kontot. Detta går ofta till så att användaren skriver in sitt användarnamn och ett lösenord. Förutom att det är besvärligt för användaren att minnas sina lösenord till alla olika konton, finns det många säkerhetsproblem med att använda lösenord för autentisering på webben. Trots det är användandet av lösenord den vanligaste formen av webbautentisering på grund av sin enkelhet.

Säkrare alternativ till lösenord existerar, till exempel används engångskoder och så kallade bankdosor till mer säkerhetskritiska tjänster på webben, men dessa lösningar är sällan enkla nog för att kunna användas dagligen till mindre kritiska tjänster. Dessa säkrare lösningar kräver även ofta att användaren har någon ytterligare utrustning, t.ex. en bankdosa, tillgänglig vid varje inloggning. Lösningarna är heller inte anpassade till att dagens användare är mobila och förflyttar sig mellan olika platser, använder oskyddade och opålitliga trådlösa nätverk och även loggar in på webbsidor från olika datorer, mobiltelefoner och surfplattor. Lösningar för autentisering som passar mobila användare i dessa opålitliga miljöer behöver vara lika enkla som lösenord, men säkrare, och inte kräva någon extrautrustning som användaren inte normalt bär med sig.

Vi har skapat en metod för att jämföra, utvärdera och designa autentiseringslösningar som passar för mobila användare i otillförlitliga miljöer. Vi har fokuserat på lösningar där en mobiltelefon används i autentiseringen, t.ex. för att lagra engångslösenord, eftersom användare idag ofta bär med sig sin telefon. Vi tar också hänsyn till att olika tjänster på webben kan behöva olika nivåer av säkerhet.

Ett populärt alternativ till att användaren loggar in på varje enskild webbsida är att istället använda *tredjepartsautentisering*. Då kan användaren logga in på flera olika webbsidor genom att autentisera sig mot en så kallad *third-party identity provider (IDP)*, en tredjepartstjänst som tillhandahåller

användarens identitet. Genom att använda en IDP behöver användaren logga in mer sällan och på färre platser. Därför kan mer komplicerade lösenord eller säkrare metoder för autentisering användas.

På de webbsidor som låter användaren logga in med en IDP kan användaren ofta välja vilken IDP de vill logga in med bland olika alternativ. Vi har undersökt vilka olika IDPer som finns och vilka kombinationer av dessa som användare kan välja mellan. Denna kartläggning har vi gjort genom att designa verktyg som automatiskt undersöker ett stort antal webbsidor. Vi har även gjort fallstudier och manuella studier över flera års tid, för att se hur olika webbsidor väljer och kombinerar de IDPer som deras användare erbjuds att logga in med.

Vid inloggning med IDP delas ofta information mellan IDPn och webbsidan som användaren vill logga in på. Denna information kan bestå av personuppgifter, personliga bilder och information som tillhör användarens vänner. Vi har därför även kartlagt hur denna information kan spridas mellan IDPer och de sidor som använder dem. Sådan informationsspridning kan leda till att användares privata information faller i fel händer och i slutänden till allvarliga problem som identitetsstöld.

Sammanfattningsvis presenterar vi bidrag för hur webbautentisering kan designas för att passa mobila användare i otillförlitliga miljöer. Vi har också undersökt hur trejdepartsaутentisering används på webben, vilka typer av webbsidor som använder sådan autentisering, vilken information dessa sidor delar och hur detta påverkar användares säkerhet och privatliv.

List of Papers

Included Papers

1. A. Vapen and N. Shahmehri, Security Levels for Web Authentication using Mobile Phones, in *S. Fischer-Hübner, P. Duquenoy, M. Hansen, R. Leenes and G. Zhang (eds.), Privacy and Identity Management for Life*, vol. 352, pp. 130–143, Springer, Boston, 2011.
2. A. Vapen and N. Shahmehri, 2-clickAuth - Optical Challenge-Response Authentication using Mobile Handsets, in *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)*, vol. 3, no. 2, pp. 1–18, April–June 2011. IGI Global, Hershey, USA.
3. A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, Third-party Identity Management Usage on the Web, in *Proc. Passive and Active Measurement Conference (PAM)*, pp. 151–162, Los Angeles, US, March 2014.
4. A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, A Look at the Third-party Identity Management Landscape, in *IEEE Internet Computing*, vol. 20, no. 2, pp. 8–25, March–April, 2016.
5. A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, Information Sharing and User Privacy in the Third-party Identity Management Landscape, in *Proc. IFIP ICT Systems Security and Privacy Protection Conference (SEC)*, pp. 174–188, Hamburg, Germany, May 2015.
6. A. Vapen, N. Carlsson, and N. Shahmehri, Longitudinal Analysis of the Third-party Authentication Landscape, in *Proc. Understanding and Enhancing Online Privacy Workshop (UEOP)*, San Diego, US, February 2016.

Other Publications

1. A. Vapen, Contributions to Web Authentication for Untrusted Computers, *Licentiate Thesis*, Department of Computer and Information Science, Linköping University, Linköping, 2011.
2. A. Vapen, D. Byers, N. Shahmehri, 2-clickAuth - Optical Challenge-Response Authentication, in *Proc. International Conference on Availability, Reliability and Security (ARES)*, pp. 79–86, Krakow, Poland, February 2010.
3. A. Vapen, N. Carlsson, A. Mahanti, and N. Shahmehri, Information Sharing and User Privacy in the Third-party Identity Management Landscape, in *Proc. ACM Conference on Data and Application Security and Privacy (CODASPY)*, pp. 151–153, San Antonio, US, March 2015. (Poster paper. Outstanding poster award!)

Acknowledgements

My deepest gratitude goes to my supervisor, Professor Nahid Shahmehri, for introducing me to computer science research, guiding me and keeping me on track. I feel that Nahid understands my struggles and she has always provided great advice, both in life and in research. I am also very grateful to my co-supervisor Niklas Carlsson for introducing me to Web measurement studies, and for all his support. Without the two of you this thesis would not have been possible.

I would also like to thank my past and present colleagues at ADIT and IDA, and the technical and administrative staff, especially Anne Moe for making life as a PhD student easier. I also thank the friends from the lunch room for all the fun discussions about work and beyond. An additional thank you to Ulf for brightening my days at work with practical jokes and research discussions, and for providing the type of advice only a good friend is allowed to give.

I thank Brittany Shahmehri for thorough proof-reading of this thesis, our ELLIIT colleagues for inspiring discussions on future technology, and our co-authors Anirban Mahanti and David Byers for their valuable insights. I also thank Linköping University as well as Lysator Academic Computer Society and various information security forums online for giving me a great education. I have also had the opportunity to supervise students in my field of research, which has been a great experience.

I thank my parents for providing a safe haven to return to when needed, and my friends for encouraging me and for listening to long stories about PhD life over dinner. My friends from the music world also deserve a thank you for having patience with me combining the different aspects of my life by writing research papers between my DJ-sets.

I would like to dedicate this thesis to Peter, for spurring my early interest in information security and inviting me into the world of buffer overflows and shell scripts, and for always being there for me.

Finally we acknowledge the partial funding of this work by ELLIIT.

Anna Vapen
Linköping, May 2016

Contents

1	Introduction	1
1.1	Motivation	1
1.1.1	Challenges with Web Authentication	2
1.1.2	Challenges with Mobility and Trust	2
1.1.3	Challenges with Third-Parties	3
1.2	Problem Formulation	4
1.3	Methodology	5
1.4	Contributions	6
1.4.1	Web Authentication for Untrusted Computers	7
1.4.2	Third-Party Identity Management Landscape	7
1.4.3	Privacy Risks in Third-Party Authentication	7
1.5	Thesis Outline	8
2	Background	9
2.1	Web Authentication	9
2.1.1	Authentication Factors	10
2.1.2	Multi-Factor Authentication	10
2.1.3	Authentication Method Examples	11
2.1.4	Password Problems	14
2.1.5	Secondary Authentication	16
2.1.6	Knowledge-based Authentication	17
2.1.7	Hardware and Biometrics	17
2.2	Authentication for Untrusted Computers	18
2.2.1	User Mobility and Trust	19
2.2.2	Multi-Level Security	19
2.3	Third-Party Authentication	20
2.3.1	Identity Management Overview	21
2.3.2	Federated Identity Management	21
2.3.3	Federation Usage Today	23
2.4	OpenID vs OAuth	24
2.4.1	Protocol Problems	24
2.4.2	Combining SSO with Data Sharing	25
2.5	Identity Management and Privacy Risks	26
2.5.1	User-Centric Authentication	27

2.5.2	Authentication, Identity Management and Beyond . .	27
3	Summary of Papers	29
3.1	Web Authentication and Untrusted Computers	29
3.2	The Third-Party Authentication Landscape	30
3.3	Third-Parties and Privacy Risks	31
4	Related Work	33
4.1	Security Levels and Evaluation	33
4.2	Authentication with Mobile Phones	34
4.2.1	Pre-Smartphone Solutions	35
4.2.2	Solutions in the Smartphone Era	35
4.3	Third-Party Authentication and Privacy	36
4.3.1	Exploring the Third-Party Landscape	36
4.3.2	Alternatives to Existing Web Authentication	37
4.3.3	Factors in Third-Party Authentication	38
4.4	Characterization of the Web	39
5	Discussion	41
5.1	Web Authentication in a Mobile World	41
5.2	Authentication vs. Identity Management	42
5.3	Measuring Web Authentication	43
5.3.1	Crawling for Authentication	43
5.3.2	Limitations in Measurements	44
5.4	Changes over Time	45
5.5	Data Classification	46
5.6	Privacy Aspects	46
6	Summary and Conclusions	49
6.1	Summary	49
6.2	Conclusions	51
6.3	Future Work	51
	References	52
	Collection of Papers	65
	Paper 1	67
	Paper 2	87
	Paper 3	115
	Paper 4	133
	Paper 5	149
	Paper 6	171

List of Figures

1.1	Examples of an untrusted computer and authentication devices	3
1.2	Overview of our methodology	5
2.1	Different types of authentication factors	11
2.2	Password-based authentication (1), with hardware aid (2), with one-time passwords (3) and with challenge-response (4).	14
2.3	Examples of common unsafe password practices.	16
2.4	Timeline showing common identity management and third- party authentication protocols.	22

List of Tables

2.1 Examples of common Web authentication methods 13

Chapter 1

Introduction

Today, users typically access their information from different devices and places. The users therefore need to be able to create, modify and share their own documents, calendars and other information, as well as access information created and shared by others. This has resulted in more and more information being stored online; e.g., on websites and in cloud-based services. With the increasing popularity of social networks, a rich set of user data is stored on the Web [1].

The personal information available on the Web is a popular target for attackers who steal personal data in order to blackmail or impersonate users [2,3]. A common problem, related to impersonation, is identity theft, leading to economical and reputational loss for the user [4].

To access their personal data on the Web, a user must be *authenticated*, meaning that they have claimed their identity and therefore have the right to access the specified resources [5]. With the increasing problems of identity theft and other privacy risks on the Web, there are many interesting challenges when designing secure, yet usable, authentication for Web users.

1.1 Motivation

Web authentication is becoming more widely used with the increased personalization of the Web, and poses specific problems and challenges. Since websites are reachable worldwide, the generation, distribution and revocation of credentials is mainly done online. Therefore, authentication must be secure, but must also be available from different devices and places.

Another problem with Web authentication is that users often have accounts on many different websites, making it difficult for the user to remember their login credentials (e.g., passwords). Remembering many passwords creates a large cognitive load for the user [6]. Therefore, users are likely to try to simplify the login process, if possible, by reusing credentials and selecting weak credentials, which are easy to remember [7]. Even if there are

many alternatives that are more secure [8], password-based authentication is by far the most common type of Web authentication. The high usage of weak authentication methods has resulted in a large worldwide attack surface and several security problems [9]. The problem with weak authentication methods being vulnerable to attacks is a well-known challenge in Web authentication. Popular solutions include (1) use of stronger and more secure authentication methods [10, 11], and (2) use of a trusted third-party to which a user authenticates to get access to multiple websites [12, 13], thus mitigating the problem with password reuse and many account details to remember. We now present a set of known challenges with Web authentication, as well as with various solutions that have been proposed.

1.1.1 Challenges with Web Authentication

Authentication methods that are more secure than passwords are not always suitable for the Web. For example, some such methods require in-person enrollment [5]. Furthermore, these methods often require additional hardware, which the user needs to carry with them, and which may be time-consuming and complicated to use. Therefore, such methods are normally used mainly for security-critical applications, such as online banking [14].

When designing and comparing authentication solutions for websites, the most secure authentication solution may not be the most suitable solution. For a website on which the user logs in often, the usability aspect is equally important. Authentication solutions that combine security with ease-of-use are, unfortunately, rare [15].

1.1.2 Challenges with Mobility and Trust

Additional equipment is especially a problem when considering *mobile users* who access websites from different devices and places. A challenge with mobile users, compared to users using the same device from the same place at each login, is that the mobile users cannot always bring and/or connect additional equipment when authenticating [14]. For authentication solutions that are both secure and feasible to use for mobile users, devices which the users already carry, such as mobile phones, are gaining popularity [16]. However, there are challenges in designing authentication solutions in which mobile phones are used as authentication devices. Mobile phones have many communication channels, and these channels can be used for strengthening security, but they also introduce new security problems. For example, malicious software can spread via different communication channels which the device provides, such as Wi-Fi, a GSM connection or Bluetooth. Because of this, authentication solutions with mobile phones become complex and therefore difficult to compare to each other.

A related problem is that mobile users are likely to use potentially *untrusted* computers and authenticate in an untrusted environment, e.g., on an open wireless network. Authentication solutions that are feasible in a

Figure 1.1: Examples of an untrusted computer and authentication devices

trusted environment may not be good options in an untrusted environment, due to the increased risk of attacks. Figure 1.1 illustrates a scenario with a computer in an untrusted environment. The figure shows how a nearby attacker can eavesdrop on the user in a public environment and capture authentication information, such as passwords. If the information can only be used once, during a limited time period, the risk for replay attacks is lower than if the information is reused, as is the case with reusable passwords [5]. Wired and wireless network connections, as well as screen radiation, can also be eavesdropped by an attacker and reveal authentication information. Hardware devices such as smart cards and USB-sticks can be used in an authentication solution to provide stronger authentication, but have the disadvantage that the user needs to remember to bring their devices when needed. Also, if using a public computer, the user may not be allowed to connect additional devices, or to install software used for authentication. Therefore, the types of hardware devices that can be used are limited. Also, the computer itself may be compromised by malicious software and therefore untrusted.

1.1.3 Challenges with Third-Parties

Besides the challenges with mobile users in untrusted environments, and with Web authentication in general, there are also interesting challenges with *third-party authentication*. Originally, third-party authentication was designed to solve the problem with users having many accounts to remember, by allowing the user to authenticate with a third-party website instead of with each website. Since the user logs in on fewer places, the third-party can potentially use stronger authentication [17].

However, since the identity provider gives access to several services, it can become a single point of failure in case of an attack. Identity providers also need to take mobility and untrusted environments into account, because if the user cannot authenticate to the identity provider, they will be locked out of a large number of their website accounts [18].

Third-party identity providers are interesting when designing Web authentication, since they are not as security critical as banking solutions, but still privacy critical. Third-parties give access to a large set of sites, which share and store a large set of potentially sensitive information [19]. A problem with third-party authentication is that only a few large websites are currently used as third-party identity providers, and these sites are not using strong authentication. Also, these providers are commonly importing and exporting a large set of user data, which is then used for different purposes such as advertising and personalization. Such sharing may cause issues related to security and privacy [20]. Furthermore, identity providers are using protocols not designed for privacy-preserving authentication, thus increasing potential security and privacy risks [13].

1.2 Problem Formulation

The Web authentication methods that are currently in use have several problems related to the wide-spread use of password-based authentication. The problems include security issues and the vulnerabilities of password-based methods in untrusted environments [8], as well as usability aspects of remembering passwords [21], and the privacy risks associated with third-party authentication.

Alternatives to passwords must meet the needs of mobile users, work well in untrusted environments and only use technology the user is likely to carry when they are not in a trusted environment, e.g. when they are not at home, work, school, etc. Such solutions exist, especially when considering the mobile phone as an authentication device. However, solutions including mobile phones are difficult to compare with each other since they vary in their use of protocols and transfer channels [16], each with their own security benefits and drawbacks.

In contrast to single site authentication, third-party authentication allows the user to access a wide range of website accounts using an account from a third-party website. However, this authentication approach often also involves information sharing between websites and can easily become time-consuming or confusing for the user. Third-party authentication also comes with its own security and privacy issues. This thesis addresses three classes of problems related to authentication in untrusted environments and using third-party providers.

First, we address the problem of mobile users authenticating in untrusted environments. For this part, we address the core question of how to attain secure and usable Web authentication for mobile users in untrusted environments. Here, we answer questions on how to evaluate and provide assurance that a Web authentication solution belongs to a specific security level. We also design an evaluation method for Web authentication which takes mobility and security into account, as well as how to evaluate and compare Web authentication solutions which use mobile phones as security devices.

Second, we consider problems and challenges of third-party authentication, including important questions such as how websites use third-party authentication. Here we consider how to sample reliable information about third-party authentication usage on the Web and develop new methods for how to identify relationships between websites and their third-parties. We also investigate how websites select and combine third-party identity providers, and how third-party authentication usage differs between different types of websites and in different regions. We then explore how third-party authentication compares to the more well-known field of third-party content delivery [122] on the Web.

Finally, we consider problems related to privacy risks in third-party authentication. Here, we investigate how information flows between third-party identity providers and the sites using them, and we also design methods to quantify privacy risks and their scope in third-party authentication. Interesting research questions include determining the differences between the risks associated with different classes of websites and the information shared between these sites. To better understand the multi-site privacy risks we also develop a structural model that allows us to investigate the structural properties of the third-party authentication landscape, and answer questions regarding current trends and how they have evolved over the years.

We will now outline our methodology for addressing the problems and questions above.

1.3 Methodology

We study Web authentication, especially with third-parties, by using a wide range of methods, focusing on evaluation, measurement and case studies. An overview of our different methodologies is shown in Figure 1.2.

Figure 1.2: Overview of our methodology

First, we proposed additional levels to the Electronic Authentication Guideline (EAG) [22], to match current and emerging authentication solu-

tions when using handheld devices. Then we designed a method for evaluation and design of authentication solutions, based on EAG and our added levels. We also designed and implemented a proof-of-concept prototype for optical authentication. The prototype was designed to be used by identity providers and other websites requiring stronger security than passwords could provide.

We then explored third-party authentication and designed tools for detecting third-party authentication on the Web. We used these tools and our novel method for logarithmic sampling of websites to take into account the heavy popularity skew on the Web [23]. We selected and crawled a large set of sites to investigate their features and their third-party usage. We used binomial hypothesis testing to identify classes of sites with statistically higher/lower third-party usage.

Our work also includes manual studies of the top 200 most popular websites in the world, for which we have done a three-year longitudinal study. We followed the same set of sites over ten measurements during three years, and we also collected snapshots of the most popular sites at each point of measurement. Thus, we were able to analyze both the evolution of a fixed set of websites and the changes in third-party usage among the most popular sites over time. Overall, the longitudinal analysis captures trends and evolution in third-party authentication. The manual studies are followed by an in-depth case study on a subset of the sites, given their properties, and focusing on privacy issues. We performed tests on combinations of third-parties as used by websites. For each possible third-party pairing, we allowed both third-parties in the pair to be used first in a sequence of tests. Also, the tests were performed both with and without creation of dedicated local accounts on the target websites. New third-party accounts were used for each test, to avoid bias with previous logins. Finally, we have performed a risk categorization based on the results of the case study and the manual studies.

1.4 Contributions

The main contributions of this thesis are divided into three different categories. First we present our methods for the evaluation and design of authentication solutions for mobile users in untrusted environments. Second, we present our work on understanding and measuring the emerging third-party identity management landscape. Here we extend the security issues discussed in the first part to cases in which there are several websites collaborating. Finally, we present our contributions regarding privacy risks in third-party authentication. In the following we present each of the classes of contributions in more detail.

1.4.1 Web Authentication for Untrusted Computers

We have designed and developed a methodology for comparing and evaluating authentication solutions which are based on the use of mobile phones as additional hardware devices for authentication [24]. The methodology can be used both as an aid when developing new authentication solutions, and for comparing existing solutions in order to choose a solution that is appropriate for a specific level of security, correlating to the levels suggested by the Electronic Authentication Guideline [22]. Our method takes into account Web authentication for mobile users and untrusted computers. We also propose more fine-grained security levels than are provided by existing guidelines to facilitate evaluation of authentication using mobile phones.

As an example of an authentication solution which makes use of the variety of communication channels and computational power provided by mobile phones, we have created an optical challenge-response solution as a proof-of-concept prototype [25, 26]. We show how authentication on the Web could combine the ease-of-use that passwords provide with the higher degree of security needed for identity providers and similar websites, which are vulnerable to identity theft because of their rich selection of user data.

1.4.2 Third-Party Identity Management Landscape

We have performed a large-scale empirical study of the current third-party authentication landscape, and shown how the use of third-party authentication differs between websites from different regions, in different popularity segments, providing different types of services, and having different site characteristics [27, 28].

For our large-scale study we designed and developed a logarithm-based sampling technique to be able to sample websites distributed over different popularity segments, together with a crawling technique adapted for Web 2.0 to crawl and find relationships between websites and their third-party identity providers on a representative subset of the 1 million most popular websites world-wide.

To put our results in context we have also compared third-party authentication with the more well-known field of content sharing. Then, we have extended our work to include privacy in third-party authentication, as described below.

1.4.3 Privacy Risks in Third-Party Authentication

Based on our measurement study above, we have analyzed which third-party identity providers and what authentication protocols are in use, how websites select and combine multiple third-parties and protocols, and which types of data a website shares with other websites in third-party authentication. Our large-scale dataset is complemented with a detailed measurement of the top 200 most popular websites. We have performed a longitudinal

evaluation of these top 200 most popular websites and their use of third-party authentication over three years [29].

We have also performed a detailed study on information merging and sharing between collaborating parties in third-party authentication. In this study, we have defined categories of data types transmitted in third-party authentication, and made a risk categorization based on the combinations of these types that occurred in our observations. We have also categorized problems related to account merging and cross-site information leaks related to these popular combinations of third-party identity providers [30].

1.5 Thesis Outline

In Chapter 2, we present the background of this work, in the areas of general Web authentication (Section 2.1) and Web authentication specifically for untrusted computers and untrusted environments in which mobile users are authenticating (Section 2.2). We also provide background information on third-party authentication (Section 2.3) and especially the differences and similarities between using the SSO protocol OpenID and the authorization protocol OAuth for third-party authentication (Section 2.4), since these are the prevalent protocols in this field. Finally, we end the background section by giving a broader perspective on identity management and privacy risks (Section 2.5).

In Section 3.1 we present summaries of our papers on authentication for mobile users in untrusted environments, and in Section 3.2, we present summaries of our papers on third-party identity management and authentication. Section 3.3 contains summaries of our papers on privacy in third-party authentication and longitudinal properties of the third-party authentication landscape. In Section 4 we summarize related work in Web authentication, especially third-party authentication, in Section 5 we discuss our findings, and in Section 6 we summarize and conclude our work, and present future work.

Chapter 2

Background

In this chapter we provide some background on Web authentication and identity management. First, we give an overview of factors, methods and protocols used in Web authentication, as well as potential security problems with existing Web authentication methods. Next, we discuss mobility and trust, explaining how untrusted computers and mobile users impact the requirement for strong, yet usable, Web authentication.

Then, we provide some background in the area of third-party authentication and identity management. We also give an overview of different, conflicting approaches to identity management, and ongoing work on developing strong identity management while preserving privacy. Thereafter, we present a comparison of the two most common protocols for third-party authentication on the Web, and discuss privacy risks in identity management, especially third-party authentication. Let's start with general Web authentication!

2.1 Web Authentication

With the increased personalization of the Web, users are creating their own spaces on the Web by logging in to personal accounts on websites, thus being able to create, save and collect information. With personal accounts, a site is also able to remember choices made by the user and adapt contents depending on the user's preferences. Examples are news sites allowing users to share and comment on articles, and social networking.

When authenticating on the Web to access a website account, there are problems which do not occur with authentication in general. The user cannot normally visit the site owner physically in order to obtain credentials and enroll in the system. Enrollment in person makes it possible to deploy strong authentication. Since enrollment in person is not feasible for the Web it is only used in highly security-critical applications such as e-banking, for which the user can visit a local authority for enrollment. For websites in

general, however, enrollment is done online [5].

We will now present a number of different *authentication factors* that together form *authentication methods*, and provide examples of which methods are commonly used on the Web, why they are popular and what security problems these methods have.

2.1.1 Authentication Factors

There is a wide variety of different methods for authentication, of which many could be adapted for the Web. An authentication method consists of at least one authentication factor combined with a protocol which outlines the steps of the authentication process. Authentication factors are commonly divided into the following three categories: knowledge factors (what the user knows), ownership factors (what the user has) and inherence factors (what the user is) [5]. Below are examples of the different factors.

- **Knowledge factors:** Secret combinations of characters, numbers, and visual information remembered by the user. Examples are passwords, passphrases, combinations of images, patterns and PIN-codes.
- **Ownership factors:** Devices and tokens held by the user, such as smart cards, mobile phones or hardware devices dedicated to authentication, e.g., online banking devices.
- **Inherence factors:** Fingerprints, retina patterns, speech or other biometric identifiers which are unique for each user.

There are also three additional categories, which are newer and not as commonly used yet. These categories are behavior factors (what the user does), friendship factors (who the user knows) and location factors (where the user is). Behavior factors can be grouped with inherence factors and are based on how the user is using a system or site, in terms of what the user does on the system, when and at what pace (e.g., how the user types on a keyboard). Friendship factors are based on the user's relationships with other users on a social network [31], and the user's ability to know people and interact with them. Location factors [32, 33], are based on the user's physical location when authenticating, and are normally used to trigger re-authentication or force the user to use stronger authentication if the user tries to authenticate from a place the user has not visited before. Another way of using location-based authentication is to ask location-related questions of a user who has forgotten their password and wants to reset it [34]. Figure 2.1 shows the different categories of authentication factors.

2.1.2 Multi-Factor Authentication

By creating authentication methods that combine factors from several different categories we get multi-factor authentication, which is considered

Figure 2.1: Different types of authentication factors

stronger than only using a single factor or several factors from the same category [22]. Two-factor authentication is the most common. Using several factors of the same type, e.g., a password and a PIN code, which are both knowledge factors, does not strengthen authentication to the same extent as choosing factors of different types. The reason is that different types of factors have different strengths and weaknesses. For example, since knowledge factors are used multiple times, they can be eavesdropped and replayed by an attacker. Ownership factors, on the other hand, can be stolen and used by the attacker. Even if the details vary, factors of the same type have both similar security problems and similar strengths [5].

However, even when selecting factors of different types which complement each other well, and when adding more factors or using more secure protocols, this increased security usually makes the process of authenticating more time consuming and complicated for the end user since it may contain more steps or require additional equipment [6]. For these reasons, multi-factor authentication is not yet wide-spread on the Web. However, existing multi-factor authentication methods for Web usage include combining knowledge factors with an ownership-based authentication factor such as specific hardware devices designed for online banking, or a mobile phone as an additional factor, e.g., in Google 2-step Verification, which is mainly used for specific situations in which the user wants to strengthen security, or for re-authentication, when the users is trying to authenticate from an unusual place [35].

2.1.3 Authentication Method Examples

An authentication method consists of one or several authentication factors and protocols. For example, both passwords and smart cards are authentication factors. The password is a knowledge factor and the smart card is an ownership factor. An authentication method can incorporate either or both of these factors. Authentication factors all have different strengths and weaknesses. Authentication methods that include the same factors can differ regarding their level of security, depending on which protocols are used.

Table 2.1 gives a high-level overview of four such cases, and Figure 2.2 highlights their differences and key features. In the following we discuss each case one at a time.

- **Case 1: Simple password authentication.** A user authenticates to a website by providing their username and a password, which is either selected by the user or provided by the website. The website stores usernames and encrypted passwords in a database, and checks a hash of the user-provided password against the stored one. If the passwords match, the user is authenticated. The passwords are static in the sense that they are reusable, and the same password can be used several times, until the user chooses to change it, or is alerted by the website to do so.
- **Case 2: Hardware aided password authentication.** A user is using a static password to access a website as in the case above, but instead of remembering the password, the user stores the password in a hardware device. Since the user does not need to remember the password, the password can be made more complex and harder to guess. The password can also be changed more often without adding a cognitive load for the user.
- **Case 3: Hardware device and one-time password.** A user uses an application in their hardware device which contains a list of passwords that can only be used once. When a password has been used, it cannot be reused. Therefore, the problem with an attacker replaying the password is mitigated. The one-time password is used in the same way as an ordinary, static password, as above.
- **Case 4: Challenge-response authentication.** A user provides a username or other identifier to the website he/she wants to access. A random number (challenge) is generated by the website and shown to the user. The user inputs the challenge into a personal hardware device, which uses a secret cryptographic key to calculate a response to the challenge. The user then types the response into a form on the website, and it is sent to the server, which has the key for each user, and thus can calculate a valid response and compare it with the response given by the user. If the responses match, the user is authenticated.

Case one presents a very common authentication method using a single factor, while case two presents a simple two-factor authentication solution. The solution in case two is more secure than the solution in case one, but only if the user selects a stronger password and/or changes it often, which is easier in case two than in case one, since the user does not need to remember the password by heart. To fully use the potential of a hardware device, consider a solution in which the hardware device generates a new, unique one-time password for each session, as in case three, and in which the device is protected with a PIN-code or other knowledge factor. In such a case, an attacker needs both the device and the PIN to impersonate the user. The challenge-response solution in case four is a variation of a one-time

Method	Factor	Description
Password	Knowledge factor	The user provides a username and password. The provided password is compared with a password stored in a database.
Password stored in hardware device	Knowledge factor and ownership factor	The user stores passwords in a hardware device. This makes it possible to use strong passwords without remembering them.
One-time password	Ownership factor	Passwords are stored in or generated by a hardware device, but each password can only be used once.
Challenge-response	Ownership factor	A challenge is generated by the authentication system. The user uses an ownership factor to calculate a valid response to the challenge.

Table 2.1: Examples of common Web authentication methods

password, but an infinite number of responses can be generated, instead of using a limited list of passwords which needs to be renewed.

As shown in the examples above, password-based authentication can be strengthened by introducing a hardware device for password storage. However, several of the weaknesses with passwords still remain, such as attackers being able to eavesdrop and replay passwords. Therefore, non-reusable credentials such as one-time passwords or challenge-response provide stronger authentication than passwords do. When using a hardware device for authentication, methods with non-reusable credentials are stronger than just storing passwords in the device, while remaining simple to use.

Perhaps the main drawback with password-based authentication (case one, bottom of Figure 2.2) is that the user may forget their password. For case 2 (middle column of Figure 2.2) we show how stronger passwords can be “remembered” by a device. The drawback with this method (bottom row) is that it only works well as long as the user has the device available. This drawback also applies to cases 3 and 4 (right-most part of Figure 2.2). Case 3 requires a one-time password list stored in a device. In case 4, a challenge-response solution is required, as illustrated by the very simple example in Figure 2.2 (right-most bottom panel). Here, the cryptographic algorithm for calculating the response is to simply add the shared secret X to the challenge. In reality, much more complex and secure calculations can be performed, since the device calculates the response for the user.

To put into context the drawbacks of using passwords for authentication, and why passwords are still in use despite their problems, we will now provide some background on password-related security problems.

Figure 2.2: Password-based authentication (1), with hardware aid (2), with one-time passwords (3) and with challenge-response (4).

2.1.4 Password Problems

Even though there are known problems related to using passwords, authentication methods incorporating passwords are the prevalent methods for Web authentication due to the familiarity, ease of use and ease of enrollment [21]. A password is a combination of characters, known to the user, but meant to be kept secret from others. A password is stronger if it is difficult for an adversary to guess. Strength can be increased by making the password longer or more random [8]. Some websites use password strength meters to help their users to select strong passwords, but these meters have been proved to need improvement [36–38]. However, remembering passwords creates a high cognitive load for users, either due to remembering many passwords or strong ones [6]. Large-scale studies on large password corpuses show that users tend to choose weak passwords for Web usage in order to be able to remember them [39]. Negative externalities from weak password practices, such as using weak passwords and reusing passwords, occur when an attacker steals a password database from a website and is able to reuse a password on other accounts belonging to the user [40]. The user habit of choosing weak passwords makes it easy for an attacker to guess the password by using lists of common passwords together with automated guessing tools [6]. Password reuse [7] raises the value of a stolen password since an attacker can try it on several websites. There have been large scale password leaks where attackers have obtained password databases from popular websites such as LinkedIn [41].

Even if passwords are normally stored in an encrypted format an attacker can use them by encrypting commonly used passwords and combi-

nations of characters and numbers the same way and comparing them with the passwords in the encrypted list [10]. Attacks using such lists will be more time-consuming for the attacker if the user has chosen a strong password, i.e. one that is either long or which contains unusual combinations of numbers, symbols and characters. A long non-dictionary word, combining letters in both upper and lower case, numbers and other characters is considered stronger than a short dictionary word, a short number combination or the name of a person or place since it takes longer to guess the password, even with automated tools [21].

When users select strong passwords, there are commonly used methods which users apply to be able to remember their passwords, such as password reuse across sites [40]. In addition to password reuse on multiple websites, users also tend to write down passwords in order to be able to handle large numbers of website accounts [21]. If a site has password policies, requiring users to select complex passwords and/or change them often, users are more likely to write them down or store the passwords in an application. Users have been shown to be more well-educated about secure password practices than previously known, but will still select the path of least resistance if possible [21]. Even skilled users have been shown to apply insecure password practices, even if these users are aware of the risks, showing that usability often overshadows security for users [42].

Besides the previously mentioned problems with password strength and randomness, there are also several attacks against passwords which will work regardless of the strength of the password, such as eavesdropping and replay attacks. Choosing a strong password makes it more difficult for an attacker to guess, but it is equally simple to capture and replay [8]. Bonneau has proposed a protocol for protecting against password guessing attacks, with the motivation that passwords will remain the most prevalent Web authentication method, and should be strengthened, as far possible, before being replaced [43]. In a later paper, Bonneau et al. also recommend user exercises in which the user practices remembering small portions of a password, until they have remembered the whole password [44]. An alternative to remembering passwords or writing them down is to use a password manager, which is a piece of software in which the user stores their passwords, thus helping the user to remember a large set of complex passwords. There are, however, known attacks against passwords managers, both Web based [45] and in general [46].

Figure 2.3 shows examples of common password problems, such as using simple and easy-to-guess passwords, users forgetting their passwords, unsafe password storage and password reuse. Independent of how users store or remember their passwords, there are situations in which users have forgotten their password and need a new one. For those situations, secondary authentication is commonly used.

Figure 2.3: Examples of common unsafe password practices.

2.1.5 Secondary Authentication

On a personal computer the user can choose the password directly and it does not need to be sent over a network. For usage in a specific, limited environment, such as in a company, passwords could be handed to users in person [22]. On the Web, however, passwords are either chosen by the user or sent over a network. A forgotten password cannot be easily replaced with a new one given to the user by an administrator in person, due to the world-wide scope. Instead, if the user forgets their password, the service sometimes e-mails a new password to the user, but then the e-mail could potentially be intercepted by an attacker [47]. E-mailing the user's current password would be even worse, since the user may be reusing the password on other websites. To avoid sending clear-text passwords, there are secondary authentication methods such as sending an e-mail with a link to the user and letting the user answer personal questions to prove their identity. These methods have flaws which often make them easier to attack than the passwords themselves [47].

Another method for password reset is to let the user answer personal questions in order to validate that the person wanting to reset the password is the correct user. These questions have been proven to be easy for an attacker to guess, since the questions are usually quite similar and the answers are often available in public records or even on the Internet, published either by the user or someone who knows the user. Examples of such information include the name of the university the user went to, the user's mother's maiden name, and the name of the user's first pet [48]. Location-based questions (e.g., "Where did you first meet your childhood best friend?") may be more difficult for an attacker to guess, depending on the question [34]. Alternatives to personal questions include answering a set of questions on likes and dislikes, either in text [49] or as visual images [50] the user selects from. However, this type of authentication scheme, with lists of likes and dislikes, has been broken [51] since it is possible to retrieve this information from social networks using data mining.

A similar problem is password reset via e-mail, which is supposed to be a more secure alternative to reset questions, but instead imposes other problems. If the user forgets their password a link is sent to the user.

By clicking the link the user is given the chance to change their password, without needing to remember the old password. However, this method opens up the possibility of phishing attacks (i.e., tricking the user into following a link to a malicious website or submitting sensitive information) since the user gets used to clicking links in e-mails from unknown senders [47].

While passwords are the most common of the knowledge-based authentication factors, there are alternatives to passwords in the area of knowledge-based authentication, as presented next.

2.1.6 Knowledge-based Authentication

There are studies on how to replace passwords with other knowledge-based factors since knowledge-based authentication is well-known to users and easy to set up [6]. One alternative to passwords is to use a passphrase, which is a sentence of several words instead of a single “word”. The strength of passphrases lies in their length rather than their randomness. A passphrase can contain common dictionary words if the phrase itself is so long that it is time-consuming to automatically guess [5]. Still, passphrases suffer from the same problems with replay attacks as passwords. There are also usability problems associated with remembering and typing long and complicated, sometimes auto-generated, passphrases [52, 53].

Another approach is to let the user remember a combination of images or positions in an image [54]. These visual patterns are shown to be easier for a user to remember than a series of characters, and more secure than passwords since the same data is not sent every time [6]. There are however, known attacks against image- and pattern-based authentication, mainly since there are obvious simple patterns that can be guessed [55].

We have now described knowledge-based authentication. There are also other authentication factors, which we will describe next.

2.1.7 Hardware and Biometrics

Besides passwords and other knowledge factors there are also biometrics (inherence factors) and ownership factors. Inherence-based authentication factors are usually not suitable in authentication methods for the Web due to the need for specific equipment to be distributed to the users [9]. Inherence-based factors are also open to several of the attacks aimed at knowledge-based factors, since they are used multiple times and can be replayed. Another problem is that, unlike a password, a biometric identifier such as a fingerprint cannot be changed [5].

When it comes to hardware devices, they must be available to the user at the time of authentication. Either the device is something the user normally carries or something small which the user can easily carry in their wallet or as a key fob. The devices also need to be simple enough for the specific application. Namely, if the application at hand (e.g., an e-commerce website) requires a level of security which justifies the additional complexity of

authenticating with something other than the usual password, the solution can be considered suitable for the specific scope, otherwise not [6].

Hardware devices are used in Web authentication when methods are needed that are stronger than passwords, and the increased cost of providing the devices and the hassle of distributing them can be justified by security requirements. A typical example is online banking [56]. There are also devices aimed at replacing passwords for everyday use [10, 57], as well as solutions in which a mobile phone is used for authentication [16, 58, 59]. Mobile phones have the advantage that users already carry them and thus do not need to carry any additional devices. Hardware devices for authentication are available in many different types with different features. To be able to review their security features and their suitability for Web use, we need to know what protocols they use. Some devices are simply password storage devices, while another, more secure way of using hardware devices for authentication, as previously shown, is to use them for generating one-time passwords that cannot be reused by an attacker since they are unique, only used once and limited to a specific time of use [60]. A variation of one-time passwords is challenge-response, in which the party to which the user is going to authenticate generates a unique challenge in the form of a number, to which the user calculates a matching response by inputting the challenge into his/her device which uses a secret, unique cryptographic key in the calculation. The response is either displayed by the device and manually typed into a form on the website by the user, or automatically transmitted between the device and the computer where the authentication takes place [14]. Automatic transmission makes it possible to use stronger cryptography and longer responses, since the user does not need to input a large amount of data manually, which helps make the solution more secure [16]. A specific solution for transferring one-time authentication information is to use sound [61], if the environment is suitable for this type of transfer (silent enough, all equipment available). We have previously shown cases and examples of commonly used Web authentication methods using one-time passwords and challenge-response to strengthen Web authentication. Authentication that is stronger than passwords is especially important to consider when a user is authenticating from an *untrusted computer*.

2.2 Authentication for Untrusted Computers

Identity management and authentication relies on entities and people trusting each other [62]. A computer or other computing device, such as a mobile phone or other handheld device, is considered untrusted if it is possible for an attacker to access the device and compromise the security of authentication. If the device is used by several people who could install software or hardware for malicious purposes onto the device, or if the device is connected to an open Wi-Fi network, it is potentially untrusted. Therefore, most computers and handhelds today are untrusted, which must be con-

sidered when authenticating from these devices, since sensitive information such as passwords may be captured by an attacker.

2.2.1 User Mobility and Trust

Users today are mobile in the sense that they use different computers and handheld devices in different places, which are potentially untrusted. The devices can be compromised with malicious software, networks can be eavesdropped and the place itself can be untrusted; e.g., a crowded area in which an attacker can see the user typing their password on a handheld device.

For the Web, authentication methods need to be secure, available to the user independent of the user's location and equipment, and easy to use. Different authentication methods meet these requirements to various extents. While methods incorporating passwords and other knowledge-based factors are simple and easy to understand, they are open to various attacks and difficult for the user to remember.

Also, Web authentication must be available on any device from which the mobile users would like to authenticate to a website. These devices may be untrusted, and there are also limitations on what additional equipment could be expected to be available on such devices. Specific equipment such as hardware tokens, Web cameras and fingerprint readers may not be available. Assuming that the user carries these additional hardware devices, there are two common problems. First, if the hardware device must be plugged into a computer not belonging to the user, the device or the data transfer between the device and the computer may be compromised, or the user may not even be allowed to plug in their own hardware. Second, the user must carry the authentication devices with them any time they need to authenticate.

Many users today are mobile, and therefore are likely to use untrusted public computers or to use computers in untrusted environments. Even devices that the user may trust may be used on an untrusted network, or in an untrusted place such as an airport, in which a nearby attacker can see sensitive information.

Besides the problems with trust in Web authentication, there are also different websites which require different levels of security in authentication. We will now describe such multi-level security.

2.2.2 Multi-Level Security

In many cases authentication becomes more complicated and time-consuming as the security of the authentication solution increases. Therefore, the most secure authentication methods may not be the best choice for an application in which fast and simple authentication is more important than security. We introduce the term *multi-level security* to indicate that there are different levels of security. In an online banking solution, a higher level of security is normally required than for authenticating to a social network. Also, an

online bank might have several services requiring different levels of authentication, e.g., allowing the user to access some services by using a password, but requiring two-factor authentication for critical services.

Besides specific security problems with authentication methods and factors, a more high-level problem is that it is difficult to compare authentication solutions with each other in order to choose a solution for a specific application, i.e. to find the correct balance between security and other factors. Authentication methods have different strengths and weaknesses depending on which factors and protocols they incorporate. There are many factors which both mitigate security problems and introduce new ones. One example is the use of hardware devices with several communication channels, such as mobile phones, as ownership factors in authentication. A mobile phone can provide additional security to an authentication method by being able to send and receive authentication credentials directly from a remote server via the phone network. At the same time, a mobile phone is also open to more attacks than a static device which cannot send and receive information over a network. Due to the wide variety of means by which to send and store authentication data it is difficult for a designer of authentication methods to select factors and protocol details in order to provide an appropriate level of security. It is also difficult for a user to understand which of two authentication methods is most secure and best suited for a specific service or website, regarding factors other than security.

Evaluating security, privacy and other aspects of authentication becomes even more complex when multiple websites cooperate and allow their users to authenticate by using another website. Such cooperation is known as third-party authentication. We next describe different types of third-party authentication.

2.3 Third-Party Authentication

Previously-shown problems with using passwords for authentication include requiring the user to remember a large number of passwords, and the time-consuming task of registering accounts at websites [12]. *Single sign-on (SSO)* was invented to avoid these problems, allowing the user to authenticate with an *identity provider*, in order to access other websites which trust the identity provider [12]. In SSO the user does not need to remember credentials for many different accounts, since the user only logs in to the identity provider. The main idea with SSO is to make it easier for users to authenticate, and since the user logs in less often and in fewer places when using SSO than when using separate services, stronger authentication can potentially be used, without compromising the user experience [13].

SSO is one type of *third-party authentication*, in which the user authenticates with a third-party website. In this thesis we explore different types of third-party authentication, including SSO. To explain different approaches to, and alternatives to, third-party authentication, we will start by defining

different approaches to *identity management* in general, which includes, but is not limited to authentication.

2.3.1 Identity Management Overview

Identity management is the theory of how to manage and store digital identities, i.e. the identifier (for example a username) and authentication credentials (passwords and other authentication information) that a user needs to authenticate to a specific site or system. Independent on the authentication method, the digital identities of the user must be stored and managed in a secure way, while still being accessible to the user at the time of authentication [17].

Identity management methods are typically classified into three categories: *user-centric identity management* [63], *federated identity management* [12] and identity management with a *trusted third-party* [17]. While user-centric authentication relies on the user managing their own identities, often using trusted hardware devices, the other two categories of identity management methods revolve around SSO and are nowadays similar to each other.

When discussing third-party authentication it is important to understand the differences and similarities between federated identity management and identity management with a trusted third-party, and how these techniques can improve user authentication. Therefore, we will now describe federated identity management.

2.3.2 Federated Identity Management

Federated identity management is when users authenticate to a website which is part of a federation of sites that form a circle of trust [12,17,64]. By authenticating to one of the sites, the user gets authenticated to the whole federation. In identity management with a trusted third-party, the idea is similar, but there is only one identity provider, a trusted website to which the user authenticates to get access to other sites which use this third-party for authentication [13].

Since the introduction of SSO on the Web, there has been a shift from federated identity management to identity management with trusted third-parties. Federations, and protocols designed especially for federated identity management, are still in use for specialized applications, but not in general. To fully understand the usage of SSO on the Web today, it is important to know the history of federated identity management.

The goals and challenges of federated identity management include creating secure SSO and maintaining trust between the collaborating parties in the federation [65]. Early initiatives for federated identity management include Security Assertion Markup Language (SAML) which is an open standard for transferring security protocol information, e.g., in SSO [12]. SAML

Figure 2.4: Timeline showing common identity management and third-party authentication protocols.

was designed by the Organization for the Advancement of Structured Information Standards (OASIS), which also supplied a suite of Web service SSO protocols (WS-*). A co-existing alternative to SAML and the WS suite was the Liberty Alliance Identity Federation Framework (ID-FF) protocol proposed by the Kantara Initiative (formerly Liberty Alliance) which is a coalition of companies and organizations involved with identity management, similar to OASIS. SAML was later used as part of the ID-FF protocol. However, in general Kantara has been focusing mostly on federated identity management for large companies and organizations, while OASIS has focused on standardization according to NIST guidelines [13]. Figure 2.4 shows a timeline containing common protocols for federated identity management and third-party authentication. The small arrows in the figure show the start of collaboration between OASIS and Liberty Alliance. In the figure we also see the newer protocols OpenID and OAuth, which will be discussed later in the thesis, in the interest of comparing when the different protocols were developed.

Today, SAML is often used as part of other SSO solutions, and for internal usage, such as for companies or organizations that have several services they want to connect. A common example is the intra-Web for a university, or larger internal federations connecting several universities. SAML has also been used for letting third-party applications authenticate to a host application (e.g., a gaming application that authenticates to a social network in which the game resides). Over the years, several security issues with the SAML protocol have been reported [66, 67]. The specific SAML implementation for Google applications was broken in 2008 [68].

There are currently several approaches to strengthening federated identity management. When security concerns were raised regarding SAML and ID-FF, these federated protocols were still seen as more secure than protocols relying on a single trusted third-party, such as Microsoft Passport. Pfizmann et al. proposed the more privacy-preserving Browser-Based

Attribute-Exchange (BBAE) protocol in 2005 to increase both security and privacy in SSO on the Web [69]. Analysis of Liberty also shows that there are security problems, though it is suitable for federations, especially when partial identity is desired [19].

Having looked back at the history of federated identity management and SSO, we will now show how federations are used today on the Web.

2.3.3 Federation Usage Today

Federated identity management and other solutions involving trusted third-parties used for authentication share the common problem that users tend to choose the simplest path when authenticated, e.g., choose a simple password to authenticate to their provider(s) used for all websites they need to authenticate to [18]. Studies also show that users do not trust third-party identity providers, but still use the providers, without knowing whether they are authenticating to a real provider or a fake one that has been set up by an attacker [70]. An overall problem, as stated by Ross Anderson, is that governments focus more on finding solutions in which users have one single identity for everything they do online. Instead, Anderson suggests that governments should set up rules for identity management, and leave the solutions to those implementing them [71]. Overall, federations are mainly used for governmental organizations and company infrastructures, rather than for Web users in general.

There is an underlying conflict between identity management with many third-parties collaborating in a federation, and one single third-party which is responsible for all authentication [72]. The latter has been seen in national identity management systems, such as the Estonian national ID cards and the Swedish BankID, but not for world-wide usage of the Web.

While the early federation protocols were not primarily designed for Web SSO, OpenID was aimed at websites that want their users to be able to login using an existing identity from a third-party identity provider. OpenID 2, which included minor updates from the first version, including extended data transfer, was seen as a user-centric initiative from the beginning, thus letting the user run their own IDP or selecting one they trusted [73]. However, OpenID today is mainly used with existing third-party identity providers, and websites may also restrict which identity providers they allow their OpenID users to authenticate with.

Overall, we have seen a shift from federated identity management to third-party identity management, which consists of third-party authentication and the storing and revoking of authentication details. Since we only focus on the actual authentication process, we will mainly use the term “third-party authentication” rather than “third-party identity management” in this thesis. Along with OpenID, which is being used increasingly as a third-party authentication protocol instead of a federation protocol, there is also the OAuth protocol, which acts as a third-party authentication

protocol and is even further from federations than OpenID. We will now compare and contrast OAuth with OpenID, since both protocols have been widely used for third-party authentication on the Web.

2.4 OpenID vs OAuth

OpenID is a loosely designed framework and rule set which is used for seamless login, in which the user authenticates to a website (the *relying party*, *RP*) by providing his/her digital identity to a third-party (the *identity provider*, *IDP*), which informs the relying party that the user is authenticated. The user does not need to trust the relying party, only the provider. Originally, in OpenID, any provider should be usable with all relying parties [17]. In practice there is often a limited set of providers for the user to choose from, among which the user tend to choose familiar ones (e.g., a social network on which the user has, or is going to create an account) [74].

Today, the OAuth authorization protocol is much more commonly used for Web authentication than OpenID [27]. While OpenID allows limited transfer of user data from IDP to RP, to be used as identifiers and similar at the RP, OAuth allows transfer of a much richer set of data in both directions between RP and IDP. OAuth also authorizes the RP to act on the user's behalf on the IDP [29]. With its rich information flow, OAuth increases the risk of privacy leaks and compromised identity in third-party identity management and authentication.

While OpenID was proposed as an SSO protocol, OAuth was designed to simplify collaboration between Web services and authorize Web applications, as well as transfer data across the Web. The SSO part of OAuth attracts users both to the RPs and the IDPs, which has helped OAuth to gain popularity as an authentication protocol [13]. However, there are well-known problems associated with both protocols.

2.4.1 Protocol Problems

Authentication with a trusted third-party, using either OpenID or OAuth, decreases cognitive load, but the third-party must be trusted [75]. The third-party may also become a single point of failure since all authentication takes place at the provider [76]. There are several known security problems related to both OpenID and OAuth. First, OpenID is vulnerable to cross-site request forgery (CSRF), which is one of the most common Web security threats [77]. OAuth, on the other hand, has been proven to be secure with regard to the protocol itself [78, 79], but several severe security weaknesses have been found in specific implementations of OAuth [80, 81], such as CSRF as with OpenID above [82]. The risk of an attacker fooling a user into entering their credentials into a fake website or following non-trustworthy links is known to increase with the use of OpenID, OAuth and similar solutions [18]. Additional problems occur when combining SSO with data sharing, which

is done in both OpenID and OAuth, with OAuth having the most support for such sharing. We will now see which kind of problems data sharing can introduce.

2.4.2 Combining SSO with Data Sharing

There are solutions that are built on having only one trusted third-party (instead of several parties in a federation), such as Facebook Connect and Microsoft Live Connect (which is the replacement for the now outdated Microsoft Passport) [6]. In Facebook Connect and Microsoft Live Connect, the main idea is not authentication in itself, rather it is to use the OAuth authorization protocol to let the user allow services on the Web to act on their behalf on social networks used by the user [83]. Authorization differs from authentication in the sense that it is used for allowing a user or a piece of software to access a specific service, without considering identity. For example, assume that a user uses a site to store photos and wants to share them on a social network on which the user has an account. If both sites are connected via OAuth, the user can give a Web application belonging to the photo site the rights to act as the user on the social network, allowing the actions the user agrees to, e.g., for the site to post photos on behalf of the user. The application is now authorized to have some privileges that belong to the user. However, with OAuth the user can now use the social network to log in to the photo site, in a single sign-on fashion.

As the examples above show, authentication and authorization are mixed and used together. Also, the difference between federated identity management and Web authentication with a trusted third-party is getting smaller. Overall, pure SSO and federations in which all parties are connected to each other are clearly in the minority [27]. In 2010, the main problem with third-party authentication and federated identity management was the problem of “billion keys, few locks”, e.g., many IDPs and few RPs supporting SSO with these IDPs [84]. Now, the situation is the opposite, with RPs using sets of a few very popular IDPs. The ROAuth recommender system by Shehab et al. [85] helps users make informed choices about which IDP to use when presented with several choices. While ROAuth is aimed at users, there are also security initiatives focused on the authentication protocols. Third-parties known to run their own protocols, such as Facebook, Google and Live (Microsoft) have increasingly started using OAuth as part of their protocols, which has spurred interest in performing security analysis of these protocols, e.g., formal analysis of Facebook Connect [86], in order to increase security.

Before the release of OAuth, and even afterwards, OpenID was the leading Web SSO protocol and various solutions for how to strengthen OpenID were proposed, such as storing credentials on SIM cards for increased security and decreasing the cognitive load [87], or combining OpenID, and later also OAuth, with the now deprecated CardSpace identity management

solution from Microsoft [88].

Currently, IDPs are moving towards the new OpenID Connect protocol, which is built heavily on OAuth 2.0 combined with the SSO technologies from OpenID [89]. OAuth's rich data sharing and authorization remains in OpenID Connect. Such sharing may have privacy implications, as we will show now.

2.5 Identity Management and Privacy Risks

A problem with identity, both digital identity and in a broader non-technical sense, is that users tend to have several identities, e.g., their work persona and the personality they express on social networks, but not in work-related situations. Multiple identities [90] and the fact that users do not want to share all aspects of their identities with all third-parties highlights the problems with federated identity management [91]. There are two important, related concepts in play here: multilateral security, which is the idea that no member of a federation should be less secure or trusted than any other member, and partial identity, which means that a user can define an arbitrary subset of personal information to share with each federation member, e.g. on one website the user may need to share their name, but on another they need only share that they are older than twenty-one [92]. A related concept is contextual identity, wherein people only want to share certain aspects of their personalities with certain people [93]. Partial and contextual identity are supported by OpenID and OAuth, but in practice are not used online today.

Overall when looking at identity, Cameron et al., in a discussion with privacy bloggers, suggested seven guidelines for identity management, with a focus on user control and consent [94]. Such guidelines are difficult to follow due to the complex information flows in third-party and federated identity management [95], and due to the complexity of customizing Web services securely [96], but can lead to interesting discussions and new ideas.

There is also plenty of advice aimed directly at users on how to avoid security problems and increase privacy. However, such advice often includes time-consuming and difficult actions, even if the risks are low. Users often understand that there is usually little danger, and do not follow the advice, as shown in user experiments [97]. Also, there are gaps in users' conceptual understanding of SSO [98]. When not understanding the privacy and security risks of SSO, federations and third-party authentication, information leaks and identity theft become more prevalent.

We will now take a look at alternatives to third-party authentication and federations by explaining user-centric authentication, which is a popular research topic, and which may evolve into a wide-spread and common type of identity management in the future.

2.5.1 User-Centric Authentication

User-centric authentication and identity management means that users control their own authentication credentials. In practice this means that the digital identities of users are managed by the users themselves e.g., remembered by the user or stored on a dedicated hardware device for authentication [10]. While there are novel examples of user-centric authentication that have been created as a more user-friendly and privacy-preserving alternative to identity federations and third-parties [10], user-centric authentication is usually the default type of authentication, originating from the time before federated identity management and trusted third-parties, normally referred to as single logons (SILOs) [76]. The difference is that user-centric authentication helps users handle multiple accounts on many SILOs, while traditional SILO authentication does not specify how to handle accounts.

User centric models as proposed by Jøsang [99] and by Procházka et al. [63] exist as alternatives to federations and third-party identity management and aim to mitigate privacy problems in identity management, while still helping users to securely authenticate to many websites without the cognitive load and other problems associated with SILO authentication.

User-centric solutions are popular among technically skilled users since they give the user the power to manage their own identity, but the broad majority of Web users are not interested in authentication in itself, they simply want to access the services protected by authentication [18]. Therefore, while user-centric authentication is a novel and interesting topic, it is still interesting to explore what users are actually using in authentication on the Web today. We will now present the topics in authentication which we have explored in this thesis.

2.5.2 Authentication, Identity Management and Beyond

In this thesis we first present problems with Web authentication in general, using the traditional SILO model. We bridge the gap between simple, yet insecure knowledge-based authentication, and hard-to-use, but more secure, ownership-based authentication and inference-based methods not suited for the Web.

Then, we move from the SILO model, with only two parties in the authentication process, to authentication with several sites involved.

We focus specifically on the ever-increasing third-party authentication landscape, which has replaced federated initiatives to a large extent. Our goal is to study existing authentication and identity management on the Web, over time, and therefore our focus is mainly OAuth-based third-party authentication. Our aim is to analyze and improve the existing network of authentication and the ad-hoc identity management that consists mainly of a mesh of OAuth relationships between RPs and their third-party IDPs.

We will now present summaries of our research papers on the above topics.

Chapter 3

Summary of Papers

In this chapter we summarize our papers by providing the abstracts. The full papers are found in the end of the thesis. The papers are divided into three categories, which are described below. The first category contains papers on Web authentication for untrusted computers, the second category contains papers on third-party authentication, and finally, the third category contains our papers on privacy risks in the third-party authentication landscape. For each paper we present the abstract.

3.1 Web Authentication and Untrusted Computers

In this section we present two papers on Web authentication for untrusted computers. The papers help address the question of how to attain secure Web authentication for mobile users in untrusted environments and/or at untrusted computers, while also achieving availability and usability. We also show how to evaluate and compare different Web authentication solutions that use mobile handsets as security devices, by extending concepts from existing guidelines for authentication and designing an evaluation method taking both untrusted computers and mobile users into account. Using our method we have shown that it is possible to evaluate and provide assurance that a Web authentication solution performs at or above a specific security level, and how security levels can be used for identifying authentication solutions which are more secure than passwords, still as easy to use and always available to the user.

Paper 1: Security Levels for Web Authentication using Mobile Phones

Mobile phones offer unique advantages for secure authentication: they are small and portable, provide multiple data transfer channels, and are nearly ubiquitous. While phones provide a flexible and capable platform, phone designs vary, and the security level of an authentication solution is influenced by the choice of channels and authentication methods. It can be a challenge to get a consistent overview of the strengths and weaknesses of the available alternatives. Existing guidelines for authentication usually do not consider the specific problems in mobile phone authentication. We provide a method for evaluating and designing authentication solutions using mobile phones, using an augmented version of the Electronic Authentication Guideline.

Paper 2: 2-clickAuth - Optical Challenge-Response Authentication using Mobile Handsets

Internet users today often have usernames and passwords at multiple websites. To simplify things, many sites support federated identity management, which enables users to have a single account that allows them to log on to many different sites by authenticating to a single identity provider. Most identity providers perform authentication using a username and password. Should these credentials be compromised, all of the user's accounts become compromised. Therefore a more secure authentication method is desirable. We have implemented 2-clickAuth, a multimedia-based challenge-response solution which uses a Web camera and a camera phone for authentication. The phone provides mobility and high availability to the solution. Two dimensional barcodes are used for the communication between phone and computer, which allows 2-clickAuth to transfer relatively large amounts of data in a short period of time. 2-clickAuth is considerably more secure than passwords while still being easy to use and easy to distribute to users. This makes 2-clickAuth a viable alternative to passwords in systems where enhanced security is desired, but availability, ease-of-use, and cost cannot be compromised. We have implemented an identity provider in the OpenID federated identity management system that uses 2-clickAuth for authentication, making 2-clickAuth available to all users of sites that support OpenID, including Facebook, Sourceforge and MySpace.

3.2 The Third-Party Authentication Landscape

In this section we present two papers on third-party authentication and identity management, including a large-scale measurement study of IDP usage on the Web, as well as other complementary measurement studies of the RP-IDP landscape, with a focus on characterizing the RPs.

Paper 3: Third-Party Identity Management Usage on the Web

Many websites utilize third-party identity management services to simplify access to their services. Given the privacy and security implications for end users, an important question is how websites select their third-party identity providers and how this impacts the characteristics of the emerging identity management landscape seen by the users. In this paper we first present a novel Selenium-based data collection methodology that identifies and captures the identity management relationships between sites and the intrinsic characteristics of the websites that form these relationships. Second, we present the first large-scale characterization of the third-party identity management landscape and the relationships that makes up this emerging landscape. As a reference point, we compare and contrast our observations with the somewhat better understood third-party content provider landscape. Interesting findings include a much higher skew towards websites selecting popular identity provider sites than is observed among content providers, with sites being more likely to form identity management relationships that have similar cultural, geographic, and general site focus. These findings are both positive and negative. For example, the high skew in usage places greater responsibility on fewer organizations that are responsible for the increased information leakage cost associated with highly aggregated personal information, but also reduces the user’s control of the access to this information.

Paper 4: A Look at the Third-Party Identity Management Landscape

Many websites act as relying parties (RPs) by allowing access to their services via third-party identity providers (IDPs), such as Facebook and Google. Using IDPs simplifies account creation, login activity, and information sharing across websites. However, different websites’ use of IDPs can have significant security and privacy implications for users. Here, the authors provide an overview of third-party identity management’s current landscape. Using datasets collected through manual identification and large-scale crawling, they answer questions related to which sites act as RPs, which sites are the most successful IDPs, and how different classes of RPs select their IDPs.

3.3 Third-Parties and Privacy Risks

Here, we present two papers on third-party authentication with a focus on the information leakage and privacy risks. We have studied the flows of information between RPs and IDPs, and performed targeted case studies to detect and analyze potential privacy problems. We also take a longitudinal

approach, studying the RP-IDP landscape and its information flows over three years.

Paper 5: Information Sharing and User Privacy in the Third-Party Identity Management Landscape

The cross-site information sharing and authorized actions of third-party identity management can have significant privacy implications for the users. In this paper, we use a combination of manual analysis of identified third-party identity management relationships and targeted case studies to (i) capture how the protocol usage and third-party selection is changing, (ii) profile what information is requested to be shared (and actions to be performed) between websites, and (iii) identify privacy issues and practical problems that occur when using multiple accounts (associated with these services). By characterizing and quantifying the third-party relationships based on their cross-site information sharing, the study highlights differences in the privacy leakage risks associated with different classes of websites, and provides concrete evidence for how the privacy risks are increasing. For example, many news and file/video-sharing sites ask users to authorize the site to post information to the third-party website. We also observe a general increase in the breadth of information that is shared across websites, and find that due to usage of multiple third-party websites, in many cases, the user can lose (at least) partial control over which identities they can merge/relate and the information that is shared/posted on their behalf.

Paper 6: Longitudinal Analysis of the Third-Party Authentication Landscape

Many modern websites offer single sign-on (SSO) services, which allow the user to use an existing account with a third-party website such as Facebook to authenticate. When using SSO the user must approve an app-rights agreement that specifies what data related to the user can be shared between the two websites and any actions (e.g., posting comments) that the origin website is allowed to perform on behalf of the user on the third-party provider (e.g., Facebook). Both cross-site data sharing and actions performed on behalf of the user can have significant privacy implications. In this paper we present a longitudinal study of the third-party authentication landscape, its structure, and the protocol usage, data sharing, and actions associated with individual third-party relationships. The study captures the current state, changes in the structure, protocol usage, and information leakage risks.

Chapter 4

Related Work

Here we present related work in the area of Web authentication for mobile users, on untrusted computers, with or without using third-parties. The first section includes related work on the concept of security levels, as well as Web authentication for untrusted computers. In the next section we focus on how to design Web authentication with a smartphone or other mobile phone as an authentication device, suitable for mobile users. We compare and contrast these solutions with our prototype 2-clickAuth, which we designed and implemented to show how mobile phones can be used to authenticate mobile users on untrusted computers. Finally, we discuss related work in third-party Web authentication, such as measurement and usability studies with focus both on security and privacy. Since our work on third-party Web authentication covers user privacy, usability factors, and protocol security we now present highlights from these different areas.

4.1 Security Levels and Evaluation

Measuring the amount or level of security in a system is difficult, and the conclusions can easily be biased by the person conducting the measurements. However, as in the related field of risk analysis and risk management, it is still valuable to be able to compare and rank authentication solutions based on how secure they are, in comparison with other benefits such as usability or technical requirements.

The NIST Electronic Authentication Guideline (EAG) [22] from 2006 inspired us to create a design and evaluation method for authentication solutions in which a mobile phone is used as a security device. The EAG contains a security level concept for authentication, however it has not been adapted to versatile authentication solutions in which many different side channels for data transfer could be used. In the later version of the EAG, from 2011, such channels are discussed [14]. However, while the EAG covers different areas of authentication and includes technical details and formal

security requirements, security is the only factor that is discussed. In our work, the aim is to combine the well-known and accepted security level concept with factors such as availability, usability and economic factors. This is intended to help developers and evaluators make the best choice among several solutions that meet the same security requirements.

At the time of creating our design and evaluation method, there was no other comparison method, to our knowledge, of authentication channels and methods made specifically for mobile phones. However, for the authentication system *Strong Authentication with Mobile Phones (SA)*, which uses a phone's SIM card in authentication, there is a comparison between the different modes of the system in which different channels are used. The comparison shows how the modes compare to each other when it comes to factors such as cost, infrastructure, security and usability [16]. Instead of cost and infrastructure, we focused on availability and usability, and the prevalent case with mobile users using untrusted computers.

There is also an authentication evaluation method in the area of IP Multimedia Subsystem (IMS), which considers several different factors such as security, simplicity and usability [59]. We use similar factors, but also include availability, since there are situations in which a mobile user cannot use, for example, the mobile phone network (e.g., GSM or 4G), or cannot install additional hardware or software, but still needs the authentication solution to be available and ready to use.

Our design and evaluation method, and the 2-clickAuth prototype, were designed when smartphones were rather new technology, and mobile phones were more of a complement to stationary computers and laptops, rather than the versatile computing devices they are today. With the rise of smartphones and access to the Internet over Wi-Fi or phone networks from everywhere, it is more common today (2016) to access websites directly from a smartphone than to use a public computer combined with a mobile phone as a security device. However, side-channels of mobile phones are increasingly used for two-factor authentication. Therefore, our design and evaluation method still holds, since it can be applied to solutions in which all parts of authentication and website usage are done from the same mobile device. We have also seen recent authentication solutions using similar concepts [100] as in our early prototype, which shows that our ideas are still applicable, even if there has been a shift in technology and hardware. We will now look at research in the area of Web authentication with mobile phones as security devices.

4.2 Authentication with Mobile Phones

There are a number of different approaches to strong authentication on the Web using mobile phones, which have been presented over the years [26].

4.2.1 Pre-Smartphone Solutions

In *Secure Web Authentication (SWA)*, the user receives a session name, both in the browser (on a computer) and as a text message (to their phone). If the names match, the user allows the session from their phone, and is now logged in on the computer [101].

The *Strong Authentication with Mobile Phone (SA)* method is a set of different solutions for challenge-response and OTP. All SA solutions rely on storing information on the user's SIM-card [16].

Confident consists of two authentication solutions where the first is visual knowledge-based authentication combined with OTP at first login, and the other is an OTP solution with the mobile phone as a side-channel [102].

Snap2Pass was released at the same time as our solution 2-clickAuth, and uses challenge-response with QR-codes in a fashion that is similar to our solution, but with the distinction that the response in Snap2Pass is sent via a side-channel over the phone network. This type of side channel can increase security, but also compromise availability, if a phone network is not available. Our solution was specifically designed to avoid side channels.

There are also other solutions, such as *Smokey (Synchronizable Mobile Key)*, which uses the mobile phone to strengthen authentication [103]. However, these solutions are adapted for internal (inside an organization) usage with local in-person enrollment.

We will now show examples of related work in the time of smartphones, which have increased capabilities as computing devices, but which also have increased security problems of their own.

4.2.2 Solutions in the Smartphone Era

SWA was proposed in 2004, SA in 2009 and Confident and Smokey both in 2010. While smartphones (i.e. mobile phones with computation capabilities, touch screen, Internet access and the possibility to install and run apps) were to some extent available in the late 90's and early 00's, the first iPhone was not released until 2007, and the first phone using the Android operating system in 2008. Large-scale Web usage from smartphones was not possible while the above-mentioned authentication solutions were developed. Internet access was slow and/or spotty, and websites were most conveniently visited from a computer. However, mobile phones (smart or not) were prevalent around 2010 and carried by most users. Therefore, using mobile phones in authentication was feasible from an availability perspective, since the user was likely to have access to a mobile phone at all times.

Today, solutions such as Confident [102] still remain, and new solutions to strengthen authentication and avoid phishing and man-in-the-browser attacks are still being proposed, such as the solution by Almeshekeh et al. in 2015 [100]. In the latter, QR codes are used to transfer authentication information as in our solution 2-clickAuth. QR codes for transferring authentication data and capturing this data with a smartphone are increasing

in popularity, and are now easier to implement since smartphone cameras and frameworks are becoming increasingly better.

Another QR-based solution is *WebTicket* (2012), using passwords transferred via QR-codes. In one version, the user prints the password (which is static) as a paper "ticket" and shows it to their Web camera. In another version, the password is displayed on the user's mobile phone display, similarly to 2-clickAuth. WebTicket makes it possible to use strong passwords since they do not need to be remembered by the user, but the method is open to eavesdropping and replay attacks. While not as secure as OTP and challenge-response solutions, it is an alternative to traditional passwords.

4.3 Third-Party Authentication and Privacy

There are different approaches to identity management and authentication for the Web. In contrast to authentication in general, the Web poses specific problems due to its world-wide scope, both regarding enrollment and the large attack surface that makes security vulnerabilities a problem. There are also usability, economic factors, availability and privacy to consider, when designing third-party authentication solutions. While identity management and third-party authentication help to solve a subset of the general Web authentication problems, new problems are introduced when connecting websites with each other for authentication purposes.

Our studies mainly focus on the existing third-party authentication landscape and the risks to user privacy, given the landscape features we have found. Such features include, but are not limited to, RP characteristics, protocol and IDP usage, and information transfer between RPs and IDPs.

We will now present related work trying to capture aspects of this existing landscape, including measurements and usability studies, as well as papers on protocol security for the major protocols we have observed in our measurement studies. Then, we present related work in the area of Web authentication, including novel third-party authentication methods, as well as other types of Web authentication which do not rely on third-parties. Finally, since Web authentication spans several fields, we present papers that combine security and privacy of Web authentication, especially in the third-party landscape, with other factors. Such cross-field aspects refer back to our previous works on evaluating Web authentication for untrusted computers.

4.3.1 Exploring the Third-Party Landscape

The third-party authentication landscape has evolved over the years, from a situation with many OpenID enabled accounts but very few RPs supporting login with an IDP [84] to a landscape dominated by OAuth and Web 2.0 services that share data between sites. Both in the current situation with few IDPs and earlier, when there were many IDPs, the risk of phishing

attacks is increasing, since users get used to being redirected between sites and providing their credentials to other sites [18]. Sun et al. [98] have shown that there are misconceptions about how third-party authentication works, which leads to many users avoiding third-party authentication, but also that users who agree to use IDPs are unsure about which information is being sent to different sites.

There are also several protocol-related security problems that enable identity theft and blackmailing [20, 80, 82]. Even if the OAuth protocol is proven to be formally secure [78], OAuth can still be implemented in an insecure way [81]. SSOScan [104] is a tool for finding vulnerabilities in the third-party authentication flow when using Facebook (with OAuth) to authenticate to RPs. A similar tool, AuthScan [105], is designed to crawl the Web for a specific set of authentication and identity management solutions (including SAML, OpenID and Facebook Connect). In our large-scale measurement of the RP-IDP landscape [27], we scanned more broadly for identity providers, while AuthScan scans for protocols. While we mapped the third-party identity management landscape to be able to calculate privacy risks, AuthScan searched specifically for known security vulnerabilities. In contrast to the focus on vulnerabilities and security, Birrell and Schneider [64] present a privacy-driven taxonomy of the design choices in the third-party identity management landscape.

From the public data available on large-scale social networks, which are often popular IDPs, it is often possible to derive potentially private information [106–108], or to correlate user data from several websites to identify a user based on behavior [109]. When connecting RPs and IDPs, non-public data also spreads between websites, and the risks increase, as we have shown in our work. Cross-site leakage and the associated privacy risks have also been studied in the context of ad services and trackers [110]. We will now present related works in the area of alternatives to existing Web authentication methods.

4.3.2 Alternatives to Existing Web Authentication

While federated identity management with Web-specific SSO solutions such as OpenID, and general protocols such as SAML, are decreasing, there is a rise in third-party identity management solutions with a single trusted third-party, especially with popular social networks serving as IDPs and OAuth as the protocol for authentication and authorization. We have chosen to study these third-party solutions because of their popularity and prevalence, but there are also competing solutions with very different approaches on how to do Web authentication.

An alternative option to third-parties and federations is to use user-centric identity management. Such authentication and identity management solutions, which use neither third-parties nor classical federated solutions, have benefits when it comes to security and privacy. In the field of user-

centric identity management solutions, users control their own credentials. Besides password storage solutions, which have security problems of their own [45,46], there are also devices aimed at replacing passwords for everyday use; e.g., the Pico token [10] and the Authentication Pad [62], later known as the OffPAD [11,57]. These personal authentication devices are designed to replace existing methods such as the wide-spread usage of passwords, but also to give the user control over their credentials, instead of using third-parties.

There is also research which focuses on the SILO model, with many user accounts which are not connected to each other. This research aims to strengthen passwords or improve the users' abilities to remember more passwords [111], and more complex passwords [44]. There are also initiatives to make password protocols more secure [43].

4.3.3 Factors in Third-Party Authentication

Other interesting related work includes Rannenberg's early work on bridging the gap between identity management and mobile phone authentication [112], as well as Xing et al.'s work on pollution attacks on social networks, which highlights the privacy problems in these services [113]. There have also been efforts in combining the OAuth protocol with other authentication protocols and factors, such as biometrics [114] and Microsoft Information Cards [88].

There have also been various initiatives on how to strengthen the security and privacy of online social networks (OSNs). Such work is related to ours, since these OSNs are commonly popular IDPs and/or RPs. Vulnerabilities and implementation problems have been reported in the SDKs of OSNs [115], and security APIs have been proposed for OSNs [116]. Anderson et al. focus on privacy in OSNs in general [117], Tarameshloo et al. propose specific privacy policies for information sharing in OSNs [118], which is similar to our work on permission agreements (app-rights), and Ko et al. also do research on this subject by studying the flows of private data between third-party applications and sites they connect to [119]. In a broader sense, Birrell and Schneider [64] have compared both existing and deprecated SSO solutions from a privacy perspective, similarly to our early work on authentication solutions. Their paper shows that independent of the solution, there are still many privacy problems left to be solved, and different solutions mitigate different problems.

Finally, Felt and Evans [120] proposed a method for only sharing a minimum of data between Facebook and applications connected to Facebook via OAuth. Their work was done in the field of internal applications built on the Facebook API, not external RP websites, but the ideas on how to minimize information sharing are still interesting. In our work we have not aimed to minimize the transferred data, but instead to map where data may spread when being leaked between RPs and IDPs.

4.4 Characterization of the Web

In our work on third-party authentication, a major part of the work is in the field of measurements and characterization of the Web. While our focus is third-party authentication relationships, there is related work in similar areas. In our large-scale characterization study of the Web [27], we compare and contrast the third-party authentication landscape with content sharing (delivery). A related study on content delivery was performed by Gill et al. [121], in that case with a focus on economic incentives of Web advertising, which is a typical form of content sharing. Krishnamurthy and Wills [122], present a study on content delivery in the form of advertising and other additional website content which is not the main type of content the site is serving, e.g., advertising on a social network. The authors show how users react to such additional content and how the users try to avoid it. Bent et al. took a broader approach when collecting data from a set of websites to study content delivery in this large-scale set [123]. The manual classification used in our work is inspired by Gill et al. [23] in which the authors characterize organizational usage of websites, and create a type classification for different websites.

Also related to our work, both our characterization study [27] and the paper in which we characterize RPs in detail [28], are studies on the structure of the Web in general such as the study by Fetterly et al. [124] on the evolution of websites over time. Others have presented longitudinal studies of video popularity [125], IPv6 adoption [126] and traceroute responses [127]. All these papers take a long-term perspective on the content and services delivered over the Web, but none of them consider the third-party authentication landscape.

Chapter 5

Discussion

In this chapter we discuss our findings and how the parts of the thesis fit together. We bridge the gap between the early papers on Web authentication and our later works on third-party authentication and privacy aspects.

5.1 Web Authentication in a Mobile World

When we started the work on Web authentication (2008), smartphones were not as prevalent as they are now (2016). For example, there were no phones running the Android operating system. However, early smartphones, small handheld tablets and pocket computers started gaining popularity, and could run software (apps) installed by the user, as well as accessing the Internet. Compared to the previous generations of mobile phones, these new devices had a wide range of short- and long-range communication channels, including a camera, which we used in the authentication prototype in our ARES paper [25] in 2010.

When we designed our optical authentication solutions with the phone camera, the phone was intended to be used as a standalone device when the user authenticated to a website on some device other than the phone (e.g., a laptop or stationary computer). There were similar solutions [128], but few which combined the benefits of using a mobile phone (instead of a dedicated device) with the simplicity of optical authentication, which allowed for strong challenge-response solutions without the user needing to type information into the phone.

Now, it is more common for a user to rely on their smartphone for most of their Web usage, either in the phone browser or in an app. There are still situations in which the user may want to involve another device, when authenticating to, for example, a computer at work, or a public computer when traveling, to add security in an untrusted environment. In such cases, the user's phone is always present and is a good choice for an authentication device. The main contributions from our first papers are mainly the compar-

ison method for authentication solutions, which can be adapted to different technologies and cases based on the situation. However, there are still new methods and solutions being designed which are similar to our original idea with 2-clickAuth, such as the solution proposed by Almeshekeh et al. [100].

As far as mobility and untrusted computers are concerned, users are more mobile than ever. Even if the users bring their own devices which they trust, the environment can be untrusted (e.g., an open Wi-Fi network). Also, even if a user uses their own device instead of a public computer, the device can be infected by malware and exploits, and therefore could also be untrusted.

5.2 Authentication vs. Identity Management

In this thesis we focus on Web authentication, both in general and with a third-party identity provider. Our focus has been on features of the authentication process itself: security, usability and privacy on a general level, as well as mobility and trust, as discussed above. We have evaluated the overall drawbacks and benefits of different authentication methods, data transfer channels and protocol and IDP usage.

Identity management is a broader field than authentication, since it includes the creation, storage and revocation of authentication credentials. We have chosen to focus especially on the authentication part of identity management, but still briefly discuss all three types of identity management in this thesis. The basic concepts of identity management, with third-parties and federations of websites, are commonly used when discussing and modeling third-party authentication. Therefore, we mention identity management in this thesis, and point out related work in the area. We also discuss SSO, since SSO was the basis of third-party authentication on the Web, before OAuth and the authentication landscape of today.

SSO, often in the context of federated identity management, is the original type of third-party authentication, with OpenID as a common protocol example. While OpenID is still used on the Web to a small extent, SAML, which was popular before OpenID, is not evident in the third-party authentication landscape we have studied. However, SAML is still used in internal cases, such as a company's internal Web infrastructure, or in federations in which university networks form their own internal federation.

While we have observed a decrease in OpenID usage, we have also observed an increasing usage of the OAuth protocol, in which a few large identity providers are used for authentication. OAuth is used the same way as authentication solutions are described in identity management with a trusted third-party. As an alternative, there is user-centric identity management, which more or less relies on removing the existing third-party structures from the Web, which does not seem likely in the near future. There is, however, a lot of interesting ongoing research in the field [10,11].

Instead of designing novel solutions, we have focused on how to analyze and be aware of the privacy and security risks of current authentication

solutions, and to help developers design and evaluate solutions which are more privacy and security preserving than existing ones.

The underlying reasons for why the RP-IDP landscape looks and evolves the way we have observed, and for why OAuth has become popular, are non-trivial to prove, but it seems to involve business benefits that are provided to both RPs and IDPs. The business model for both RPs and IDPs is increasingly built based on a rich set of data being shared between the two sites, and on RPs being able to promote their services by posting to the IDPs. Therefore, even with new SSO protocols such as Mozilla Persona, and with OpenID Connect (which relies on OAuth), authorization protocols will likely still be more popular than SSO-only protocols, since they provide capabilities for additional personalization by their users, as well as business opportunities for RPs and IDPs.

When capturing the structure, protocol usage and evolution of the RP-IDP landscape, as we have done, the design of the actual measurements and characterization of websites are also important parts of our work.

5.3 Measuring Web Authentication

The Web is rapidly changing, and popular websites and services come and go. Therefore, snapshots of the Web can be interesting for describing a current phenomenon, without giving a broader picture of Web usage in general. We wanted to develop and apply measurement methodologies that would be valid and useful even in the future. Therefore, we focused mostly on developing methodologies for measurement, such as our logarithmic sampling combined with click-based crawling of websites. Even if the Web and individual site relationships change, we expect that the Pareto-like distribution of website popularity, and the phenomenon of a few popular websites being much more popular than the rest will still hold [23]. Looking ten years back in time, before third-party authentication became largely used, the popular websites were portals, Webmail and chat services and online communities (according to the statistics on Alexa.com). These sites resembled the social networks of today, and were early adopters of personalization and providing personal website accounts for their users. It seems like these types of social sites have kept their popularity over the years and it is still relevant to study how they are used. We will now explain how we performed the crawling of websites, and how we selected which sites to crawl.

5.3.1 Crawling for Authentication

Measuring third-party Web authentication can be more complex than simple text matching, since only searching for potential IDP login links and OAuth connections may give false positives when encountering constructs similar to third-party authentication. The most common example is when a website let its users use another site (typically a social network) to comment on articles

or “like” contents. Such actions require authentication to the other website, but do not log the user in to an account on the original website. Therefore, such usage of third-parties is not counted as third-party authentication in our results.

We used a crawling methodology in which we designed a Selenium-based “clicker” which crawled a pre-defined set of websites and clicked on clickable elements. In contrast to text matching, a clicker can react to interactive Web 2.0 elements and therefore catch more RP-IDP connections. The clicker also incorporated matching on common OAuth and OpenID link structures. To avoid false matches on other OAuth usage, we also match on typical login-link constructions. However, there are many ways of designing an authentication interface on a website, and it is therefore difficult to catch all RP-IDP connections.

We also created alternative crawlers in order to investigate a more straightforward text-matching approach. Even though this approach turned out to produce an overly high number of false matches, it spurred an interest in more closely investigating the relations between content sharing and third-party authentication, since many content-sharing relationships were falsely classified as RP-IDP relationships. The content-sharing landscape is more well-known and better explored than the third-party authentication landscape, and this allowed for an interesting comparison of the two landscapes.

The pre-defined set of sites we crawled with our tools was selected by using our method for logarithmic sampling of websites. Based on the Pareto-distribution of website popularity, we wanted to capture the most popular websites, but also less popular websites, distributed among the 1 million most popular websites. Our logarithmic sampling gave us 35,619 sampled sites to crawl. We also collected a set of complementary datasets to gain more in-depth information on the sites which are RPs. The sets contain information about byte volume of the sites, and numbers of internal and external links. To better understand regional bias in IDP selection, we collected information on the approximate location of the IDP based on (i) a whois service based on where sites are registered, (ii) an online geo-location service located in the US to map the location of the IDP servers, and (iii) statistics provided by Alexa estimating the region in which the site’s primary user audience is located. While this data allowed us to classify the RPs along many dimensions, we acknowledge that many alternative metrics could have been collected, and alternative methods to classify and categorize websites could have been applied.

5.3.2 Limitations in Measurements

With the increasing popularity of mobile devices that are able to run applications (apps), many of these apps connect to each other in an RP-IDP fashion. However, in this work, we only consider websites and their specific problems. We also do not consider the internal usage of OAuth, where a

website allows users to run third-party plug-ins within the website. Our idea for this thesis is to study the evolution of Web authentication, rather than connected applications.

For some cases it is difficult to decide if a website is an RP or not, even after manually verifying that the website does not have a false IDP-connection and the website clearly allows login with an identity from another website. An example is when a website consists of different parts which are connected to each other with OAuth, such as in the case where users of a Web portal can log in with their portal username on a video service run by the portal, for example. If the different parts have different domain names we distinguish between them, otherwise not. However, we consider obvious same-owner relationships to be internal, regardless of protocol usage. A similar example is when an IDP is bought by another IDP, but with the login links still pointing towards the old IDP. The user gets redirected to the new IDP when trying to login, but at a glance or for a text-matching tool, it still looks like the old IDP is in use. Problems also occur when IDPs are removed by the RP, or go out of business, but the RP GUI is not redesigned, thus still indicating that the IDP is in use. Temporary downtime of websites, both RPs and IDPs, also complicate measurements. Finally, there are websites which change their domain names, which may lead to broken login links for IDPs and popularity changes for RPs (e.g., the service is still popular, but the old domain drops in popularity as users start using the new domain name). The most notable example in our sets is when Chinese website Sina.com.cn moved their popular social network Weibo to its own domain Weibo.com. Another example is Russian social network Vkontakte.ru changing its name to Vk.com and seemingly dropping in popularity, while the service was actually becoming even more popular.

5.4 Changes over Time

Another interesting aspect of measuring the Web is that it is constantly changing, both because of new technological inventions and also because of user demands and economical factors. The RP-IDP landscape has changed from having many IDPs and few RPs [84], to its current structure which is the opposite. While being IDPs gave websites a chance to be noticed (free “advertising” due to their visibility at RPs) and new potential users, RPs would get fewer registered users, since the users would log in with the IDP instead. Now, with the rise of OAuth, being an RP gives a website a rich set of user data which can be used for targeted advertising and personalization. The RP can also post on the IDP, thus promoting itself to new potential users. Being an IDP is still beneficial for the previously-mentioned reasons. In addition, the IDP now also has access to posted material sent from the RP, which creates more data traffic.

By performing longitudinal studies, both of the overall RP-IDP landscape, as well as permission agreements between RPs and IDPs, we have

observed trends over several years, and identified an overall bias towards less privacy-preserving third-party authentication on the Web.

5.5 Data Classification

One of the aspects we have considered in our longitudinal study is what data is being shared between the RP and the IDP. We classified the data as basic data (B) such as public identifiers and other non-sensitive information, personal data (P), created content (C) or data belonging to the user's friends (F).

When we defined risk classes based on combinations of data types being shared, we built them from observed combinations. The risk classes still hold, even if some RPs and IDPs in our dataset have changed their APIs and specific app-rights. We have seen that B-data only is decreasing, as IDPs often bundle information of B and P-type into blocks of information which RPs always import when using that specific IDP.

Besides data, IDPs often allow RPs to perform actions on the IDP on behalf of the user. Whether or not an IDP allows this does not affect the risk class in our risk classification. However, when combining several IDPs, multi-hop leaks can more easily occur when actions are involved. When discussing risk, we bundle all actions, but when discussing multi-hop leaks, we therefore split actions into write and update/remove, since the latter only changes information on one IDP, while write actions can spread information in multiple steps.

Besides multi-hop leaks, there are also other privacy-related problems when connecting multiple IDPs to one RP. Our studies show that RPs do not handle multiple IDPs well. For example, the user may end up with several RP accounts (one per IDP), temporary and limited RP-accounts when using an IDP for login, or an RP-account on which information from several IDPs is merged in a way the user may not approve of, potentially revealing sensitive information.

5.6 Privacy Aspects

We have focused on user privacy in third-party authentication and calculated potential privacy risks in the RP-IDP landscape. It is worth noticing that we make the assumption that the authentication is not compromised in any way, and that the only information being transferred is the information which the user has agreed to share. Thus, we are not considering security breaches, only leaks that occur when several data flows, accepted by the user, collide and/or interfere with each other.

We use the term leaks in the sense that information is shared in several steps without the user being aware of the complete flow of information.

However, we only consider information that the user has consented to sharing. Our idea is that since RPs and IDPs are connected into complex graphs, and/or structures with a few IDPs being used by many RPs, the user cannot easily get an overview of all possible leakage paths. Even if the user accepts two different information flows, the user is not likely to calculate the potential effect of the combination of these two flows.

Our leakage calculations are based on users having connected accounts with several IDPs and RPs. Of course the problems become smaller if users do not connect accounts. However, users today normally do not connect accounts for the sake of SSO, but rather to be able to share information between popular social networks (IDPs) and content creation sites (RPs, such as photo and music services). Therefore, users may choose to access an RP with several IDPs, not for simplifying login, but to connect the different accounts and share information. However, such multi-IDP scenarios also come at a privacy cost. For example, in our experiments we have seen that RPs do not handle multi-IDP scenarios well, and may merge account information from different IDPs in a way which can compromise user privacy.

We do not try to weigh risks based on the number of actual users of specific IDPs and RPs. We only determine if an IDP is being used by an RP and discuss the potential risks that users associated with such RP-IDP relationships may experience. However, for our discussion we do assume that highly popular social networks have many users using them as IDPs, in contrast to specialized pro-IDPs.

Overall, we have observed an increase in the usage of the less privacy-preserving OAuth protocol, which allows for sharing a rich set of data between RPs and IDPs. We have also observed a few large OAuth-IDPs becoming increasingly more popular as IDPs, while also being highly popular social networks with rich user data to share. We have seen these popular IDPs being combined with each other in pairs or triples, often with at least one of the IDPs allowing write-actions. Such use of actions, as allowed by the OAuth protocol, may lead to multi-hop data leaks. The increased entanglement of the Russian Web, and the still nested structure of the Chinese Web, also have a high number of potential data leaks. With users being increasingly mobile, using their devices in untrusted environments, there is a need for stronger authentication to protect user privacy, especially at the IDP level. Perhaps the recent move by some top IDPs to offer their users two-factor authentication will help here.

Chapter 6

Summary and Conclusions

In this chapter we present a summary of our work, our conclusions and a discussion of future work.

6.1 Summary

We have studied Web authentication in untrusted environments, especially for users which are mobile in the sense that they use different devices in different places, to access their website accounts. In untrusted environments, security measures must be designed and implemented so that they are in place, even if the device being used is untrusted, or the user is in an insecure location (which may be a crowded place prone to eavesdropping, or a place in which the user may be fooled into using an untrusted Wi-Fi network). While one solution to providing secure authentication for untrusted computers is to equip the user with a dedicated authentication device, these solutions rely on users to carry these devices and have them available each time the user wants to authenticate. We have designed and implemented a proof-of-concept prototype of an authentication solution which uses a device the user is likely to carry, their mobile phone, and which combines security with other factors that are important in untrusted environments with mobile users.

Websites have different security requirements, and it may be difficult to compare different authentication solutions to find a solution which fits the requirements of a specific type of site. We wanted to know how to evaluate and compare Web authentication solutions to be able to ensure that a solution reaches a specific level of security. We used the security level concept from NIST, but extended the concept to adapt the levels to solutions in which authentication tokens with many communication channels, such as mobile phones, were taken into account. We then designed and developed a methodology for comparing and evaluating authentication solutions which are based on the use of mobile phones as additional hardware devices for authentication.

Identity providers in third-party authentication are a specific type of Web service which may need secure authentication while also being easily accessible for mobile users, since these sites are used for providing authentication to users of other sites and services on the Web. We performed a large-scale empirical study of the current third-party authentication landscape, to determine which websites are using which type of third-party Web authentication and how they handle security and privacy. We collected information, both longitudinal and large-scale, on how the use of third-party authentication differs between websites of different types, target audience, actual audience and country of origin. We designed and developed a novel logarithm-based sampling technique together with a crawling technique adapted for Web 2.0, to crawl and find relationships between websites and their third-party identity providers. To put our results in context we have also compared third-party authentication with the more well-known field of content sharing. We discovered that in third-party authentication the usage is more biased towards using a few large providers and relying on the security of these providers. There is also a large bias towards using authentication providers from the same region (e.g., a part of the world with the same language), compared with how content providers are selected.

We also discovered that the few large providers that are in use mainly rely on the OAuth protocol, which allows authorization of applications to act on behalf of the user, as well as sharing of a rich set of user data between sites. We performed case studies on how to minimize privacy risks in third-party authentication, based on which protocols are used and what data is being shared. We discovered that combinations of specific identity providers can be dangerous to user privacy and therefore made in-depth studies on these combinations of sites and the data they transfer. We have defined categories of data types transmitted in third-party authentication, and designed a risk categorization based on combinations of these types that are known to be in use. We have also categorized problems related to account merging and cross-site information leaks that are related to these popular combinations of third-party identity providers.

When modeling the third-party authentication landscape, we quickly realized that the majority of relationships are captured by a bipartite graph structure, with a small layer of identity providers and a large layer of all the sites relying on these providers. In our model we also allow for intra-layer edges to capture the relationships of hybrid nodes. By modeling the identity landscape we have been able to determine how privacy risks may spread between websites and we are now able to recommend how to use and combine third-party authentication in a privacy-preserving way. We have also made a longitudinal evaluation of the top 200 most popular websites worldwide and these sites' usage of third-party authentication over the past three years. Overall, we have observed a growing trend of websites using a few very large social networks as their identity providers in third-party authentication.

6.2 Conclusions

Web authentication is an ever-changing field. But, while new technology and methods arise, simple and usable practices will still remain widespread. Still, there is a need for strong authentication methods suitable for specific websites and services. We have explored the area of Web authentication in terms of combining security with factors such as usability and privacy. Existing frameworks for evaluating the security of existing authentication, must include other factors besides security, and must be adapted for the security level of the application.

Third-party authentication makes the field of Web authentication even more complex, especially when comparing and evaluating different solutions. While third-party solutions are often beneficial to users, they also introduce various new security and privacy risks, some of which may be evaluated vastly differently by different users.

To capture the structure of the RP-IDP landscape we have developed a methodology for logarithmic sampling of websites, combined with a click-based crawler, which evaluates websites of different types and of varying popularity. We have also studied the changes in IDP and protocol usage as well as data-sharing between sites. By classifying the data into different types we were able to create a risk classification, showing that many RP-IDP relationships create a medium-high risk to the user's privacy by sharing more data than is actually used by the RP, e.g. for filling out account details for the user. While users may benefit from data sharing, in terms of personalization and seamless integration of social networks, sensitive data belonging to the user may end up on websites where the user did not intend to reveal this particular data.

Over the past ten years several large websites serving as social networks and identity providers have dropped in popularity or been forced to change their business models drastically, due to how they handled user data, security, and privacy. How different websites handle third-party authentication may have a similar effect on how the Web landscape will change over the next ten years.

6.3 Future Work

The future work of this thesis includes the following directions:

Security levels and mobility: With new emerging Web services, and an increased usage of third-party authentication, security levels not only need to be adapted to the security requirements of individual websites or services, but on the whole structure of RPs and IDPs connected to the website. Also, while users are becoming increasingly mobile, they are likely to use their own devices rather than public computers. In work environments the bring-your-own-device concept is gaining popularity which brings new challenges to the design of authentication solutions for different purposes.

Simple and secure authentication: Passwords are still very common in authentication, even though they have many security problems. It is a challenge both to design novel solutions which can compete with passwords and to improve the existing password-based authentication methods. With mobile phones getting more capable as computing devices, they can be used in authentication solutions based on concepts that are similar to our 2-clickAuth prototype [25].

RP-IDP measurements: Our measurement studies could be followed up in the future, using a similar methodology and a larger data set, to see how the landscape has evolved over a longer time. It would also be of interest to see studies from the user’s point of view on account merging when allowing multiple IDPs. There are several usability studies in similar fields [70, 85], but none at the time of writing on account merging, which we have explored from the user’s point of view.

IDP usage in the wild: The actual usage of IDPs, measured by the number of users actually using a specific IDP with an RP, would complement our results. Such a study could rely on network traffic to measure parts of the usage of specific IDPs and give insight into how users tend to use and combine IDPs [129].

Risk modeling: Finally, our work on modeling the RP-IDP landscape as a bipartite graph and characterizing risk spreading between websites could be extended and applied to other phenomena on the Web, for example content sharing, fingerprinting and tracking, which may have similar properties to the RP-IDP landscape. Comparing the structures within the graph may allow for better head-to-head comparisons of these aspects of the current Web.

References

- [1] Kwei-Jay Lin. Building Web 2.0. *IEEE Computer*, 40(5):101–102, May 2007.
- [2] David Dagon, Tom Martin, and Thad Starner. Mobile phones as computing devices: The viruses are coming! *IEEE Pervasive Computing*, 3(4):11–15, October–December 2004.
- [3] George Lawton. Is it finally time to worry about mobile malware? *IEEE Computer*, 41(5):12–14, May 2008.
- [4] Esma Aïmeur and David Schönfeld. The ultimate invasion of privacy: Identity theft. In *Proc. Intl. Conf. on Privacy, Security and Trust (PST)*, pages 24–31, Montreal, Canada, July 2011.
- [5] Matt Bishop. *Introduction to computer security*. Addison-Wesley Professional, 2004.
- [6] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Proc. IEEE Symposium on Security & Privacy*, pages 553–567, San Francisco, US, May 2012.
- [7] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The tangled web of password reuse. In *Proc. Network and Distributed System Security Symposium (NDSS)*, pages 23–26, San Diego, US, February 2014.
- [8] Joseph Bonneau and Sören Preibusch. The password thicket: technical and market failures in human authentication on the Web. In *Proc. Workshop on the Economics of Information Security (WEIS)*, pages 230–237, Cambridge, US, June 2010.
- [9] Cormac Herley, Paul C. van Oorschot, and Andrew S. Patrick. Passwords: If we’re so smart, why are we still using them? In *Proc. Financial Cryptography and Data Security (FC)*, pages 230–237, Accra Beach, Barbados, February 2009.

-
- [10] Frank Stajano. Pico: No more passwords! In *Proc. International Security Protocols Workshop (SPW)*, pages 49–81, Cambridge, UK, March 2011.
 - [11] Kent Are Varmedal, Henning Klevjer, Joakim Hovlandsvåg, Audun Jøsang, Johann Vincent, and Laurent Miralabé. The OffPAD: Requirements and usage. In *Proc. Network and System Security (NSS)*, pages 80–93, Madrid, Spain, June 2013.
 - [12] Simon S. Y. Shim, Geetanjali Bhalla, and Vishnu Pendyala. Federated identity management. *IEEE Computer*, 38(12):120–122, December 2005.
 - [13] Lucy Lynch. Inside the identity management game. *IEEE Internet Computing*, 15(5):78–82, September–October 2011.
 - [14] William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, and W. Timothy Polk. Electronic authentication guideline. Technical report, NIST, December 2011.
 - [15] Charlott Lorentzen, Markus Fiedler, Henric Johnson, Junaid Shaikh, and Ivar Jørstad. On user perception of web login - a study on QoE in the context of security. In *Proc. Australasian Telecommunication Networks and Applications Conference (ATNAC)*, pages 84–89, Auckland, New Zealand, October–November 2010.
 - [16] Do van Thanh, Ivar Jørstad, Tore Jønvik, and Do van Thuan. Strong authentication with mobile phone as security token. In *Proc. IEEE Intl. Conf. on Mobile Adhoc and Sensor Systems (MASS)*, pages 777–782, Macau, China, October 2009.
 - [17] Eve Maler and Drummond Reed. The venn of identity. *IEEE Security & Privacy*, 6(2):16–23, March–April 2008.
 - [18] Rachna Dhamija and Lisa Dusseault. The seven flaws of identity management. *IEEE Security & Privacy*, 6(2):24–29, March–April 2008.
 - [19] Susan Landau, Hubert Lê Van Gông, and Robin Wilton. Achieving privacy in a federated identity management system. In *Proc. Financial Cryptography and Data Security (FC)*, pages 51–70, Accra Beach, Barbados, February 2009.
 - [20] Rui Wang, Shuo Chen, and XiaoFeng Wang. Signing me onto your accounts through Facebook and Google: A traffic-guided security study of commercially deployed single-sign-on Web services. In *Proc. IEEE Symposium on Security & Privacy*, pages 365–379, San Francisco, US, May 2012.
 - [21] Dinei Florêncio and Cormac Herley. A large-scale study of Web password habits. In *Proc. Intl. Conf. on World Wide Web (WWW)*, pages 657–666, Banff, Canada, May 2007.

- [22] William E. Burr, Donna F. Dodson, and W. Timothy Polk. Electronic authentication guideline. Technical report, NIST, 2006.
- [23] Phillipa Gill, Martin Arlitt, Niklas Carlsson, Anirban Mahanti, and Carey Williamson. Characterizing organizational use of Web-based services: Methodology, challenges, observations, and insights. *ACM Transactions on the Web (TWEB)*, 5(4):19:1–19:23, October 2011.
- [24] Anna Vapen and Nahid Shahmehri. Security levels for web authentication using mobile phones. In Simone Fischer-Hübner, Penny Duquenoy, Marit Hansen, Ronald Leenes, and Ge Zhang, editors, *Privacy and Identity Management for Life*, volume 352 of *IFIP Advances in Information and Communication Technology*, pages 130–143. Springer Berlin Heidelberg, 2011.
- [25] Anna Vapen, David Byers, and Nahid Shahmehri. 2-clickAuth optical challenge-response authentication. In *Proc. Intl. Conf. on Availability Reliability and Security (ARES)*, pages 79–86, Krakow, Poland, February 2010.
- [26] Anna Vapen and Nahid Shahmehri. 2-clickAuth: Optical challenge-response authentication using mobile handsets. *Int. J. Mob. Comput. Multimed. Commun.*, 3(2):1–18, April 2011.
- [27] Anna Vapen, Niklas Carlsson, Anirban Mahanti, and Nahid Shahmehri. Third-party identity management usage on the Web. In *Proc. Passive and Active Measurement Conference (PAM)*, pages 151–162, Los Angeles, US, March 2014.
- [28] Anna Vapen, Niklas Carlsson, Anirban Mahanti, and Nahid Shahmehri. A look at the third-party identity management landscape. *IEEE Internet Computing*, 20(2):8–25, March–April 2016.
- [29] Anna Vapen, Niklas Carlsson, Anirban Mahanti, and Nahid Shahmehri. Information sharing and user privacy in the third-party identity management landscape. In *Proc. IFIP ICT Systems Security and Privacy Protection Conference (SEC)*, pages 174–188, Hamburg, Germany, May 2015.
- [30] Anna Vapen, Niklas Carlsson, and Nahid Shahmehri. Longitudinal analysis of the third-party authentication landscape. In *Proc. Understanding and Enhancing Online Privacy Workshop (UEOP)*, San Diego, US, February 2016.
- [31] Serge Egelman. My profile is my password, verify me!: The privacy/convenience tradeoff of Facebook Connect. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 2369–2378, Paris, France, April–May 2013.

-
- [32] Vincent Lenders, Emmanouil Koukoumidis, Pei Zhang, and Margaret Martonosi. Location-based trust for mobile user-generated content: Applications, challenges and implementations. In *Proc. Workshop on Mobile Computing Systems and Applications (HotMobile)*, pages 60–64, Napa Valley, California, February 2008.
- [33] Dorothy E Denning and Peter F MacDoran. Location-based authentication: Grounding cyberspace for better security. *Computer Fraud & Security*, 1996(2):12–16, 1996.
- [34] Alina Hang, Alexander De Luca, Matthew Smith, Michael Richter, and Heinrich Hussmann. Where have you been? Using location-based security questions for fallback authentication. In *Proc. Symposium On Usable Privacy and Security (SOUPS)*, pages 169–183, Ottawa, Canada, July 2015.
- [35] Google. Two-step verification from google. <http://www.google.com/landing/2step/>, 2013. Accessed: May 20, 2013.
- [36] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. How does your password measure up? The effect of strength meters on password creation. In *Proc. USENIX Security*, pages 65–80, Bellevue, US, August 2012.
- [37] Claude Castelluccia, Markus Dürmuth, and Daniele Perito. Adaptive password-strength meters from Markov models. In *Proc. Network and Distributed System Security Symposium (NDSS)*, San Diego, US, February 2012.
- [38] Xavier De Carné De Carnavalet and Mohammad Mannan. From very weak to very strong: Analyzing password-strength meters. In *Proc. Network and Distributed System Security Symposium (NDSS)*, pages 23–26, San Diego, US, February 2014.
- [39] Joseph Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Proc. IEEE Symposium on Security & Privacy*, pages 538–552, San Francisco, May 2012.
- [40] Sören Preibusch and Joseph Bonneau. The password game: Negative externalities from weak password practices. In *Proc. Intl. Conf. on Decision and Game Theory for Security (GameSec)*, pages 192–207, Berlin, Germany, November 2010.
- [41] David Goldman. More than 6 million linkedin passwords stolen. <http://money.cnn.com/2012/06/06/technology/linkedin-password-hack/index.htm>, 2012. Accessed: May 20, 2013.

- [42] Ijlal Loutfi and Audun Jøsang. Passwords are not always stronger on the other side of the fence. In *Proc. Workshop on Usable Security (USEC)*, San Diego, US, February 2015.
- [43] Joseph Bonneau. Getting Web authentication right: A best-case protocol for the remaining life of passwords. In *Proc. Intl. Conf. on Security Protocols (SP)*, pages 98–104, Cambridge, UK, March 2011.
- [44] Joseph Bonneau and Stuart Schechter. Towards reliable storage of 56-bit secrets in human memory. In *Proc. USENIX Security*, pages 607–623, San Diego, US, August 2014.
- [45] Zhiwei Li, Warren He, Devdatta Akhawe, and Dawn Song. The emperor’s new password manager: Security analysis of web-based password managers. In *Proc. USENIX Security*, pages 465–479, San Diego, US, August 2014.
- [46] David Silver, Suman Jana, Dan Boneh, Eric Chen, and Collin Jackson. Password managers: Attacks and defenses. In *Proc. USENIX Security*, pages 449–464, San Diego, US, August 2014.
- [47] Robert W. Reeder and Stuart Schechter. When the password doesn’t work: Secondary authentication for websites. *IEEE Security & Privacy*, 9(2):43–49, March–April 2011.
- [48] Joseph Bonneau, Mike Just, and Greg Matthews. What’s in a name? evaluating statistical attacks on personal knowledge questions. In *Proc. Financial Cryptography and Data Security (FC)*, pages 98–113, Tenerife, Spain, January 2010.
- [49] Markus Jakobsson, Erik Stolterman, Susanne Wetzels, and Liu Yang. Love and authentication. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 197–200, Florence, Italy, April 2008.
- [50] Markus Jakobsson and Hossein Siadati. Improved visual preference authentication. *Proc. Workshop on Socio-Technical Aspects in Security and Trust (STAST)*, pages 27–34, June 2012.
- [51] Payas Gupta, Swapna Gottipati, Jing Jiang, and Debin Gao. Your love is public now: Questioning the use of personal information in authentication. pages 49–60, Hangzhou, China, May 2013.
- [52] Richard Shay, Pg Kelley, and Saranga Komanduri. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *Proc. Symposium on Usable Privacy and Security (SOUPS)*, Washington, US, July 2012.
- [53] Glen Nielsen and Michael Vedel. Improving usability of passphrase authentication. Technical report, DTU, 2009.

-
- [54] Robert Biddle, Sonia Chiasson, and Paul C. van Oorschot. Graphical passwords. *ACM Computing Surveys*, 44(4):1–41, August 2012.
 - [55] Ziming Zhao, Gail-Joon Ahn, Jeong-Jin Seo, and Hongxin Hu. On the security of picture gesture authentication on the security of picture gesture authentication. In *Proc. USENIX Security*, pages 383–398, Washington, US, August 2013.
 - [56] Atanas Filyanov, Jonathan M. McCune, Ahmad Reza Sadeghiz, and Marcel Winandy. Uni-directional trusted path: Transaction confirmation on just one device. In *Proc. Intl. Conf. on Dependable Systems and Networks (DSN)*, pages 1–12, Hong Kong, June 2011.
 - [57] Audun Jøsang, Christophe Rosenberger, Laurent Miralabé, Henning Klevjer, Kent A. Varmedal, Jérôme Daveau, Knut Eilif Husa, and Petter Taugbøl. Local user-centric identity management. *Journal of Trust Management*, 2(1):1–28, 2015.
 - [58] Audun Jøsang. Identity management and trusted interaction in internet and mobile computing. *IET Information Security*, 8(2):67–79, April 2013.
 - [59] Charlott Eliasson, Markus Fiedler, and Ivar Jørstad. A criteria-based evaluation framework for authentication schemes in IMS. In *Proc. Intl. Conf. on Availability, Reliability and Security (ARES)*, pages 865–869, Fukuoka, Japan, March 2009.
 - [60] Leslie Lamport. Password authentication with insecure communication. *Communications of the ACM*, 24(11):770–772, November 1981.
 - [61] Nikolaos Karapanos, Claudio Marforio, Claudio Soriente, and Srdjan Čapkun. Sound-proof: Usable two-factor authentication based on ambient sound. In *Proc. USENIX Security*, pages 483–498, Washington, US, August 2015.
 - [62] Audun Jøsang, John Fabre, Brian Hay, James Dalziel, and Simon Pope. Trust requirements in identity management. In *Proc. Australasian Workshop on Grid Computing and E-research (ACSW Frontiers)*, pages 99–108, Sydney, Australia, January–February 2005.
 - [63] Michal Procházka, Daniel Kouřil, and Luděk Matyska. User centric authentication for Web applications. In *Proc. Intl. Symposium on Collaborative Technologies and Systems, (CTS)*, pages 67–74, Chicago, US, May 2010.
 - [64] Eleanor Birrell and Fred B. Schneider. Federated identity management systems: A privacy-based characterization. *IEEE Security & Privacy*, 11(5):36–48, September–October 2013.

- [65] Jostein Jensen. Federated identity management challenges. In *Proc. Intl. Conf. on Availability, Reliability and Security (ARES)*, pages 230–235, Prague, Czech Republic, August 2012.
- [66] Wolfgang Hommel and Helmut Reiser. Federated identity management: Shortcomings of existing standards. In *Proc. Symposium on Integrated Network Management (IM), Managing New Networked Worlds*, Nice, France, May 2005.
- [67] Juraj Somorovsky, Andreas Mayer, Jörg Schwenk, Marco Kampmann, and Meiko Jensen. On breaking SAML: Be whoever you want to be. 2012.
- [68] Alessandro Armando, Roberto Carbone, Luca Compagna, Jorge Cuelar, and Llanos Tobarra. Formal analysis of SAML 2.0 web browser single sign-on: Breaking the SAML-based single sign-on for Google apps. In *Proc. ACM Workshop on Formal Methods in Security Engineering*, pages 1–10, Alexandria, US, October 2008.
- [69] Birgit Pfitzmann and Michael Waidner. Federated identity-management protocols - where user authentication protocols may go. In *Proc. International Workshop on Security Protocols*, pages 153–174, Cambridge, UK, April 2005.
- [70] San-Tsai Sun, Eric Pospisil, and Konstantin Beznosov. What makes users refuse Web single sign-on? An empirical investigation of OpenID categories and subject descriptors. In *Proc. Symposium on Usable Privacy and Security (SOUPS)*, Pittsburgh, US, July 2011.
- [71] Ross Anderson. Can we fix the security economics of federated authentication? In *Lecture Notes in Computer Science (7114 LNCS), Postproc. of Workshop on Security Protocols (SPW)*, pages 25–32, Cambridge, UK, March 2011.
- [72] David P. Kormann and Aviel D. Rubin. Risks of the passport single signon protocol. *Computer Networks*, 33(1-6):51–58, June 2000.
- [73] David Recordon and Drummond Reed. OpenID 2.0: A platform for user-centric identity management. In *Proc. ACM Workshop on Digital Identity Management (DIM)*, pages 11–16, Alexandria, US, October – November 2006.
- [74] Bart van Delft and Martijn Oostdijk. A security analysis of OpenID. In *Proc. Identity Management (IDMAN)*, pages 73–84, Oslo, Norway, November 2010.
- [75] Abhilasha Bhargav-Spantzel, Anna C. Squicciarini, and Elisa Bertino. Trust negotiation in identity management. *IEEE Security & Privacy*, 5(2):55–63, March 2007.

-
- [76] Gergely Alpár, Jaap-Henk Hoepman, and Johanneke Siljee. The identity crisis. security, privacy and usability issues in identity management. *arXiv preprint arXiv:1101.0427*, 2011.
- [77] San-Tsai Sun, Kirstie Hawkey, and Konstantin Beznosov. Systematically breaking and fixing OpenID security: Formal analysis, semi-automated empirical evaluation, and practical countermeasures. *Computers and Security*, 31(4):465–483, June 2012.
- [78] Suhas Pai, Yash Sharma, Sunil Kumar, Radhika M. Pai, and Sanjay Singh. Formal verification of OAuth 2.0 using Alloy framework. In *Proc. Intl. Conf. on Communication Systems and Network Technologies (CSNT)*, pages 655–659, Katra, India, June 2011.
- [79] Suresh Chari, Charanjit Jutla, and Arnab Roy. Universally composable security analysis of OAuth v2.0. Technical report, 2011. Cryptology ePrint Archive, Report 2011/526.
- [80] San-Tsai Sun. Simple but not secure: An empirical security analysis of OAuth 2.0-based single sign-on systems. Technical report, 2012.
- [81] San-Tsai Sun and Konstantin Beznosov. The devil is in the (implementation) details. In *Proc. ACM Conference on Computer and Communications Security (CCS)*, pages 378–390, Raleigh, US, October 2012.
- [82] Ethan Shernan, Henry Carter, Dave Tian, Patrick Traynor, and Kevin Butler. More guidelines than rules: CSRF vulnerabilities from noncompliant OAuth 2.0 implementations. In *Proc. Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, pages 239–260, Milan, Italy, July 2015.
- [83] Barry Leiba. OAuth web authorization protocol. *IEEE Internet Computing*, 16(1):1–47, January–February 2012.
- [84] San-Tsai Sun, Yazan Boshmaf, Kirstie Hawkey, and Konstantin Beznosov. A billion keys, but few locks: The crisis of Web single sign-on. In *Proc. Workshop on New security Paradigms (NSPW)*, pages 61–71, Concord, US, September 2010.
- [85] Mohamed Shehab, Said Marouf, and Christopher Hudel. ROAuth: Recommendation based open authorization categories and subject descriptors. In *Proc. Symposium on Usable Security and Privacy (SOUPS)*, Pittsburgh, US, July 2011.
- [86] Marino Miculan and Caterina Urban. Formal analysis of Facebook connect single sign-on authentication protocol. In *Proc. Student Research Forum of the Intl. Conf. on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, pages 22–28, Nový Smokovec, Slovakia, January 2011.

- [87] Ivar Jørstad, Tor Anders Johansen, Elias Bakken, Charlott Eliasson, and Do van Thanh. Releasing the potential of OpenID & SIM. In *Proc. Intl. Conf. on Intelligence in Next Generation Networks (ICIN)*, pages 1–6, Bordeaux, France, October 2009.
- [88] Haitham S. Al-Sinani. Integrating OAuth with information card systems. In *Proc. Intl. Conf. on Information Assurance and Security (IAS)*, pages 198–203, Melaka, Malaysia, December 2011.
- [89] OpenID Foundation. Get an OpenID. <http://openid.net/get-an-openid>, 2009. Accessed: May 10, 2010.
- [90] Ernesto Damiani, Sabrina De Capitani di Vimercati, and Pierangela Samarati. Managing multiple and dependable identities. *IEEE Internet Computing*, 7(6):29–37, November 2003.
- [91] Christine Satchell, Graeme Shanks, Steve Howard, and John Murphy. Identity crisis: user perspectives on multiplicity and control in federated identity management. *Behaviour & Information Technology*, 30(1):51–62, January 2011.
- [92] Sebastian Clauß and Marit Köhn top. Identity management and its support of multilateral security. *Computer Networks*, 37(2):205–219, 2001.
- [93] Monica Chew and Sid Stamm. Contextual identity: Freedom to be all your selves. In *Proc. Web 2.0 Security & Privacy Workshop (W2SP)*, San Francisco, US, May 2013.
- [94] Kim Cameron. The laws of identity. Technical report, May 2005.
- [95] Abdelmounaam Rezgui, Athman Bouguettaya, and Mohamed Y. El-toweissy. Privacy on the web: Facts, challenges, and solutions. *IEEE Security & Privacy*, 1(6):40–49, November–December 2003.
- [96] Željko Obrenović and Bart Den Haak. Integrating user customization and authentication: The identity crisis. *IEEE Security & Privacy*, 10(5):82–85, September–October 2012.
- [97] Cormac Herley. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Proc. Workshop on New Security Paradigms (NSPW)*, pages 133–144, Oxford, UK, September 2009.
- [98] San-Tsai Sun, Eric Pospisil, Ildar Muslukhov, Nuray Dindar, Kirstie Hawkey, and Konstantin Beznosov. Investigating users’ perspectives of Web single sign-on: Conceptual gaps and acceptance model. *ACM Transactions on Internet Technology (TOIT)*, 13(1):2:1–2:35, November 2013.

-
- [99] Audun Jøsang. Authentication assurance with identity management models, November 2010. Keynote presentation at IFIP IDMAN’10.
 - [100] Mohammed H. Almeshekeh, Mikhail J. Atallah, and Eugene H. Spafford. Enhancing passwords security using deceptive covert communication. In *Proc. IFIP ICT Systems Security and Privacy Protection (SEC)*, pages 159–173, Hamburg, Germany, May 2015.
 - [101] Min Wu, Simson Garfinkel, and Rob Miller. Secure web authentication with mobile phones. In *Proc. DIMACS Workshop on Usable Privacy and Security Software*, Piscataway, US, July 2004.
 - [102] Confident Technologies Inc. Intuitive and secure image-based authentication solutions. <http://www.confidenttechnologies.com>, 2010. Accessed: May 11, 2010.
 - [103] Johannes Hassmund. Securing credentials on untrusted clients. Technical report, 2010.
 - [104] Yuchen Zhou and David Evans. SSOScan: Automated testing of Web applications for single sign-on vulnerabilities. In *Proc. USENIX Security*, pages 495–510, San Diego, US, August 2014.
 - [105] Guangdong Bai, Jike Lei, Guozhu Meng, and Sai Sathyanarayan Venkatraman. AUTHSCAN: Automatic extraction of web authentication protocols from implementations. In *Proc. Network and Distributed System Security Symposium (NDSS)*, San Diego, US, February 2013.
 - [106] Abdelberi Chaabane, Gergely Acs, and Mohamed Ali Kaafar. You are what you like! Information leakage through users’ interests. In *Proc. Network and Distributed System Security Symposium (NDSS)*, San Diego, US, February 2012.
 - [107] Reza Zafarani and Huan Liu. Connecting users across social media sites: A behavioral-modeling approach. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 41–49, Chicago, US, August 2013.
 - [108] Marco Pennacchiotti and Ana-Maria Popescu. Democrats, republicans and Starbucks aficionados: User classification in Twitter. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Diego, US, August 2011.
 - [109] Oana Goga, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. Exploiting innocuous activity for correlating users across sites. In *Proc. Intl. Conf. on World Wide Web (WWW)*, pages 447–458, Rio de Janeiro, Brazil, May 2013.

- [110] Delfina Malandrino, Andrea Petta, Vittorio Scarano, Luigi Serra, Raffaele Spinelli, and Balachander Krishnamurthy. Privacy awareness about information leakage. In *Proc. ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 279–284, Berlin, Germany, November 2013.
- [111] Dinei Florêncio, Cormac Herley, and Paul C. van Oorschot. Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. In *Proc. USENIX Security*, San Diego, US, August 2014.
- [112] Kai Rannenberg. Identity management in mobile cellular networks and related applications. *Information Security Technical Report*, 9(1):77–85, January 2004.
- [113] Xingyu Xing, Wei Meng, Dan Doozan, Alex C Snoeren, Nick Feamster, and Wenke Lee. Take this personally: Pollution attacks on personalized services. In *Proc. USENIX Security*, pages 671–686, Washington, US, August 2013.
- [114] Ibrahim Abd-elatif Gomaa, Gouda I. Salama, and Ibrahim F. Imam. Biometric OAuth service based on finger-knuckles. *Proc. International Conference on Computer Engineering and Systems (ICCES)*, pages 170–175, 2012.
- [115] Rui Wang, Yuchen Zhou, Shuo Chen, Shaz Qadeer, David Evans, and Yuri Gurevich. Explicating SDKs: Uncovering assumptions underlying secure authentication and authorization. In *Proc. USENIX Security*, pages 399–414, Washington, US, August 2013.
- [116] Michael Backes, Matteo Maffei, and Kim Pecina. A security API for distributed social networks. In *Proc. Network and Distributed System Security Symposium (NDSS)*, pages 35–51, San Diego, US, February 2011.
- [117] Jonathan Anderson, Claudia Diaz, Joseph Bonneau, and Frank Staiano. Privacy-enabling social networking over untrusted networks. In *Proc. ACM Workshop on Online Social Networks (WOSN)*, pages 1–6, Barcelona, Spain, August 2009.
- [118] Ebrahim Tarameshloo, Philip W.L. Fong, and Payman Mohassel. On protection in federated social computing systems. In *Proc. of ACM Conference on Data and Application Security and Privacy (CODASPY)*, pages 75–86, San Antonio, US, March 2014.
- [119] Moo Nam Ko, Gorrell P. Cheek, Mohamed Shehab, and Ravi Sandhu. Social-networks connect services. *IEEE Computer*, 43(8):37–43, August 2010.

-
- [120] Adrienne Felt and David Evans. Privacy protection for social networking APIs. In *Proc. Web 2.0 Security & Privacy Workshop (W2SP)*, Oakland, US, May 2008.
- [121] Phillipa Gill, Augustin Chaintreau, Dina Papagiannaki, and Pablo Rodriguez. Follow the money: Understanding economics of online aggregation and advertising. In *Proc. ACM SIGCOMM Conference on Internet Measurement (IMC)*, pages 141–148, Barcelona, Spain, October 2013.
- [122] Balachander Krishnamurthy and Craig E. Wills. Cat and mouse: Content delivery tradeoffs in Web access. In *Proc. World Wide Web (WWW)*, pages 337–346, Edinburgh, Scotland, May 2006.
- [123] Leeann Bent, Michael Rabinovich, Geoffrey M. Voelker, and Zhen Xiao. Characterization of a large Web site population with implications for content delivery. In *Proc. World Wide Web (WWW)*, pages 505–536, Edinburgh, Scotland, May 2006.
- [124] Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener. A large-scale study of the evolution of Web pages. In *Proc. World Wide Web (WWW)*, pages 669–678, Budapest, Hungary, May 2003.
- [125] Youmna Borghol, Siddharth Mitra, Sebastien Ardon, Niklas Carlsson, Derek Eager, and Anirban Mahanti. Characterizing and modeling popularity of user-generated videos. In *Proc. IFIP International Symposium on Computer Performance, Modeling, Measurements and Evaluation (IFIP Performance)*, pages 1037–1055, Amsterdam, Netherlands, October 2011.
- [126] Jakub Czyz, Mark Allman, Jing Zhang, Scott Iekel-Johnson, Eric Osterweil, and Michael Bailey. Measuring IPv6 adoption. In *Proc. ACM Conference on SIGCOMM*, pages 87–98, Chicago, US, August 2014.
- [127] Nevil Brownlee. On searching for patterns in traceroute responses. In *Proc. Passive and Active Measurement (PAM)*, pages 67–76, Los Angeles, US, March 2014.
- [128] Guenther Starnberger, Lorenz Frohofer, and Karl M. Goeschka. QR-TAN: Secure mobile transaction authentication. In *Proc. Intl. Conf. on Availability, Reliability and Security (ARES)*, pages 578–583, Fukuoka, Japan, March 2009.
- [129] Hassan Metwalley, Stefano Traverso, Marco Mellia, Stanislav Miskovic, and Mario Baldi. The online tracking horde: A view from passive measurements. In *Proc. Workshop on Traffic Monitoring and Analysis (TMA)*, pages 111–125, Barcelona, Spain, April 2015.

Collection of Papers

Papers

The articles associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-127304>

Dissertations

Linköping Studies in Science and Technology

Linköping Studies in Arts and Science

Linköping Studies in Statistics

Linköping Studies in Information Science

Linköping Studies in Science and Technology

- No 14 **Anders Haraldsson:** A Program Manipulation System Based on Partial Evaluation, 1977, ISBN 91-7372-144-1.
- No 17 **Benkt Magnhagen:** Probability Based Verification of Time Margins in Digital Designs, 1977, ISBN 91-7372-157-3.
- No 18 **Mats Cedwall:** Semantisk analys av processbeskrivningar i naturligt språk, 1977, ISBN 91- 7372-168-9.
- No 22 **Jaak Urmi:** A Machine Independent LISP Compiler and its Implications for Ideal Hardware, 1978, ISBN 91-7372-188-3.
- No 33 **Tore Risch:** Compilation of Multiple File Queries in a Meta-Database System 1978, ISBN 91- 7372-232-4.
- No 51 **Erland Jungert:** Synthesizing Database Structures from a User Oriented Data Model, 1980, ISBN 91-7372-387-8.
- No 54 **Sture Hägglund:** Contributions to the Development of Methods and Tools for Interactive Design of Applications Software, 1980, ISBN 91-7372-404-1.
- No 55 **Pär Emanuelson:** Performance Enhancement in a Well-Structured Pattern Matcher through Partial Evaluation, 1980, ISBN 91-7372-403-3.
- No 58 **Benkt Johnsson, Bertil Andersson:** The Human-Computer Interface in Commercial Systems, 1981, ISBN 91-7372-414-9.
- No 69 **H. Jan Komorowski:** A Specification of an Abstract Prolog Machine and its Application to Partial Evaluation, 1981, ISBN 91-7372-479-3.
- No 71 **René Reboh:** Knowledge Engineering Techniques and Tools for Expert Systems, 1981, ISBN 91-7372-489-0.
- No 77 **Östen Oskarsson:** Mechanisms of Modifiability in large Software Systems, 1982, ISBN 91- 7372-527-7.
- No 94 **Hans Lunell:** Code Generator Writing Systems, 1983, ISBN 91-7372-652-4.
- No 97 **Andrzej Lingas:** Advances in Minimum Weight Triangulation, 1983, ISBN 91-7372-660-5.
- No 109 **Peter Fritzson:** Towards a Distributed Programming Environment based on Incremental Compilation, 1984, ISBN 91-7372-801-2.
- No 111 **Erik Tengvald:** The Design of Expert Planning Systems. An Experimental Operations Planning System for Turning, 1984, ISBN 91-7372- 805-5.
- No 155 **Christos Levcopoulos:** Heuristics for Minimum Decompositions of Polygons, 1987, ISBN 91-7870-133-3.
- No 165 **James W. Goodwin:** A Theory and System for Non-Monotonic Reasoning, 1987, ISBN 91-7870-183-X.
- No 170 **Zebo Peng:** A Formal Methodology for Automated Synthesis of VLSI Systems, 1987, ISBN 91-7870-225-9.
- No 174 **Johan Fagerström:** A Paradigm and System for Design of Distributed Systems, 1988, ISBN 91-7870-301-8.
- No 192 **Dimitër Driankov:** Towards a Many Valued Logic of Quantified Belief, 1988, ISBN 91-7870-374-3.
- No 213 **Lin Padgham:** Non-Monotonic Inheritance for an Object Oriented Knowledge Base, 1989, ISBN 91-7870-485-5.
- No 214 **Tony Larsson:** A Formal Hardware Description and Verification Method, 1989, ISBN 91-7870-517-7.
- No 221 **Michael Reinfrank:** Fundamentals and Logical Foundations of Truth Maintenance, 1989, ISBN 91-7870-546-0.
- No 239 **Jonas Löwgren:** Knowledge-Based Design Support and Discourse Management in User Interface Management Systems, 1991, ISBN 91-7870-720-X.
- No 244 **Henrik Eriksson:** Meta-Tool Support for Knowledge Acquisition, 1991, ISBN 91-7870-746-3.
- No 252 **Peter Eklund:** An Epistemic Approach to Interactive Design in Multiple Inheritance Hierarchies, 1991, ISBN 91-7870-784-6.
- No 258 **Patrick Doherty:** NML3 - A Non-Monotonic Formalism with Explicit Defaults, 1991, ISBN 91-7870-816-8.
- No 260 **Nahid Shahmehri:** Generalized Algorithmic Debugging, 1991, ISBN 91-7870-828-1.
- No 264 **Nils Dahlbäck:** Representation of Discourse-Cognitive and Computational Aspects, 1992, ISBN 91-7870-850-8.
- No 265 **Ulf Nilsson:** Abstract Interpretations and Abstract Machines: Contributions to a Methodology for the Implementation of Logic Programs, 1992, ISBN 91-7870-858-3.
- No 270 **Ralph Rönquist:** Theory and Practice of Tense-bound Object References, 1992, ISBN 91-7870-873-7.
- No 273 **Björn Fjellborg:** Pipeline Extraction for VLSI Data Path Synthesis, 1992, ISBN 91-7870-880-X.
- No 276 **Staffan Bonnier:** A Formal Basis for Horn Clause Logic with External Polymorphic Functions, 1992, ISBN 91-7870-896-6.
- No 277 **Kristian Sandahl:** Developing Knowledge Management Systems with an Active Expert Methodology, 1992, ISBN 91-7870-897-4.
- No 281 **Christer Bäckström:** Computational Complexity of Reasoning about Plans, 1992, ISBN 91-7870-979-2.
- No 292 **Mats Wirén:** Studies in Incremental Natural Language Analysis, 1992, ISBN 91-7871-027-8.
- No 297 **Mariam Kamkar:** Interprocedural Dynamic Slicing with Applications to Debugging and Testing, 1993, ISBN 91-7871-065-0.
- No 302 **Tingting Zhang:** A Study in Diagnosis Using Classification and Defaults, 1993, ISBN 91-7871-078-2.
- No 312 **Arne Jönsson:** Dialogue Management for Natural Language Interfaces - An Empirical Approach, 1993, ISBN 91-7871-110-X.
- No 338 **Simin Nadjm-Tehrani:** Reactive Systems in Physical Environments: Compositional Modelling and Framework for Verification, 1994, ISBN 91-7871-237-8.
- No 371 **Bengt Savén:** Business Models for Decision Support and Learning. A Study of Discrete-Event Manufacturing Simulation at Asea/ABB 1968-1993, 1995, ISBN 91-7871-494-X.
- No 375 **Ulf Söderman:** Conceptual Modelling of Mode Switching Physical Systems, 1995, ISBN 91-7871-516-4.
- No 383 **Andreas Kägedal:** Exploiting Groundness in Logic Programs, 1995, ISBN 91-7871-538-5.
- No 396 **George Fodor:** Ontological Control, Description, Identification and Recovery from Problematic Control Situations, 1995, ISBN 91-7871-603-9.
- No 413 **Mikael Pettersson:** Compiling Natural Semantics, 1995, ISBN 91-7871-641-1.
- No 414 **Xinli Gu:** RT Level Testability Improvement by Testability Analysis and Transformations, 1996, ISBN 91-7871-654-3.

- No 416 **Hua Shu:** Distributed Default Reasoning, 1996, ISBN 91-7871-665-9.
- No 429 **Jaime Villegas:** Simulation Supported Industrial Training from an Organisational Learning Perspective - Development and Evaluation of the SSIT Method, 1996, ISBN 91-7871-700-0.
- No 431 **Peter Jonsson:** Studies in Action Planning: Algorithms and Complexity, 1996, ISBN 91-7871-704-3.
- No 437 **Johan Boye:** Directional Types in Logic Programming, 1996, ISBN 91-7871-725-6.
- No 439 **Cecilia Sjöberg:** Activities, Voices and Arenas: Participatory Design in Practice, 1996, ISBN 91-7871-728-0.
- No 448 **Patrick Lambrix:** Part-Whole Reasoning in Description Logics, 1996, ISBN 91-7871-820-1.
- No 452 **Kjell Orsborn:** On Extensible and Object-Relational Database Technology for Finite Element Analysis Applications, 1996, ISBN 91-7871-827-9.
- No 459 **Olof Johansson:** Development Environments for Complex Product Models, 1996, ISBN 91-7871-855-4.
- No 461 **Lena Strömbäck:** User-Defined Constructions in Unification-Based Formalisms, 1997, ISBN 91-7871-857-0.
- No 462 **Lars Degerstedt:** Tabulation-based Logic Programming: A Multi-Level View of Query Answering, 1996, ISBN 91-7871-858-9.
- No 475 **Fredrik Nilsson:** Strategi och ekonomisk styrning - En studie av hur ekonomiska styrsystem utformas och används efter företagsförvärv, 1997, ISBN 91-7871-914-3.
- No 480 **Mikael Lindvall:** An Empirical Study of Requirements-Driven Impact Analysis in Object-Oriented Software Evolution, 1997, ISBN 91-7871-927-5.
- No 485 **Göran Forslund:** Opinion-Based Systems: The Cooperative Perspective on Knowledge-Based Decision Support, 1997, ISBN 91-7871-938-0.
- No 494 **Martin Sköld:** Active Database Management Systems for Monitoring and Control, 1997, ISBN 91-7219-002-7.
- No 495 **Hans Olsén:** Automatic Verification of Petri Nets in a CLP framework, 1997, ISBN 91-7219-011-6.
- No 498 **Thomas Drakengren:** Algorithms and Complexity for Temporal and Spatial Formalisms, 1997, ISBN 91-7219-019-1.
- No 502 **Jakob Axelsson:** Analysis and Synthesis of Heterogeneous Real-Time Systems, 1997, ISBN 91-7219-035-3.
- No 503 **Johan Ringström:** Compiler Generation for Data-Parallel Programming Languages from Two-Level Semantics Specifications, 1997, ISBN 91-7219-045-0.
- No 512 **Anna Moberg:** Närhet och distans - Studier av kommunikationsmönster i satellitkontor och flexibla kontor, 1997, ISBN 91-7219-119-8.
- No 520 **Mikael Ronström:** Design and Modelling of a Parallel Data Server for Telecom Applications, 1998, ISBN 91-7219-169-4.
- No 522 **Niclas Ohlsson:** Towards Effective Fault Prevention - An Empirical Study in Software Engineering, 1998, ISBN 91-7219-176-7.
- No 526 **Joachim Karlsson:** A Systematic Approach for Prioritizing Software Requirements, 1998, ISBN 91-7219-184-8.
- No 530 **Henrik Nilsson:** Declarative Debugging for Lazy Functional Languages, 1998, ISBN 91-7219-197-x.
- No 555 **Jonas Hallberg:** Timing Issues in High-Level Synthesis, 1998, ISBN 91-7219-369-7.
- No 561 **Ling Lin:** Management of 1-D Sequence Data - From Discrete to Continuous, 1999, ISBN 91-7219-402-2.
- No 563 **Eva L Ragnemalm:** Student Modelling based on Collaborative Dialogue with a Learning Companion, 1999, ISBN 91-7219-412-X.
- No 567 **Jörgen Lindström:** Does Distance matter? On geographical dispersion in organisations, 1999, ISBN 91-7219-439-1.
- No 582 **Vanja Josifovski:** Design, Implementation and Evaluation of a Distributed Mediator System for Data Integration, 1999, ISBN 91-7219-482-0.
- No 589 **Rita Kovordányi:** Modeling and Simulating Inhibitory Mechanisms in Mental Image Reinterpretation - Towards Cooperative Human-Computer Creativity, 1999, ISBN 91-7219-506-1.
- No 592 **Mikael Ericsson:** Supporting the Use of Design Knowledge - An Assessment of Commenting Agents, 1999, ISBN 91-7219-532-0.
- No 593 **Lars Karlsson:** Actions, Interactions and Narratives, 1999, ISBN 91-7219-534-7.
- No 594 **C. G. Mikael Johansson:** Social and Organizational Aspects of Requirements Engineering Methods - A practice-oriented approach, 1999, ISBN 91-7219-541-X.
- No 595 **Jörgen Hansson:** Value-Driven Multi-Class Overload Management in Real-Time Database Systems, 1999, ISBN 91-7219-542-8.
- No 596 **Niklas Hallberg:** Incorporating User Values in the Design of Information Systems and Services in the Public Sector: A Methods Approach, 1999, ISBN 91-7219-543-6.
- No 597 **Vivian Vimarlund:** An Economic Perspective on the Analysis of Impacts of Information Technology: From Case Studies in Health-Care towards General Models and Theories, 1999, ISBN 91-7219-544-4.
- No 598 **Johan Jenvall:** Methods and Tools in Computer-Supported Taskforce Training, 1999, ISBN 91-7219-547-9.
- No 607 **Magnus Merkel:** Understanding and enhancing translation by parallel text processing, 1999, ISBN 91-7219-614-9.
- No 611 **Silvia Coradeschi:** Anchoring symbols to sensory data, 1999, ISBN 91-7219-623-8.
- No 613 **Man Lin:** Analysis and Synthesis of Reactive Systems: A Generic Layered Architecture Perspective, 1999, ISBN 91-7219-630-0.
- No 618 **Jimmy Tjäder:** Systemimplementering i praktiken - En studie av logiker i fyra projekt, 1999, ISBN 91-7219-657-2.
- No 627 **Vadim Engelson:** Tools for Design, Interactive Simulation, and Visualization of Object-Oriented Models in Scientific Computing, 2000, ISBN 91-7219-709-9.
- No 637 **Esa Falkenroth:** Database Technology for Control and Simulation, 2000, ISBN 91-7219-766-8.
- No 639 **Per-Arne Persson:** Bringing Power and Knowledge Together: Information Systems Design for Autonomy and Control in Command Work, 2000, ISBN 91-7219-796-X.
- No 660 **Erik Larsson:** An Integrated System-Level Design for Testability Methodology, 2000, ISBN 91-7219-890-7.
- No 688 **Marcus Bjärelund:** Model-based Execution Monitoring, 2001, ISBN 91-7373-016-5.
- No 689 **Joakim Gustafsson:** Extending Temporal Action Logic, 2001, ISBN 91-7373-017-3.
- No 720 **Carl-Johan Petri:** Organizational Information Provision - Managing Mandatory and Discretionary Use of Information Technology, 2001, ISBN 91-7373-126-9.
- No 724 **Paul Scerri:** Designing Agents for Systems with Adjustable Autonomy, 2001, ISBN 91 7373 207 9.
- No 725 **Tim Heyer:** Semantic Inspection of Software Artifacts: From Theory to Practice, 2001, ISBN 91 7373 208 7.
- No 726 **Pär Carlshamre:** A Usability Perspective on Requirements Engineering - From Methodology to Product Development, 2001, ISBN 91 7373 212 5.
- No 732 **Juha Takkinen:** From Information Management to Task Management in Electronic Mail, 2002, ISBN 91 7373 258 3.

- No 745 **Johan Åberg:** Live Help Systems: An Approach to Intelligent Help for Web Information Systems, 2002, ISBN 91-7373-311-3.
- No 746 **Rego Granlund:** Monitoring Distributed Teamwork Training, 2002, ISBN 91-7373-312-1.
- No 757 **Henrik André-Jönsson:** Indexing Strategies for Time Series Data, 2002, ISBN 917373-346-6.
- No 747 **Anneli Hagdahl:** Development of IT-supported Interorganisational Collaboration - A Case Study in the Swedish Public Sector, 2002, ISBN 91-7373-314-8.
- No 749 **Sofie Pilemalm:** Information Technology for Non-Profit Organisations - Extended Participatory Design of an Information System for Trade Union Shop Stewards, 2002, ISBN 91-7373-318-0.
- No 765 **Stefan Holmlid:** Adapting users: Towards a theory of use quality, 2002, ISBN 91-7373-397-0.
- No 771 **Magnus Morin:** Multimedia Representations of Distributed Tactical Operations, 2002, ISBN 91-7373-421-7.
- No 772 **Pawel Pietrzak:** A Type-Based Framework for Locating Errors in Constraint Logic Programs, 2002, ISBN 91-7373-422-5.
- No 758 **Erik Berglund:** Library Communication Among Programmers Worldwide, 2002, ISBN 91-7373-349-0.
- No 774 **Choong-ho Yi:** Modelling Object-Oriented Dynamic Systems Using a Logic-Based Framework, 2002, ISBN 91-7373-424-1.
- No 779 **Mathias Broxvall:** A Study in the Computational Complexity of Temporal Reasoning, 2002, ISBN 91-7373-440-3.
- No 793 **Asmus Pandikow:** A Generic Principle for Enabling Interoperability of Structured and Object-Oriented Analysis and Design Tools, 2002, ISBN 91-7373-479-9.
- No 785 **Lars Hult:** Publika Informationstjänster. En studie av den Internetbaserade encyklopedins bruksegenskaper, 2003, ISBN 91-7373-461-6.
- No 800 **Lars Taxén:** A Framework for the Coordination of Complex Systems' Development, 2003, ISBN 91-7373-604-X.
- No 808 **Klas Gäre:** Tre perspektiv på förväntningar och förändringar i samband med införande av informationssystem, 2003, ISBN 91-7373-618-X.
- No 821 **Mikael Kindborg:** Concurrent Comics - programming of social agents by children, 2003, ISBN 91-7373-651-1.
- No 823 **Christina Ölvingson:** On Development of Information Systems with GIS Functionality in Public Health Informatics: A Requirements Engineering Approach, 2003, ISBN 91-7373-656-2.
- No 828 **Tobias Ritzau:** Memory Efficient Hard Real-Time Garbage Collection, 2003, ISBN 91-7373-666-X.
- No 833 **Paul Pop:** Analysis and Synthesis of Communication-Intensive Heterogeneous Real-Time Systems, 2003, ISBN 91-7373-683-X.
- No 852 **Johan Moe:** Observing the Dynamic Behaviour of Large Distributed Systems to Improve Development and Testing - An Empirical Study in Software Engineering, 2003, ISBN 91-7373-779-8.
- No 867 **Erik Herzog:** An Approach to Systems Engineering Tool Data Representation and Exchange, 2004, ISBN 91-7373-929-4.
- No 872 **Aseel Berglund:** Augmenting the Remote Control: Studies in Complex Information Navigation for Digital TV, 2004, ISBN 91-7373-940-5.
- No 869 **Jo Skåmedal:** Telecommuting's Implications on Travel and Travel Patterns, 2004, ISBN 91-7373-935-9.
- No 870 **Linda Askenäs:** The Roles of IT - Studies of Organising when Implementing and Using Enterprise Systems, 2004, ISBN 91-7373-936-7.
- No 874 **Annika Flycht-Eriksson:** Design and Use of Ontologies in Information-Providing Dialogue Systems, 2004, ISBN 91-7373-947-2.
- No 873 **Peter Bunus:** Debugging Techniques for Equation-Based Languages, 2004, ISBN 91-7373-941-3.
- No 876 **Jonas Mellin:** Resource-Predictable and Efficient Monitoring of Events, 2004, ISBN 91-7373-956-1.
- No 883 **Magnus Bång:** Computing at the Speed of Paper: Ubiquitous Computing Environments for Healthcare Professionals, 2004, ISBN 91-7373-971-5.
- No 882 **Robert Eklund:** Disfluency in Swedish human-human and human-machine travel booking dialogues, 2004, ISBN 91-7373-966-9.
- No 887 **Anders Lindström:** English and other Foreign Linguistic Elements in Spoken Swedish. Studies of Productive Processes and their Modelling using Finite-State Tools, 2004, ISBN 91-7373-981-2.
- No 889 **Zhiping Wang:** Capacity-Constrained Production-inventory systems - Modelling and Analysis in both a traditional and an e-business context, 2004, ISBN 91-85295-08-6.
- No 893 **Pernilla Qvarfordt:** Eyes on Multimodal Interaction, 2004, ISBN 91-85295-30-2.
- No 910 **Magnus Kald:** In the Borderland between Strategy and Management Control - Theoretical Framework and Empirical Evidence, 2004, ISBN 91-85295-82-5.
- No 918 **Jonas Lundberg:** Shaping Electronic News: Genre Perspectives on Interaction Design, 2004, ISBN 91-85297-14-3.
- No 900 **Mattias Arvola:** Shades of use: The dynamics of interaction design for sociable use, 2004, ISBN 91-85295-42-6.
- No 920 **Luis Alejandro Cortés:** Verification and Scheduling Techniques for Real-Time Embedded Systems, 2004, ISBN 91-85297-21-6.
- No 929 **Diana Szentivanyi:** Performance Studies of Fault-Tolerant Middleware, 2005, ISBN 91-85297-58-5.
- No 933 **Mikael Cäker:** Management Accounting as Constructing and Opposing Customer Focus: Three Case Studies on Management Accounting and Customer Relations, 2005, ISBN 91-85297-64-X.
- No 937 **Jonas Kvarnström:** TALplanner and Other Extensions to Temporal Action Logic, 2005, ISBN 91-85297-75-5.
- No 938 **Bourhane Kadmiry:** Fuzzy Gain-Scheduled Visual Servoing for Unmanned Helicopter, 2005, ISBN 91-85297-76-3.
- No 945 **Gert Jervan:** Hybrid Built-In Self-Test and Test Generation Techniques for Digital Systems, 2005, ISBN: 91-85297-97-6.
- No 946 **Anders Arpteg:** Intelligent Semi-Structured Information Extraction, 2005, ISBN 91-85297-98-4.
- No 947 **Ola Angelsmark:** Constructing Algorithms for Constraint Satisfaction and Related Problems - Methods and Applications, 2005, ISBN 91-85297-99-2.
- No 963 **Calin Curescu:** Utility-based Optimisation of Resource Allocation for Wireless Networks, 2005, ISBN 91-85457-07-8.
- No 972 **Björn Johansson:** Joint Control in Dynamic Situations, 2005, ISBN 91-85457-31-0.
- No 974 **Dan Lawesson:** An Approach to Diagnosability Analysis for Interacting Finite State Systems, 2005, ISBN 91-85457-39-6.
- No 979 **Claudiu Duma:** Security and Trust Mechanisms for Groups in Distributed Services, 2005, ISBN 91-85457-54-X.
- No 983 **Sorin Manolache:** Analysis and Optimisation of Real-Time Systems with Stochastic Behaviour, 2005, ISBN 91-85457-60-4.
- No 986 **Yuxiao Zhao:** Standards-Based Application Integration for Business-to-Business Communications, 2005, ISBN 91-85457-66-3.
- No 1004 **Patrik Haslum:** Admissible Heuristics for Automated Planning, 2006, ISBN 91-85497-28-2.

- No 1005 **Aleksandra Tešanovic:** Developing Reusable and Reconfigurable Real-Time Software using Aspects and Components, 2006, ISBN 91-85497-29-0.
- No 1008 **David Dinka:** Role, Identity and Work: Extending the design and development agenda, 2006, ISBN 91-85497-42-8.
- No 1009 **Iakov Nakhimovski:** Contributions to the Modeling and Simulation of Mechanical Systems with Detailed Contact Analysis, 2006, ISBN 91-85497-43-X.
- No 1013 **Wilhelm Dahllöf:** Exact Algorithms for Exact Satisfiability Problems, 2006, ISBN 91-85523-97-6.
- No 1016 **Levon Saldamli:** PDEModelica - A High-Level Language for Modeling with Partial Differential Equations, 2006, ISBN 91-85523-84-4.
- No 1017 **Daniel Karlsson:** Verification of Component-based Embedded System Designs, 2006, ISBN 91-85523-79-8.
- No 1018 **Ioan Chisalita:** Communication and Networking Techniques for Traffic Safety Systems, 2006, ISBN 91-85523-77-1.
- No 1019 **Tarja Susi:** The Puzzle of Social Activity - The Significance of Tools in Cognition and Cooperation, 2006, ISBN 91-85523-71-2.
- No 1021 **Andrzej Bednarski:** Integrated Optimal Code Generation for Digital Signal Processors, 2006, ISBN 91-85523-69-0.
- No 1022 **Peter Aronsson:** Automatic Parallelization of Equation-Based Simulation Programs, 2006, ISBN 91-85523-68-2.
- No 1030 **Robert Nilsson:** A Mutation-based Framework for Automated Testing of Timeliness, 2006, ISBN 91-85523-35-6.
- No 1034 **Jon Edvardsson:** Techniques for Automatic Generation of Tests from Programs and Specifications, 2006, ISBN 91-85523-31-3.
- No 1035 **Vaida Jakoniene:** Integration of Biological Data, 2006, ISBN 91-85523-28-3.
- No 1045 **Genevieve Gorrell:** Generalized Hebbian Algorithms for Dimensionality Reduction in Natural Language Processing, 2006, ISBN 91-85643-88-2.
- No 1051 **Yu-Hsing Huang:** Having a New Pair of Glasses - Applying Systemic Accident Models on Road Safety, 2006, ISBN 91-85643-64-5.
- No 1054 **Åsa Hedenskog:** Perceive those things which cannot be seen - A Cognitive Systems Engineering perspective on requirements management, 2006, ISBN 91-85643-57-2.
- No 1061 **Cécile Åberg:** An Evaluation Platform for Semantic Web Technology, 2007, ISBN 91-85643-31-9.
- No 1073 **Mats Grindal:** Handling Combinatorial Explosion in Software Testing, 2007, ISBN 978-91-85715-74-9.
- No 1075 **Almut Herzog:** Usable Security Policies for Runtime Environments, 2007, ISBN 978-91-85715-65-7.
- No 1079 **Magnus Wahlström:** Algorithms, measures, and upper bounds for Satisfiability and related problems, 2007, ISBN 978-91-85715-55-8.
- No 1083 **Jesper Andersson:** Dynamic Software Architectures, 2007, ISBN 978-91-85715-46-6.
- No 1086 **Ulf Johansson:** Obtaining Accurate and Comprehensive Data Mining Models - An Evolutionary Approach, 2007, ISBN 978-91-85715-34-3.
- No 1089 **Traian Pop:** Analysis and Optimisation of Distributed Embedded Systems with Heterogeneous Scheduling Policies, 2007, ISBN 978-91-85715-27-5.
- No 1091 **Gustav Nordh:** Complexity Dichotomies for CSP-related Problems, 2007, ISBN 978-91-85715-20-6.
- No 1106 **Per Ola Kristensson:** Discrete and Continuous Shape Writing for Text Entry and Control, 2007, ISBN 978-91-85831-77-7.
- No 1110 **He Tan:** Aligning Biomedical Ontologies, 2007, ISBN 978-91-85831-56-2.
- No 1112 **Jessica Lindblom:** Minding the body - Interacting socially through embodied action, 2007, ISBN 978-91-85831-48-7.
- No 1113 **Pontus Wärnestål:** Dialogue Behavior Management in Conversational Recommender Systems, 2007, ISBN 978-91-85831-47-0.
- No 1120 **Thomas Gustafsson:** Management of Real-Time Data Consistency and Transient Overloads in Embedded Systems, 2007, ISBN 978-91-85831-33-3.
- No 1127 **Alexandru Andrei:** Energy Efficient and Predictable Design of Real-time Embedded Systems, 2007, ISBN 978-91-85831-06-7.
- No 1139 **Per Wikberg:** Eliciting Knowledge from Experts in Modeling of Complex Systems: Managing Variation and Interactions, 2007, ISBN 978-91-85895-66-3.
- No 1143 **Mehdi Amirijoo:** QoS Control of Real-Time Data Services under Uncertain Workload, 2007, ISBN 978-91-85895-49-6.
- No 1150 **Sanny Syberfeldt:** Optimistic Replication with Forward Conflict Resolution in Distributed Real-Time Databases, 2007, ISBN 978-91-85895-27-4.
- No 1155 **Beatrice Alenljung:** Envisioning a Future Decision Support System for Requirements Engineering - A Holistic and Human-centred Perspective, 2008, ISBN 978-91-85895-11-3.
- No 1156 **Artur Wilk:** Types for XML with Application to Xcerpt, 2008, ISBN 978-91-85895-08-3.
- No 1183 **Adrian Pop:** Integrated Model-Driven Development Environments for Equation-Based Object-Oriented Languages, 2008, ISBN 978-91-7393-895-2.
- No 1185 **Jörgen Skågeby:** Gifting Technologies - Ethnographic Studies of End-users and Social Media Sharing, 2008, ISBN 978-91-7393-892-1.
- No 1187 **Imad-Eldin Ali Abugessaisa:** Analytical tools and information-sharing methods supporting road safety organizations, 2008, ISBN 978-91-7393-887-7.
- No 1204 **H. Joe Steinhauer:** A Representation Scheme for Description and Reconstruction of Object Configurations Based on Qualitative Relations, 2008, ISBN 978-91-7393-823-5.
- No 1222 **Anders Larsson:** Test Optimization for Core-based System-on-Chip, 2008, ISBN 978-91-7393-768-9.
- No 1238 **Andreas Borg:** Processes and Models for Capacity Requirements in Telecommunication Systems, 2009, ISBN 978-91-7393-700-9.
- No 1240 **Fredrik Heintz:** DyKnow: A Stream-Based Knowledge Processing Middleware Framework, 2009, ISBN 978-91-7393-696-5.
- No 1241 **Birgitta Lindström:** Testability of Dynamic Real-Time Systems, 2009, ISBN 978-91-7393-695-8.
- No 1244 **Eva Blomqvist:** Semi-automatic Ontology Construction based on Patterns, 2009, ISBN 978-91-7393-683-5.
- No 1249 **Rogier Woltjer:** Functional Modeling of Constraint Management in Aviation Safety and Command and Control, 2009, ISBN 978-91-7393-659-0.
- No 1260 **Gianpaolo Conte:** Vision-Based Localization and Guidance for Unmanned Aerial Vehicles, 2009, ISBN 978-91-7393-603-3.
- No 1262 **AnnMarie Ericsson:** Enabling Tool Support for Formal Analysis of ECA Rules, 2009, ISBN 978-91-7393-598-2.
- No 1266 **Jiri Trnka:** Exploring Tactical Command and Control: A Role-Playing Simulation Approach, 2009, ISBN 978-91-7393-571-5.
- No 1268 **Bahlol Rahimi:** Supporting Collaborative Work through ICT - How End-users Think of and Adopt Integrated Health Information Systems, 2009, ISBN 978-91-7393-550-0.
- No 1274 **Fredrik Kuivinen:** Algorithms and Hardness Results for Some Valued CSPs, 2009, ISBN 978-91-7393-525-8.
- No 1281 **Gunnar Mathiason:** Virtual Full Replication for Scalable Distributed Real-Time Databases, 2009, ISBN 978-91-7393-503-6.

- No 1290 **Viacheslav Izosimov:** Scheduling and Optimization of Fault-Tolerant Distributed Embedded Systems, 2009, ISBN 978-91-7393-482-4.
- No 1294 **Johan Thapper:** Aspects of a Constraint Optimisation Problem, 2010, ISBN 978-91-7393-464-0.
- No 1306 **Susanna Nilsson:** Augmentation in the Wild: User Centered Development and Evaluation of Augmented Reality Applications, 2010, ISBN 978-91-7393-416-9.
- No 1313 **Christer Thörn:** On the Quality of Feature Models, 2010, ISBN 978-91-7393-394-0.
- No 1321 **Zhiyuan He:** Temperature Aware and Defect-Probability Driven Test Scheduling for System-on-Chip, 2010, ISBN 978-91-7393-378-0.
- No 1333 **David Broman:** Meta-Languages and Semantics for Equation-Based Modeling and Simulation, 2010, ISBN 978-91-7393-335-3.
- No 1337 **Alexander Siemers:** Contributions to Modelling and Visualisation of Multibody Systems Simulations with Detailed Contact Analysis, 2010, ISBN 978-91-7393-317-9.
- No 1354 **Mikael Asplund:** Disconnected Discoveries: Availability Studies in Partitioned Networks, 2010, ISBN 978-91-7393-278-3.
- No 1359 **Jana Rambusch:** Mind Games Extended: Understanding Gameplay as Situated Activity, 2010, ISBN 978-91-7393-252-3.
- No 1373 **Sonia Sangari:** Head Movement Correlates to Focus Assignment in Swedish, 2011, ISBN 978-91-7393-154-0.
- No 1374 **Jan-Erik Källhammer:** Using False Alarms when Developing Automotive Active Safety Systems, 2011, ISBN 978-91-7393-153-3.
- No 1375 **Mattias Eriksson:** Integrated Code Generation, 2011, ISBN 978-91-7393-147-2.
- No 1381 **Ola Leifler:** Affordances and Constraints of Intelligent Decision Support for Military Command and Control – Three Case Studies of Support Systems, 2011, ISBN 978-91-7393-133-5.
- No 1386 **Soheil Samii:** Quality-Driven Synthesis and Optimization of Embedded Control Systems, 2011, ISBN 978-91-7393-102-1.
- No 1419 **Erik Kuiper:** Geographic Routing in Intermittently-connected Mobile Ad Hoc Networks: Algorithms and Performance Models, 2012, ISBN 978-91-7519-981-8.
- No 1451 **Sara Stymne:** Text Harmonization Strategies for Phrase-Based Statistical Machine Translation, 2012, ISBN 978-91-7519-887-3.
- No 1455 **Alberto Montebelli:** Modeling the Role of Energy Management in Embodied Cognition, 2012, ISBN 978-91-7519-882-8.
- No 1465 **Mohammad Saifullah:** Biologically-Based Interactive Neural Network Models for Visual Attention and Object Recognition, 2012, ISBN 978-91-7519-838-5.
- No 1490 **Tomas Bengtsson:** Testing and Logic Optimization Techniques for Systems on Chip, 2012, ISBN 978-91-7519-742-5.
- No 1481 **David Byers:** Improving Software Security by Preventing Known Vulnerabilities, 2012, ISBN 978-91-7519-784-5.
- No 1496 **Tommy Färnqvist:** Exploiting Structure in CSP-related Problems, 2013, ISBN 978-91-7519-711-1.
- No 1503 **John Wilander:** Contributions to Specification, Implementation, and Execution of Secure Software, 2013, ISBN 978-91-7519-681-7.
- No 1506 **Magnus Ingmarsson:** Creating and Enabling the Useful Service Discovery Experience, 2013, ISBN 978-91-7519-662-6.
- No 1547 **Wladimir Schamai:** Model-Based Verification of Dynamic System Behavior against Requirements: Method, Language, and Tool, 2013, ISBN 978-91-7519-505-6.
- No 1551 **Henrik Svensson:** Simulations, 2013, ISBN 978-91-7519-491-2.
- No 1559 **Sergiu Rafiliu:** Stability of Adaptive Distributed Real-Time Systems with Dynamic Resource Management, 2013, ISBN 978-91-7519-471-4.
- No 1581 **Usman Dastgeer:** Performance-aware Component Composition for GPU-based Systems, 2014, ISBN 978-91-7519-383-0.
- No 1602 **Cai Li:** Reinforcement Learning of Locomotion based on Central Pattern Generators, 2014, ISBN 978-91-7519-313-7.
- No 1652 **Roland Samlaus:** An Integrated Development Environment with Enhanced Domain-Specific Interactive Model Validation, 2015, ISBN 978-91-7519-090-7.
- No 1663 **Hannes Uppman:** On Some Combinatorial Optimization Problems: Algorithms and Complexity, 2015, ISBN 978-91-7519-072-3.
- No 1664 **Martin Sjölund:** Tools and Methods for Analysis, Debugging, and Performance Improvement of Equation-Based Models, 2015, ISBN 978-91-7519-071-6.
- No 1666 **Kristian Stavåker:** Contributions to Simulation of Modelica Models on Data-Parallel Multi-Core Architectures, 2015, ISBN 978-91-7519-068-6.
- No 1680 **Adrian Lifa:** Hardware/Software Codesign of Embedded Systems with Reconfigurable and Heterogeneous Platforms, 2015, ISBN 978-91-7519-040-2.
- No 1685 **Bogdan Tanasa:** Timing Analysis of Distributed Embedded Systems with Stochastic Workload and Reliability Constraints, 2015, ISBN 978-91-7519-022-8.
- No 1691 **Håkan Warnquist:** Troubleshooting Trucks – Automated Planning and Diagnosis, 2015, ISBN 978-91-7685-993-3.
- No 1702 **Nima Aghaee:** Thermal Issues in Testing of Advanced Systems on Chip, 2015, ISBN 978-91-7685-949-0.
- No 1715 **Maria Vasilievskaya:** Security in Embedded Systems: A Model-Based Approach with Risk Metrics, 2015, ISBN 978-91-7685-917-9.
- No 1729 **Ke Jiang:** Security-Driven Design of Real-Time Embedded System, 2016, ISBN 978-91-7685-884-4.
- No 1733 **Victor Lagerkvist:** Strong Partial Clones and the Complexity of Constraint Satisfaction Problems: Limitations and Applications, 2016, ISBN 978-91-7685-856-1.
- No 1734 **Chandan Roy:** An Informed System Development Approach to Tropical Cyclone Track and Intensity Forecasting, 2016, ISBN 978-91-7685-854-7.
- No 1746 **Amir Aminifar:** Analysis, Design, and Optimization of Embedded Control Systems, 2016, ISBN 978-91-7685-826-4.
- No 1747 **Ekhiot Vergara:** Energy Modelling and Fairness for Efficient Mobile Communication, 2016, ISBN 978-91-7685-822-6.
- No 1748 **Dag Sonntag:** Chain Graphs – Interpretations, Expressiveness and Learning Algorithms, 2016, ISBN 978-91-7685-818-9.
- No 1768 **Anna Vapen:** Web Authentication using Third-Parties in Untrusted Environments, 2016, ISBN 978-91-7685-753-3.

Linköping Studies in Arts and Science

- No 504 **Ing-Marie Jonsson:** Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour, 2009, ISBN 978-91-7393-478-7.
- No 586 **Fabian Segelström:** Stakeholder Engagement for Service Design: How service designers identify and communicate insights, 2013, ISBN 978-91-7519-554-4.
- No 618 **Johan Blomkvist:** Representing Future Situations of Service: Prototyping in Service Design, 2014, ISBN 978-91-7519-343-4.
- No 620 **Marcus Mast:** Human-Robot Interaction for Semi-Autonomous Assistive Robots, 2014, ISBN 978-91-7519-319-9.
- No 677 **Peter Berggren:** Assessing Shared Strategic Understanding, 2016, ISBN 978-91-7685-786-1.

Linköping Studies in Statistics

- No 9 **Davood Shahsavani:** Computer Experiments Designed to Explore and Approximate Complex Deterministic Models, 2008, ISBN 978-91-7393-976-8.
- No 10 **Karl Wahlin:** Roadmap for Trend Detection and Assessment of Data Quality, 2008, ISBN 978-91-7393-792-4.
- No 11 **Oleg Sysoev:** Monotonic regression for large multivariate datasets, 2010, ISBN 978-91-7393-412-1.
- No 13 **Agné Burauskaite-Harju:** Characterizing Temporal Change and Inter-Site Correlations in Daily and Sub-daily Precipitation Extremes, 2011, ISBN 978-91-7393-110-6.

Linköping Studies in Information Science

- No 1 **Karin Axelsson:** Metodisk systemstrukturering- att skapa samstämmighet mellan informationssystem-arkitektur och verksamhet, 1998, ISBN-9172-19-296-8.
- No 2 **Stefan Cronholm:** Metodverktyg och användbarhet - en studie av datorstödd metodbaserad systemutveckling, 1998, ISBN-9172-19-299-2.
- No 3 **Anders Avdic:** Användare och utvecklare - om anveckling med kalkylprogram, 1999, ISBN-91-7219-606-8.
- No 4 **Owen Eriksson:** Kommunikationskvalitet hos informationssystem och affärsprocesser, 2000, ISBN 91-7219-811-7.
- No 5 **Mikael Lind:** Från system till process - kriterier för processbestämning vid verksamhetsanalys, 2001, ISBN 91-7373-067-X.
- No 6 **Ulf Melin:** Koordination och informationssystem i företag och nätverk, 2002, ISBN 91-7373-278-8.
- No 7 **Pär J. Ågerfalk:** Information Systems Actability - Understanding Information Technology as a Tool for Business Action and Communication, 2003, ISBN 91-7373-628-7.
- No 8 **Ulf Seigerroth:** Att förstå och förändra systemutvecklingsverksamheter - en taxonomi för metautveckling, 2003, ISBN91-7373-736-4.
- No 9 **Karin Hedström:** Spår av datoriseringens värden – Effekter av IT i äldreomsorg, 2004, ISBN 91-7373-963-4.
- No 10 **Ewa Braf:** Knowledge Demanded for Action - Studies on Knowledge Mediation in Organisations, 2004, ISBN 91-85295-47-7.
- No 11 **Fredrik Karlsson:** Method Configuration method and computerized tool support, 2005, ISBN 91-85297-48-8.
- No 12 **Malin Nordström:** Styrbar systemförvaltning - Att organisera systemförvaltningsverksamhet med hjälp av effektiva förvaltningsobjekt, 2005, ISBN 91-85297-60-7.
- No 13 **Stefan Holgersson:** Yrke: POLIS - Yrkeskunskap, motivation, IT-system och andra förutsättningar för polisarbete, 2005, ISBN 91-85299-43-X.

No 14

Benneth Christiansson, Marie-Therese Christiansson:

Mötet mellan process och komponent - mot ett ramverk för en verksamhetsnära kravspecifikation vid anskaffning av komponent-baserade informationssystem, 2006, ISBN 91-85643-22-X.