Diss. ETH No. 20702

User-Centered Security Mechanisms for Protecting Information Sharing in the Cloud

A dissertation submitted to
ETH Zurich

for the degree of
Doctor of Sciences

presented by
Iulia Ion
M.Sc. in Computer Science, International University in Germany
born October 25, 1983
citizen of Romania

accepted on the recommendation of
Prof. Dr. Friedemann Mattern, examiner, ETH Zurich
Prof. Dr. Marc Langheinrich, co-examiner, University of Lugano
Prof. Dr. Srdjan Čapkun, co-examiner, ETH Zurich
Prof. Dr. Lujo Bauer, co-examiner, Carnegie Mellon University

2012

Abstract

End-users have become accustomed to the ease with which cloud-based systems allow them to exchange messages, pictures, and other files with colleagues, friends, and family. This convenience, however, typically comes at the expense of disclosing this (often highly personal) information to the service provider in the process. Furthermore, users have little control over which third-parties—e.g., storage providers, unauthorized friends, hackers, advertisement companies, and governmental agencies—access their data.

Several studies have identified security and privacy as the biggest concerns for companies when adopting cloud-based solutions, but not much is known about end-users' attitudes and practices. Given the high amount of personal information that users often disclose on such platforms, detractors claim that users care little or not at all about their privacy. To disprove such beliefs, we conducted an extensive cross-cultural study. Our results show that consumers have strong privacy concerns, trust local storage more than the cloud when storing sensitive data, and are only partially aware of the risks they face in the cloud.

Based on this initial study, we identify the need for novel, user-centered security mechanisms to help non-technical users protect the information they share in the cloud. A number of systems have been proposed to limit the service providers' access to this information, yet these systems typically require trusted servers, are platform specific (e.g., work for Facebook only), or fail to hide that confidential communication is taking place. In this thesis, we present a novel system that enables users to share data over any web-based cloud storage platform, while both protecting the confidentiality of the communicated data and hiding the fact that the exchanged data is confidential. We provide a proof-of-concept implementation of our system in the form of a publicly available Firefox plugin, and demonstrate the viability of our approach through a performance evaluation.

To bootstrap secure communications in systems like the one we propose, current solutions leave it as an exercise for the user to manually verify key material (e.g., public key fingerprints) through offline channels with potentially hundreds of online contacts. Instead, in our system, we take advantage of users' encounters and we verify keys automatically through a secure, direct connection between users' mobile devices. The usability of the device pairing protocol used to establish the secure connection is crucial, as overly complex mechanisms might prompt users to choose a lower security level, or lead them to abandon security altogether. To this end, we conducted a comparative usability study of existing device pairing methods. Unlike previous work, our study places pairing tasks in specific real-life situations. Our results disprove the commonly held belief that users always choose the easiest method. Instead, users prefer different methods in different situations, depending on their time constraints, relationship to the interacting partner, social conventions appropriate for the place, and perceived security needs and guarantees.

Kurzfassung

Benutzer haben sich an die Leichtigkeit gewöhnt, mit der Cloud-basierte Systeme es ermöglichen, Nachrichten, Bilder und andere Dateien mit Kollegen, Freunden oder der Familie auszutauschen. Allerdings geschieht diese Bequemlichkeit typischerweise auf Kosten der Bekanntgabe dieser (oft sehr persönlichen) Informationen gegenüber dem Dienstanbieter. Ausserdem haben Benutzer wenig Kontrolle darüber, welche Drittparteien—z. B. Speicheranbieter, unautorisierte Freunde, Hacker, Werbefirmen oder Regierungsbehörden—auf ihre Daten zugreifen.

Verschiedene Studien haben Sicherheit und Datenschutz bei der Einführung Cloud-basierter Lösungen als die größten Vorbehalte von Unternehmen eingeschätzt, jedoch ist wenig über das Verhalten und die Praktiken der Endbenutzer bekannt. Angesichts der grossen Menge an persönlichen Informationen, die Benutzer unbedarft auf solchen Plattformen ablegen, behaupten Kritiker, dass Benutzer sich nur wenig oder überhaupt nicht um ihre Privatsphäre kümmern. Um solche Annahmen zu prüfen, haben wir eine umfassende interkulturelle Studie durchgeführt. Unsere Ergebnisse zeigen, dass die Benutzer erhebliche Sorgen hinsichtlich ihrer Privatsphäre haben, sie der lokalen Speicherung mehr als der Cloud vertrauen um sensible Daten zu speichern, und sie nur teilweise die Risiken kennen, denen sie in der Cloud ausgesetzt sind.

Basierend auf dieser Studie identifizieren wir den Bedarf an neuen, benutzerfokussierten Sicherheitsmechanismen, um nichttechnische Anwender beim Schutz der Informationen, die sie in der Cloud teilen, zu unterstützen. Eine Reihe von Systemen ist in der Literatur vorgeschlagen worden, um den Zugang der Dienstanbieter zu diesen Informationen zu begrenzen, jedoch benötigen diese Systeme typischerweise vertrauenswürdige Server, sind plattform-spezifisch (z. B. funktionieren nur für Facebook) oder verbergen die vertrauliche Art der Kommunikation nicht. In dieser Arbeit stellen wir ein neues System vor, das es den Benutzern ermöglicht, Dateien über beliebige Web-basierte

Cloud-Speicherplattformen zu teilen und gleichzeitig die Vertraulichkeit der übermittelten Daten zu schützen und die vertrauliche Art der Kommunikation zu verschleiern. Wir liefern eine Proof-of-Concept-Implementierung unseres Systems in Form eines öffentlich zugänglichen Firefox-Plugins und demonstrieren die Machbarkeit unserer Lösung durch eine Leistungsevaluation.

Um sichere Kommunikation einzuleiten, verlangen aktuelle Lösungen von den Benutzern, Schlüsselmaterial (z. B. Public-Key-Fingerprints) über Offline-Kanäle für möglicherweise Hunderte von Online-Kontakten manuell zu verifizieren. Im Unterschied dazu nutzen wir in unserem System das physische Zusammentreffen von Nutzern, um Schlüssel automatisch über eine sichere, direkte Verbindung zwischen den Mobilgeräten der Nutzer zu überprüfen. Die Gebrauchstauglichkeit der verwendeten Methode zur Gewährleistung einer sicheren Verbindung ist von entscheidender Wichtigkeit, da komplexe Mechanismen Benutzer veranlassen können, eine niedrigere Sicherheitsstufe zu wählen oder die Sicherheit vollkommen aufzugeben. Zu diesem Zweck haben wir eine vergleichende Benutzbarkeitsstudie existierender Gerätepaarungs-Protokolle durchgeführt. Im Gegensatz zu früheren Studien platziert unsere Studie die Aufgaben in spezifische realitätsbezogene Situationen. Unsere Ergebnisse widerlegen die allgemein verbreitete Meinung, dass die Benutzer immer die einfachste Methode wählen. Stattdessen bevorzugen Anwender in verschiedenen Situationen unterschiedliche Methoden, abhängig von ihren zeitlichen Einschränkungen, der Beziehung zum interagierenden Partner, angemessenen gesellschaftlichen Konventionen bezogen auf den jeweiligen Ort sowie empfundenen Sicherheitsanforderungen und erwarteten Sicherheitsgarantien.

Acknowledgements

I was extremely fortunate to be given complete freedom to discover and pursue my research interests during my PhD. For this unique opportunity, I will be forever indebted to my supervisor, Prof. Friedemann Mattern. Throughout my entire journey, Prof. Mattern had unshattered confidence in my abilities to succeed and encouraged me to pursue my own ideas despite obstacles ahead, while inspiring me through his extraordinary vision. Nevertheless, Prof. Mattern raised the bar high and constantly challenged me to think of big problems by asking tough questions. I will always remember and admire his integrity and generosity, and remain thankful for this unique experience. This journey made me grow professionally and personally in ways that will impact my life beyond this PhD.

Further, I would like to thank my committee members for their guidance. I am grateful to Prof. Marc Langheinrich for the support and advice he has given me throughout my PhD. Prof. Langheinrich was my closest mentor and advisor, always open to collaborating and providing guidance. I relied on Prof. Srdjan Čapkun to tell me when I was "thinking in too small of a box," validate research ideas and give me the confidence to pursue impactful projects. My visits to his office fueled me with clarity and enthusiasm. Finally, I was fortunate to have Prof. Lujo Bauer in my thesis committee and benefit from his feedback and expertise. I enjoyed working with Prof. Bauer and learning from him during my internship at Carnegie Mellon University (CMU).

Profound gratitude goes to Prof. Ponnurangam Kumaraguru, alias PK. Probably no other person influenced the direction of my PhD as much as him. PK took a chance on me back during my internship at CMU, and taught me how to design proper usability studies. My work as a PhD student was much more enjoyable thanks to him. Furthermore, by checking up on my progress regularly, PK eased the burden of self-motivation to carry large projects through.

I am also grateful to Rene Mayrhofer for initiating me in device pairing protocols, which bootstraped my research work.

My internship at CMU with Prof. Lorrie Cranor in the CyLab Usable Privacy and Security Lab was a turning point in my PhD. I want to express my deepest gratitude to Prof. Cranor and all her team for the excellent collaborations and for repeatedly giving me very valuable advice. My extreme gratitude also goes to Diana Smetters, my host during my internship at Google, for giving me the clarity to shape the projects needed to finish this PhD. Special thanks go to Jerry Hoff for great inspiration and sparking an idea that became central to my thesis.

I would also like to thank all the collaborators and students with whom I worked throughout this time. In particular, I want to heartily acknowledge Niharika Sachdeva for her dedication and all the sleepless nights she spent online working with me. The privacy study would not have been the same without her hard work.

My deepest thanks go to my colleagues at ETH Zurich from the Distributed Systems Group, to my friends in Zurich, and to the countless people to whom I turned for advice and support. Thanks go to Dominique Guinard for being the best office mate and friend one could wish for; to Silvia Santini who looked out for me from day one like a protective sister; and to Christof Roduner who gave me the uncensored story of what a PhD is like. I am grateful also to Christina Pöpper, Patrick Schaller, Mohammad Torabi Dashti, Ghassan Karame, and Viktor Galliard for interesting security discussions, advice and support.

Special thanks go to my friend Calicrates Policroniades. Cali prepared me better than anybody could have prepared a newbie for this journey; yet to my shame I fell in the same traps as anybody else.

Few PhD students are lucky enough to have a sibling who understands the challenges of doing a PhD, let alone one who is going through the same experience at the same time. My sister, Mihaela, has been an enormous support during my downs and provided great inspiration through her ups. My boyfriend, Harshpreet, gave me unique encouragements and helped me put things in perspective. Finally, my utmost gratitude goes to my parents, for teaching me the importance of good education. A lifetime won't be enough to repay them for their love, support, and sacrifices.

Contents

1. Introduction

"If the user can't use it, it doesn't work." – Susan Dray

1.1. Motivation

1.1.1. User-Centered Security Design

In its very early days, the Internet was only used by technically savvy people. In contrast, nowadays, few users have formal computer training and education. The Internet World Stats estimates that in 2012 one in three people in the world (that is 2.3 billion) are using the Internet [81]. Back in 2000, the Internet counted only 360 million users—not even half number of active Facebook users today [80]. Complex tasks, such as managing security settings, which were traditionally carried out by trained computer administrators, must now be conducted by regular users who have little or no computer science education.

Furthermore, users own an ever increasing number of devices, most of which never see a professional administrator and must be self-managed. No longer restricted to desktop computers and laptops, users enjoy continuous Internet connectivity through their mobile devices, such as smart-phones and tablet computers. The Cisco Global Mobile Data Traffic Forecast predicts that the number of mobile-connected devices will exceed the number of people on earth by the end of 2012 [31]. Additionally, mobile devices today have become far more powerful and complex than the early computers.

A recent incident revealed the extent to which some Internet users lack understanding of even basic concepts, such as URL-based website navigation. On February 10th 2010, a ReadWriteWeb blog discussing the Facebook login [115] became popular and ranked high in the Google Search results. In the next days, thousands of users wanting to access their Facebook account—and who normally rely on Google Search

to navigate to Facebook—ended up on the ReadWriteWeb blog and mistook the page for a Facebook redesign [116]. Perhaps confused by the Facebook connect button, by the end of the first day hundreds of users had typed their Facebook credentials in the comment field box of the blog. While accompanying comments like *"this new Facebook is terrible. I can't find the login! It used to have all my friends"* might be taken with a hint of a smile by computer savvies, they should instead serve to raise serious warning flags for system designers about the knowledge gap between expert and novice or average users. Consider for a moment that the blog had instead been a malicious website that cloned the Facebook page to steal user credentials.

Designing secure systems that are usable by all Internet users today is challenging, even more so because security is a secondary task and almost never the users' final goal [175]. Security systems are often cumbersome. As a result, users end up bypassing security mechanisms and sacrificing security for usability. For example, Adams and Sasse found that, due to low motivation and poor understanding of the threats, users circumvent password security policies [3]. As Edward Felten, director of the Princeton's Center for Information Technology Policy, said, *"Given a choice between dancing pigs and security, users will pick dancing pigs every time"* [78].

In a digital world in which security threats abound [38], security advice users receive is becoming crushingly complex. Consequently, users have no chance of keeping up. To give just one example, the US-CERT best practices guidelines currently contain 56 tips spanning diverse topics including firewalls, wireless networks and portable devices, each of which is at least one page long [160]. Security experts should keep in mind that users' time is expensive: Herley estimates an hour of time for all the users in the United States of America to be worth $2.6 billion [72]. To devise effective security mechanisms, this thesis follows *user-centered security* approaches, a concept introduced by Zurko and Richards in 1996 to refer to systems that have usability as their primary goal [175].

1.1.2. The Rise of Consumer Cloud Storage

Recent developments and trends in cloud computing promise to partially relieve users from the burden of performing security tasks. Users

can now leave some security management tasks, such as installing updates and performing anti-virus scanning, to experts of cloud computing companies they trust with their data [79]. This change towards cloud storage also brings other significant benefits, such as continuous availability of data anytime, anywhere, and sharing data easily with friends or family. Users currently store personal documents as attachments in webmail accounts, collaborate on documents and spreadsheets in Google Docs, synchronize data across their computers through the cloud, and store personal pictures and communications in Facebook and Flickr. A survey by the Pew Research Center estimates that 69% of all Internet users had either stored data online or used a web-based software application by 2008 [76]. In this thesis, we refer to web-based data storage and sharing platforms intended for private users, such as Facebook, Flickr, Gmail, or Dropbox, as *consumer cloud storage* systems, as previously defined by Hu et al. [77].

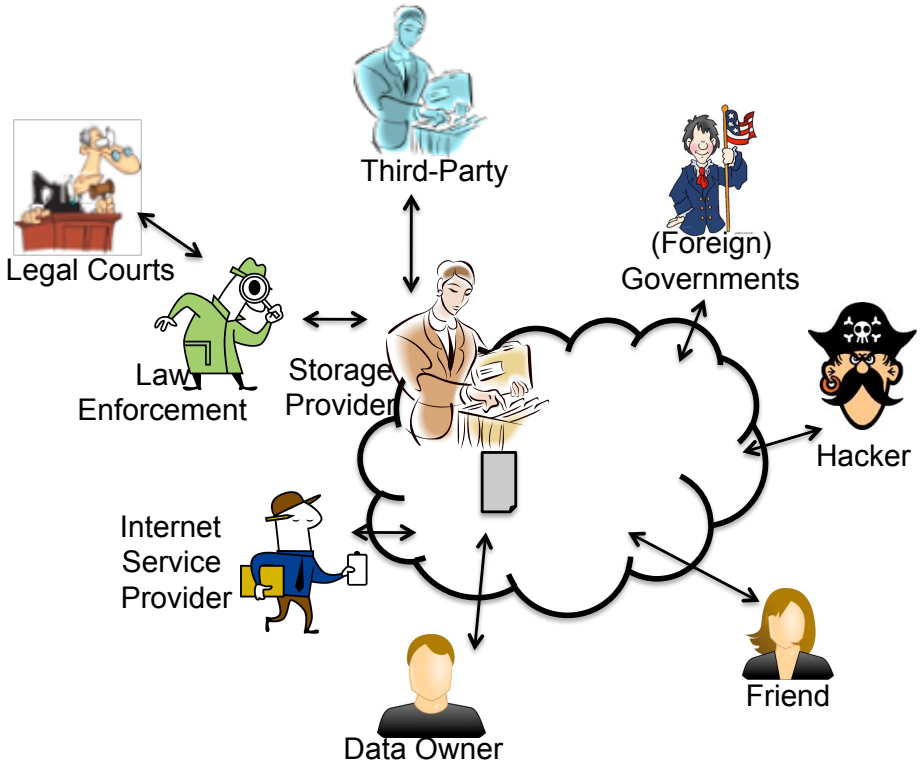


Figure 1.1.: Many stakeholders affect end-users' privacy when digital data is shared online.

Countless security breaches and privacy violations on users' data show that, instead of solving security problems, cloud computing in fact raises new challenges. Recent years have seen a wide range of security issues that threaten consumer privacy, such as the 2009 cyber at-

tack on Google [65] and the 2009 Google Docs bug that made private documents briefly public [163]. As shown in Figure 1.1, consumers' private data stored in the cloud is exposed to privacy violations from a number of different parties. For example, despite assurance to the contrary, several social networks (including Facebook and MySpace) have been leaking personal and identifiable information about users to advertisers [95, 149]. Currently, privacy laws are not fully effective in enforcing data protection. A recent study estimates that 45% of the large organizations in UK breached data protection laws in 2011 [127].

Consumer data stored in the cloud is not exposed just to hackers and companies, but also to abuse by governmental agencies. Governments have repeatedly demanded that companies install backdoors in security solutions or build local servers to facilitate surveillance [92, 144]. Unlike with local storage, in the cloud users do not typically know when their data is being accessed by other parties. The notice requirement for stored communications in the United States, for instance, is satisfied by notifying only the storage provider—and not the user!—of government access [144].

1.1.3. Consumer Privacy Concerns in the Cloud

Considering the apparent carelessness with which consumers disclose personal information on cloud sharing platforms and social networks, it has been debated whether users are aware of the involved risks and whether they are concerned about privacy at all. In 1999, Scott McNealys, the CEO of Sun Microsystems, shocked the media by undermining privacy expectations in the digital world with the statement: *"You have zero privacy anyway. Get over it"* [148]. Ten years later, Eric Schmidt, the CEO of Google, said that *"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place"* [51].

Strong outcry against privacy invasive practices, such as Facebook Beacon [119] and Google Buzz [167], seem to suggest that, to some extent, consumers do treasure their privacy. Furthermore, surveys have repeatedly ranked privacy high among consumer concerns. For example, in 1999, 52.8% of the respondents of the 10th WWW User Survey conducted by the Graphic, Visualization, & Usability (GVU) Center at Georgia Tech [90] declared themselves "very" concerned with security,

26.7% were "somewhat" concerned, and 19% identified privacy as "the most important issue facing the Internet." Furthermore, despite media reports and anecdotal evidence that seems to suggest otherwise, Hoofnagle et al. [74] showed that even young adults are concerned about their privacy.

However, quantifying privacy needs is not an easy task. Individual privacy concerns often differ widely, may depend on different factors, such as user education [70, 138] and cultural background [41], and might change over time [10]. A basic categorization introduced by Westin in 1967 groups people according to their privacy concerns into three categories: "fundamentalist," "pragmatists," and "largely unconcerned" [169]. More than 75% of users are typically to be found in the first two categories, which are actively concerned about privacy issues.

Existing data shows that awareness of privacy risks impacts online behavior and the adoption of new technologies, as was the case in the adoption of RFID tags [147] and the altering of user purchasing behavior on privacy-intrusive websites [157]. In another study, Acquisti and Gross [2] investigated users' privacy perceptions and concerns regarding the personal data they disclose on Facebook. The authors found that many users had misconceptions about the visibility of their profiles and about the online community's actual size and composition. Furthermore, the study showed that priming users about Facebook's information practices could partially alter their behavior.

Unfortunately, most of the time users are not aware of the privacy risks they face. For example, a recent study showed that most consumers do not know that they are being subject to targeted advertisement practices [114]. To have users demand better data protection from cloud storage providers, Hu et al. [77] and Soghoian [144] plead for raising users' awareness on the security risks in the cloud.

The data practices of Internet websites are partially reflected by the terms of service and privacy policies they publish. However, users find it hard to understand the contractual agreements and are often left unprotected against claims by service providers. Existing studies show that reading these documents is time consuming and often requires college-level education [9], or several years of graduate school [139]. Another study estimated that, if Internet users read just once per year the privacy policies at websites they visit, they would each spend over

200 hours reading privacy policies every year [113].

Security and privacy studies for cloud computing have so far been mainly restricted to enterprise adoption of cloud services [28, 30, 62, 68]. Such studies conclude that companies are storing only their less sensitive data in the cloud, and that security concerns are the main reason impairing cloud adoption [30, 134]. However, not much is known about end-users' concerns and practices. We believe that understanding users' expectations of privacy is essential both in devising appropriate laws and regulations and in designing privacy compliant systems. Therefore, we identify the need for a study eliciting consumers' privacy needs, expectations, and behavior regarding the cloud.

1.1.4. Security Mechanisms for the Cloud

Current technical solutions are not successful in guaranteeing protection of user data in the cloud. Encryption provided by the service provider is not a solution to defend against abuse by the service provider itself [144]. If the encryption key is generated by (or with the help of) the provider's software, even if the key generation takes place on the client side, the user must trust the provider not to leak the key. As a solution, Soghoian [144] suggests the use of open source software and diligent fingerprinting of local installation files to avoid later backdoor insertion. In this context, one viable solution is for users to encrypt their data independently, before uploading it to the cloud.

A similar challenge to data protection in the cloud has been faced by email communications over the past 20 years. Pretty Good Privacy (PGP) was designed by Phil Zimmermann in 1991 and provides cryptographic protection for emails and attachments based on public key cryptography [57]. Intended as a "cryptographic tool for the masses," PGP does not rely on any central certification authority (CA) that everybody trusts. Instead, PGP creates a Web of Trust: users generate and distribute their public keys, and sign each other's public keys.

However, PGP failed to see an extended adoption among non-technical users. Whitten and Tygar [171] identified usability problems with PGP 5.0, and showed that users have trouble setting up encrypted email communications because they do not understand the difference between public and private keys and the role of certificates [171]. This is due

perhaps to the fact that there is no intuitive model that explains the security properties of public key infrastructures [14]. Data security mechanisms for non-technical users should, therefore, make setup easy and key distribution transparent, to abstract away from complex, technical concepts.

More recently, a number of systems have been proposed to limit the access of unauthorized parties to user data. Yet these solutions typically just shift the trust users currently place in platforms providers to other parties by introducing trusted servers, are platform specific (e.g., work for Facebook only), or fail to hide that confidential information is being exchanged [17, 105, 106, 154]. We, therefore, identify the need for a novel system that enables users to share data over any web-based platform, while both protecting the confidentiality of the communicated data and hiding the fact that confidential information is being exchanged.

1.1.5. Secret Key Exchange off the Cloud

In security systems in which users encrypt data locally and then share it online through third-party platforms, typically users must first securely exchange public keys or agree on shared secret keys. Kapadia [89] discusses the two main approaches for users to reliably distribute their public keys to recipients: (1) rely on a trusted third party to certify that the received public key is bound to the given identity, or (2) manually verify the authenticity of the received key by checking the fingerprint of the key through an out-of-band channel that ensures data authenticity. Kapadia argues that most everyday users do not have mutually trusted certification authorities, and therefore they should verify the fingerprints [89]. Furthermore, obtaining certificates from certification authorities is typically too difficult, expensive and time consuming even for power users. Based on user comments, Gutmann estimates that "it takes a skilled technical user between 30 minutes and 4 hours work to obtain a certificate from a public CA that performs little to no verification" [71].

One viable solution consists of having users perform key exchange through trusted, out-of-band channels, such as by taking advantage of users' mobility and personal mobile devices to perform key exchange or verification. In 1996, Ellison [48] proposed establishing identity without

certification authority: "If two parties are in personal contact periodically, they can exchange their public keys in person (or alternatively they can exchange secure hashes of their keys)." Nicholson et al. [121] proposed using the SMS network to verify the fingerprints of keys exchanged over the Internet. Another work proposed exploiting users' everyday mobility plus the capabilities of an overlay network to resend hashes from diverse access points [122]. To verify the key, the system multicasts the key fingerprint to a subset of peers in an overlay network who forward the fingerprint to the destination.

In key distribution solution based on personal contact, securely connecting users' mobile devices is of crucial importance. Out-of-band channels are commonly used to securely connect two mobile devices that share no a priori context over a wireless link; such methods are referred to as device pairing protocols [101, 112, 151, 164]. In the absence of actual wires, an out-of-band channel is used to verify the authenticity of the wireless link. An example is the popular Bluetooth pairing method of displaying a 4–8 digit number on one device, and having the user enter the same number on the other device [24]. The usability of such methods is very important, as complex mechanisms might raise the probability of human error, and can make users disable security. We identify the need for a comparative study on devices pairing protocols.

1.2. Contributions

This thesis investigates *usable security mechanisms for protecting information sharing in the cloud.* First, we elicit, through an in-depth, cross-cultural study, users' privacy needs and expectations regarding the cloud, as well as current practices and concerns. Second, we design a system that allows non-technical users to set access control rules and protect the data they share on web-based data sharing platforms. In our system, we leverage the availability of mobile personal devices to perform secure key exchange between users during personal encounters. To this end, we investigate through a user study the usability of different methods proposed to secure spontaneous, short-range interactions between mobile devices. Our work provides insight into users' perceptions of security and privacy, and proposes novel, user-centered mechanisms for managing the security of user data and communications.

The main contributions of this thesis are:

1. We show—through 36 semi-structured interviews in Switzerland and India, and 402 responses in an online survey—that end-users are highly concerned about the privacy of their data and communications in the cloud, and that they desire novel privacy protection mechanisms more effective than those currently available.
2. We design and implement a novel system that enables users to share data over any web-based cloud storage platform. Our system protects the confidentiality of the communicated data and hides the fact that confidential data is being exchanged.
3. We evaluate the usability of different approaches to securely connect mobile devices and exchange encryption keys, and elicit users' mental models regarding the security of different device pairing methods.

In the following, we describe in more detail each of these contributions.

1.2.1. Elicit End-users' Privacy Concerns Regarding Cloud Storage

Several studies identified security and privacy as major reasons of concern in cloud adoption for companies [30, 37, 134, 144, 153], but no study investigated end-users' attitudes and practices. Not much is known about consumers' privacy beliefs and expectations for cloud storage, or about users' assumptions on contractual terms and conditions. To this end, we conducted 36 in-depth interviews in Switzerland and India (two countries with different privacy perceptions and expectations); and followed up with an online survey with 402 participants in both countries. In this study, we explored users' privacy attitudes and beliefs regarding their use of cloud storage systems.

Our results show that privacy requirements for consumer cloud storage differ from those of companies. Users are less concerned about some issues, such as guaranteed deletion of data, country of storage, and storage outsourcing, but still hesitate to use cloud storage services. Our results further show that end-users consider the Internet intrinsically insecure and prefer local storage rather than the cloud for storing sensitive data. However, users desire better security and claim that

they are willing to pay for on-line services that provide strong privacy guarantees. Participants in our study had misconceptions about most guarantees their cloud storage providers offer. For example, users believed that their provider is liable for data loss, does not have the right to view and modify user data, and cannot disable user accounts. Finally, our results show that cultural differences greatly influence users' attitudes and beliefs, such as their willingness to store sensitive data in the cloud and their acceptance that law enforcement agencies monitor user accounts. Our observations can help in improving users' privacy in cloud storage systems. Given our findings, we believe there is a need for novel security mechanisms that enable users to better protect their privacy in cloud storage.

1.2.2. Design a Novel Security System to Protect Information Sharing in the Cloud

While several systems have been proposed to limit the service providers' and unauthorized parties' access to user data, they typically require additional trusted servers, are platform specific (e.g., work for Facebook only), or fail to hide that confidential information is being exchanged [17, 105, 106, 154]. In this thesis, we propose a novel system that enables users to share data over any web-based platform, while both protecting the confidentiality of the communicated data and hiding the fact that confidential information is being exchanged. Our system encrypts the data the user wants to post on the communication platform, and stores the encrypted data on another storage platform. Instead of storing encrypted data on the communication platform, our system posts fake, genuinely looking data that contains a secret pointer to the location of the encrypted data. This secret pointer can only be extracted by intended recipients.

We provide a proof-of-concept implementation of our system in the form of a publicly available Firefox plugin, and demonstrate the viability of our approach through a performance evaluation. We show the extensibility of our solution by devising website-specific rules for receiving protected messages on Facebook, Gmail, and Twitter. In our system, a mobile application assists the user in performing and verifying key exchange with contacts. To perform key verification during physical encounters, users must first securely connect their mobile devices by

running a secure device pairing protocol.

1.2.3. Evaluate the Usability of Device Pairing Protocols

Recent years have seen a proliferation of secure device pairing methods that try to improve both the usability and security of today's de-facto standard, the PIN-based authentication. Evaluating the usability of device pairing protocols is a challenging task. Most comparative laboratory studies have focused on completeness, trying to find the single best method among the dozens of proposed approaches—one that is both rated the most usable by study participants, and which provides the most robust security guarantees [88, 93, 96, 97, 98]. This search for the "best" pairing method, however, fails to take the variety of situations into account in which such pairing protocols may be used in real life.

In this thesis, we conduct a comparative study that explicitly situates pairing tasks in a number of more realistic situations. Our laboratory studies interviewed 25 participants who had to learn four representative device pairing methods, and explain which of the methods they would use in given, real-life situations. Our results indicate that people do not always use the easiest or most popular method. Instead, they prefer different methods in different situations, based on the sensitivity of data involved, their time constraints, and the social conventions appropriate for a particular place and setting. Our study also provides qualitative data on factors influencing the perceived security of a particular method, the users' mental models surrounding security of a method, and their security needs.

1.3. Thesis Outline

This thesis is structured as follows: Chapter 2 focusses on consumer privacy studies and on cloud computing security concerns. We present the methodology and demographics of the interview studies and online survey we conducted, and show our main findings regarding current user practices, perceived privacy in the cloud, and awareness of terms and conditions. Chapter 2 is based on the publication titled "Home is Safer than the Cloud! Privacy Concerns for Consumer Cloud Storage,"

presented at the *Symposium of Usable Privacy and Security (SOUPS 2011)*, held in Pittsburgh, PA [83]. I would like to give special thanks to my co-author Niharika Sachdeva who conducted the interview studies in India and collaborated in analyzing the results. Furthermore, I would like to thank my co-authors and advisors Prof. Ponnurangam Kumaraguru, from IIIT-Delhi, India, and Prof. Srdjan Čapkun from ETH Zurich for their guidance. The design of the study and research questions as well as the presentation of the results were, however, carried out independently.

In Chapter 3 we present a novel, user-centered system for protecting information sharing in the cloud. We introduce our security protocol and follow up with its security analysis. We then present the implementation challenges, and prove the feasibility of our approach through a performance evaluation. This chapter is based on the technical report number 767 titles "For Some Eyes Only: Protecting Information Sharing in the Cloud," which was published at ETH Zurich in June 2012. I would like to thank my co-author Filipe Beato for his advice regarding the implementation of the prototype and collaborating on system design and security analysis. The development of the project idea and the implementation efforts were, however, carried out independently. Prof. Bart Preneel from ESAT/ COSIC – KULeuven and IBBT, Leuven, Belgium, Prof. Srdjan Čapkun from ETH Zurich, and Prof. Marc Langheinrich from University of Lugano (USI), Switzerland, provided guidance and feedback.

In Chapter 4 we present the methodology and results of our comparative usability study on device pairing protocols. This chapter is based on the publication titled "Influence of User Perception, Security Needs, and Social Factors on Device Pairing Method Choices," published at the *Symposium of Usable Privacy and Security (SOUPS 2010)*, held in Seattle, WA [82]. The paper is co-authored with three outstanding advisors: Prof. Marc Langheinrich from the University of Lugano (USI), Switzerland, Prof. Ponnurangam Kumaraguru from IIIT-Delhi, India, and Prof. Srdjan Čapkun from ETH Zurich, who provided guidance during study design and results analysis. The interview studies, interpretation and presentation of the results were done independently. Chapters 2, 3, and 4 contain text and figures from the respective publications.

Finally, in Chapter 5 we provide a summary of our contributions. Tak-

ing a step back, we also discuss some of the open challenges and interesting future directions that we believe will affect privacy and security in the next decades.

2. Home is Safer than the Cloud! Privacy Concerns for Consumer Cloud Storage

2.1. Introduction

Based on a recent survey by Pew Research Center, experts predict that, in the next decade, cloud computing will become more dominant for end-users than desktop computing [8]. A 2011 survey by Hosting concludes that cloud storage drives the growth of cloud computing [7]. Data is moving from user-owned desktops and laptops to dedicated online storage systems, e.g., Dropbox [44], Google Docs [67]. As mentioned in Chapter 1, in this thesis we focus on cloud storage systems intended for private users, also known as *consumer cloud storage* systems [77].

Cloud storage poses novel security and privacy threats, which may slow down or impair its adoption. Security and privacy analysis so far have mostly focussed on enterprise cloud adoption [28, 30, 62, 68]. However, clouds equally impact end-users' privacy and expose users private documents to hackers (e.g., 2009 Google cyber attack [65], bugs in access control enforcement systems [163]), or to governments [144]. While companies and governments may afford to hire trained security consultants, end-users lack the necessary resources and security education to investigate the data practices of cloud storage providers. The data confidentiality, integrity, and availability risks are partly reflected by the terms of service and privacy policy of consumer cloud storage companies. It is common practice for free consumer cloud storage services not to offer any service guarantees, to assume no liability for any data loss, and to reserve the right to disable accounts without reason or prior notification. Furthermore, storage providers may change or stop providing the service at any time. Given that users don't usually read the

terms of service and privacy policies, it is unclear how many are actually aware of these conditions. Cloud reliability questions have been raised when 150,000 Gmail users and 17,000 Hotmail users found decades of personal email and documents deleted from their accounts [6].

Understanding users' expectations of privacy is essential in devising appropriate laws and regulations. Governments have in the past demanded that companies install backdoors in security solutions and build local servers to facilitate surveillance [92, 144]. Users do not typically know when the data they stored in the cloud is being accessed by other parties. For example, in the United States only the storage provider must be informed of government access to user data, not the user [144]. The issues of surveillance and notice requirement have only recently come to media's attention, when Twitter disclosed the US government subpoena to turn over user data, including IP addresses, for a number of people connected to Wikileaks [142]. Privacy activists argue that consumers expect privacy in the cloud [69], while law enforcement agencies in United States, to which most cloud storage providers are subject to, stipulate that "a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties" [143].

In this chapter, we analyze users' expectations of privacy in the cloud and their awareness of the terms of service agreement with cloud storage providers. We investigate how practices and concerns towards cloud storage differ from those of local storage. Through 36 interview studies with users of cloud storage systems, such as Dropbox, Google Docs or webmail services, we gathered qualitative data that helps elicit users' current attitudes and concerns related to the security and privacy of their data. We test the conclusions derived from the interviews with 402 participants in an online survey. We formulate the central research question as follows: (1) *What do users think about the security and reliability of cloud storage?* (2) *What are users' privacy concerns in cloud storage?* (3) *How do privacy concerns influence users' behavior?* and (4) *How do privacy concerns differ among cultures?*

Privacy studies so far have been mostly focused on the United States. Instead, we compare privacy attitudes towards cloud storage in India and Switzerland, two countries with substantial cultural differences, as shown by Hofstede's cultural dimensions [58]. Switzerland has an individualistic society and India a collectivist one. Switzerland has an Individualism Index (IDV) value of 68 and India of 45; the US value is

91, the European one is 61 and the world average IDV is 43 [59]. Indians accept that power and control in society are distributed unequally, whereas people of Swiss nationality expect an equal distribution. This is shown by Hofstede's Power Distance Index (PDI). India has a PDI value of 77, which is very high compared to the Switzerland's 34 and the world average of 56.5. For reference, the European average is 45 and the US value is 40.

Furthermore, the Swiss Federal Constitution guarantees the right to privacy, but the Constitution of India does not explicitly recognize it. While in Switzerland, privacy is regulated through the Swiss Federal Data Protection Act established in 1992 and amended in 2008 [54, 133], in India there is no general data protection law [117]. However, the Indian government did pass the Information Technology Act (IT Act 2000), later amended in 2008 [141], and there are current efforts to introduce a data protection bill [91]. We expect the topic of privacy to get more momentum of discussion in India, especially with the introduction of the Unique Identification (UID) [159] and National Intelligence Grid [120] projects.

In this chapter we make three main contributions. First, we study users' attitudes and beliefs with respect to their privacy in cloud storage systems. We observe that, despite security expertise and guarantees provided by storage providers, users still consider local storage safer than the cloud, because they believe that nothing on the Internet is safe. Users would, therefore, rather rely on physically protecting devices storing their digital data. Nevertheless, a strong feeling of security in the cloud emerges from the belief that nobody would be interested in seeing their data, because "I am not important," "not famous," or "not criminal." Second, our results also show that users believe to have more rights and protection than the contractual terms with the cloud storage provider actually grants them. Users are typically unaware of the terms and conditions, and in fact assume to have higher availability, integrity, ownership guarantees and privacy protection in the cloud than they actually have. Still, when prompted, they said they would pay for better privacy in their cloud storage account. Third, we study privacy concerns and expectations in populations from two distinctive cultural backgrounds and observe that their cultural differences affect their privacy concerns and expectations in the cloud. We found significant attitude differences between Swiss and Indians: par-

ticipants from Switzerland store less sensitive data in the cloud than Indian participants, and are more aware of the lack of guarantees. Furthermore, while Swiss participants considered government monitoring of data stored in the cloud a fundamental infringement of their privacy rights, participants from India regarded this surveillance activity as a necessary measure for combating terrorism.

The rest of this chapter is structured as follows: Section 2.2 gives an overview of previous consumer privacy studies and of cloud computing security concerns, Section 2.3 describes the methodology and demographics of our interview studies and online survey, and Section 2.4 presents our main findings regarding current user practices, perceived privacy in the cloud, and awareness of terms and conditions. Finally Section 2.5 presents the conclusions and implications of our study results.

2.2. Related Work

For companies, security and privacy concerns are the main issues impeding cloud adoption; as a result, major cloud adopting corporations are mostly putting only the less sensitive data in the cloud [30, 37, 134]. Many studies evaluate enterprise security risks and cloud computing adoption [86] and devise security guidelines and best practice recommendations [85], or propose instruments to assess the cloud's security [126] and provide insurance for stored data [77]. A study by the Data Security Council of India investigated how companies in India deal with security risks when adopting cloud computing [37]. Most companies mitigate risks by negotiating legal terms with the cloud provider that explicitly share liability in case of security breaches and unavailability of data. Such risk mitigating approaches are currently not available in consumer cloud storage.

Several studies have analyzed the terms of usage and conditions laid down by cloud storage providers, as well as relevant national and international data protection laws [144, 153]. However, no study has explored in depth users' understanding and expectations of privacy guarantees for cloud storage. A study by the Pew Research Center surveyed levels of privacy concerns in American Internet users [76]. In the survey, 63% of participants said they would be very concerned if the

cloud storage provider retained copies of files which they tried to delete. Forty nine percent of participants said it would be an issue of concern if the provider gave their files to law enforcement agencies when asked. It is not yet clear if, and to what extent, users are aware of such issues, their expectations of privacy and how these concerns would alter their behavior towards online storage services.

Hu et al. [77] evaluated four cloud storage systems: Mozy, Carbonite, Dropbox, and CrashPlan. None of these systems offered any guarantees for data integrity and availability, nor assumed any liability in case of data security breaches or data loss. Although generally viewed as safe backup solutions, online storage systems are far from the perfect solution users envisage. Hu et al. suggest that special tools are needed to make users *aware* of existent risks and to create *demand* for better data protection and privacy solutions from cloud storage companies.

More focused on the legal issues of data confidentiality, Soghoian [144] makes a detailed analysis of threats to personal data in Web 2.0 technologies. His work emphasizes the legal and technical issues of which users should be aware. Currently, inadequate data protection mechanisms expose users to hackers and excessive government access. Not only do Web 2.0 companies have no incentives to provide better data protection as part of their free services, but their business models rely on gathering large amounts of private information which can potentially be used for targeted advertisement. Soghoian claims that users are highly unaware of the privacy risks to which they are being exposed, but so far no empirical data has been collected to support such statements.

A number of studies on Internet privacy attitudes and social networks have been conducted. Westin designed special indices to classify people as “fundamentalist” and “pragmatists,” denoting people of high and medium privacy concerns. Only around 20-25% of people are “unconcerned" [100, 169]. Hoofnagle and King [75] investigated Californians’ privacy perceptions and expectations in the online world, and found that users do not read privacy policies. Furthermore, users assume that, if a website has a privacy policy, it treats data in a privacy-compliant manner and it does not sell user data to third-parties. In social networks, Acquisti and Gross [2] found that users have misconceptions about the visibility of their profiles on Facebook, and that priming about Facebook’s information practices can alter users’ behavior.

Most existing privacy studies are targeted at consumers in the United States. However, there is a need for a global, technical, and legal framework for privacy protection. For this reason, it is necessary to understand consumers' privacy behavior and differences across different nations. Few studies so far have looked at privacy expectations in India and in Europe. Kumaraguru and Cranor [99] showed that Indians exhibit an overall lack of awareness of privacy issues and less concern about privacy than Americans do. In a more recent study, Patil et al. [125] compared privacy attitudes of knowledge workers in India and the United States. While their results confirmed that privacy concerns in India are lower than those in the United States., in some regards, Indians unexpectedly expressed higher interpersonal privacy concerns compared to their American colleagues. Bellman et al. [19] showed that cultural differences and national regulation influence Internet privacy concerns.

To fill the gap in understanding users' perceptions, in this study we explore users' beliefs about the rights and privacy protection they enjoy in cloud storage. In particular, we investigate issues such as the right of the storage provider to disable accounts at any time and with any reason, and the lack of guarantees for permanent deletion of data.

2.3. Methodology

To explore users' privacy practices and expectations, we conducted 36 semi-structured, in depth interview studies—16 in Zurich, Switzerland and 20 in Delhi, India. Next, we designed an online survey which was filled by 402 participants to confirm our interview findings. In this section, we describe the methodology of the interview and online studies, and present the demographics of our participants.

2.3.1. In-depth Interviews

Interview sessions involved one participant at a time and were run by one moderator. They took place either in our offices or in the participant's home or office. Figure 2.1 depicts the setup of a study session. Interviews were mostly conducted in English, but also in German and Hindi. They lasted between 45 and 120 minutes (M=80min,

Figure 2.1.: Study session in the home of one of our participants in Delhi. The sessions were audio recorded for future analysis.

SD=20min). The sessions were audio recorded for future analysis.

Two moderators, one living in Zurich and one in Delhi, were involved in carrying out the interviews. This ensured that the moderator understood the participant's culture and could later provide explanations for differences in attitudes between Europe and India. For example, events that had been featured in local press and specific services available in the region were mentioned during the interviews. Interviews in Delhi started once those in Zurich were completed. To ensure consistency of methodology and focus, the Zurich moderator travelled to India and took part in the first seven interviews in Delhi. In the course of these interviews, the Indian moderator's role changed from passive observer to main discussion leader.

We started the discussions by asking participants about the electronic devices they use and about the types of data they store on these devices and in the cloud. During the interviews, the moderator never used the term "cloud" unless the participant used it first, which almost never happened. We asked participants what attachments they have in their webmail accounts, what documents they email to themselves, as well as questions about their picture albums on social networking sites, blogs, and personal documents in dedicated storage systems, such as Dropbox

and Google Docs. A complete list of interview questions can be found in Appendix A. Next, we asked participants whether they currently store and if they would store in the cloud or on their personal computers (1) digital copy of *passports* or other ID documents, (2) *financial files*, (3) *health history information*, and (4) *password lists*. To avoid bias, we did not inquire about security and privacy concerns until the participant opened up the discussion.

Next, we asked participants what they thought their rights were regarding country of storage, outsourcing data storage, unauthorized modification, guaranteed deletion of data, liability in case of data loss, and account disabling. To investigate their beliefs, for each of these categories we showed participants a printed paper with three or four variations of statements that appear in the Google, Google Docs or Dropbox privacy policies. We inquired about which statement participants assumed is the stated one. In doing so, we tried to understand how much privacy participants thought they had in the cloud, as well as how safe and confidential they considered their data to be from hackers, company employees, police, and governments.

The interviews involved collecting data about participants, such as password practices, where they store sensitive information, their attitudes towards police, government surveillance, and practices regarding storage of pirated music and movies. To conduct such interviews, we were not required, neither in Switzerland nor in India, to go through an approval process similar to the one for the Institutional Review Board (IRB) in the United States. However, authors of this study have previously been involved in studies with IRB approvals, and have applied similar practices in this work. Prior to the interview, each participant was shown a printed consent form, which he or she had to read and sign, if they were comfortable with it. The form stated that an audio recording would be taken and that collected data would be anonymized and used only for the purpose of this research. Furthermore, participants were informed that they could withdraw from the interview at any point and request the deletion of the audio recording. One participant chose to stop the interview and requested the deletion of the audio file after 15 minutes, as we were asking questions about sensitive personal digital data, such as passport copy and password list. We deleted the audio file as requested and have not used the data in our analysis. Table 2.1 summarizes the demographics of interview participants.

	Zurich N=16	**Delhi N=20**
Gender		
Male	7	12
Female	9	8
Age		
<25	8	12
25 - 30	3	3
30 - 39	1	3
40 - 49	4	2
Education		
High School	3	3
Bachelor's	8	7
Master's	5	10
Heard of encryption	6	10
Leave laptop or wallet in the car	3	8
Save credit card info on websites	6	2
Helped fix a computer	7	14
Have created a web page	4	4
Store pictures online	8	20
Use a cloud storage service	8	3

Table 2.1.: Demographics of interview participants.

2.3.2. Online Study

To confirm our interview findings, we posted an online questionnaire on SurveyMonkey [152] which probed on privacy attitudes regarding the cloud. Some questions were multiple choice; these questions were constructed from frequent answers we obtained during the interviews. Other questions asked respondents to specify on a Likert scale from 1 to 4 how much they agree with certain statements; an N/A option was also provided. To filter users who only clicked through, we included a question that tested whether participants read the question description or not. On average, the survey took 23 minutes to complete (excluding the largest 15 values). We discuss the recruitment process for our participants in the following section.

2.3.3. Participants

Interviews. We recruited participants through flyers posted in the city and at universities, through online advertisements on website hosted by ETH Zurich, on mailing lists, and through word of mouth. To avoid a biased sample, the advertisement did not mention privacy nor security, and said only that we are looking for people who use online platforms to store data. In particular, we mentioned that participants should be using a webmail account, such as Gmail, Yahoo Mail, or Hotmail, or share pictures online through Picasa Web or Flickr. During recruitment, we preferred Dropbox and Google Docs users and rejected IT experts and computer science students. We offered a monetary reward of 20 Swiss Francs (aprox. USD 17) to participants in Zurich, and 250 Indian Rupees (aprox. USD 6) in India.

Of the 20 people interviewed in Dehli, 19 reported their nationality as Indian and one as Estonian. Of the 16 people interviewed in Zurich, 13 were Europeans (4 Swiss, 4 Germans, 2 Italians, 2 Serbians, one Austrian), one American, one Chinese and one Indian. Professions varied with 10 participants in business and sales, 7 in social sciences and linguistics, 5 in natural sciences like chemistry and biology, 4 in engineering, 4 in art and design, 2 in finance, and 2 computer scientists, one economist, and one urbanist.

Online survey. Participants were recruited through Facebook postings, student mailing lists, and word of mouth. In Delhi, 100 forms

were distributed as hardcopy in several universities and later collected. To incentivize participation, we offered three $100 Amazon vouchers given to random participants at the end of the study. We had 450 respondents from which we dropped 48 based on the test question. Table 2.2 shows the demographics of the remaining 402 participants. Of these, 189 had Indian nationality, 132 were Swiss and the other 47 were Europeans. Of the total 402, 182 participants lived in India and 160 in Switzerland.

2.3.4. Data Analysis

We transcribed all audio interview recordings into English. For each question in the interviews, the interview moderators identified trends and grouped answers in a few categories. If the moderators did not agree that the participant unequivocally understood the question, the answer was discarded. Throughout the interviews we received many "I don't know" answers, which we generally exclude from reporting in the results section. Finally, we formed hypotheses for the survey about current practices, perceived and expected privacy, and cultural differences based on observed trends and aggregated answers.

To analyze differences among various groups of respondents—e.g., Swiss vs. Indians, computer scientists vs. non-computer scientists—, we used the two-sample Wilcoxon rank-sum (Mann-Whitney U) test for the Likert scale questions. For the multiple choice questions, we applied Fisher's exact test for each of the possible answers to determine if a certain group (e.g., Swiss or Indians) is more likely to provide the respective answer. For the Likert scale results, we discarded neutral (N/A) responses from the analysis.

2.4. Results

In this section, we present the main findings of our study. We start by reporting on current practices, such as what kind of data users store in the cloud, and continue by presenting users' mental models. Section 2.4.2 describes perceived privacy and privacy expectations of consumers, and Section 2.4.3 discusses users' understanding of key conditions stipulated in the terms of service. We refer to interview partic-

	Swiss N=132	**Indians N=190**	**All N=402**
Gender			
Male	70	55	60
Female	30	45	40
Age			
18 -24	60	66	60
25 - 34	34	22	30
35 - 44	4	7	5
>45	2	5	5
Education			
High School	47	30	35
Bachelor's	30	40	35
Master's	16	25	25
PhD	4	4	5
Computer Scientists	60	21	36
Computer Skills			
Novice	2	3	3
Intermediate	26	53	40
Proficient	42	33	38
Expert	30	11	19
Platforms Used			
Google Docs	48	70	60
Dropbox	51	17	34
FolderShare	1	14	8
Gmail	61	91	77
Yahoo Mail	13	60	40
Hotmail	34	10	22

Table 2.2.: Demographics of online survey participants; values presented as percentages.

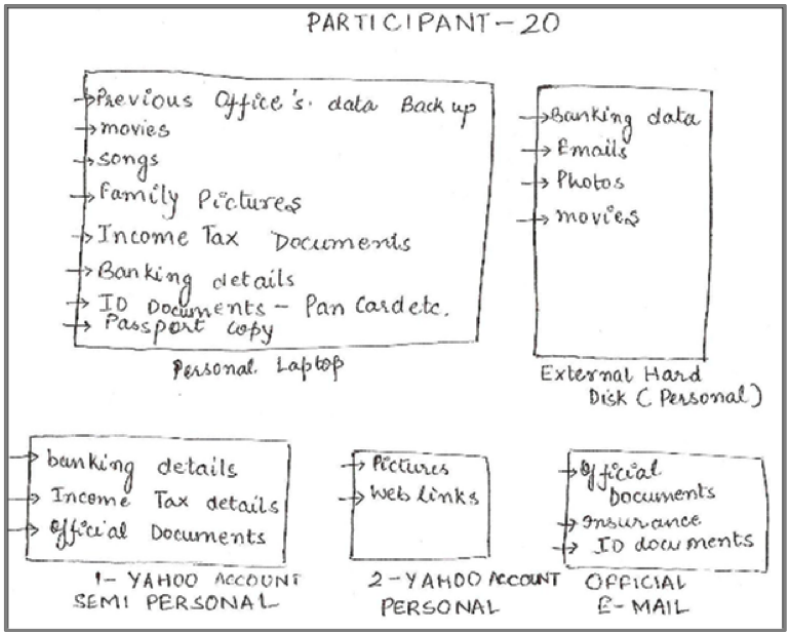


Figure 2.2.: Participants had several webmail accounts, even with the same provider, for separate purposes: private, business use, spam, backup. As shown above, D20 stores different types of data in her accounts: from ID documents in her business webmail to pictures and web-links in her personal one.

ipants in Zurich as Z1, Z2, Z3, ..., Z16, and to participants in Delhi as D1, D2, D3, ..., D20.

2.4.1. Current Practices

Six participants in Zurich and 2 in Delhi used dedicated cloud storage systems such as Google Docs, Dropbox or FolderShare. These systems were used mostly for work and collaborative projects, e.g., in school assignments, surveys. For personal data, participants made heavy use of webmail accounts. Z10 said she would rather store sensitive documents in her Gmail account rather than in Google Docs, because *"email feels more like your private space."* Participants emailed documents to themselves to synchronize data between computers, to backup important files, and to have documents available when needed.

Two participants in Zurich and 7 in Delhi said they have *"folders"* in their webmail account, referring to email labels. Figure 2.3 shows a participant's inbox. Most participants (14 in Zurich, 18 in Delhi) had several webmail accounts to differentiate between private, "official" and newsletter/spam use. Furthermore, some participants (7, all in Delhi)

had more than one account with a single webmail provider. Figure 2.2 shows D20's data distribution on local devices and the cloud, as drawn and discussed during the interview. Participants stored pictures, school or project work, official letters, CVs, music files, videos, passport copies, tax, and financial files in the cloud.

Complementary to email, participants made heavy use of USB and external hard drives to synchronize and backup data. For example, Z4 said that, when she creates a Word document, she stores it *"in My Documents, I back it up on a USB stick, and email it to myself for backup."* USB sticks were used not only as a data transportation device (e.g., to share files with colleagues or synchronize between computers), but also for permanent storage.

The online survey confirms that users do not use the cloud as a main storage unit. As shown in Figure 2.4, 83% of respondents somewhat or strongly agreed with *"I tend to keep a backup of all data I store on the Internet"* (M=1.62, SD=0.82, N=285, where 1 is strongly agree and 4 is strongly disagree). However, Swiss agreed stronger (M=1.5, SD=0.89, N=124) than Indians (M=1.71, SD=0.72, N=185), as shown by the Wilcoxon rank-sum test (z=1.67, $p<0.050$). Participants mentioned that the most annoying part about losing access to their email account would not be the loss of data, but the hassle of informing their contacts of a new email address.

Just like companies, participants were storing only the less sensitive data in the cloud. For example, Z13 said: *"If I will download a file for free, pirated, I will not put it of course on my Yahoo account. I would keep it on the laptop."* Z14 agreed: *"Would you write a diary on Google Docs, would you trust them with your secrets? I guess not."* However, what was considered *"sensitive"* differed among participants, and nationalities. Z9 considered health history more sensitive than the passport because *"a passport I show the policeman; my health card I show the doctor. The doctor is one."* Figures 2.5 shows interview participants' willingness to email some types of sensitive files to themselves. No participant said they would store sensitive data in their webmail account rather than storing it on their computer. Furthermore, for really sensitive data, like bank and tax statements, print-outs were preferred to electronic copies. For example, Z1 would not keep an electronic copy of financial files: *"I don't trust myself. Sometimes my computer is kind of hectic. It happened that I sent some files to wrong*

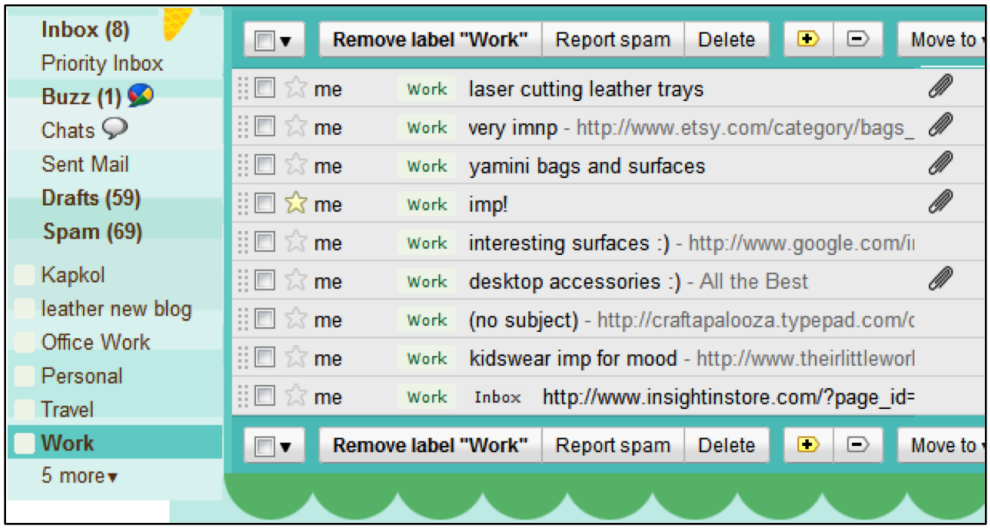


Figure 2.3.: Most participants used their webmail accounts as a cloud storage platform. D6 regularly emails documents and links to herself and then stores them in specific "folders," by setting email labels. She lives in Delhi, is a leather designer and has 11 "folders" in her Gmail account. (Screenshot presented with the participant's written permission.)

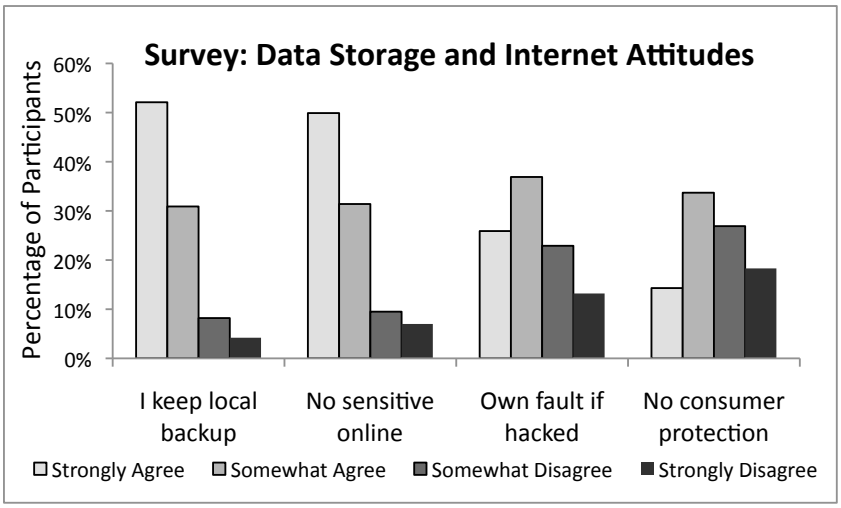


Figure 2.4.: Users keep local backups of data they store in the cloud and try to keep sensitive data away from the cloud. Most feel it is their fault if they store sensitive data in the cloud and it gets hacked, and that there is no legal protection authority they can turn to.

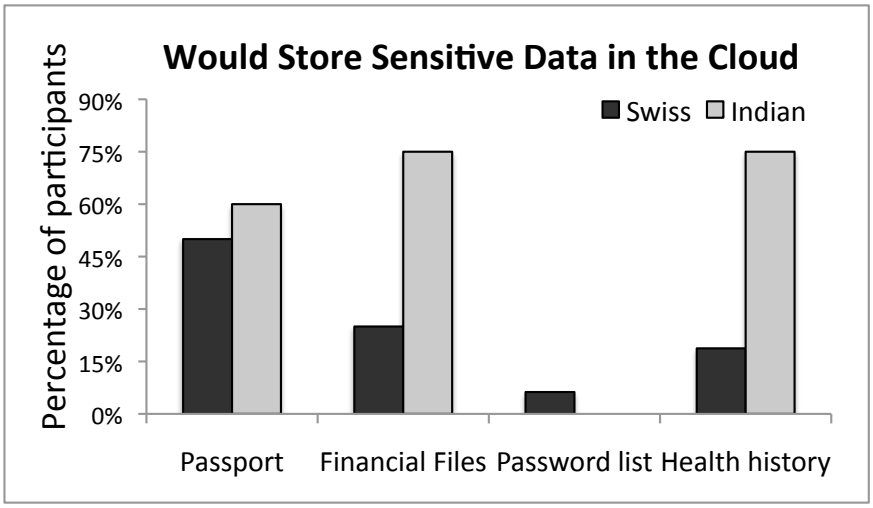


Figure 2.5.: Indian participants were more willing to store sensitive data in the cloud than Swiss participants.

people." The online survey confirmed that users prefer to keep sensitive data on local storage. 81% of respondents (M=1.71, SD=0.90, N=287) somewhat agreed or strongly agreed with: *"I try not to store important, sensitive documents on the Internet, and instead keep them offline, on my personal computers."*

We noticed several differences between study participants in Zurich and Delhi, which were later confirmed by the online survey. While Indian participants did not consider health information sensitive data, European participants were very reluctant to even store it in digital format. During our interviews, 15 people in Delhi and only 3 in Zurich said they would store financial documents in the cloud. In the online survey we asked participants to rate on a Likert scale from 1 to 7 the sensitivity of the data they have stored in the cloud. Figure 2.6 shows that Indians reported to have stored more sensitive data than Swiss (Wilcoxon rank-sum test, $z=4.23$, $p<0.001$).

Unconcerned with identity theft, and unaware of the value of a digital copy of one's ID, participants (3 in Zurich, 1 in Delhi) considered that a passport copy is not sensitive: *"It is a copy. I think that important is the original."* Overall, users were less willing to store a password list than a copy of passport in the cloud, though this is in big part due to perceived need. Some users have been pushed into emailing digital copies of official documents, with which they were not conformable, by other party's requirements, e.g., when applying for a job. In Delhi, 5 users reported to store password lists on their mobile phones, which was

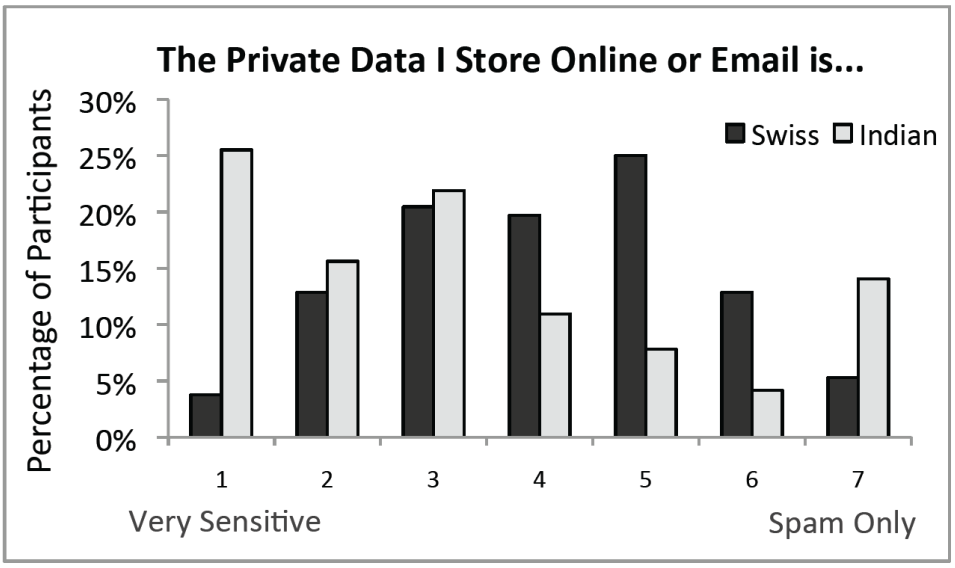


Figure 2.6.: The online survey confirmed the interviews: Indians reported to have stored more sensitive data in the cloud than the Swiss.

considered more trusted, and more accessible than a computer. Some users said they do not need to store a password list because they reuse combinations of several passwords. Only very few said they remember passwords.

2.4.2. Perceived Privacy

In this section, we discuss users' perceived privacy for the cloud, in particular their perception on who else, except for themselves, might be able to access their data, and what guarantees they think current technical solutions provide. Several participants in the study said privacy means *"that nobody else has access to my data."*

Participants' understanding of the cloud architecture was rather limited. They unanimously believed that their storage providers keep one, two, or maximum five copies of their documents. D8 said: *"I think the server needs one copy only, because from any computer in the world I can access this copy."* D17 said: *"They have so many users, [...] they would need so much space to keep multiple copies."* For Z10 *"it would be weird if they stored a backup,"* because that would mean *"they are keeping a copy for themselves."* D8 agreed: *"If it is secure and nobody can access them, why should they make more copies? One would be enough."*

Our results suggest that users consider the Internet highly insecure and feel responsible for protecting their sensitive documents themselves, rather than relying on cloud storage providers. Z16 said: *"It is still my responsibility what I upload or what I send and where it is stored."* We summarized such responses in a statement which we then included as a Likert scale in the online survey. As shown in Figure 2.4, 63% of respondents agreed that: *"If people put their private data on the Internet and it gets hacked, it is their own fault. They should know that nothing is really safe on the Internet"* (M=2.24, SD=0.98, N=291).

Almost all interview participants voiced concerns about the safety of storing documents online, many even before we prompted them about the possible sensitivity of the documents. According to statements by our participants, common perception as well as extensive media coverage on the subject shaped their beliefs. Some participants seemed to believe that digital data cannot be contained, because the Internet is *"everywhere."* They could not imagine that it might be technically possible to have online data stored in a single country. Similarly, some believed that, once uploaded on the Internet, digital copies remain there forever. According to Z16, there is a nice saying: *"The Net will not forget."*

Anybody Can See My Data, If They Want To

We asked participants who else, except for themselves, might be able to see their cloud stored data. Several participants said *"anybody"* could see it. Z15 said: *"I know that when I store data [in the cloud], the data is really for more people than myself."* We inquired about *hackers*, *storage providers* and *governments*.

Hackers. Participants unanimously believed that it would be *"easy"* or *"really easy"* for a hacker to get their data from the cloud. Only one (in Zurich) said that it would be *"hard, but not impossible"* for a hacker to break into their account and another (in Delhi) believed that *"Google cannot be [hacked, because] they have Russian army to protect their data, but Facebook and Twitter have been [hacked]."* For example, Z11 said: *"If he is a good hacker, he can do everything."* According to D19, any measures to protect online data are useless, because ultimately *"there are supernatural hijackers who are sitting there, who can dig everything away."* Z1 agreed: *"They can even get access to the websites*

of governments. Why shouldn't they [be able to access my account], if they really want to."

Storage provider. Except for one, all interview participants were aware that their storage providers can access their data. When asked why his provider would need to see the data, Z2 said: *"To arrange it. If they are keeping an account, then they look after it."* D3 was not convinced that there is a valid reason behind access: *"They come up with all stupid, stupid excuses: security reason, we need to see it."* Except for one participant who said that it might be that every employee of the company can access customer documents, people said that only *"some"* employees would have access, most often quoted being system administrators or *"security people."* Several participants said that internal *"policies"* impede other employees to access user data, or the fact that accessing customer data is *"taboo"* within the organization. Only one user had *"never thought about it. [...] Is an account accessible just from the user or, for example Gmail or Google [her storage provider], can have access as well? Now I am getting scared."* D7 said: *"They can but they don't."*

Governments. Only two interview participants, both in Zurich, said that the police or government cannot access their account. Z4 believed that they could not because only she knows her password, and Z5 because she is *"a normal citizen, in the sense of not criminal. [...] The state cannot access my bank account also, so I suppose it is more or less the same."* For D5, even with a paid account, *"at the end of the day, there is no guarantee. Like the bank account, if a governmental agency wants, they can access your information."* The affirmative answers varied from *"Yes, it would be very easy," "they [the government] must have a direct access"* or *"a program,"* to *"only through Google."*

But I am not Interesting to Them

Although participants believed hackers, storage providers and governments could theoretically view their data, none showed great concern about it. In practice, people did store sensitive data in the cloud and considered that the risk of somebody actually viewing *their* data was minimal or nonexistent. Few participants believed unauthorized access might have already taken place. The main reason given was: *"I am a normal person," "not famous," "not criminal,"* and *"not as interesting*

as Obama." Such attitudes were stated by 10 participants in Zurich and 4 in Delhi. Z1 said: *"I am not interesting to them [government], because I am just a little boy somewhere in Switzerland."* Z4 agreed: *"I am a student, I don't know why a hacker should access my account."* For governments, only a couple of participants mentioned that automatic monitoring might occur, but then again: *"I don't write bomb, bin Laden."*

Not storing valuable data online kept the hackers away. Z0 said: *"It is very unlikely that they [hackers] want to see my documents, as long as I don't store financial documents, access codes or passwords online. If I store my bank account access, yes, they would be interested."* Similarly, D10 said: *"There are too many documents and too few people in Gmail, [...] so not many manage to see my Gmail documents. [...] But in future, if I hold a good position, then they may."* It is also a matter of time. D2 said: *"I don't think anybody has that much time. Why would someone be interested?"* To participants, attacks on the Internet are targeted; viruses targeting a bulk of random user computers or accounts are not considered by the users.

Home is Safer than the Cloud

We asked interview participants where they considered their data to be safer, in the cloud or on local storage. Participants felt that availability is better online, *"in case my computer crashes,"* but for sensitive documents they strongly preferred to keep these offline. The ultimate protection against hackers is unplugging the Internet cable. For example, Z11 said: *"Hackers can access the data when we are online, not offline."* Z13 said there is a higher risk if the data is saved online compared to her own computer: *"They can try to enter on my account also if my laptop is closed, so they have more time."*

Physical protection of data stored locally, i.e., by locking the disk in the cupboard, is still better than online protection of documents. Z3 said: *"There are many people online; at home it is put away."* Z5 would not store her passport copy in the cloud because *"I look after my laptop [...] and I take care of it. But on Google Docs, I just have to depend on people that program security."* Even though she knows that Google has more experts, she *"would still keep the copy on my computer.[...] It feels here and more accessible."* The USB stick is even safer than the

laptop: *"I keep it always with me. Somebody has to really kidnap me to have the USB stick."* Even if she believed that it would be easier for a hacker to break into her computer than into Google systems, Z6 still considered her laptop safer than the cloud: *"Google has experts to deal with hackers, I have no one to help me,"* but *"professional hackers want to hack big companies, organizations, not individuals, because there is more value in that."* Similarly, D4 said: *"people know where Google Docs are, but it's difficult to find which connection I am using, where I am sitting!"*

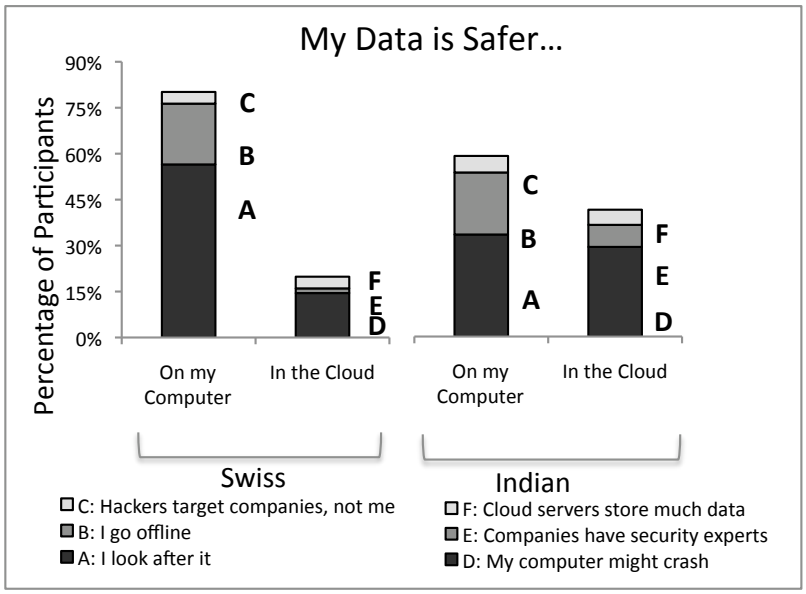


Figure 2.7.: Indians considered cloud storage almost as secure as local storage. Swiss trust much more their computer than the cloud.

In Zurich, 2 participants said the risk is the same *"if my computer is connected to the Internet,"* and none said higher on the laptop. In Delhi, however, 13 participants said the risk is higher offline and 4, online. The online survey also showed differences between attitudes of participants Swiss and Indian nationalities. We asked respondents to rank six given reasons on why local or cloud storage might be safer. Figure 2.7 compares the choices made by survey participants of Swiss and of Indian nationalities. 69% of all respondents said local storage is safer, and 31% that the cloud is. The highest rated reason why it is safer was A: *"On my computer, because I can physically protect my data,"* with 44%. Next reason was availability D: *"Online, because my computer may crash."* Interestingly, only 5% chose E: *"Online, because big companies have security experts,"* an argument often stated as major cloud advantage in enterprise usage.

A significantly higher percentage of Swiss respondents (82%) compared to Indians (60%) considered local storage safer than the cloud ($p<0.001$, Fisher's exact test). One might argue that participants' background, not just nationality, might be a factor influencing the difference in perception between Swiss and Indians. The percentage of computer scientists in the Swiss group was significantly higher than in the Indian group (60% vs. 21%, see Table 2.2). We refute this by noting that the difference persisted among the groups of Swiss computer scientists and Indian computer scientists ($p<0.010$), as well as Swiss non-computer scientists and Indian non-computer scientists ($p<0.006$).

Fisher's exact test showed no significant differences between the scores for Indian, non-computer scientists group and Indian, computer scientists ($p=1.0$). Similarly, we obtained no significant difference between Swiss non-computer scientists and computer-scientists ($p=1.0$). A failure to see a difference between computer scientists and non-computer scientists might be also attributed to the young age of participants. We note that computer scientists in our study are mainly students; their attitudes towards cloud security might be different from those of experienced professionals. Differences might be noticeable in a future study among higher age groups.

Government Surveillance

Throughout our interviews, participants from India showed very different attitudes towards government surveillance compared to participants from Europe. For example, we asked participants whether it is their right to protect the privacy of their data and communications, followed by whether everybody should be able to, and then by: even terrorists? In Zurich, 6 participants said everybody should have the right, including terrorists. In Delhi, 11 people said that terrorists should not have the right to privacy and only 3 said that everybody should. For example, Z13 said: *"There are terrorists, but it is not because of them all the people cannot have their privacy.[...] I think this is an excuse to control everything."* Z14 agreed: *"Who defines who is terrorist?"* Only one person answered that *"the police from all states"* should be able to access any data. Participants in Delhi showed a much stronger acceptance of government surveillance. They felt that *"national security comes first."*

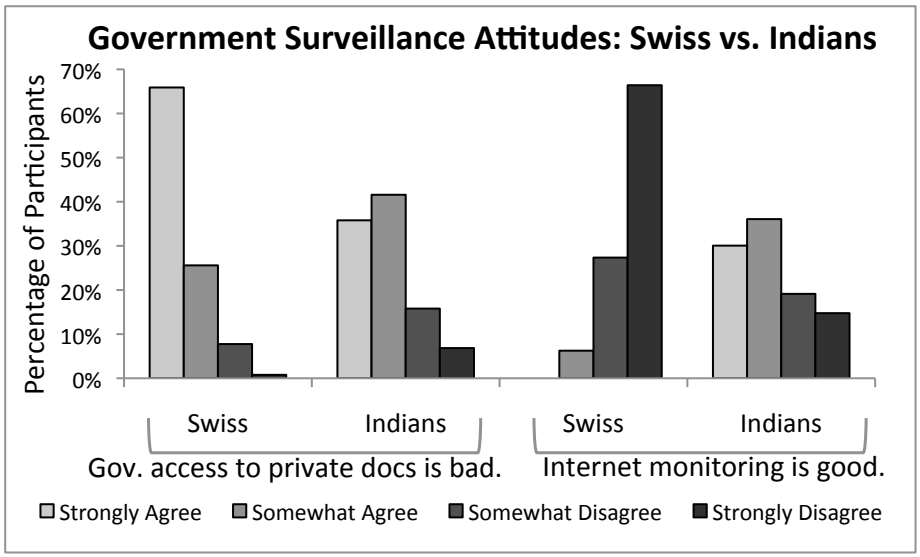


Figure 2.8.: Indians are more acceptant than Swiss of government surveillance over data stored in the cloud.

Furthermore, we asked interview participants if a communication technology currently exists, through which they could talk to a friend, for example over the Internet, and nobody, not even the government, could listen in to their communication. Overall 13 people said such a technology is technically possible, and 13 said it is not. While among Zurich participants, the general trend was that this technology is not currently being deployed for surveillance and security reasons, in Delhi people felt that such a technology should not exist, because it would be misused: *"then terrorists will enjoy themselves."*

In the online survey, we asked respondents to rate on a 4 point Likert scale, with 1 for strongly agree, two statements which we received from our European and Indian participants in the interviews. Figure 2.8 shows the answers of respondents of Swiss and of Indian nationalities. For the statement *"If the government had access to every document users store on the Internet, that would be a major violation of individual privacy,"* Swiss (M=1.43, SD=0.67, N=129) agreed stronger than Indians (M=1.94, SD=0.89, N=190): Wilcoxon rank-sum test, z=4.96, $p<0.001$. For the statement *"It is good if the government monitors every Internet communication and all user accounts. National security comes first,"* Indians(M=2.18, SD=1.03, N=193) agreed stronger than the Swiss (M=3.60, SD=0.61, N=128): Wilcoxon rank-sum test, z=10.56, $p<0.001$.

I Would Pay for Privacy

During the interviews we asked 8 Dropbox users and some non-users to identify the statement that appears in the Dropbox privacy policies among three statements. None chose the correct variant from the tree possible choices: *"Dropbox may sell, transfer or otherwise share some or all of its assets, including your Personal Information, in connection with a merger, acquisition, reorganization or sale of assets or in the event of bankruptcy."* All participants said that this should not be, and non-users said they would not open an account with such a company. Motivated by this finding, we used the online survey to see if respondents would be willing to pay for privacy guarantees.

In the online survey respondents had to choose between two companies with which they could open an online data storage account. Company A offered the service for free, but said that they may sell documents of personal information. Table 2.3 shows the results of the online survey. 79% of respondents agreed that they would pay 20 USD per year for Company B, whose policy says that they will not sell any personal information. This amount would be enough to buy twelve 2 GB of increased redundancy storage on Amazon S3 for a year.[1] We did not notice a strong difference between Swiss and Indian respondents: 81.6% of Indians and 78.5% of Swiss chose Company B. Even though the statement used is much stronger than the Dropbox policy, which may sell only in connection with a merger, our survey does show a strong user response towards privacy protection.

During our interviews, we asked participants if they would be interested to purchase insurance for their cloud-stored data the same way they have for cars and houses, so that they receive some compensation in case a hacker breaks in and they lose their data. Half of the interview participants, split evenly among Delhi and Zurich, said they were interested. For example, D6 said he would pay 1000 Indian Rupees (approx. 20 USD) per year for data insurance, while D10 said he would pay 60 USD per year. Z5 would pay 50 Swiss Francs per year and Z6 would pay *"several hundred Swiss Francs."* Others said what matters is the data, and they would instead prefer investing in an additional backup system.

[1] Amazon S3 charges $0.125 per GB per month: http://aws.amazon.com/s3/pricing/. Dated on September 2, 2012.

Table 2.3.: Willing to pay for privacy: *"Which company would you choose to store your data and why?"*

CompanyA: free, may sell user data	
– It is free.	3.0%
– I don't have sensitive data anyway.	11.4%
– I never know what they do with my data.	6.5%
	Total: 20.9%
CompanyB: costs $20, won't sell data	
– I value my privacy.	37.3%
– If the price was lower.	9.7%
– If they are trustworthy.	32.1%
	Total: 79.1%

2.4.3. Terms and Conditions

Unsurprisingly, our results confirm that users do not read the terms of service and privacy policies. D14 said: *"It's massive! It's just in five Arial font and it's massive! It's ten pages!"* A few participants said they skim through the text. Although they do not read them, participants believed strongly that these documents are legally binding and valid contracts in court. Z8 said: *"It is your fault if you did not read it."* D14 said: *"You should be smart enough not to do all that stuff [store confidential customer information]. And if you've done it, then welcome to the world, wherein you had said, 'I accept'. So, if you have accepted it, you have to take it."* Only one user said that some of the things the company claims in the terms might not be legal.

Several participants said they do not read these documents because *"I don't think they [terms of service] would have an impact."* However, the terms of service and privacy policy documents explain conditions such as: Google has the right to disable the account at any time and without notice, to read, delete and modify their data; Dropbox may sell user data, the storage provider assumes no liability in case of data loss. We explored in detail users' awareness of these terms.

Country of Storage and Storage Outsourcing

We asked participants where they thought their online data was being stored and whether the country of storage was important to them. We then asked them to imagine that their storage provider contracted a third party company to store their data. This is, for example, the case with Dropbox, who is using Amazon S3 to store user files [45]. Only 7 out of 36 interview participants said the country of storage is important or they care about storage outsourcing. Another 4 participants said it might matter if they were storing more sensitive data. Reasons given for not caring were *"I trust the company," "maybe if I had more sensitive data,"* but also *"as long as security is guaranteed"* and *"if they [the third party company] have the same privacy policy."* All participants said the data is safer in their own country, except for one Indian who said in India there are more hackers. Only 2 participants, both in Zurich, mentioned country-specific data protection laws to be a factor in data security. Fourteen participants said they care if their data storage is outsourced by their storage provider, and 19 said they should be informed when that happens. Two participants said they would close their account if data storage was outsourced, and one that he would sue the company.

Unauthorized Modification

We showed participants a slide with three variants of the policy *"Google reserves the right (but shall have no obligation) to pre-screen, review, flag, filter, modify, refuse or remove any or all Content from any Service."* One variant did not grant the right to modify, and one said *"except for personal documents in user accounts."* Only 4 participants chose the variant currently stated in the terms. Like 13 other participants, Z10 said: *"They should have the right to review, but I don't think they should have the right to modify."* Similarly, D5 said: *"No, not modify it, but this if it's illegal then they [Google] can delete it."*

Participants felt a strong ownership right over the data, even if stored in the cloud. They accepted that their data might not remain confidential and that the provider might choose to delete it, but expected full data integrity. For example, D17 said: *"It's my personal data they have to respect this thing"* or *"they are my personal documents [...] even*

if I put them on Google." Z13 thinks she still owns the copyrights of whatever data she uploads: *"They give me a space on their system. They don't say put the stuff here and everything gets mine."* D3 said that the documents she stores in the cloud: *"are my things. [...] It was created by me. [..] They might delete it, they might remove it, but they cannot make changes themselves."* For security reasons, many participants accepted that the provider might need to look at their data. For example, Z12 said: *"I can understand that Google wants to be able to look at the data that is stored. In case it is criminal data, they could inform the police or delete it."*

We summarized participants' views in four choices in the online survey. Table 2.4 shows the results for Swiss and Indian nationalities. From all respondents in the survey, only 8% answered "Yes;" 77.3% were using Gmail and 50% named Gmail as their main email account. We applied the Fisher's exact test for each of the multiple choice answers of the survey question; except for the willingness to answer "I don't know," we observed no significant difference between Swiss and Indian participants.

Table 2.4.: Unauthorized modification: *"Does your webmail provider have the right to see or modify the documents you have as attachments in your email account?"*

Response	**Swiss**	**Indians**
No.	22.3%	28.4%
They can see, but not modify my files.	12.2%	26.8%
They have the right to see and modify only in criminal or terrorists cases.	7.2%	21.1%
Yes.	10.1%	6.8%
I don't know.	48.2%	16.8%

Guaranteed Deletion of Data

At the time of the study, the Google Docs policy stated that *"residual copies of your files may take up to 30 days to be deleted from our active servers and may remain in our offline backup systems for up to an additional 60 days."* We asked participants to identify the correct statement among three other variations: one saying that data can never be deleted, one saying it gets deleted within 24 hours, and one that it

gets immediately deleted. The correct variant was chosen by 14 participants. Five participants said the data never gets deleted and 4 said deleted data resides for 24 hours. The most mentioned source of information was the media: *"Probably there are traces still there. I heard in the media, television, advertisement in journal."* No participant said the data would be immediately deleted.

Few participants felt strongly that when they delete data it should get deleted. For example, Z8 said: *"It is the private right that when it is deleted it actually is, and if somebody uses it nevertheless it is infringing my privacy."* Others said they would care about copies of sensitive data such as online banking transactions, but not about advertisement emails. Overall, participants did not show great concern: *"If somebody is storing important stuff like ID, official documents, then they should be deleted immediately."* Other participants regarded this as a good feature. D10 believed that the data would still remain on the cloud *"because it is a very good system. If I delete my document, there must be some technology through which I can retrieve my data back."*

Table 2.5.: Guaranteed deletion of data: *"When you delete a file stored on the Internet or an email in your webmail account, what do you think happens?"*

Response	**Swiss**	**Indians**
The file gets permanently deleted.	2.9%	15.3%
Some copies still exist for a few weeks.	34.5%	38.9%
Copies are kept, for security reasons.	36.7%	25.3%
I don't know.	25.9%	20.5%

We followed up these findings in our online survey. As Table 2.5 shows, very few respondents believed data gets immediately deleted. Only 15% of Indians and 3% of Swiss chose: *"The data gets permanently deleted, just as when I deleted it from my computer"* (significant according to the Fisher's exact test, $p<0.001$).

Account Disabling

At the time of the study, the Google's terms of service stated: *"You acknowledge and agree that if Google disables access to your account, you may be prevented from accessing the Services, your account details or any files or other content which is contained in your account."* Other

storage providers followed a similar policy, e.g., *"Dropbox reserves the right to terminate Free Accounts at any time, with or without notice."* Eight people in Zurich and 7 in Delhi said that their service provider (mostly referring to Google) has the right to disable their account. Seven participants in Zurich and 4 in Delhi said they do not. For example, D2 said: *"There is some trust that we have put in, they should take care of that. Not without my consent."* Doing so *"is not ethically correct."* D5 said they do not have the right *"because all my data is there, they should inform me before."*Other participants said Google may disable their account, but only with prior notice, only with a reason of *"if you do not access it anymore."* Several participants said they would sue Google if they disabled their account. Many participants accepted security reasons such as *"if I am a terrorist"* or *"use it for criminal purposes," "if I have done something and that is against their rule,"* or *"if they get complaint from other people."* Paying for the service was not always regarded as a guarantee of having more rights. For example, Z14 said: *"you paid for the storage, not the privacy."*

Table 2.6.: Account disabling: *"Does your webmail provider have the right to disable your account?"*

Response	**Swiss**	**Indians**
Yes, at any time, without advanced notice and without explanation.	34.5%	15.3%
Yes, but only with advanced notice and a valid reason.	21.6%	48.4%
Only if I use it for criminal purposes.	10.1%	13.7%
No.	4.3%	8.4%
I don't know.	25.2%	12.6%

We asked the same question in the online survey, with four multiple choice answers. Table 2.6 shows the responses; these confirm limited user awareness. The Fisher's exact test confirmed that Swiss are more aware than Indians of the fact that their storage provider has the right to disable their account without advanced notice and without explanation ($p<0.001$). Indians, on the other hand, assume that this can happen only with advance notice and a valid reason ($p<0.001$). The difference persisted between Swiss non-computer scientists and Indian non-computer scientists ($p<0.001$ and $p<0.02$ respectively).

Data Loss Liability

We asked participants what their rights would be if their storage provider lost some of their data (e.g., due to accidental deletion or server crash). The terms of service of all important online storage and Webmail companies dismiss any liability for data loss. For example, *"Google [...] shall not be liable to you for [...] the deletion of, corruption of, or failure to store, any content [...] whether or not Google has been advised of or should have been aware of the possibility of any such losses arising."* Participants had diverse views on companies' liability and their rights. For instance, D14 said: *"They're already giving you a service. [...] if you're stupid enough to keep your important documents there as a storage device and not use your external hard disks and stuff, then it's not their liability."* D5 disagreed: *"I didn't ask them to give a free service, they decided this. [...] They should pay me a large sum."*

Five participants in Zurich believed that the storage provider would be liable to them in case of data loss, and 4 participants said they would not have any rights. In Delhi, 14 participants said the storage provider is liable, and 5 participants said the consumer would have no rights to claim compensation. Several participants said they would not care about money, because the data is lost anyway, that even if they had rights the company would not pay, or that there is nobody they could contact to make the claim. A few said they would sue the company. We investigated these issues further in the online survey. Table 2.7 shows the results.

Table 2.7.: Data loss liability: *"If your webmail provider lost some of the data you store with them, what would your rights be?"*

Response	**Swiss**	**Indians**
They should pay me for the damages.	10.8%	15.3%
If it is a free service, I have no rights, otherwise they have to pay me.	27.3%	48.4%
I have no rights, even if paid-for service.	20.9%	13.7%
Don't care, my data is lost anyway.	8.4%	7.2%
I don't know.	33.8%	12.6%

Our results show that Indians are more prone than Swiss to expect liability from their service provider. They are more likely to expect the provider to pay them for damages (Fisher's exact test, $p<0.001$),

whereas Swiss are less prone to believe that they do not have any rights, even if it is a paid-for service ($p<0.003$).

Interview participants said that, if their account was hacked, disabled, or if the provider lost some of their data, they would: *"change the password," "delete all my emails," "close my account,"* or *"write to Google and ask why."* Some participants (8 in Delhi and 1 in Zurich) said they would sue the company if they felt their rights had been infringed. Others said they would not because their data is not that important, lawyers are expensive, or they don't have the time. Generally, participants did not know with whom to file a complaint: *"I don't know whom I should go to. [..] I don't think you can contact anybody."* Participants felt stronger about complaining in case of account disabling than in case of unauthorized data modification. If his account got disabled, D15 said: *"I will shut down my computer,"* because it must have been because of *"virus attack or some hijack."* No participant said they would go to police if their account had been hacked. Some said there is no possibility to complain, or that they do not know how to contact the company. Most were not aware of laws or agencies protecting their rights, but *"would like to have some laws so that I can complain."* D3 mentioned *"cyber client court"* and Z4 the *"postal police."* We followed up in the online survey. Figure 2.4 shows that 58% percent of all respondents agreed with the statement *"There is no such thing as consumer protection service or police on the Internet, whom I could turn to, if I felt that my rights were violated"* ($M=2.53$, $SD=0.98$, $N=279$). The Wilcoxon rank-sum test showed no statistical difference between Swiss and Indian participants.

2.5. Conclusions

In this chapter, we explored end-users' privacy expectations and assumptions for cloud storage, their awareness of risks, terms, and conditions. We conducted 36 in-depth interviews in Switzerland and India, and followed up with an online survey with 402 participants. Our results suggest that users make heavy use of free webmail accounts as cloud storage drives. However, instead of relying on the cloud as a main storage unit, users keep local backups of cloud-stored data. Study participants had a strong belief, fueled by media stories and hacker stereotypes, that the Internet is intrinsically insecure. The loss

of control over where their data is stored, and inability to physically protect it prevent them from storing sensitive data in the cloud. Our results suggest that users' mental model of cloud storage providers is very different from banks. Unlike money (people trust banks to protect their savings), personal documents are still perceived to be safer at home, regardless of how many security experts the cloud storage providers hire.

Unlike data stored locally, consumers accept that cloud-stored data might be viewed by other parties, such as hackers, cloud storage providers, or by law enforcement agencies. However, they believe that this privacy breach would only happen to famous people or criminals, not to them. Users don't read privacy policy and terms of service documents, and believe they have more rights and guarantees than what these actually grant them. For example, an alarmingly high percentage of users in our study were unaware that their storage provider reserves the right to modify user data and disable user accounts at any time. Our results suggest that consumers assume to have the same ownership rights over their data if stored in the cloud as if stored on their personal devices.

Clearly, there is a great mismatch between users' expectations of privacy and the actual rights and guarantees they enjoy for their data in the cloud. To foster business and cloud adoption and to protect consumers, regulation bodies and cloud storage companies alike should try to close this gap by meeting users' expectations and/or educating consumers on the risks they face. Possible measures to take include: (1) changing the content and the presentation of privacy policies and terms of service agreements to make it easier for users to read and understand; (2) offering better visibility into security settings by adopting stronger authentication mechanisms such as two-factor authentication, access log visualization, etc; and (3) accounting for internationalization. The later involves going beyond just translating the service interface and privacy policy. Companies should keep in mind that users from different countries may have different privacy expectations and understandings of privacy guarantees offered by the cloud storage system.

Our results show that cultural differences and local events influence users' expectation and perception of cloud storage privacy. Furthermore, our results imply that certain countries place a much greater emphasis over individual privacy, whereas others prioritize national

security over privacy—differences which companies and international cloud privacy bodies should keep in mind when designing global policies and services. For example, Swiss respondents were more aware of the lack of guarantees and stored less sensitive data in the cloud than Indians. While Indians considered government monitoring of users accounts to be a good thing because *"national security comes first,"* to Swiss government surveillance was a great violation of individual privacy. This is not surprising considering the two countries' political situations and cultural attitudes towards privacy. First, Switzerland is considered a safe haven of stability, whereas India is increasingly dealing with terrorist attacks and violence. Second, while privacy is deeply rooted into the Swiss culture, in India the social and family structures place little importance on privacy. Differences in perceptions of guarantees and privacy in the cloud suggest that the cloud storage policy and system level designers cannot expect one-size-fit-all solution that can accommodate different cultures.

Participants in our study were mostly young. Although young people are a major group target for consumer cloud storage systems, they are not representative of the entire world population. However, young people tend to be more technically-savvy than the general population, and likelier to use such cloud storage systems and understand how they work. The general population is, therefore, likely to have an even stronger mistrust in the cloud and a higher misunderstanding of the privacy guarantees it offers than our study participants. Future work could look into privacy attitudes and differences among higher age groups, and compare awareness of privacy policies among technical and non-technical users.

Furthermore, future work should explore consumer perception of international laws and regulation, as well as data protection authorities they could turn to. Finally, novel, usable mechanisms are needed to educate users and provide them with visibility and control over personal data in the cloud. In the next chapter, we we propose such a solution. Our system allows users to protect their data before storing it in the cloud.

3. For Some Eyes Only. Protecting Information Sharing in the Cloud

3.1. Introduction

Online sharing platforms enable a new communication and data-management paradigm in the cloud. Users disclose intimate thoughts on Facebook, blog about their political views, upload holiday pictures to Google Plus, and publish their current activities on Twitter. According to estimations by the social media blog 'The Social Skinny,' over one billion Facebook posts, 175 million Tweets, and 10 years worth of YouTube videos are being uploaded by users every day [128]. This rise in online sharing activity has prompted increasing privacy and security concerns among consumers [52]. By publishing their private information on a range of public or semipublic platforms, consumers get exposed to unauthorized disclosure of their data. Unauthorized access can happen, for instance, if hackers break into user accounts (e.g., 2009 Google cyber attack [66]), platform providers grant advertisement companies or governmental agencies access to user data without the user's consent [144], bugs in the access control enforcement system allows unauthorized online contacts to view user data [163], or platform providers might share user data with third parties for economical reasons [132].

Several solutions have been proposed to protect user data from unauthorized usage. One solution requires the user to encrypt the data before uploading it to the online sharing platform, and distribute encryption keys to authorized recipients only. For instance, PGP encryption allows users to protect email communication and attachment documents. However, this approach breaks the ease of sharing data on dedicated platforms such as online social networks, video or picture

sharing websites, which manage users' network of friends and support viewing the data in the browser.

Other solutions displace the trust users put in online platform providers by creating sharing systems owned and hosted by users. Such systems require users to run their own web server or sharing platform to host their data. For example, Diaspora [47] is a private social network that runs on servers owned and operated by the user. Such approaches, however, force the user to trade the usability of popular data sharing platforms for better privacy protection. This tradeoff might come with the expense of losing the interaction with potentially less privacy-concerned friends. Even apparently successful systems like Diaspora have a much lower user base than popular sharing systems used today. For instance, while Twitter has 140 million users [158] and Facebook over 900 million users [53], Diaspora's user base is only 1.8 million [40]. To view protected pictures, user's friends would instead of accessing Facebook, need to access the user's personal web server, or start using Diaspora or other similar services.

Instead, we desire a technical solution that allows users and their friends to have a similar experience on the platforms they normally use for sharing (e.g., Facebook), but protect against unauthorized usage of the data they store in the cloud. Some systems have been proposed that attempt to protect user data from unauthorized usage while still allowing consumers to use their platforms of choice. However, these solutions either require the existence of a trusted, third party server to handle user data and encryption keys, are platform specific (e.g., work for Facebook only), or do not hide the fact that confidential data is being exchanged.

In this chapter, we propose a system that works for web-based cloud storage platforms, does not require the user to run dedicated infrastructure or place trust in another third party, and hides from unauthorized recipients that confidential communications is taking place. In doing so, we are inspired by what Boyd calls *social steganography* [25]. Boyd found that teenagers sometimes post messages on Facebook that seem innocent to parents (e.g, song lyrics), but carry hidden meaning for friends. Similarly, our system allows users to post innocent looking pictures, files, or status updates that will transparently be replaced with real information for selected recipients in the user's network. Note that while *steganography* typically refers to concealing a message or file

within another message or file, our system hides a pointer to the protected data, not the data itself.

As a proof of concept, we implemented our system as a plugin for the Firefox browser. Despite the vastly different nature of websites, the underlying HTML elements used to construct user interfaces for uploading text input, pictures, and other documents are the same on all platforms. Our plugin is thus able to support virtually any web-based data-sharing platform with the help of platform-specific XML-based definition files that allow it to seamlessly replace dummy postings with hidden values.

In this chapter we make the following contributions. First, we propose a system for protecting data on online sharing platforms through strong user-side encryption. Seconds, we introduce a novel mechanism inspired from social steganography techniques to hide the fact that encrypted communication is being transmitted. Third, we demonstrate the feasibility of our approach through a proof of concept implementation in the form of a publicly available browser plugin.

The remainder of this Chapter is structured as follows. Section 3.2 gives an overview of the main idea, presents our threat model, and describes the goals of our system. Section 3.3 defines the components and protocol we use, and Section 3.4 presents the security analysis of our protocol. Then, Section 3.5 presents our implementation approach and performance evaluation. We discuss possible solutions to make our approach resistant to data-mining techniques for detecting protected messages in Section 3.6, review related work in Section 3.7, and conclude in Section 3.8.

3.2. Overview and Goals

Consider a user who wishes to upload a protected wall post, status update, or a picture to an online social networking site. While the user wants to take advantage of the communication channel offered by the platform, he also wants to ensure that only a specific set of authorized recipients can access it, keeping the platform provider and unauthorized parties oblivious. To this end, the user could just post the encrypted version of the content. However, some sharing platforms impose length limitation and are not able to display encrypted pictures. Thus, our

system stores the encrypted data on a different storage service and a different cleartext, fake data on the sharing platform. To enable authorized friends to retrieve the encrypted data, our system stores a pointer to its location on a different Internet mapping service.

More specifically, our system performs the following operations. At first, it replaces the user's real posting (i.e., text or a picture) on the website with fake data that looks like another genuine message—either automatically or with the user's help. The user's real data is encrypted for a user-defined set or recipients and stored in a user-selectable, arbitrary public storage service (e.g., Dropbox, or the user's own server), which returns a URL to the encrypted content. Next, in order to keep the storage location private the system applies a pseudo-random function to the posted fake data (i.e., a keyed hash) and computes a lookup-key. In a final step, this lookup-key is then used to store the (encrypted) URL of the encrypted file in another user-selectable, arbitrary URL lookup service (e.g., TinyURL). On the recipient side, the authorized recipients using the system while accessing the sharing platform, perform the reverse version of this process, in an automatic, transparent manner.

Threat model. We consider an attacker that has control over the communication channels used by the user to store and share data. However, we assume that the attacker does not know the secret keys of the users, and cannot control the user computing environments, such as their browsers and computers, and any device used in the protocol.

Goals. Our system targets information sharing through cloud storage systems that present a web interface. The system should support sharing most types of data on such online platforms. All content published by the user should be kept confidential by means of cryptographic techniques. However, for a good usability, the operations should be simple and the cryptographic techniques transparent. Only authorized recipients should be able to read and verify the integrity of the protected data. However, once a recipient gets access to the protected content, the recipient could redistribute it. Our main goal is to make it so that a casual observer is not able to infer when hidden information is being transmitted.

3.3. The System

To unlink innocent-looking data stored on the communication platform from the encrypted data covertly exchanged, our system makes use of the following main services:

Online Sharing Platform (*SP*) is any online communication platform for storing and sharing digital content (e.g., Facebook, Flickr, Gmail). Such platforms usually requires registered login, keep and manage the user's list of contacts, and might be often accessed by the user's friends.

Storage Service (*SS*) allows storing user data in the cloud and accessing this data through a browser (e.g., Dropbox, SugarSync). We assume that *SS* requires users to register before storing data. We assume that each file f stored in *SS* is accessible through a unique URL, which we denote url_f, and that anybody who knows url_f can retrieve the file without authenticating. Nevertheless, only the account owner can modify and delete stored data.

Hashmap Directory (*HD*) is a web-based service that stores short strings mapping (*index, value*) pairs, such as URL shortener services TinyURL or Bit.ly. Given an index, it allows anybody to retrieve the value. The service does not accept duplicate indices and places a restriction on the length of both the index and value strings (e.g., 30–140 characters). We assume that stored entries do not expire and cannot be deleted. We also assume that *HD* accepts any anonymous requests to store and retrieve entries, and places no limit on the number of entries a single user can make.

3.3.1. Transmitting a Protected Message

In our system, every user U owns a public/private key pair (pk_U, sk_U). Let's consider two users, Alice and Bob who want to exchange protected messages on the platform *SP*. We assume that Alice and Bob have exchanged and verified their public keys pk_A and pk_B. These keys will be used to encrypt m. We also assume that Alice and Bob have run a key agreement protocol and agreed on two shared keys (k_{AB}, hk_{AB}), where k_{AB} is a symmetric encryption key, and hk_{AB} is used to compute a pseudo-random function which we denote *PRF* [63]. We will discuss

key management in detail in Section 3.3.2.

Send Protected Text

Figure 3.1 shows in detail the messages exchanged by the involved parties. First, to ensure message integrity, Alice signs m with her private key sk_A and obtains $\sigma_m = Sign_{sk_A}(m)$. Then, Alice encrypts m, σ_m with Bob public key pk_B to obtain $c = E_{pk_B}(m, \sigma_m)$. We abbreviate encryption with public key pk_B as $E_{pk_B}(\cdot)$ and decryption using the private key sk_B as $D_{sk_A}(\cdot)$. Furthermore, we abbreviate symmetric encryption using the shared secret key k_{AB} as $ENC_{k_{AB}}(\cdot)$ and decryption as $DEC_{k_{AB}}(\cdot)$.

Next, Alice uploads the ciphertext c to SS. Subsequently, Alice notes the URL url_c under which c can be retrieved. Note that SS allows anyone to access c through the URL url_c without requiring authentication. Next, Alice chooses a dummy text d that looks like a genuine message she wants to transmit to Bob. Alice applies PRF to d using the shared key hk_{AB} and computes $h_d = PRF_{hk_{AB}}(d)$. Then she encrypts url_c using a symmetric algorithm to obtain $el_c = ENC_{k_{AB}}(url_c)$. We use symmetric encryption to compute el_c, which produces a ciphertext smaller than public key encryption, because we assume that HD poses a length limitation on the registered content. Finally, Alice registers el_c under the index h_d with HD and publishes the dummy text d on the online sharing platform SP. Since HD only accepts unique index h_d entries, d must not have been used by Alice to communicate with Bob before, under the same key hk_{AB}.

Read Protected Text

Figure 3.2 shows the steps needed to retrieve a protected message. Bob first reads the dummy text d published by Alice on the platform SP. He then tries to see if there is a hidden message m delivered with d. To verify this, Bob first computes $h_d = PRF_{hk_{AB}}(d)$ and then queries HD to see if there is any value registered under h_d. If no value is registered, Bob concludes that d is a genuine, plaintext message from Alice with no protected content behind it. Otherwise Bob receives the encrypted link el_c from HD and decrypts it to obtain $url_c = DEC_{k_{AB}}(el_c)$. Knowing url_c, Bob retrieves c from SS. Next, Bob decrypts c using his private

key sk_B to obtain the initial message $(m, \sigma_m) = D_{sk_B}(c)$. Knowing σ_m, Bob verifies the integrity of the message using pk_A. If the integrity verification fails, Bob concludes that Alice is not the actual sender or that an attacker tampered with c. Otherwise, he considers m a protected message transmitted by Alice.

Send Protected Image or File

Assume that instead of the text m, Alice wants to share a secret file. To share, for example, a secret image i, Alice follows the same protocol as for text but with a slight variation. First, Alice encrypts i with pk_B and stores the resulting $c = E_{pk_B}(m, \sigma_i)$ in SS. Next, Alice chooses a dummy image d. Because online sharing platforms often perform image processing techniques such as image compression on uploaded pictures, Alice and Bob cannot use a pseudorandom function on the dummy image; doing so might result in different h_d values on the sender and receiver side. For this reason, Alice chooses a random value w_d and hides it in the image d as a secret watermark using a secret derived from hk_{AB}. She then computes $h_d = PRF_{hk_{AB}}(w_d)$. Finally, as for text, Alice registers the encrypted link el_c under the index h_d with HD, and publishes the dummy message d on SP.

To receive the protected image i, Bob extracts the secret watermark w_d from the dummy image d, computes $h_d = PRF_{hk_{AB}}(w_d)$, and queries HD to retrieve the encrypted link el_c. He then decrypts el_c to obtain $url_c = D_{k_{AB}}(el_c)$ and retrieves c stored on SS at location url_c. Finally, Bob decrypts c using sk_B, obtains the initial image i and verifies the signature σ_i.

Group Communications

Assume that Alice wants to share m with a group of contacts G, not just with Bob. All the recipients in G know that Alice is the sender of the protected message based on information provided by the sharing platform, but they should not find out the identity of other recipients in G. In a straightforward solution, Alice could perform the steps described earlier for each recipient U in G, using shared keys (k_{AU}, hk_{AU}). However, this solution results in poor performance for large groups of recepients, because Alice must encrypt the message m for each recipi-

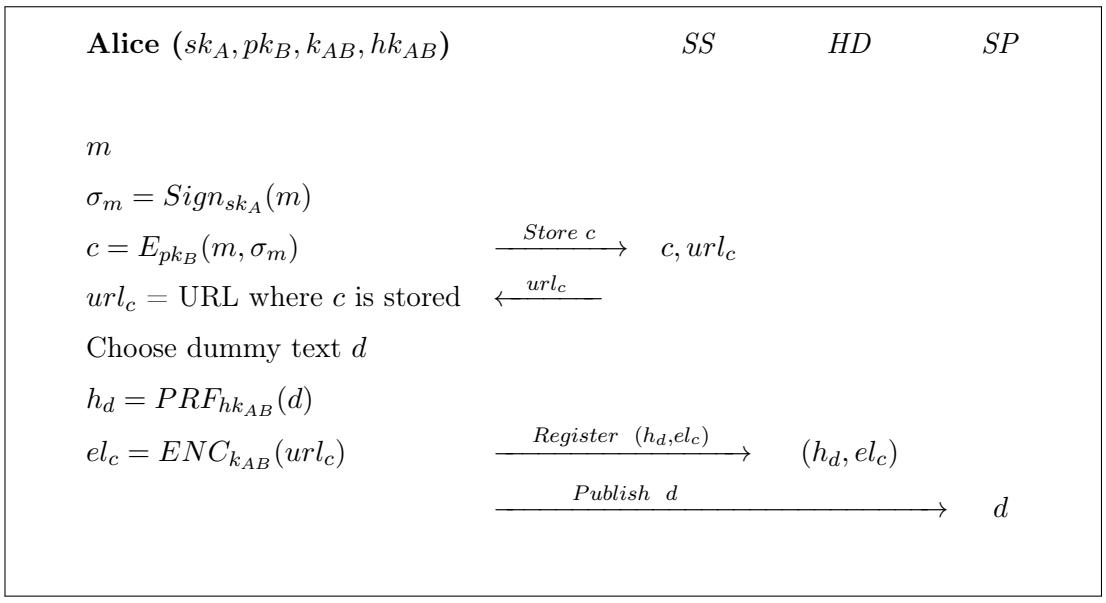


Figure 3.1.: Sending a protected message; (pk_B, sk_B) is Bob's public/private key pair, k_{AB} is the symmetric encryption key shared by Alice and Bob, and hk_{AB} the key for the pseudo-random function PRF.

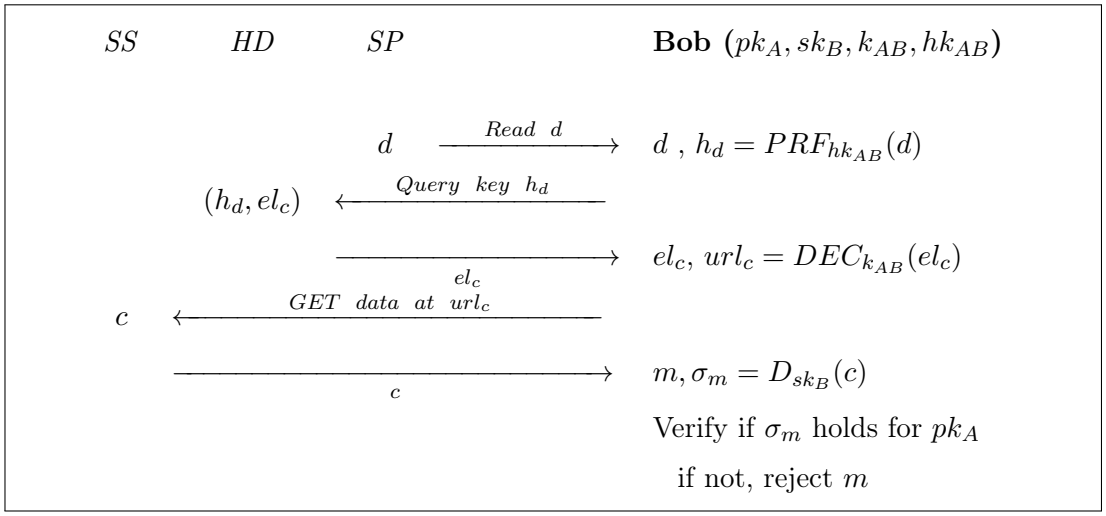


Figure 3.2.: Retrieving a protected message.

ent, and then compute and register different (h_d, el_c) values with *HD*. To obtain better performance, Alice can encrypt the ciphertext only once, using an anonymous broadcast encryption scheme [15, 103].

Depending on the desired trade-off between protocol security guarantees, on the one hand, and system performance and scalability, on the other, the shared keys (k_G, hk_G) used by Alice to transmit protected messages to the group G could be the same for all the contacts in a specific group, or different for each contact. Assume Alice shares different keys (k_{AU}, hk_{AU}) with each contact U. To publish protected content,

she must compute different hd_U and el_{c_U} values for each recipient U and register each (hd_U-el_{c_U}) pair with HD. Note that all el_{c_U} actually decrypt to the same url_c. Maintaining different keys per contact provides stronger security guarantees (e.g., in case some keys are compromised or leaked), but requires creating more HD entries, which ultimately affects performance and increases network traffic.

Access Right Revocation

Assume that Alice shared m with a group G of users that includes Bob. She now wants to remove Bob's access, but have the data available m to other recipients. Note that it might be impossible for Alice to remove or modify a message m already published on SP (e.g., emails that have already reached the recipients' inbox, some forum entries, etc). We distinguish between two cases, depending on whether Alice used different shared keys k_{AU} and hk_{AU} with each recipient U in the group G or group keys k_G and hk_G to transmit m. We assume that Alice stored the ciphertext c at the same location url_c for all recipients. A basic solution to revoke Bob's access to m has Alice simply re-encrypt m to the altered recipient list that excludes Bob, and update the ciphertext c stored on SS.Bob is still in possession of the shared keys k_{AB} and hk_{AB} or the group keys k_G. If he still has access to d (or knows d from a previous access), Bob can find out url_c and therefore retrieve c, by simply following the steps for retrieving a protected message under d. Therefore, Bob could still obtain c, and thus prove the existence of a protected message. However, since c is no longer encrypted with his public key pk_B, Bob cannot decrypt c and find out the content of m.

If Alice is not able to delete d or remove Bob's access from SP and the $(index, value)$ entries in HD are permanent, she must follow the following steps to impede Bob from proving the existence of m. First, Alice must delete c from SS, thus invalidating url_c and store the ciphertext at a new location url'_c. Furthermore, Alice must establish new keys (k'_{AU}, hk'_{AU}) or (k'_G, hk'_G) with the other recipients in G. She should then create new ($h_d = PRF_{hk'_G}(d)$, $el_c = ENC_{k_G}(url'_c)$) entries on HD the same d using the new keys. These entries should point to the new location url'_c. If, however, Alice encrypted and stored the ciphertext c in different locations url_{c_U} for each recipient U, Alice needs only to remove url_{c_B} since this operation does not affect other recipients in G.

3.3.2. Key Management

In this section we discuss the aspects related to key management. In particular, we have a look at possible approaches for contacts to perform key exchange and agreement, as well as key revocation. Furthermore, we discuss migrating private keys and contact lists across different personal devices.

Key Exchange and Agreement

Prior to being able to use the system and exchange protected messages, Alice and Bob must first exchange and verify their public keys pk_A and pk_B. Furthermore, Alice and Bob must agree on the shared keys (k_{AB}, hk_{AB}). Then, Alice adds Bob to her contact list and stores the keys $\{pk_B, k_{AB}, hk_{AB}\}$ locally on her machine. To perform key exchange and agreement, Alice and Bob must follow the following steps:

1. Exchange public keys. Alice and Bob can exchange pk_A and pk_B over a private channel such as email or by publishing them on a public or semi-public platform such as social networking sites. Publishing the keys over a public platform accessed by her contacts and automatically retrieving them from there offers better scalability than performing one-on-one exchange with each of her contacts. By using multiple *SP* platforms, one can separate the exchange of cryptographic material from the account used to transmit protected messages, thus hiding clues that encrypted information might be exchanged in the future. Note that this channel is considered untrusted and might be subject to man-in-the-middle attacks.

2. Verify public keys. Because a malicious attacker who might have mounted a man-in-the-middle attack during step 1, Alice and Bob must make use of a trusted out-of-band channel (e.g., QR codes, GSM network, Bluetooth communications) to verify the fingerprints of pk_A and pk_B. We consider this channel harder (e.g., the SMS network) or impossible (e.g., direct capturing of QR codes) to compromise. To hide the exchange of public keys from a suspicious observer, Alice and Bob could even resort to the out-of-band channel to perform step 1. If we operate under a honest-but-curious attacker, we might choose to skip or postpone this step to a later stage, for instance, until after the start of the protected communication. This is desirable if, for instance, Alice

and Bob did not yet have the chance to establish a secure channel over which to perform the key verification (e.g., meet in person, exchange phone numbers), but want to start communicating immediately.

3. Generate shared keys. Having exchanged and verified their public keys, Alice and Bob can run a key exchange or agreement protocol over the untrusted Internet channel to establish the shared keys k_{AB} and hk_{AB}. For example, the shared secret key agreement can be performed using the sigma protocol [94], an extended version of the authenticated Diffie-Hellman key agreement protocol. As long as the private keys sk_A and sk_B are not compromised, revocation of k_{AB} and hk_{AB} can take place by merely publishing signed (and possibly encrypted) messages over any agreed-upon Internet communication platform. For example, Alice and Bob could perform the sigma protocol by transmitting messages on Facebook signed with their private keys sk_A and respectively sk_B to ensure that no attacker interfered during the key agreement protocol.

Key Revocation

If the key k_{AB} gets compromised, an attacker could compute h_d and find out el_c. If, additionally, the attacker also knows hk_{AB}, he can then decrypt el_c to obtain url_c. Because from url_c he can obtain c, the attacker is able to prove the transmission of a protected message. However, as long as he does not know the private key sk_B, he cannot decrypt c to find out the the content of m.

If the shared keys k_{AB} and hk_{AB} get compromised, Alice must re-run the key agreement protocol in step 3 with Bob to obtain new (k'_{AB}, hk'_{AB}) values. If Alice's private key sk_A gets compromised, she must inform her contacts, re-run the key exchange and agreement protocol, and re-encrypt previous content with the new keys.

Key Migration

Assume Alice has generated her public/private key pair (sk_A, pk_A) on her personal laptop, but now wants to be able to view protected messages on her work desktop as well. To be able to decrypt protected messages from another device, Alice must have (pk_A, sk_A) and the shared secret keys (k_{AU}, hk_{AU}) available on all her device. Therefore, she must

securely migrate the secret keys to her work desktop. Note that it suffices for Alice to transfer her key pair (pk_A, sk_A) to the new device through an out-of-band channel, and then synchronize her encrypted contact list and shared keys between several devices by posting encrypted and signed messages in *SS* or another online storage platform. For weaker protection, but arguably more usability, instead of her public/private key pair, Alice could carry only a strong passphrase to the new device through a confidential out-of-band channel. The passphrase could then be used in a key derivation function (KDF) to generate a symmetric key. This key is then used to encrypt Alice's sk_A and the keys (k_{AU}, hk_{AU}) (e.g., using a symmetric authenticated encryption, such as AES in CCM-mode [170]), and then decrypt them on a new device.

3.4. Security Analysis

We analyze the resilience of our system against a number of attacks. We show that none of the services used by our system can find out the content of m or provide its existence, whether they work independently or collaborate. Furthermore, we show that an attacker cannot carry out impersonation attacks and discuss possible approaches to defend our system from traffic analysis attacks.

Sharing Platform. Because Alice published d on *SP*, *SP* knows d. However, *SP* does not know the key hk_{AB}. Therefore, it cannot compute $h_d = PRF_{hk_{AB}}(d)$. Consequently, *SP* cannot query *HD* to find out the value el_c registered under the index h_d. Hence, *SP* cannot find out m or find out whether there is a hidden message m behind d. However, *SP* can erase or alter d , which would result in a denial of service for the communication between Alice and Bob. Under suspicion, *SP* might be able to replace d with a previous message transmitted by Alice d', which could hide a protected message m'. In this case, Bob would verify m' as a message originating from Alice, but *SP* could not know or choose the value of m'.

Hashmap Directory. *HD* knows h_d and el_c, but it does not know the identity of the users who posted and access the entry. *HD* does not know k_{AB}, therefore it cannot decrypt el_c to obtain $url_c = DEC_{k_{AB}}(el_c)$, the location of the ciphertext c. Also, although *HD*

know h_d, it cannot extract the value of the fake message d (this follows immediately from the properties of a pseudo-random function, $h_d = PRF_{hk_{AB}}(d)$). Furthermore, *HD* cannot tell if h_d corresponds to any given d because it does not know the key hk_{AB} in the case of text or the key variable for extracting the secret watermark in the case of images. As a result, *HD* cannot identify m and c corresponding to (h_d, el_c).

Storage Service. *SS* has access to the encrypted data c, and possibly all other ciphertext stored by Alice. However, since *SS* does not know the private key sk_U of any authorized recipient, it cannot decrypt c to find out $m = D_{sk_U}(c)$. Furthermore, even though *SS* knows url_c, it cannot compute $el_c = DEC_{k_{AB}}(url_c)$ because it does not know k_{AB}. Given an entry el_c on *HD*, *SS* cannot verify if el_c decrypts url_c. Therefore, *SS* cannot find out if c is linked to a message d stored by Alice on another platform *SP*. However, *SS* can tamper with c, remove it or replace it with a different ciphertext c', previously generated and posted by Alice.

Collusion. We consider that in special circumstances *SP*, *HD*, and *SS* might collude and share user information among themselves or with other parties for profit or legal obligations. We show that, although any attacker with access to d stored on *SP* can tell that Alice and Bob are communicating, he cannot tell that there is a hidden message behind d. An attacker cannot link d published on *SP* to the ciphertext c stored in *SS* because he cannot compute h_d.

However, *SP*, *HD*, and *SS* might keep logs, record users' IP addresses and requests, which could be used by an attacker to match requests made by the same user. By matching IP addresses or the timing of the requests, an attacker could conclude that a message d on *SP*, a ciphertext c on *SS* and an entry (h_d, el_c) originate from the same user. Thus the attacker might be able to infer the existence of a protected message transmitted with d, though he cannot find out the content of m. We acknowledge that a highly motivated attacker (e.g., governments) might be able to link information across all protocol parties (*SP*/*SS*/*HD*). In practice, however, such attacks may be non-trivial for commercial and/or political reasons, as these parties may be competitors or located in different countries. To avoid such attacks, one could hide IP addresses by running our system on top of Tor [155] and defend against timing attacks by introducing random noise and delay

in user requests.

Impersonation attacks. An attacker with read and write access to all messages posted and received by Alice on *SP* (e.g., *SP* themselves, a hacker who got a hold of Alice's account, or a governmental agency who requests access from *SP*) could try to create fake messages d and convince Bob that a hidden message m actually comes from Alice. Since the attacker does not know the secret k_{AB} (which is only know by Alice and Bob), he cannot create a valid d, h_d pair. Given a d published by the attacker on Alice's behalf, Bob will query *HD* for h_d and conclude that there is no hidden message when nothing is returned. If, however, *HD* is malicious and colludes with the attacker, it could fake an entry by returning a chosen value el_c for any h_d submitted by Bob. However, since *HD* does not know k_{AB}, it cannot compute a valid el_c that decrypts to an url_c, but could mount a replay attack if in possession of a previous value el'_c posted by Alice for Bob. If *SS* also colludes and returns a given c' for any request el'_c originating from Bob, the attacker has the chance to deliver a chosen ciphertext c' to Bob. Knowing Bob's public key pk_B, the attacker could compute $E_{pk_B}(m')$. However, the attacker cannot trick Bob to believe this message comes from Alice because he cannot generate a valid signature $\sigma = Sign_{sk_A}(m')$.

3.5. Implementation

We implement our system as a Firefox plugin, basing our implementation on the Scramble! open source project [135]. Part of our plugin is implemented in Java, therefore the user must have Java Applet support enabled to run our plugin. We use AES-CCM for symmetric (authenticated) encryption, HMAC-SHA-256 as pseudorandom function, and the OpenPGP standard [27] for broadcast encryption. We make use of Dropbox as a Storage Service (*SS*) and TinyURL as the Hashmap Directory (*HD*). Ultimately, the user could select from a list of available storage platforms. We use TinyURL because it allows the user to choose a custom short URL to map to. TinyURL could be interchanged with similar URL shortening services, publishing services or online blogs that can store a public list of index-value pairs.

During the first installation for user U, the plugin creates a new Dropbox account with a random username; subsequently it generates an

OpenPGP public/private key pair (pk_U, sk_U) and the shared keys k_G and hk_G for his group of friends G. All encrypted user data is later stored in the `Public` folder of the Dropbox account and is accessible through a public URL.

3.5.1. Sharing Protected Text

The user invokes the plugin while the mouse cursor is inside the input area, e.g., by a mouse right click menu. The plugin extracts the message m typed in by the user in the input field and manipulates the HTML page to replace m with a chosen fake text d. In our implementation, the user must enter d. We discuss approaches on how to automatically generate good fake messages in Section 3.6.

Support for any text input fields. Websites are becoming richer and more complex, making use of complex Javascript calls and HTML code. As a result, text entry is no longer restricted to just a few HTML elements such as `<input type='text'>` and `<textarea>`. For example, in Gmail, the input area for composing email messages is in fact an editable `<html>` element within an `<iframe>`. Our plugin can handle special text input types. It identifies the HMTL node containing the entered user input through the `document.popupNode` Firefox API call. It then obtains the inserted text from its `.value` attribute or `.innerHTML`, depending on the HTML node type.

Support for rich text formatting. Web-based sharing platforms increasingly encourage users to edit and annotate documents, and write HTML rich emails and blog entries. As a consequence, separating user-generated content from the page source is becoming more challenging. Our plugin aims to protect user data without loss of website functionality. For example, the email reply together with the initial secret email are tightly coupled with Gmail's specific HTML webpage email header. When clicking the "Send" button, it is crucial to avoid reposting the initial secret message m (which is being displayed on the page, but is unknown to Gmail who only knows d). To achieve this, in our implementation the whole message thread, including the Gmail reply headers and the tags for rich HTML formatting are encrypted, and replaced with a new dummy text. idden text messages that need to be retrieved on a displayed page at any given time.

3.5.2. Sharing Protected Images

Unlike text input, which can be implemented through a variety of means, file upload in the browser takes place exclusively through an `<input type='file'>` HTML element. When a webpage is loaded, the plugin identifies all `file input` elements and registers `change` event listeners for all of them. Consequently, when the user selects a file to upload, the plugin gets notified first and prompts the user whether to protect the file. Note that this file protection mechanism can be applied to any file type. In our implementation, the plugin retrieves images from different Flickr pages, given a start URL and XPath-based webpage parsing and navigation rules. In a real setting, one might want to be careful about copyright issues.

For watermarking images, we use the DCT-watermark library [56]. Unlike steganography, good image watermarks are resistant to typical image compression, some cropping and scaling techniques. To ensure that only intended recipients can retrieve the watermark from the image, we use a secret derived from the encryption key k to embed and extract the watermark. We noticed that success rate is dependent on the used images. Based on our experiments, the DCT-watermark library needs on average two tries to successfully embed a 20-digit long watermark on a random Flickr picture (success rate 54%, N=595 pictures, t_{avg}=0.3s). We are not currently aware of other libraries who might perform better. However, a better chosen pool of pictures might lead to better success rates.

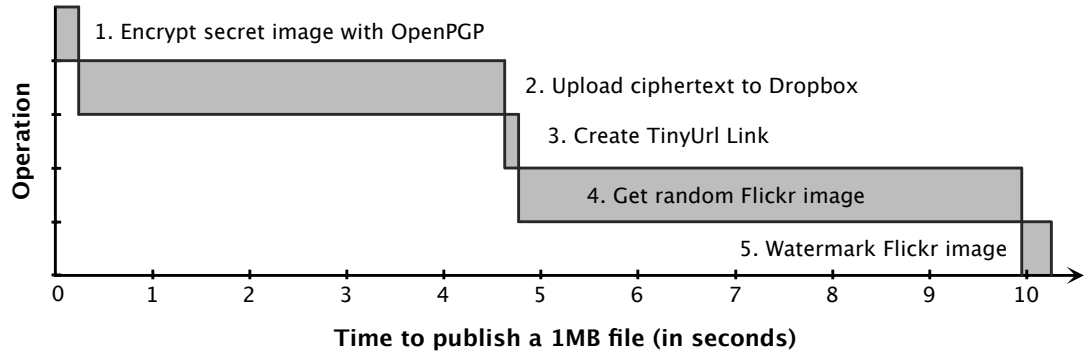


Figure 3.3.: Steps needed to publish a protected file. For optimization, steps could be run in parallel or be precomputed.

Finally, the plugin automatically updates the file selection in the `<input>` field to the watermarked image. For security reasons, websites and Javascript code, including Firefox plugins, are by default not

allowed to change the value of a file HTML `<input>` element. To this end, we sign our plugin with a trusted certificate, and request higher security privileges needed.

Unlike text, we display secret images through a pop-up window. For security reasons, images stored locally cannot be embedded in a webpage hosted remotely by simply manipulating the value of the `src` attribute on an HTML `<img>` tag (see the strict origin security policy [150]). This ultimately helps raise more user awareness on data protection levels. Alternatively, decrypted images could be hosted and retrieved from a local web server and displayed in line.

3.5.3. Extensible Page Parsing Rules

For our implementation to be online sharing platform independent, we make use of simple XML rules that define where and on which pages the browser should expect hidden data. Adding support for one more communication platform comes down to adding the XML specification files. To specify the page structure on a generic form, we make use of XPath queries [173]. XPath is a language used to navigate through elements and attributes in an XML document which uses path expressions to select nodes or node-sets. We use XPath queries to identify (sender, message) pairs on a page. Figure 3.4 shows an example. The `region` query is used to restrict the search on the page to a single section containing published messages. The execution of the `sender` and `data` subqueries is then restricted to the identified region. Next, the identified sender is matched against contacts from the address book, which can contain email addresses, nicknames and user IDs. To show the universal applicability of our solution, we defined parsing rules for any type of communication over Gmail, Facebook and Twitter. Such XPath-based rules need to be updated if web interfaces change. For the full specifications, see Appendix B.

3.5.4. Key Management and Distribution

There are two main approaches for end-users to distribute public keys: (1) rely on a mutually trusted certification authority, or (2) manually verify the authenticity by checking the fingerprints of the public keys through an out-of-band channel [89]. However, most everyday users do

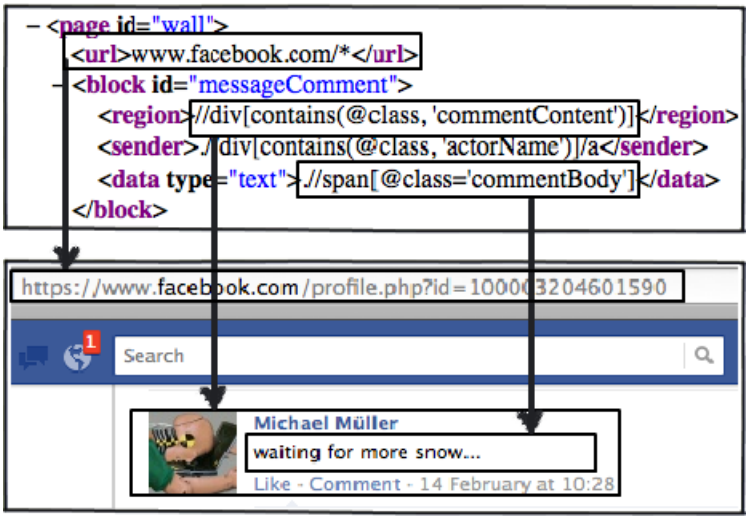


Figure 3.4.: The plugin identifies the dummy messages candidates based on webpage-specific XPath parsing rules.

not have mutually trusted certification authorities (CA) [89]. Furthermore, obtaining certificates from certification authorities is too difficult, expensive and time consuming. It takes even for power users 30 minutes to 4 hours to obtain a certificate from a public CA that performs little to no verification [71]. Therefore, our system perform key exchange and verification through trusted, out-of-band channels.

The plugin makes cryptographic operations, including key generation, management and distribution, transparent to users, thus avoiding the pitfalls of previous systems [171]. The plugin distributes public keys by publishing them on users' Facebook profiles embedded in a QR code image. Our plugin can be easily extended to distribute keys and run key agreement and cryptographic communication protocols over any other platform, by adding external JAR files that are automatically loaded at runtime.

To support key management and verification, we implemented an Android mobile application with two out-of-band key verification methods: SMS and phone-to-phone QR code scanning. The mobile application holds the user's public/private key pair (which are transferred from the computer through a QR code). Furthermore, the plugin can encrypt and sign verified contact keys and upload them to the Dropbox account from where they are synchronized with the browser plugin on other user devices. Finally, the power of the presented solution would not be complete if limited to PCs only. All the functionality of the plugin can be easily ported to a mobile platform, for example in the

form of stand-alone mobile applications for different platforms. (Unfortunately, plugin development for Firefox on the Android platform currently lacks Java Applet support.)

3.5.5. Performance Evaluation

A smooth user experience is essential in the success of any security solution; otherwise users will sacrifice security for usability. To calculate its performance, we run the plugin on a MacBook Pro laptop with an Intel Core i5 2.4GHz processor and 4GB of memory over a wireless network. The plugin has a memory consumption of 70MB. We measured the time needed to retrieve and display hidden messages on a Facebook page from the time the page is loaded in the browser. Note that only messages with senders in the contact list are candidates for protected communication. Processing a Facebook page with one hidden message (out of two candidates) took on average 0.9s, (N=10, SD=0.2s). Displaying a page with 10 hidden messages (out of 11 candidates) took 6s (N=10, SD=0.6s). On average, retrieving a hidden text message took 0.5s (N=25, SD=0.07s), and processing a message that does not hide any communication took 0.06s (N=25, SD=0.004s). Posting a hidden message took on average 0.67s (N=10, SD=0.1s). Therefore, two users talking over a protected chat message system would experience a delay of approx. 1 second. For our plugin, the time to display a page increases linearly with the number of hidden messages.

Figure 3.3 displays the time needed to execute each step of our implementation, in order to securely send a 1MB file to 100 contacts who share group shared keys k_G and hk_G. We present here only the extra security steps that must be preformed by our plugin, in comparison to the normal browser experience. The computation intensive tasks, file encryption (1) and image watermarking (5), take very little time compared to network operations, uploading the encrypted file to Dropbox (2) and retrieving a random image from Flickr (4). Note that given the OpenPGP symmetric type of encryption, the encrypted file has approximatively the same size as the initial file. Uploading an encrypted 1MB file to Dropbox took on average 4.4s (N=20, SD=0.6s). Figure 3.5 shows that the time needed to encrypt a file increases linearly with the number of contacts and file size, but remains relatively low, below 2 seconds for a 100MB file and 500 contacts. Finding and saving a Flickr

image took on average 5.2s, of which 3.8s were needed to download and parse the starting webpage. Once the URL was identified, saving an image locally took on average only 0.9s (N=50). Creating a TinyURL mapping the secret watermark to the encrypted Dropbox link took only 0.1s (N=20, SD=0.01s).

While executing all steps sequentially could account for slow browser response time and ultimately poor usability, implementation optimizations can make the process seem instantaneous. For example, a pool of Flickr pictures could be retrieved and stored locally beforehand. Since for files and images the TinyURL index h_d is not a secret derived from a dummy text chosen by the user, but rather from a randomly generated string, even image watermarking could be pre-executed. Similarly, uploading the encrypted file could happen in parallel to other operations and finish after the upload of the watermarked image.

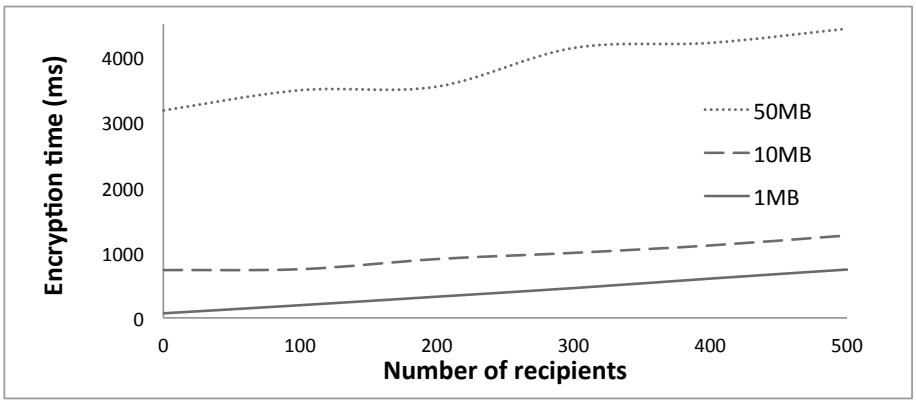


Figure 3.5.: Encryption time increases slowly, remaining bellow one second for a 1MB file shared with 500 contacts.

3.6. Semantics and Mining Attacks

Suspicion of using our system could cause trouble to users in countries with totalitarian regimes, or simply refusal of service from platform providers with business models exclusively based on targeted advertising. It might, therefore, be desirable that no single or colluding services can tell which users communicate using protected messages. As discussed in Section 3.3, our solution provides complete data confidentiality against any attacker having access to one or all of SP, HD, SS. However, as data mining and profiling techniques are becoming more advanced and possibly even used by oppressive governments to identify

activists, it is crucial to ensure that users cannot be singled out based on semantic analysis of dummy messages and steganographic data. In this section we make a few initial considerations on how dummy messages could be automatically generated.

It should be hard or impossible for an outside observer (e.g., platform provider, governmental agency) to tell with high probability, through mass-targeted or user-targeted mining, that users are protecting their communication. Automatically generated messages must, therefore, be consistent with past user behavior. Ultimately, to ensure less detectability, users can compose the dummy messages themselves. For good usability, however, the plugin should make a good suggestion. Furthermore, it is likely that users are not good at coming up with diverse dummy messages either.

Resistance to machine detection. Previous work on detecting automated posts on Twitter and social networks mainly focused on spam and used simple detection techniques that would not work against our solution. For example, Benvenuto et al. [20] use behavior attributes, such as average hash tags per tweet, number of tweets received, account age, number of followers per number of followees, and fraction of tweets with URLs. Zhang et al. [174] analyze timestamps and observe that humans post messages at random times of the day, whereas bots post at specific minutes of the hour. Since in our system the user would always be the one initiating the posting, such techniques would not be effective to identify dummy messages. Automatic publishing of noise messages, however, should take into account such considerations.

Consistent user behavior. Constantinides et al. [32] have shown that user behavior on social networks follows well defined trends. The authors found that user profiles cluster into four main categories, depending on their usage patterns on Facebook: *Beginner*, *Habitual*, *Outstanding*, and *Expert*. The authors then quantify the likelihood for a certain type of user to engage in a certain activity. For instance, searching for people online, sending private messages and updating profiles are popular among all users, while reporting about products used and commenting about advertising are mostly done by *Expert* users. A plugin user might be easily identified if, for instance, all the public posts on his Facebook wall are about sharing current activities, while all the steganographic messages are TinyURL links (e.g., Scramble! [18]) or sentences from Wikipedia (e.g., FaceCloak [106]).

A solution based on topic models. To make sure dummy text messages are consistent with the previous topics in users' communications, one could use text document analysis techniques, such as constructing and applying topic models [23]. In particular, we propose using the Latent Dirichlet Allocation (LDA), a generative model in which any specific document is viewed as a mixture of topics. Each topic is characterized by a distribution of words. LDAs can be used to learn and define topics from an existing set of documents or communication. Other models consider correlation [22] and hierarchy between topics [29]. Previous work already applied topic models to social networks [111] and images [166].

Possible implementation. To generate valid dummy text messages, one might want to restrict the social network activities identified by Constantinides et al. [32] to those that can be used to transmit a social steganographic message by our plugin: discuss what people do, communicate news or issues, share mood, share links about interesting web sites, report about current activities, and report about brands or products. Based on such predefined types of communication and their expected frequencies, a heuristic could be defined that, for each posting attempt, takes as input a pseudorandom number and determines what kind of dummy text message to generate. In addition, to disguise the usage of steganographic messages the plugin could insert noise communication.

The Machine Learning for Language Toolkit [108] a Java-based software for document classification and topic modeling could be easily integrated with our plugin and used to analyze past user communication. Dummy messages consistent with identified topics could be generated through different means. For instance, the set of keywords in a predicted topic could be used to retrieve sentences from the web through a Google search. Steganographic replies could be generated with the use of online Turing test chats (e.g., Touring Hub [156]), or aggregated among several users and outsourced as tasks to human workers on Mechanical Turk [118].

Good usability. As described above, generating semantically correct sentences, which must be consistent topic-wise with user profiles, is not a trivial task for a computer. The problem becomes even more challenging on long steganographic communication threads, which might be visible to other users, not just to the intended recipients (e.g., pub-

lic posts on Facebook). Furthermore, the existence of steganographic messages poses usability challenges. More research is needed to find out how well users cope with friends replying to fake posts. Finally, adequate interfaces should help users keep track of the two worlds: the steganographic messages and the hidden message thread.

3.7. Related Work

Some solutions have been proposed to increase privacy on different online sharing platforms like exclusively on social networks or webmail platforms. Existing solutions either (1) do not hide that communication is confidential [18, 12], (2) protect only certain kind of data (e.g., profile information or private messages) [106], (3) require the existence of dedicated infrastructure or a trusted third-party, or (4) are restricted to a specific platform. Other solutions propose novel, privacy-friendly architectures meant to replace existing platforms [1, 34, 35, 39, 84]. Instead, our solution enables consumers to keep using current system while protecting their privacy.

Many research solutions and commercial products have been proposed to encrypt messages and files exchanged over webmail platforms [18, 36, 42, 55, 105, 107]. However, these solutions do not hide that communications are encrypted. Scramble! [18] uses cryptographic mechanisms to enforce access control rules and ensure confidentiality of sensitive information. Our plugin inherits the concept of using OpenPGP for group communication from Scramble!. It can replace encrypted text with a TinyURL link that points to a server storing the ciphertext, but the link and it's content is public. Furthermore, Scramble! does not offer any support for files, which is a central contribution of our system.

Pashalidis et al. [124] protect users' privacy by making messages, tweets or short emails difficult to parse for machines. Their system replaces text users post online with pictures containing the distorted text, similar to CAPTCHA systems. However, the content of the message can still be read by unauthorized human recipients. Furthermore, their system limits the size of messages that can be shared, alters the user experience by making the text more difficult to read, and does not deal with document exchange of any kind.

In the context of social networks, Conti et al. [33] propose Virtual Private Social Networks to protect some static user profile information on social networks: name, picture, and current city. The authors do not consider messages and files. The user posts fake information online and distributes his real data unencrypted in an XML file to his friends over email. This data is then matched using regular expressions and automatically replaced by the browser, similar to how our plugin offers extensibility through XPath queries. Yeung et al. [11] also take a decentralized approach. Each user has a trusted server which stores his data, has knowledge of social network specific functionality and applications (e.g., photo tagging, personal wall), and enforces access control rules based on cross-platform specifications. When trying to access data on different platforms, the user's friends are redirected to the trusted server which must handle the access control rules.

FaceCloak [106] is similar to our approach, but limited to Facebook profile data and messages. The secret information is encrypted and stored on a dedicated server, while fake information (e.g., random sentences from Wikipedia) is posted on Facebook. Just like FaceCloak, our system substitutes real information with dummy one. In FaceCloak, the fake information and a key that only the sender and all his contacts know serve to compute the index under which the third-party server stores the ciphertext. In contrast, our solution works solely with already available services on the Internet and does not require dedicated infrastructure. In addition, our scheme supports per-group communications and file exchange on any platform. Furthermore, by separating the storage service from the Hashmap Directory, we protect against fake data creation in case steganography keys are leaked. This can happen if a malicious user leaks the shared group steganography keys, but the encryption keys of other users (i.e., their private keys) remain uncompromised. If one has control over the FaceCloak server and access to the user's index key, he could swap ciphertexts and create successful plaintext swapping.

StegoWeb [21] is implemented as a browser bookmarklet, i.e., as a simple program that can be executed by clicking on a bookmark in the browser. The user must rely on a trusted third-party server to perform the encryption and data steganalysis. StegoWeb does not use public key cryptography. Instead, for each piece of data shared with each recipient, both the sender and the receiver must enter a shared

passphrase. By supporting public key encryption, our solution offers more security and better scalability. Oren and Wool [123] propose a system which increases webmail privacy by hiding the email content with text steganography and then splitting the output in two parts. The user must send the two parts over two different email accounts. This solution requires no key distribution, but protects only against a weak attacker who has access to one of the two webmail servers.

SecreTwit [136] uses text and image steganography to transmit secret Twitter messages (e.g., by appending whitespaces at the end of the tweet). The size of messages that can be transmitted purely through steganography is limited, because the size of the hidden data must be much smaller than that of the carrier message. Therefore, an approach purely based on steganography causes a serious limitation on the type and the volume of data users can transmit. For example, picture sharing platforms often compress and resize images, which could not hide a high-quality picture. By transmitting only a pointer to the location of the secret data instead of the data itself, we pose no limitation on the size and type of protected information. For text, we provide the user with complete freedom on how to compose the dummy messages, thus making it less likely to being identified as unusual communication. Furthermore, our approach is robust against basic image manipulation techniques applied by online sharing platforms.

Adkinson-Orellana et al. [5] encrypt documents stored in Google Docs and enable simultaneous editing of encrypted documents among a group of people by intercepting the HTTP requests for AJAX calls and encrypting/decrypting transmitted document content. This approach requires knowledge of a platform-specific AJAX protocol and does not hide the nature of encrypted communication. However, by dealing with document editing, this work is orthogonal to ours and could be extended to provide a viable solution for group editing of protected data in our system.

3.8. Conclusions

While users are lured into storing their personal data in the cloud and sharing it with others over an ever wider range of free services (e.g., webmail, social networks, photo sharing platforms), they have little

control over what third parties can access their data (e.g., hackers, advertisement companies, governmental agencies). In this chapter, we proposed a system that allows users to protect data they share online, but also hide the fact that confidential information is being exchanged from unauthorized recipients. Our system does not rely on dedicated infrastructure or trusted servers. While the secret data is encrypted and stored in the cloud, dummy data that looks like genuine files or user communication is uploaded on the sharing platform. Users share secret keys which they use to discover the hidden data behind the dummy one. They find out the encrypted location of the secret data through a public indexing service like TinyURL. Our system accounts for easy specification of location of expected hidden messages on HTML paged through simple XML rules based on XPath parsing queries. These rules need to be updated if websites interfaces change drastically. We provide a proof of concept implementation of our system in the form of a Firefox plugin that focusses on protecting text messages and images.

Our solution does not hide the exchange of communication between two parties, thus leaving such transactional data open to law enforcement agencies. However, an attacker who does not have access to users' secret keys cannot detect the exchange of confidential communication. Our system preserves most website functionality including text and image display. Nevertheless, because data is encrypted and not actually stored on the online sharing platform, our solution does cause a loss of functionality on certain types of systems, e.g., Google Spreadsheets. For most online sharing platform functionality though, techniques such as encrypted search promise to be a viable solution [165]. Further research should investigate techniques to impede websites from sniffing the data while being entered by the user in the browser. One possible solution is to have the plugin disable Javascript which the user types in the message. Furthermore, a future study should evaluate the detectability of image watermarks in different watermarking algorithms. If an attacker can identify users who post watermarked pictures, he might be able to narrow down the consumers who user our system.

While our proof-of-concept implementation deals only with plaintext and image exchange, our solution can be similarly used to implement protected exchange of any data type, including video files and data documents. Further work could implement sharing protected video files through platforms like YouTube. Based on research by Boyd [25], we

believe users can cope well with the usage of steganographic messages. However, further research is needed to test the usability of our plugin and devise adequate user interfaces to help them distinguish between regular messages and protected communication. Most importantly, further research should look into having users specify recipient rules for each website or page. The browser should learn users' preferences in terms of when, which data should be protected for which recipients, and automatically apply those data protection policies. Dummy pictures used for watermarking could come from a local folder with users' personal pictures, be automatic repostings of Facebook pictures in which the user was tagged from his friends' profiles, or general photos from public websites (e.g., search queries on Flickr, Google Image Search), possibly corrupted to be harder to match against originals. Finally, future work should investigate techniques to generate sound dummy messages and data.

In the following chapter, we address the usability of key exchange methods in our system. In particular, we conduct a comparative analysis of device pairing protocols, which can be used to exchange and verify public keys in our system, during causal user encounters. To this end, we give special consideration to eliciting users' mental models of proposed security protocols, their preference for certain methods in different real-life situations, and the influence of social factors on their decisions. Instead of focussing solely on key exchange scenarios, we broaden the scope of our study to gain a better understanding of user behavior, which could be also applied to other applications involving secure device pairing protocols as well.

4. Influence of User Perception, Security Needs, and Social Factors on Device Pairing Method Choices

4.1. Introduction

With the increasing proliferation of mobile devices–mobile phones, PDAs, netbooks, and tablet PCs—the need to spontaneously connect two devices over a wireless link has become prominent. Apart from exchanging business cards and appointments, spontaneous wireless links can be used to send files to Bluetooth-enabled printers and to make electronic payments in busses, train stations, and coffee shops. To authenticate spontaneous wireless device communication, several secure device pairing protocols that allow device authentication in the absence of a centralized security infrastructure have been proposed.

With no wires to verify actual connection, users cannot be sure what device they connected their wireless link to. The basic approach of spontaneous pairing protocols is thus the use of an "out-of-band" channel, i.e., a secondary information channel that can be used to verify the authenticity of the primary wireless link. An example for such an out-of-band channel is the popular Bluetooth pairing method of displaying a number on one device, and having the user enter it on the other [24]. Consequently, the usability of such methods is of crucial importance, as complex mechanisms might raise the probability of human error, might prompt users to choose a lower security level, or lead them to abandon security altogether.

Existing usability studies that tried to compare the various pairing methods proposed so far [88, 93, 96, 97, 98] mostly focussed on covering a high number of protocols and protocol variants. Consequently, prior

work rarely investigated the use of such methods in real-life situations, but instead used a single generic task to determine the best method overall regardless of the purpose of connecting those devices and the physical and social situation. Furthermore, prior studies predominantly recruited male study participants—mostly university students and researchers, often with technical backgrounds. Last but not least, the higher number of investigated methods implied a significant cognitive load on participants, sometimes resulting even in 30 to 50 individual pairing tasks that each participant had to go through in a single session [88, 98].

Instead, we decided to conduct a more explorative study, in order to determine the usability of proposed pairing methods *in specific situations* and to elicit the needs and the underlying mental models of users with respect to their security considerations in device pairing scenarios. We explicitly recruited participants with diverse, non-technical backgrounds. To limit their cognitive load, we restricted the study to four carefully chosen device pairing protocols, which spanned a wide range of channels (visual, audio, tactile) and required different degrees of user involvement (from completely passive to fully active). After having learned those four methods, participants were asked to choose among them in the context of three distinct pairing tasks, each one with a different real-world situation as a motivation. Last but not least, while previous studies only investigated device authentication, we also incorporate device *identification* into the pairing process, because choosing the device to connect to is often the most frequent and time consuming part of the process.

Our results show that device pairing methods are more than just the means to connect two devices: devices and methods used represent *people*, may make owners seem more professional (e.g., in a newly established business relationship), provide a playful moment between friends, or even act as an "ice breaker" when meeting someone new. The proper pairing method can reassure device owners that they handled a payment transaction well, or that they acted responsibly with customer data. Even more, methods evoke strong emotions: they are "annoying," "drive crazy," and even make users "fall in love" with them.

4.2. Related Work

The last few years saw a number of comprehensive studies that evaluated many of the hitherto proposed device pairing methods [88, 93, 96, 97, 98]. Our work differs from these studies in three important points: (1) we explored user preferences not in terms of pure pairing speed, but by investigating particular situations and their corresponding social factors; (2) we reduced mental load on participants by testing only four representative pairing methods; (3) we recruited participants with diverse, non-technical backgrounds and aimed for a more balanced gender composition.

Early comparative usability studies such as Suomalainen et al. [151], Valkonen et al. [162], and Uzun et al. [161] involved only simple methods based on string and number entry or comparison. The main emphasis was on measuring completion time and determining the error rate of methods. Qualitative data was not gathered, and the task given to participants was a generic pairing task that did not model any real-world situations.

In one of the most comprehensive studies, Kumar et al. [96] tested 14 variations of 8 basic methods, resulting in almost 50 individual test cases that each participant had to perform. Participants were mostly "technology-savvy" university students, with 70% male participants. While the authors argued that "if highly-motivated and technology-savvy young people do not react well to a given method, the same method will perform a lot worse with average users," our results suggest that non-technical participants do like newer methods, which performed less well in their study. Perhaps non-technical users are more excited about "what technology can do" or perhaps methods shunned by the technology-savvy fit better into their mental model for how security is provided.

In "Serial hook-ups" by Kobsa et al. [93], participants were told to imagine that they had just bought a new phone and when they return home they want to pair it to the old one. Study participants had to try 11 diverse methods, based on video, audio channel, button presses, comparison based. The authors proposed three "best" methods, based on the availability of displays: PIN-comparison or image-comparison for devices with a display, and (automatic and semi-automatic) audio-

based comparison for devices without a display [93]. The study does not give insight into *why* users thought that a particular method would be more secure than another.

Kainda et al. [88] tested 14 methods, but placed a stronger emphasis on the trade-off between usability of a method and its susceptibility to security failures. While users also preferred numeric comparison methods for their usability, the authors point out that numeric PIN entry, which requires the user to enter a number displayed on the screen into the partner device, is much less prone to accidentally confirming non-matching numbers and thus should be preferred, even if it ranks lower. The study did provide participants with a scenario—making an electronic payment to another device—yet it did not explore how this influenced the participant's choice of method. According to the usability rating used, the method Barcode—which involves using the phone's built-in camera to make a photo of a barcode displayed on the other device—was classified as unusable. It is unclear whether this was simply a result of the low reliability of the employed barcode decoder. In our study, the barcode-based method was considered more secure than other methods and were thus relatively popular in payment scenarios.

A technical report by Kumar et al. [98] specifically explores scenarios involving two users. Their results show that people are unwilling to hand over their phone to strangers. This work confirms our belief that pairing methods must be explored in more realistic social settings.

All of the studies discussed above have only focused on *authenticating* the connection. They do not consider the additional step of *device identification*, i.e., pairing in the presence of other (potentially pairable) devices. Having to make a choice between several available device significantly affects the pairing process—both in terms of time spent and the perceived security of the process. Our study incorporates both identification and authentication methods.

Rashind and Quigley [129] compared five methods for device identification: shaking or bumping devices, simultaneous button pressing, "stitching," and touching both devices at the same time. Even though the focus of their study had not been on security, users raised privacy concerns and worried about the risks of undesired intrusions. The authors used storyboards to show participants different usage scenarios

and found that both the purpose of pairing and the social context were important to users when choosing a method. This is very much in line with our own findings, though Rashind and Quigley did not explore the actual impact of these factors, nor users' perception on the security level of a method.

4.3. Methodology

Designing proper usability studies that ensure a fair and comprehensive comparison of device pairing methods is a challenging task. First, the designer has to consider a large number of methods that have been proposed, the situations in which they apply and the type of devices they were intended for. The mental load on the participants should be considered; researchers have to pay careful consideration to set the number of methods and options such that the user can learn and evaluate them appropriately. For this reason, we restricted the number of test methods to four methods that span a wide range of auxiliary channels and interaction models. Second, there are currently no consistent implementations for all these methods. Software development frameworks for mobile devices are still far behind those for desktop systems; many methods require special libraries that are not robust or freely available (e.g., barcode decoder). Finally, the nature of wireless communications makes device pairing techniques intrinsically different from standard Internet security solutions and therefore hard to grasp even for technical people.

In the study, the designer must place the protocols in some context, but keep it simple enough for users to understand; equally, the users should not be trained more than they would be in real life. We adhere to these principles in our study. We explore which security levels users keep in given situations, when they are willing to use security, and how much time and effort they want to spend on pairing. We therefore designed each of the four methods—*Select the device* with PIN entry, *Take a picture*, *Listen up*, and *Push the button*—to run under three security levels: *Not secure*, *Secure* and *Very secure*. The *Not secure* level is equivalent to running only device discovery or device identification, without authenticating the device nor securing the communication. The *Secure* and *Very secure* levels imply both identification and authentication and differ in the amount of information

transmitted over the auxiliary channel. Each level builds on the previous one, takes slightly longer time to complete, and most often requires an increased user effort.

4.3.1. Selected Methods

The four methods offer different automation degrees by involving the user to varying extents in the connection process. The methods also span a wide range of channels (visual, audio, tactile), and degrees of user involvement (from completely passive to very active). In all considered methods two communication channels are used: a primary communication channel and an auxiliary (out of band) channel. Here, the auxiliary channel is an authentic (typically low-throughput) channel that allows the exchange of Short Authenticated Strings between the devices. The methods primarily differ in the way they implement auxiliary channels. Their security depends on the size of the authentication string [60, 61, 102]; it has been shown [164] that, given appropriate protocol constructs, the use of short (20 bit) strings is sufficient to provide strong security guarantees. In the following we describe each of the methods and how they implement the different security levels.

Select the device is based on the Bluetooth Simple Device Pairing protocol [24], and entails device selection from a list and PIN entry; this method is a de facto standard for device pairing today. If the user selects an incorrect device he will connect to an unintended party and if he types in the wrong PIN the connection will fail. For the *Not secure* level, the user's device searches during four seconds for available devices and displays the list. The user chooses the name of the device to which he wants to connect from the list. For the level *Secure*, the user's device additionally displays a 6 digit PIN (equivalent to 20 bits of data) and for *Very secure*, a 9 digit PIN. The user types in the PIN into the other device. This method differs from the others because switching to a secure level involves adding a new kind of interaction (typing in the PIN vs. selecting from a list). For the other three methods, the interaction type remains the same, but the completion time and number of (repetitive) tasks the user has to perform increases as the security levels increase.

Take a picture is based on Seeing-is-Believing by McCune et al. [112], namely using the phone camera to take a picture of a barcode displayed

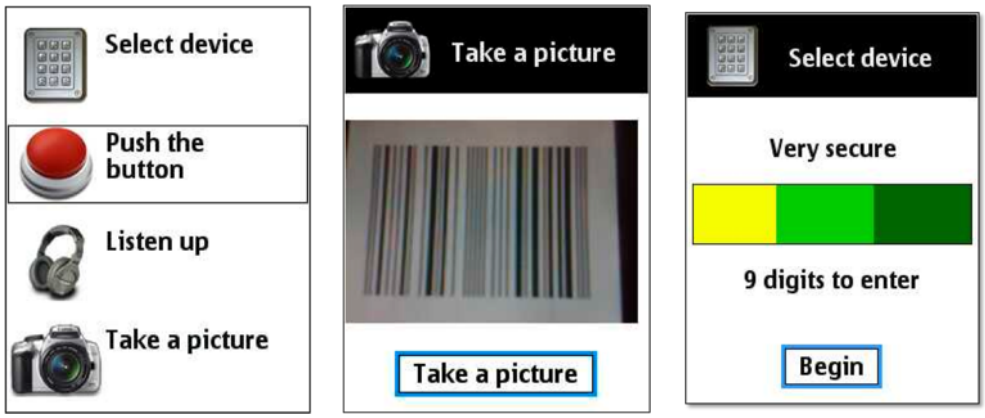
	Not Secure	Secure	Very Secure	Units
Select device	0	6	9	digits of PIN
Take picture	1	2	3	barcode pics
Listen up	3	6	9	seconds of melody
Push button	3	6	9	button presses

Table 4.1.: Each of the four methods has three security levels, which correspond to different completion times and degrees of user involvement.

by the partner device. Assuming the user does not accidentally take a picture of another barcode, and that the barcode is successfully recognized, connecting to an unintended party is not possible. For the *Not secure* level, the barcode contains the 48 byte Bluetooth MAC address of the device. Identifying devices through barcode pictures is a well established procedure, used even in systems for physical access control [16]. For *Secure*, the user must take a picture of an additional barcode displayed by the device and for *Very Secure* the user takes three barcode pictures. The additional barcodes encode the authentication string. The security guarantees of these method vary depending on the encoding capacity of the barcode.

Listen up is based on Loud-and-Clear [64] and the newer HAPADEP [146], and uses the audio channel for data transmission. It has the highest degree of automation among all the methods considered in our study, and places very little strain on the user. For *Not secure* the partner device plays a 3 seconds melody, which encodes its Bluetooth MAC. The user's device records, decodes and extracts the MAC, and establishes the connection. It is hard to estimate how many bits could be encoded in a 3 seconds audio transmission, but even in the very likely case in which the entire MAC address does not fit, transmitting the first or last 12 bytes and then matching these against the devices discovered or supplementing it with wireless messages would still provide a reliable enough implementation. For *Secure* and *Very secure* the melody lasts 6, respectively 9 seconds. The additional seconds are used to transfer authentication strings.

Push the button is inspired from Button Enabled Device Authentication by Soriente [145]. The user's device makes short vibrations. With every vibration, the user pushes a button on the other device. The *Not*



(a) Choose method (b) Take a picture (c) Security level

Figure 4.1.: Application screenshots. (a) choosing a method, (b) taking a picture of the 1D barcode displayed by the other device (c) choosing the *Very secure* level for the *Select the device* method, which entails entering a 9 digit PIN.

secure level requires 3 button presses and with each of them, messages are being broadcasted, either by both devices or by one. Received packages are matched against the time intervals at which the button was pressed. If several connections are being established at the same time in the same place, interference might occur. In general though, we expect the method to be reliable enough, similar to the Bump application for exchanging phone number through a central server [26]. Currently, Bluetooth does not support message broadcast, but it is reasonable to assume that in the near future wireless spontaneous communications will be broadcast enabled (e.g., through the upcoming Wi-Fi Direct standard [172]). An alternative would be a WLAN infrastructure to which both devices are connected. Similar interaction concepts were proposed in SyncTab [130], Network-in-a-box [13], WiFi Setup, and are available in several products on the market. For the *Secure* level the user must perform 6 button presses and for *Very secure*, 9. The additional presses are used for transmitting the authentication string, which could provide 9 and 18 bits of entropy (if we assume that the interval between two presses can be used to transmit 3 bits, like in the original paper [145]). This method requires increased user attention and is time consuming, due to the low information entropy of the channel. Table 4.1 summarizes the options for security levels for each method.

We implemented mock-ups of the four chosen protocols in Python for

Symbian S60. Instead of a 2D barcode we used a 1D barcode and the BaToo decoding library [4]. For *Listen up* we took an audio file sample from the original HAPADEP implementation. For Push the button we allowed a 500ms user reaction time (the time the user has to push the button on the other device once the first one vibrates), higher than the 300ms proposed by the original authors, to minimize failure rates. The two devices used in the study were a Nokia N95 as the user's device and a Nokia N96 phone as the partner device. Figure 4.1(a) shows the application screen for choosing one of the four methods, 4.1(b) taking a picture of the 1D barcode and 4.1(c) setting the *Very secure* level for the Select the device method.

4.3.2. Tasks

During the study, participants were given three hypothetical situations and asked which method (and which security level) they would choose. The moderator read the task description from the study script. In the following, we present the task descriptions the participants received.

Task 1: Print a document. *Imagine that you work for a consulting company. You are at the airport and will soon board the plane. You will fly to London to visit your client. You have saved your client's confidential financial report on your mobile phone. In the waiting area there is a printer. Connect your mobile phone to the printer, so you can send the financial report wirelessly to the printer. Pretend the display of the other device represents the printer's display.*

The goal of this task was to see how critical users perceived the document and how they perceived the security threat. We asked them questions to understand which criteria users used, and whether the nature of the environment influenced their choice.

Task 2: Make a payment. *In London, you will also visit a good friend. Before you board the plane, you want to buy him a bottle of whisky in the duty-free shop. You hear the announcement that your flight's boarding has just begun. Connect your mobile phone to the payment terminal to pay for the bottle.*

Through this task we tried to evaluate whether users perceive a higher security threat when paying compared to printing, the effect of the time pressure on their choice and whether they are generally more concerned

about protecting private than business data.

Task 3: Send electronic business cards. *You are now in London, at your client's site. At a conference, you meet the CEO of another company, who is interested in doing business with you. You want to exchange electronic business cards with him. Use your mobile phone and connect wirelessly to his phone.*

The goal of the task was to see whether users differentiate between different sensitivity levels of data, type of environment, and whether the business nature of the setting influences their choice.

Table 4.2 summarizes the three chosen tasks, the devices participants had to connect, the data that was to be transmitted, the place where the connection was hypothetically performed and the amount of time pressure participants were theoretically facing.

	Devices to connect	**Data to send**	**Place**	**Time pressure**
Task 1	Phone & printer	Confidential financial report	Airport lounge	Some
Task 2	Phone & payment terminal	Credit card information	Duty-free shop	High
Task 3	Phone & CEO's phone	Business card	Business event	Low/None

Table 4.2.: The three chosen tasks simulate diverse real-world situations.

4.3.3. Session Structure

Sessions lasted between 50 and 110 minutes with an average of 70 minutes and a variance of 240 minutes, they involved one participant at a time and were run by one moderator. Figure 4.2 shows the outline of a session. In the "Introduction" phase, users were asked to fill in the background questionnaire. To ensure that no bias is being created, throughout the study we used a script to introduce the purpose of the study and explain the methods. We recorded each session using a video camera placed behind the participant. We took an additional audio recording with a laptop.

To motivate the study, we told participants that the methods they would learn could, for instance, be used to send a friend some pictures

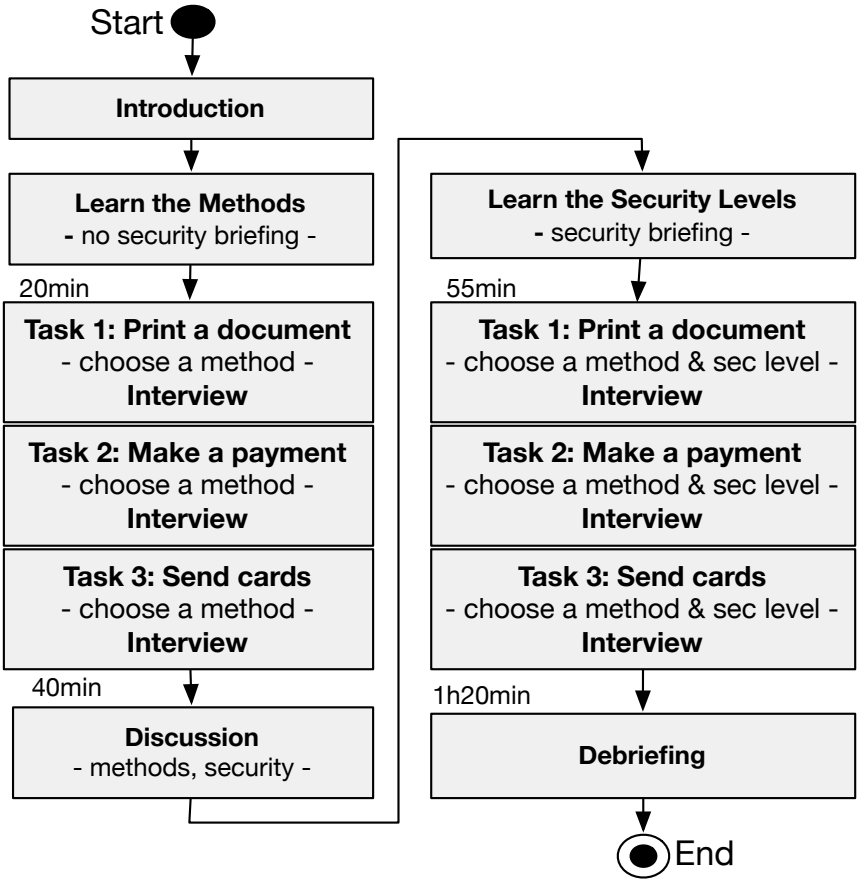


Figure 4.2.: In the first part of the study we did not mention security. Participants learned the four methods as in the *Not secure* variant and performed the tasks. In the second part they had to choose both a method and a security level for the tasks. Typical time taken to reach the point of the study is presented above the rectangles.

while sitting together in a restaurant. There could easily be dozens mobile phones in the restaurant, so the role of the methods is to ensure that the pictures will not arrive at a neighboring table by mistake. If users know that the purpose of the study is security related, their behavior might change. Therefore, during the first stage of the study (the left side of the Figure 4.2), we did not mention security. During the "Learn the Methods" phase participants were introduced to the *Not secure* variant of each method. The name of the level and the existence of different security levels were not mentioned. To avoid bias, we introduced the methods in pseudo-random order, overall covering all 24 possible permutations.

To teach the methods, the moderator read step by step detailed instructions and waited for the participant to execute each one before moving on. This process simulates the user buying a new phone with usage instructions on the methods and possibly running the methods once in the shop under the sales person's guidance. A minimum amount of explanation was given on how the methods work. For example, for Take a picture and *Listen up* participants were told that the barcode and the melody contain messages which their phone decodes. For Push the button we said the two devices synchronize each other through the button presses. If the participant failed to execute a method the moderator would start over again until the participant felt comfortable with the method. Our pilot studies showed that it is important for the participant to successfully execute each method on their own until they succeed, otherwise, they will avoid it throughout the study and consider it too hard. Finally, participants were asked to run all the methods again by themselves. The learning phase took between 15 to 40 minutes.

The moderator then read the task description and asked the participant to choose the method she or he would use in real life to establish the connection. As each task was presented, the participant was shown a picture of the potential situation (an airport lounge, a duty-free shop, business people talking at an event) to help set the mood. After each task a small semi-structured interview followed. If peculiar answers or inconsistencies emerged, further questions were asked to explore the answers. Special care was given to ensure the participant understood the methods (within the limits of the script information) and that peculiar or incorrect believes emerged from user's general perceptions and secu-

rity mental models, not from lack of clarity on the tasks and methods. If, during the task phase, the moderator realized that the participant had not properly understood a method, she went back to the learning phase and explained the method again. However, to avoid biases, no further details except for what the script contained were provided, even if the participant was inquiring. At the end of all three tasks, the participant was asked to speak freely about his general impression of the methods. We further asked whether he worried about security and how this influenced his choice.

The second part of the study, depicted on the right side of the Figure 4.2, was security oriented. In the "Learn the Security Levels" phase, the participant was told that, the way they had used the methods until now, anybody with the proper tools could listen in on their communication read and possibly modify the data transmitted. For each method, following step by step instructions, the participant learned the three security levels. Screens similar to Figure 4.1(c) allowed the user to select the desired security level. At the end of this phase, the participant was asked to run on their own all methods with the *Secure* level and then again with the *Very secure* level.

Finally, the participant was asked to think about each of the three task again and decide which method and which security level they would use. To refresh her or his memory, shorter tasks descriptions were re-read. As in the first part of the study, at the end of each task a semi-structured interview was carried to understand the participant's choices and preferences. A larger set of questions explored the perceived threat level and how different factors, such as the data being sent and the social setting, would influence the choice. The session ended with a free discussion.

4.3.4. Participants

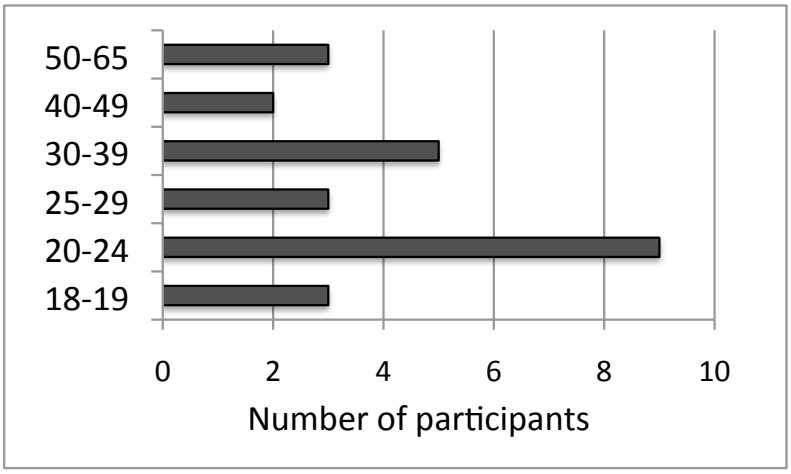
We recruited 25 participants, 15 women and 10 men, through an online job advertisement website hosted by ETH Zurich but regularly visited by people not affiliated with the university. None of the participants had studied computer science nor had taken security or advanced computer courses. Professions and study areas varied widely, with no more than 2 participants from the same field, and included secretaries, housewives, veterinary medicine students, a cook, economists, a sales

Figure 4.3.: Study session: The moderator (on the right) reads instructions and task description from the script. The participant (on the left) pairs the two devices. A video camera is recording the session.

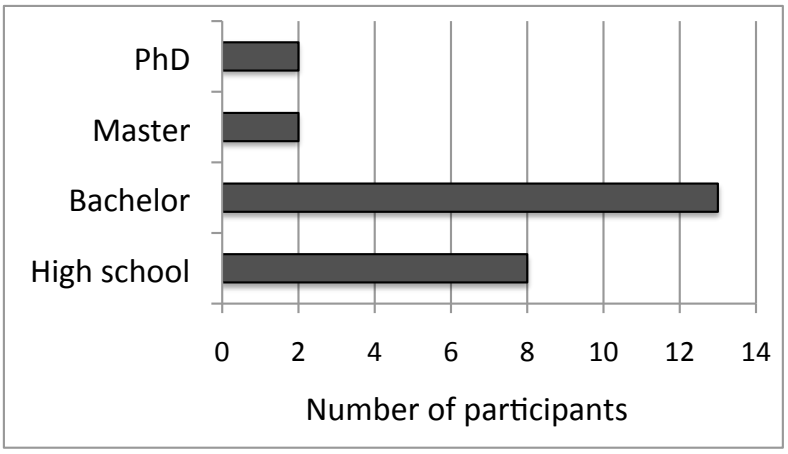
person, lawyers, psychology students, journalist, etc. Figure 4.4 summarizes the demographics. All participants reported owning a mobile phone. Seventeen participants reported that their mobile phone was Bluetooth or WLAN capable, 14 said that they do not use Bluetooth nor WLAN regularly on their device, 7 use it once every few months or once per month, 3 weekly and one several times a day. Several participants said they had never used such "advanced" phones before and many said they "don't know much about technology." The study was conducted in German in our offices in Zurich and involved one participant at the time. The complete study script, in German, is included in Appendix C.

4.3.5. Data Analysis

We transcribed all audio recordings into English. For each question in the interviews, we tried to identify trends and place answers in a few big categories. Most questions explored the preferred method and/or security level in a given situation. On a first pass, we categorized answers first on the chosen method and secondly on the reason for choosing the method. Next, we observed patters across different questions and



(a) Age



(b) Education

Figure 4.4.: (a) shows that many participants were young, but higher age groups were also represented; (b) presents the education levels equivalent to US degrees. Fifteen participants were female and 10 male.

tasks. The perceived security and usability of the methods emerged in different places throughout the session. Finally, higher level conclusions such as mental models, perceived security, the need for control, and the role of social context emerged through associations and combinations of all of the above.

4.4. Results

In this section we present participants' choice for methods, perceived security and mental models, and draw conclusions on influencing factors. We refer to participant 1 as P1, participant 2 as P2, and so on. We start by presenting high-level take-aways, discuss method prefer-

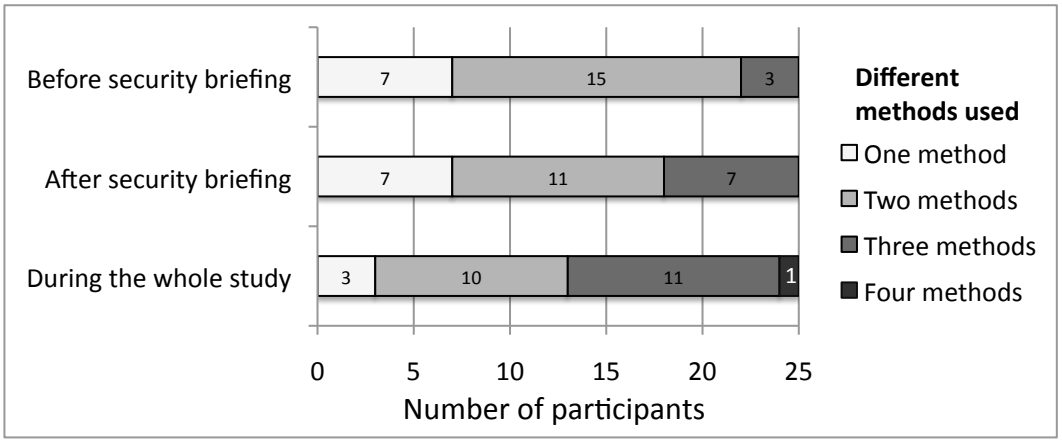


Figure 4.5.: Participants preferred different methods in different situations. For example, only seven people chose the same method for all tasks before the security briefing.

ence for each task and decision factors, then present perceived security, mental models and the role of social factors. The results of our study are purely qualitative. We do report the number of participants who fall into a given category, but we do not imply statistical significance.

Non-technical users do like newer methods. Previous studies mostly recruited participants with technical background and concluded that users prefer simple methods like number comparison instead of the newer ones, inferring that newer methods will perform even worse with non-technical users [96]. In our study, while the most popular method was indeed *Select the device*, on average half of the people preferred another method in any given task. If designed and explained well, non-technical users do embrace non-standard methods: *"Take a picture is cool"* (P8). The methods are *"reliable, fast, uncomplicated"* (P12); *"interesting and exemplary"* (P20); *"really cool, especially Listen up, that is really great"* (P25).

Different users prefer different methods. We found no single method that fits all users. In terms of personal preference, opinions differed widely. Some users said *Push the button* is *"funny"* and *"cool,"* while others said it is *"silly," "annoying"* and *"cumbersome."* The most controversial methods were by far *Take a picture* and *Listen up*. Some participants excluded *Listen up* because of the sound while others thought the method is very practical and the sound would not bother. P16 said about *Take a picture*: *"it would drive me crazy if somebody would want to do that to my phone"* and P6 has *"fallen in love with it."*

Same user prefers different methods in different situations. Three participants explicitly stated that *"in different situations different methods are applicable."* Figure 4.5 shows the number of participants that used different methods for the three tasks in the study. For example, only seven people used the same method in all three tasks in the first part of the study. In terms of security levels, five users chose the same security level for all tasks, all of which used *Very secure*. No user chose three different security levels, which might be an indication that users are more willing to vary the method used than the security level.

Same user prefers the same method in the same situation. Although very diverse and fine-grained, participants' choices for methods were not aleatory. For each task, we asked participants questions of the kind "would you use another method, if, for example, you had to print another document?" Almost unanimously the answers were *"no, if it works, I would always use this one"* or *"once good always good."* The few answers of the kind *"yes, would use another method"* were almost always followed by a condition: *"if a less sensitive document [were to be printed]"* or *"if not in a hurry."* Figure 4.8(a) shows the answers for each task. Only three participants said they would use an alternative method for paying, seven for printing and seven for exchanging business cards.

To show users' diverse preferences for the methods, and how many factors play a role, we give the following policy example. Figure 4.6 depicts some of the decision rules mentioned by more users.

Example policy. P6 (male, 19 years old) chose *Select the device* for printing, because *"I can see the list"* and *Take a picture* for paying, because it gave him a double assurance: *"By taking a picture, I actively recognize whether this is the device I want."* He thinks *Listen up* is less secure than *Take a picture* because *"my device could make a mistake"* and *"I would send my payment data to somebody else. [...] If another person's device rings I cannot walk over, take his phone and say, 'Sorry, I have to delete my data from your device.' He would say, 'Are you crazy? What are you doing with my phone, somebody was calling me, what do you want?' I find this kind of risky, even when I hope that the mobile phone always chooses the right device."* But for exchanging business cards he would, nevertheless, use *Listen up*: *"Now we are at a meeting, if the CEO is there he can see whether he received something*

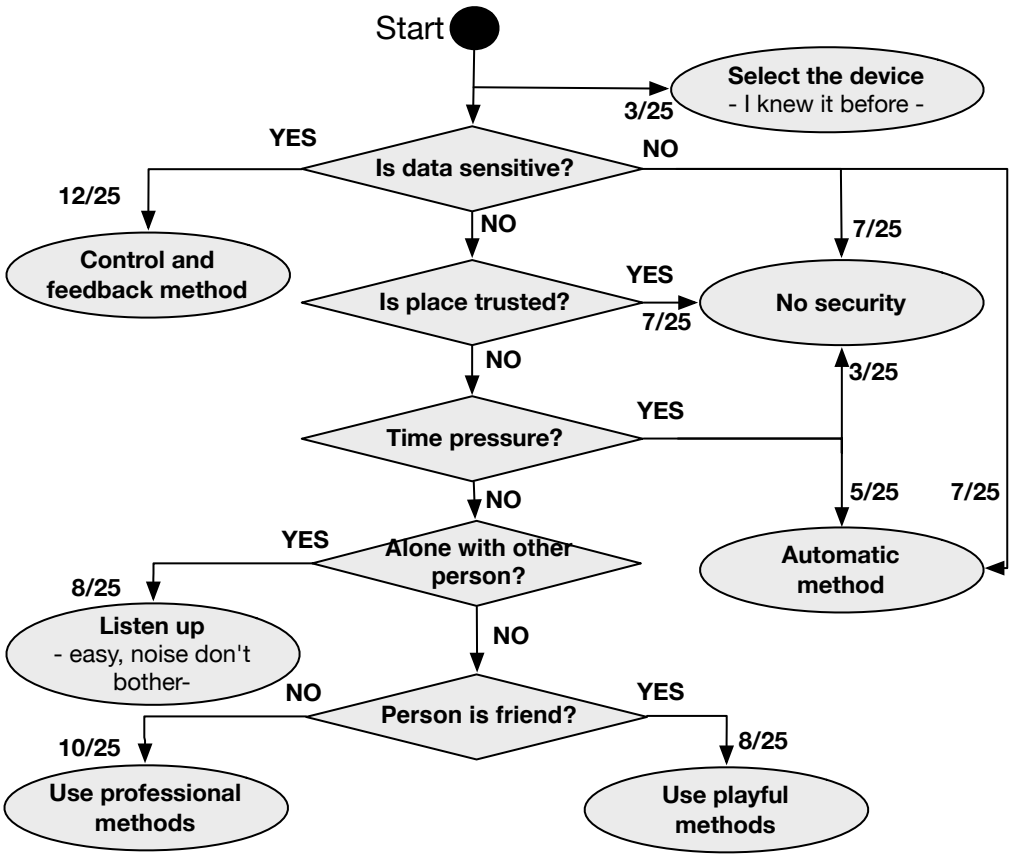
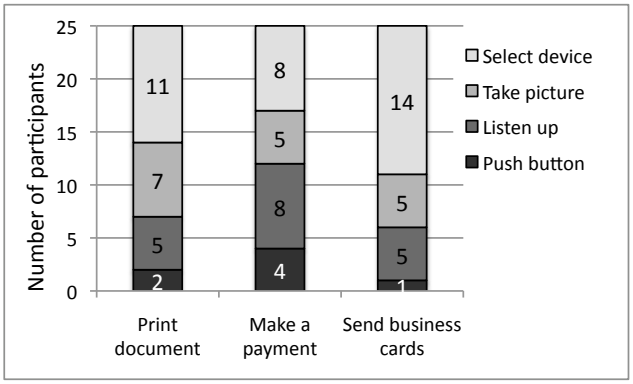
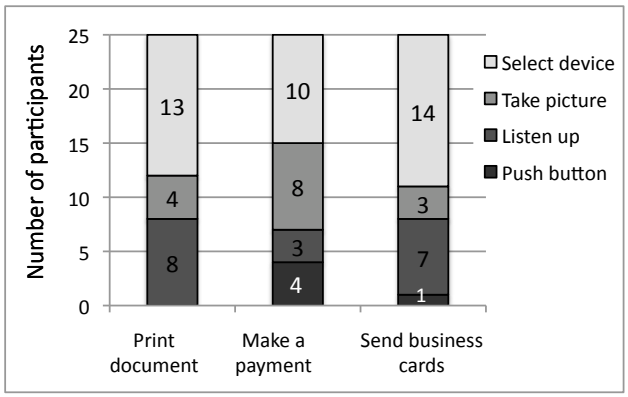


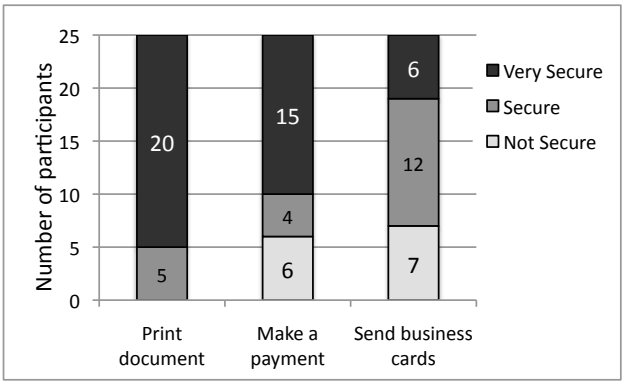
Figure 4.6.: Participants showed fine-grained decision process based when choosing a method. For example, 7 out of the 25 users said that in a trusted place they would not use security.



(a) Method choice before security briefing



(b) Method choice after security briefing

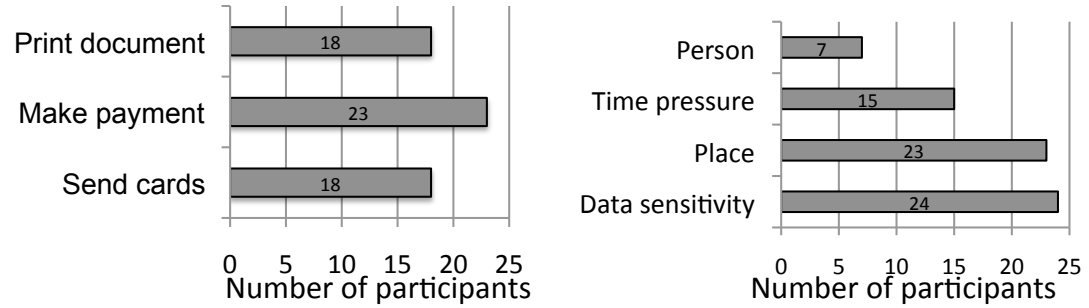


(c) Security levels choice

Figure 4.7.: (a) displays method choice for the three tasks in the first part of the study, before security was mentioned. We then introduced the security levels. Participants had to perform the same three tasks again, choosing (b) the preferred method and (c) the preferred security level.

from me. [...] It would not be so bad if somebody else received my business card, because that is not something extremely personal. In this case the connection needn't be double-verified." After the security briefing, P6 said he would use a lower security level if he were printing his own tax document because it is not so sensitive. In the office or at home, he feels *"generally safer"* because he is alone, and therefore would choose *"one security level less."* To pay he would still use *Take a picture*, with *Secure*, and to exchange business cards *Listen up*, also *Secure*.

In the following section, we summarize the results and impressions of the 25 participants and outline some of the main factors influencing their choice.



(a) Same method again (b) Main decision factors

Figure 4.8.: (a) For each task, users said they would use the same method again if encountering the same situation. This suggests that user's choice, although fine-grained, is not aleatory. (b) Users chose methods based on the sensitivity of data, the place where the connection is established, the time pressure they are in, and the person handling the other device.

4.4.1. Preferences and Decision Factors

Previous studies tried to identify the preferred method and rate easiness of use. Our results show that users do not always use the easiest or fastest method, nor the one they like best. For example P11 said *"Push the button annoys me"* but he would use it for printing a sensitive document *"even if I don't like it,"* because the method seemed secure, it gave him a sense of control: *"there I have a direct influence on the devices, I synchronize them myself."*

Before security briefing. Figure 4.7(a) displays the participants' choice of methods, in the first part of the study. To print, 11 people

chose *Select the device*, 7 *Take a picture*, 5 *Listen up* and 2 *Push the button*. *Listen up* became more popular for paying because it was perceived as fast (users were in a hurry to catch the flight) and *Push the button* as well because, even if tedious, it was perceived as secure, which is very important when dealing with money. Only 6 people did not mention security as a selection criteria during the first two tasks, all of which chose *Select the device* every time. *Take a picture*, *Push the button*, and occasionally *Select the device* were regarded as secure because it made users feel in control. All the participants who chose *Take a picture* or *Push the button* for printing or paying said they did so for security reasons. People who chose *Listen up* said they did so because it is fast and/or easy. Reasons for choosing *Select the device* were more diverse: easy, fast, somewhat secure, "I knew it before," or "I am certain it works." For exchanging business cards, users once again tended more to *Select the device* (14 participants), which generally was considered professional and most adequate in business settings.

After security briefing. Figure 4.7(b) displays the preferred method and Figure 4.7(c) the chosen security levels, in the second part of the study. For printing a document, 20 people used *Very secure* and 5 *Secure*. For paying, fifteen people used *Very secure*, four *Secure* and 6 *Not secure*. The reason for lowering the security level for payment was mostly the hypothetical time pressure in the task and for exchanging business cards the low sensitivity of data. When keeping security high in task 3, some users said they wanted to seem responsible in front of the CEO, would like to keep security by default or worried that there is always a risk.

For each task, we asked participants to sort five criteria used to choose a method in their order of importance. According to the average ratings, *security* ranked first for printing, followed by *ease of use*, *speed*, *professional look-and-feel*, and finally by *fun*. For paying speed became the second factor, while for exchanging business cards speed and ease of use were both ranked first.

The varying differences in completion times for security levels for the four methods was a reason for switching to another method. Table 4.3 depicts the completion time for P6. For example, *Listen up Very secure* took 17 seconds, only 8 seconds more than *Not secure*. *Push the button*, however, took 16 seconds more for *Very secure*, compared to *Not secure*. The least number of people, 6 out of 25 (compared to

12 and 13 for the first two tasks), changed their chosen method for exchanging business cards after the introduction of the security levels, which might indicate the weight of social factors in this situation.

	Not secure	Secure	Very secure
Select the device	14	29	29
Take a picture	24	28	34
Listen up	9	15	17
Push the button	20	22	36

Table 4.3.: Completion times for participant 6 in seconds, for each combination of methods and security levels.

There was a tension between users' tendency for a default method and security level, and their tendency to adapt to various data protection requirements. Interestingly enough, these tendencies were at odds even for the same participant. After having said *"Why are there three security levels? I would always use the highest one,"* P24 nevertheless said he would use the *Secure* level (i.e., only the second highest level) for printing his own tax document: *"My tax data is not so secret. I have an average salary."*

Overall, users varied both the security level and the method used depending on a wide range of factors: the sensitivity of data being transmitted, the place where the transaction was made, the time pressure, the person operating the other device, the social setting, people present, noise level, and perceived security threat. Figure 4.8(b) depicts the main factors and the number of participants using them as decisive criteria.

Data sensitivity. An overwhelming twenty-four out of the twenty-five people used sensitivity of data as a criteria in their choices, e.g., when exchanging business cards or printing their own tax document. Surprisingly though, people did not only vary the security levels based on the sensitivity of data, but also (and maybe more or equally often) the method used. P16 said: *"If it is about money the method has to be extremely secure, and it can also be more tedious. It is a completely different situation than before [when printing], where it can be easy, or when you have nothing to lose."* P15 would use *Listen up* to print her own tax document: *"in the worst case, it is not so bad if it goes somewhere else. [...] But for the financial report of my client, I would*

not want to risk that. An error could occur; I could believe that the sound comes from this device but it would not be so."

Place. Figure 4.9 shows the number of people for whom place was a decisive factor. Summed up, 23 people used place as a criteria in at least one situation. For example, when printing at home instead of in the airport, 9 participants would use the same method but less security. To pay at night in a gas station, when no other customers are waiting in line, 8 participants would make a different choice than in the airport. If instead of at the conference, they would be exchanging business cards with the CEO in the office, in the first part of the study, 9 participants said they would use a different method, 8 of which opted for *Listen up*. For 6 participants, a coffee place requires a different choice than the conference.

Time pressure. Fifteen participants mentioned time pressure as a decisive factor. Six people said they would have used a higher security level, had they not been in a hurry to board the plane, 4 of which had used the *Not secure* level. P20 felt under time pressure when printing in the airport and said he would use a higher security level to print in the office because he would have more time. Unsurprisingly, when in a hurry, 9 participants selected a faster method and/or a lower security level. Furthermore, under stress, 4 participants preferred less attention demanding methods. P17 said: *"if you are under stress you are careless."* P5 worried that *"because of the rush I could not take pictures so well"* and P14 said that she *"could make a mistake when typing in the number."*

Person. Seven participants said they would choose a different method to exchange addresses with a friend than when exchanging business cards with the CEO.

We discuss more the reasoning and security mental models that make places and situations more or less risky in the following section.

4.4.2. Perceived Security

In the first part of the study, 19 people mentioned security as a choice criteria and 16 said they had worried about it. (Even if they used security as criteria, some people said they didn't worry about security because they were reassured by the use of an adequate method.) How-

ever, what participants worried about was not cryptographic protocols nor malicious attackers. Instead, they worried about connecting to the wrong device by mistake and how to avoid errors. P14 said about printing: *"If I chose the right device, then I am not concerned that somebody else would get the document, even if I use no security."*

Four participants believed *Select the device* provided high security assurance because they could *"see the name"* and then they could be sure nothing bad would happen. Seven people said *Select the device* is not secure, but only one person worried that somebody might try to impersonate the printer; the other 5 were concerned with accidentally choosing the wrong device, having more devices with the same name in the room, or that in real life they would not know the name of the device. During the study, the name of the partner device was displayed on its screen. *Select the device* was regarded more secure in the office than in a public place by 5 participants, not because of a lower risk from attackers, but because *"in my own office I would definitely recognize the devices"* (P18) or *"if I have set-up the printer myself, I then know exactly which one it is. Maybe I even used it several times. I don't really feel insecure"* (P25).

Device naming was confusing for many participants, even after the learning phase. Most of these users thought the name of the device was "Nokia N96" because the model "N96" was printed on the device, above the screen. P11 tried to infer the name: *"I thought the printer is in the Lounge. That's why I chose this Lounge printer."* For the paying task, there was an accidental misspelling in the name of the device users had to choose: the other phone's display said "Dutty-free A" and the user's device showed "Duty-free A" in the list. Only one participant out of the 15 that went through this screen observed the name mismatch. This confirms that people are very likely to tolerate some sort of spelling mistakes during the identification phase. Feedback and verification measures is therefore of extreme importance. For three participants, typing in the PIN was valuable as double confirmation that they indeed chose the right device.

Five participants said that automatic methods are secure, because the user cannot make a mistake. Afraid that she might select the wrong device or type in the wrong number, P20 used *Listen up* for paying. When using *Take a picture*, P10 said: *"It seems the most secure to me, with this method I think an error is not possible."*

For some users, perceived security was more important than the predefined security levels. Unaware of the role of authentication string, P20 believed that *Take a picture – Not secure*, is more secure than *Listen up*, *Very secure*: *"In my opinion it doesn't change much for one or two pictures. [...] I think it is secure enough, even with a picture."* We asked participants how concerned they are about somebody seeing their credit card data during the wireless transmission, on a Likert scale from 1 to 7, where 7 is very concerned. P20 rated *Take a picture – Not secure* with 2 and *Listen up – Very secure* with 7. About *Push the button*, P15 said: *"how does this increase security if I press three or six times? It's no extra step, no double confirmation. What I find good there is that something new happens, there is a new aspect."* P19 said about *Listen up "from the security point of view, it doesn't matter to me if it is 3 or 6 seconds."* Understanding what makes people perceive a method as secure is of crucial importance in designing systems.

Seven participants explained that a method is secure if it provides control, feedback, and double assurance, and allows operating both devices. When dealing with sensitive data, users would go through the trouble of using a tedious method as long as it fulfills these requirements. Although considered tedious, hard to perform, and slow, *Push the button* was preferred by some participants in security relevant situations because is is interactive, it seemed very precise, and provided control over both devices. *Take a picture* was also regarded as secure, because of the double confirmation and control.

Control also means the ability to cancel the connection at any point. P15 worried about *Listen up* and the lack of stop and cancel functionality of the prototype: *"How can I stop this? If I hear it comes from another phone? Could I stop it?"* Other participants worried they could not distinguish from which device the sound would come from. P15 said: *"I don't know if it comes from the purse, from a phone, or from the payment terminal."*

Four participants said that if they put in an extra effort for security they feel at peace. P2 said *"If I use security I have the feeling that I did something about it, so I am less worried."* Twelve participants said they would use more security than they consider necessary to be on the safe side or that they would have security enabled by default. P12 said for printing: *"Better a bit more secure than too little."* Similarly, although P17 believed the home is more secure than a public place,

she still used the highest security: *"I would set the settings to that and then change it rarely. Simply because it is set to this."*

Seven participants said that, when dealing with sensitive data, they prefer methods where they have control and double feedback, but that for less sensitive data or under time pressure, automatic methods are better. P22 said: *"When you have to pay, it is better to have feedback but it can also be more tedious. It is a completely different situation than before, where it can be easy, or when you have nothing to lose."* Similarly, P16 said: *"For everything that is not sensitive data I would immediately use Listen up. [...] I think that you need to have interaction here, to have the feeling that it is secure."* It is not enough for designers to create the most secure and the easiest method, if it does not also inspire users trust. The designer might need to introduce, even artificial if necessary, explicit steps to provide for a feeling security.

4.4.3. Mental Models

Even when properly accounting for perceived security and mapping this to the actual security guarantees of the protocols, designers need to be aware of user mental models, user requirements, and their implications. For example, 4 people said that when handling sensitive data in a public place the method should be discrete. P19 said: *"If it is really confidential, then other people don't need to be aware that I try to setup a connection."* With *Listen up*, *"people would wonder what I do with the music, what is that. [...] If it sings then people will perhaps look, throw a look at the financial report."* When printing sensitive data, 3 participants chose a method because it required physically interacting with the printer. P15 said: *"It is a confidential document, and therefore I have to take some precautions, that I am next to the printer when it gets printed."*

Our interviews revealed several mismatches between people's mental models and current systems designs and operations. As credit card information gets used more and more for small payments in daily life, it is crucial to convey to users the importance of properly securing *every* transaction in mobile payment applications. Alarmingly enough, 6 out of the 25 participants, even very well educated and security-concerned people, said they would use less or no security when buying a pack of cigarettes than when buying a bottle of whisky, because

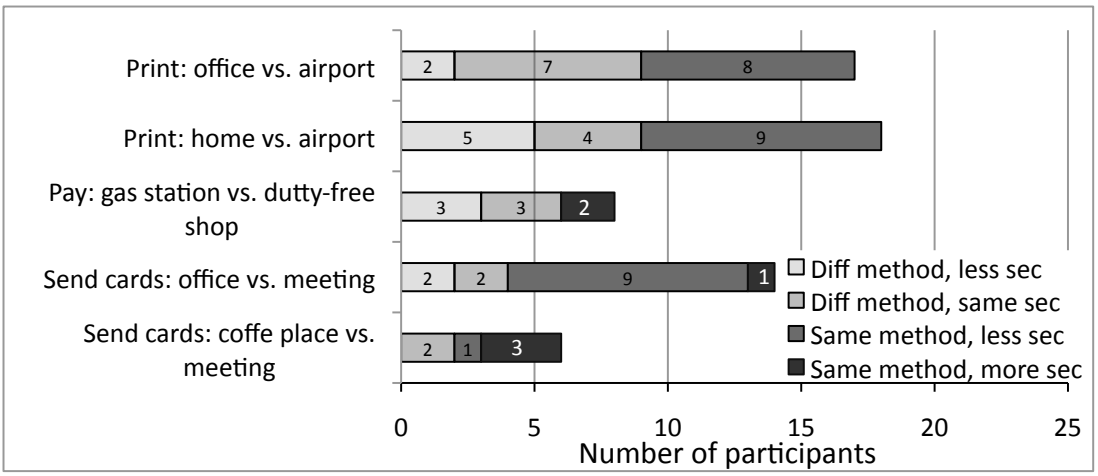


Figure 4.9.: The place where the connection is being established is a decisive factor in choosing the preferred method.

the price is lower. Although we did not specify whether credit card information or electronic cash gets transmitted wirelessly, subsequent questions referred to credit card information. P17 even said: *"even though I assume it is the same data that is being transferred, it is less money and one thinks it is not so bad."*

Keeping users alert about security in time is a challenge that security designers should keep in mind. If nothing bad ever happens, even security aware users are very likely to lower their guard. P22 never hands out her credit card in a restaurant: *"That is the biggest mistake,"* but even though she thinks *Listen up* is not secure enough for paying, she admits that eventually she would no longer use *Select the device* which she considers secure: *"I would be weak and select* Listen up*, because I will have gotten to trust it."* To avoid such cases, security sensitive applications like mobile payments should enforce security by default, and not give users a choice to opt out.

Three participants thought devices are predestined to fulfill specific purposes and cannot act as other types of devices. For P7, *Select the device* is not appropriate when connecting to a phone, because there might be more phones with the same name, but printers are less common devices, so then the method is good. Although one of the most diligent and security concerned participants, P15 chose not to enable security for payment, because she was convinced that she cannot transfer money from her phone to the phones of people around. This is a very dangerous assumption, given that sniffers and protocol implementations are possible. Relay attacks on in-shop credit card

payments have been proven [43].

Two participants worried that data can be stored and reused, but only if in a non-obfuscated format. P24 believed that the *Push the button* is more secure than *Take a picture*: *"when you push [...] there are several steps that one maybe cannot as easily trace back like with a picture that one can recall. Push the button is more secure because you cannot trace it back."* Even compared to *Select the device "it is more discrete, more hidden in the device."* P7 said: *"for the printer I would maybe worry that the data is saved somewhere and then it could be printed out again. For paying I worry less about this."* Countless incidents of in-store credit card cloning dismiss this assumption.

When connected to an unattended device, participants generally wanted interaction and control over both devices. However, if the partner device is operated by a person, this requirement diminished, because the other person could act as a feedback provider, confirming that the data arrived at the right place. P15 said *"If I can coordinate this with the second person, I am certain that no other person can take my data."* This can be a dangerous assumption, since the protocol is just as vulnerable to eavesdropping and man-in-the-middle attacks. Since the scenarios we explored covered only less sensitive data exchange with a human party, it would be interesting to see if users' concern increases if, instead of the business cards, sensitive data was being transmitted.

When being alone or in a trusted place, participants generally felt safer. P23 said: *"If there are so many people here, you don't feel so protected anymore."* Also, when alone, the probability of connecting to an unintended party is lower. For printing, 13 participants said there is less risk in the office. Ten would not use security if printing in the office instead of in the airport. Extending the concept to paying at night alone in a gas station could have undesired consequences in the presence of eavesdropping, unattended devices. P14 said that in the gas station he would not enable security for paying. *"I would be sure I pay to the intended party, because there is nobody else around."*

Participants often refrained from using a method because they had not understood it or felt it didn't "make sense." We witnessed the value of explanation and education, which constitutes a big challenge in the real world when introducing new methods. Five people preferred a

method in a given situation because it seemed "appropriate," it resembled something they knew, like debit card payments or passwords. Some participants had even more surprising criteria: P17 chose *Take a picture* to exchange business cards *"because business cards are more visual. And it goes better with something optical."* If instead she was exchanging an mp3 file would have used *Listen up*, and to exchange a financial report with the CEO, *Select the device*.

4.4.4. Social Factors

Our results show that designers should pay careful attention to ensure the methods comply with social conventions; if not users might compromise security for social compliance. For example, P16 said: *"Listen up would be more secure, but it draws more attention than it should."* Furthermore, lowering the security level in the office is not necessarily because of lower risk: *"it is more quiet, and if we are both there, it would feel awkward if it it rings too long."* Social factors influence requirements for interaction models, ease of use, speed, and security and were used by twenty users as a reason in their choices. Fourteen people said they would be embarrassed to use one of the methods in the social setting: 10 with *Push the button*, 3 with *Listen up*, 2 *Take a picture*, and one if typing the PIN for *Select the device*.

The method used is critical for building a good business relationship. Eight people said they would use a different method with a friend than with the CEO. All of them decided to change from *Select the device* to other methods if pairing with a friend's phone. P5 said: *"If it is somebody that I know then I would either use Take a picture, or put the phones together and use Listen up, because I am closer to him, he is not such a big boss."* P7 said: *"when I know the other person well, I think any method would be appropriate. When it is somebody important or whom I do not know, I would take the most professional method: Select the device with typing in the PIN."*

P16 used *Push the button* to pay for the feeling of control and used *Listen up* with the CEO: *"The other methods would be too personal, if I now have to press around on his phone. It would be silly to have to tell him he has to press the button 3 times when my phone vibrates, or if I would have to push the button on his phone, because it is too personal, too close. I want to build a good relationship with him. If*

you don't know a person too well you don't want to go like a bull at a gate. Take a picture is just as inappropriate. [...] When it is about a business contact, I would like it to be the easiest for him, and for the situation: the method that could least go wrong." On the other hand, *"if it is friends or acquaintances or my parents or whatever, then I don't care. They know me and I know them, so it doesn't matter which method I use."*

Several participants said that, in a more relax environment or among friends, the methods can provide a playful moment. P16 said: *"It depends if the other person knows it already, but for example, if he doesn't know these methods, I would have the demo effect with Take a picture: 'Look, it works!"'* P19 thinks that even with the CEO the method *"maybe plays a role to establish contact. If we have a bit of fun together, it will remain in his memory."* At the conference he would use *Select the device*, but in the office *Take a picture*, and even *Push the button* could be appropriate: *"Maybe it even has something that connects us, an ice breaker."* With a friend he would normally use *Take a picture* because *"it is more intimate"* and joked about how he would maybe even use *Push the button "to annoy somebody, like my grand mother, because she cannot do it."*

Depending on the social context, even the speed requirements of the method vary. When establishing a new personal contact, e.g. in a business relationship, the method should be faster than normal, easier and not disruptive. For exchanging business cards, P21 used a lower security level: *"it would not be the most pleasant when establishing personal contact to spend such time in this technique [Take a picture, Very secure]. So it should work relatively fast. Additionally I don't have to concentrate very much and I can nevertheless continue engaged in conversation with the discussion partner while I establish the connection."*

To make a good impression and protect the CEO's data, users sometimes seemed even to exaggerate the security requirements. P9 chose *Listen up* and middle security to pay, but *Select the device* and highest security to exchange business cards: *"Out of respect towards the CEO. I wouldn't want his data to arrive to somebody else but me."* P8 also used *Very secure*: *"it shows that you worry about the data security of somebody else, which could further strengthen the business relationship."* In fact, only 7 people disabled security in task 3.

A funny anecdote was provided by P24, who would use *Take a picture* to exchange business cards. At the conference the *Secure* level is enough, because business cards are not so important, but in the office he would use *Very secure*: *"In the office the CEO is next to me and maybe he sees that I use the highest security. He probably expects that. At the conference there are also other people. He is more attentive when there are no other people around. And he sees that I choose highest security."* In the coffee place he would again use the middle security level. *"The CEO sits on the other side of the table. He doesn't necessarily see this."*

We asked participants to rate the sensitivity of the data contained on the business cards. P9 said there is a difference between his business cards and the CEO's and rated the CEO's with 5–6 on the Likert scale (7 is the highest) and his own with only 3–4. *"Maybe he has his private address written there, which nobody should have."* P13 would rate the CEO's cards as extremely sensitive, 7 on the Likert scale, if his private number would be on them. She used the highest security level for exchanging the cards *"I hope that the CEO does not give everybody his business cards, but just to me."*

Users used higher security when dealing with somebody else's data to make a good impression, but also out of a sense of responsibility towards other people's data. Eleven participants said that business or confidential data is more important than private data or were extremely concerned with protecting the CEO's business card.

Twelve people said they would use a lower security level and maybe even a "less secure" method to print their own tax document instead of the financial report. Fourteen participants said the tax report is less important: *"My tax document is mine, private, but what concerns the company does not belong to me. That I do for the company. So I have more responsibility"* (P14). P17 said *"I think it has less priority because it is something personal, and if it is a customer's, a business contract, you have to be twice as careful."* P16 also thinks *"tax document data is no longer so sensitive as the financial report, because the financial report concerns other people too, while the tax document just me. So it has more consequences, because I would damage other people too, if I were not secure."* P21 says *"I am accountable in front of the CEO when I handle his confidential data. So I take the highest security."* Losing money is comparable to losing somebody's trust for P23, in case she

handled the data irresponsibly. Eight participants said the financial report is more important even than the payment.

Finally, the right method is very dependent on the social situation. When printing a document in the airport, P16 thinks that *"it would be totally ridiculous if I wanted to take a picture [of the printer]. Even Push the button is a bit foolish. The PIN is professional."* P10 said: *"noise is a criteria. In the meeting I cannot use Listen up."* P17 also wants a silent method: *"At a dinner you meet and talk to people, vibrations and sounds are not appropriate."* P9 agrees: *"When there are other people present, I think it is better to be discrete. If I am alone with the CEO then I would use Listen up, otherwise Select the device."* But in the airport *Listen up* is appropriate and he chooses also the highest security, because *"the airport is always noisy, so the music wouldn't bother."*

In the office, seven users switched to *Listen up* when printing. P18 said *"If I'm alone the sound can't come from somewhere else."* However, only two people switched to *Listen up* in the office for exchanging business cards. This might be due to the higher weight of the social factors: choosing a method that would be appropriate for interacting with the CEO. Different decision factors have different priority for different people.

4.5. Conclusions

We conducted a laboratory user study with 25 participants to investigate the usability of device pairing methods in different real-life situations and the security needs users perceive. We tested four methods that span a wide range of auxiliary channels (visual, audio, tactile) and require different levels of user involvement (from very active to fully passive). Our results show that users do worry about security, but not in terms of malicious attackers or data encryption. It is not protection against man-in-the-middle attacks and eavesdroppers what makes a method secure in their perception. Instead, a method is perceived to be secure if it reassures users through double confirmation and control that everything went as planned and they indeed connected to the intended device.

Users prefer different methods in different situations. For example,

when dealing with sensitive data, control and feedback is needed and, when handling less sensitive data or under time pressure, automatic methods are preferred. Furthermore, social factors influence greatly method requirements. For example, when connecting to a friend's phone the method can be playful, but with a newly met person in a business environment professionality is required. Similarly, in the office, at home, in a public place, or in a meeting different methods and security levels are desired. We investigated factors influencing users' choice for a method in different real-life situations and detailed on perceived security and mental models.

Laboratory studies cannot infer with full confidence real-world behaviour. In the three tasks, we tried to depict real-life situations as clearly as possible and hope that provided answers are consistent with real-world behavior. Strong correlation between reported previous security concerns during real-life situations and participants' security concern during the study might be a reassuring fact. Future work should investigate user behavior in the wild, over a longer period of time. For example, an initial study could deploy device paring methods among the participants of a one-week conference and see which methods people use for exchanging business cards.

Users' method preferences were influenced to some extent by the reliability of the software and mock-ups used, a pitfall that any user study will encounter. For example, sometimes the barcode decoding library did not focus on the first try. This made some users believe that *Take a picture* is not as reliable as *Listen up*, which always worked due to the mock-up nature of the prototype. Nevertheless, *Take a picture* seems to have awakened most enthusiasm in users.

Were we to conduct the study again, we would refrain from using the term "PIN" in the description of the *Select the device* method. Three users associated the number with the PIN of credit and debit cards, two of which were confused when, upon typing-in on the partner device, the number was displayed on its screen. One of these users indicated that for printing, unlike for paying, she would like to see the number displayed on the screen. It would be interesting to see if users maintain this association for payment tasks but not for others in the absence of the "PIN" naming.

In future work, we would also like to test whether users have a lesser

requirement for control over the process if connected to a device operated by a person (e.g., the CEO's phone) than to an unattended printer or payment terminal, regardless of the sensitivity of the data being exchanged. Furthermore, we would like to test our results and conclusions against other pairing methods and see if control and confirmation are still the major factor in user security perception or if methods such as distance bounding protocols intrinsically inspire more trust. One participant mentioned during the study that when putting the phones together he automatically feels safer.

4.6. Guidelines for Developers

Creating a technically secure and highly usable method is not always sufficient to meet users' needs. The method should also comply with users' security perception and be appropriate for the specific social situation.

1. **Map perceived security to method guarantees.** Designers should create methods whose actual security guarantees are consistent with users' perceived security. To achieve this, it might be necessary to introduce redundant step, controls, cancel buttons, and double confirmations.

2. **Include security by default.** We detected several mismatches between users' mental models and systems design, which prove the need to include security by default when dealing with sensitive data involving other parties also, such as a customer entrusting a confidential financial report or a bank issuing a credit card. Also, our results show users' willingness to have security enabled by default.

3. **Support several methods.** Some users liked very much *Take a picture* and disliked *Listen up*, and others did exactly the opposite. To account for diverse personal preferences, mobile devices should support a set of different pairing methods.

4. **Account for social factors.** No single method is adequate for all situations. Users are likely to bypass security before breaking social norms. Designers should provide appropriate methods for professional environments, public or private places, for interaction

with friends or strangers. The user could, for instance, choose between several variants: meeting mode, quiet room mode, professional mode, play/fun mode, etc.

5. Conclusions

This thesis proposed a novel, user-centered mechanisms that help non-technical users protect the privacy of their data and communications in the cloud. More specifically, we provided three contributions:

- We conducted an exploratory user study with 36 in-depth interviews in Switzerland and India, followed by an online questionnaire with 402 respondents. Our study results show that end-users have strong concerns about the privacy of their data in the cloud; they prefer storing sensitive private data locally, and feel that current technical solutions for ensuring privacy in the cloud do not meet their needs.

- We proposed a novel system that allows users to share data over any web-based cloud sharing platform, while protecting both the confidentiality of the communicated data and hiding the confidential nature of the communication from platform providers and any unintended recipients. We implemented a prototype of our system in the form of a Firefox plugin, and showed the technical feasibility of our approach through a performance evaluation.

- We conducted a comparative usability study of proposed device pairing protocols. Such protocols can be used by users of our system to securely exchange encryption keys by establishing a direct wireless communication between their mobile devices when they meet in person. Our study results show that users have a complex decision-making process regarding which pairing method they prefer in a given real-life situation. Users prefer different methods in different situations, depending on a number of factors. Furthermore, we elicited users' mental models of the security of different device paring methods and showed the importance of accounting for social factors in designing usable security mechanisms for mobile interactions.

This last chapter is structured as follows. Section 5.1 summarizes the research questions and findings of this thesis. Section 5.2 outlines future work and limitations of our findings. Finally, we end with Section 5.3 by discussing our findings and the applicability of our research in the light of upcoming technological developments.

5.1. Summary

5.1.1. Privacy Concerns in Consumer Cloud Storage

In Chapter 2 we explored end-users' privacy concerns and attitudes about consumer cloud storage. Most privacy studies so far have mostly focused on the United States. However, cloud services are being increasingly used by more people around the world, especially in developing countries. Therefore, to devise privacy-aware solutions and global regulatory frameworks, a good understanding of privacy concerns across different cultures is necessary. In our study, we focused on India and Switzerland—two countries with very different privacy perceptions and expectations, where people place different importance on individual privacy within families and society. We conducted 16 in-depth interviews in Switzerland and 20 in India, and then tested derived hypothesis in an online survey with 402 participants.

Strong privacy concerns. Our results show that users are less concerned about some issues, such as guaranteed deletion of data, country of storage and storage outsourcing, than enterprise administrators. Nevertheless, despite security expertise and guarantees given by storage providers, users consider local storage safer than the cloud. They have an intrinsic belief that nothing on the Internet is safe, and that once they store some data online, they lose control over what happens to it. Users have a good awareness of the fact that other parties, including cloud storage providers and law enforcement agencies, might view the data they store online. For these reasons, when it comes to sensitive data, users prefer to store their data on local devices and protect these using physical security, rather than storing their data in the cloud and relying on data encryption solutions.

The optimist bias. Inspite of awareness of privacy risks and mistrust of Internet-based technologies, users have a feeling of reasonable secu-

rity in the cloud. This feeling of security emerges from the belief that nobody would be interested in seeing *their* data, since they themselves are not "important," "famous," or "criminal." This discrepancy between general awareness of risks and failure to act upon it could be explained by what psychologists call the *optimist bias* [137]. Sharot found that people tend to significantly underestimate the probability of something bad actually happening to them, but are more realistic in assessing those same risks when applied to somebody else. This biased optimism creates a false sense of security, which can make people reluctant to take appropriate security measures.

Unaware of lack of guarantees. In our study, we also explored users' awareness of the contractual terms governing their relationship with cloud storage providers. Users do not read the Terms of Service and Privacy Policy documents; we explored users' beliefs and assumptions on what these conditions stipulate. Since these terms are unilaterally decided by companies, users feel they have no choice but to agree to the terms if they want to use the services. Our results show that users believe they have more rights and protection regarding a number of issues, such as the availability, integrity, and data ownership guarantees, than they actually have. Based on our results, we believe that, if users trust the service provider, they are willing to pay in exchange for better privacy guarantees for their sensitive personal data in the cloud.

Cultural differences. Finally, major part of our study focused on eliciting differences in attitudes between participants from Switzerland and India. Our results show that cultural differences can affect privacy concerns and expectations in the cloud. For example, we found that participants from Switzerland store less sensitive data in the cloud than participants form India. Furthermore, Swiss are more aware of the lack of guarantees, such as the fact that storage providers have the right to disable their account without advanced notice or explanation. Indians are more likely to assume that their data in the cloud will be instantly deleted once they press the "Delete" button, and are more likely to expect the provider to pay them for damages in case of data loss. Furthermore, our study results show a very strong difference between the two populations regarding acceptance of surveillance technologies. While Swiss consider government monitoring of cloud-stored data a fundamental privacy infringement, Indians regard it as a necessary step in combating terrorism.

5.1.2. Protecting Information Sharing in the Cloud

To alleviate privacy concerns in the cloud, we proposed in Chapter 3 a system that gives users control over who can access their data. Existing solutions that try to protect user data from the eyes of nosy platform operators while still allowing the use of the platform's services, usually require the existence of a trusted, third-party server, or are cloud storage platform specific (e.g., work for Facebook only). In contrast, our solution does not require any dedicated infrastructure or trusted third-party servers, and hides the confidential nature of communications from platform providers. In Chapter 2, we have seen that users make use of a number of different platforms to store and share data in the cloud. The system we proposed in Chapter 3, therefore, supports sharing over any web-based cloud storage platform.

Furthermore, existing solutions do not hide the confidential nature of communications from the platform provider, something we consider of particular importance. Therefore, instead of posting clearly encrypted communication on these platforms, our system allows users to post "innocent" looking pictures, files, and status updates, which will be transparently replaced with "real" information for selected recipients in the user's network. This is a technique inspired by practices that users currently take to protect their data. Boyd [25] found that teenagers sometimes post messages on Facebook that seem innocent to parents (e.g, song lyrics), but carry hidden meaning for friends. We implemented a proof-of-concept of our system as a plugin for the Firefox browser.

Unlike enterprise users, end-users cannot rely on globally trusted third-parties to provide them and their contacts with credentials and help distribute encryption keys. Consequently, end-users must, using specialized tools, perform key generation and verification themselves—a problem that has often been identified as the biggest challenge in the setup and large-scale adoption of encryption solutions for end-users. To protect against man-in-the middle attacks during key distribution, we complement our plugin with a mobile application that helps users verify exchanged public keys through trusted, off-the-cloud mobile device interactions during personal encounters. To wirelessly connect users' mobile devices and securely exchange key material, we make use of device pairing protocols. Since complex pairing methods might prompt

users to sacrifice security, the usability of such methods is of crucial importance.

5.1.3. Usability Analysis of Device Pairing Protocols

In Chapter 4 we analyzed the usability properties of the most prominent methods to securely connect two mobile devices over a wireless channel. We chose four device pairing protocols which cover a wide range of channels (visual, audio, tactile), and require different degrees of user involvement (from completely passive to fully active). Prior studies comparing the usability of proposed protocols failed to investigate the use of these methods in real-life situations. These studies tried to identify a single best method, regardless of the purpose for connecting the devices, and of the physical and social situations. Instead, we asked participants to choose a method in the context of different real-world conditions: connecting devices under or without time pressure; and dealing with different environments, devices, people, and degrees of data sensitivity.

Different methods in different situations. Our results show that users prefer different methods in different situations, according to a deterministic decision process that considers a diversity of factors, including perceived security and the social situations. Furthermore, our results show that, in given situations, the method people would use in real life is not necessarily the easiest, fastest, nor the one they like best, but rather the one that inspires most trust. It is not enough for designers to create the most secure and the easiest method, if it does not also fit users' mental models of security and inspire trust.

Perceived security. Contrary to beliefs that users will always choose "dancing pigs" over security [78], in real-life situations, users act responsibly with their dat. In specific situations in which they feel they need high security, users would even chose methods they find "annoying," if these methods feel more secure and provide a better feeling of control. Our results show that users worried about security when wirelessly connecting their devices, though not in terms of malicious attackers, protection against man-in-the-middle attacks or data encryption. Instead, pairing methods were regarded as more secure if they offered double confirmation and control. Furthermore, our results show that users bypass good security mechanisms if they do not trust and un-

derstand them or if they mistakenly believe that enough assurance is provided by an insecure method.

Social factors. Finally, as computer technology "weaves itself into the fabric of every day life" [168], system designers should account for social conventions and situation-specific needs. In our study, we saw how device-pairing methods can embarrass users or help them build a good business relationship. Our results show that device-pairing methods have a strong social impact: devices and methods used may make owners seem more professional (e.g., in a newly established business relationship), provide a playful moment between friends, or even act as an ice breaker when meeting somebody new.

5.2. Limitations and Future Work

In this section we discuss the limitations of our work and propose future research directions.

Reevaluate privacy concerns. In Chapter 2 we presented a study that explored users' privacy concerns when storing data online. In Chapter 3, we proposed a system that allows users to protect the data they share over any web-interfaced cloud storage system from untrusted parties and platform providers. A follow-up study similar to the one presented in Chapter 2 should be conducted to assess the effectiveness of our technical solution in alleviating users' concerns. In particular, the study could analyze if users who use our system are more comfortable sharing sensitive data online than users who do not benefit from such data protection mechanisms. To truly evaluate the effectiveness of the system we proposed, our Firefox plugin prototype should be brought to the reliability standards of industrial software products and be evaluated in a long-term user study. Such a study could also analyze if and how users' privacy behavior changes in time.

Learn recipients & preferences. Users will adopt such security systems only if these systems do not impose a significant usability trade-off. In our prototype from Chapter 3, before uploading data to the cloud, users must manually select the intended recipients from their contact list. Future work should investigate automatic means to infer target recipients. The plugin could then automatically apply, for each message, website and data, the protection mechanisms to the correct,

intended recipients. In some cases, the list of intended recipients could be retrieved from the current HTML page. For example, if the user is sending an email through a webmail platform, the plugin could match the email addresses in the recipients field with contacts in the address book. Similarly to page-specific message-parsing rules, XPath-based recipient rules could be specified for each web platform to automatically set intended recipients. Complementarily, users could specify access control policies beforehand, or the system could automatically learn such rules based on past user choices. Furthermore, future research should explore how well users cope with friends replying to fake posts and design adequate user interfaces to help them deal with the two worlds: the fake messages and the hidden message thread.

Study key verification in the wild. Laboratory studies can only infer with a certain confidence how users would behave in real-life and what security, and privacy measures they would take. In Chapter 4 we studied users' attitudes related to device pairing protocols. However, the tasks in the study were not carried out in the the context of the system we proposed in Chapter 3. The Android application for public key verification we presented in Chapter 3 should be further extended with a context-awareness module that allows users to adapt the choice of a device pairing method based on the person with which they verify keys and the setting in which the key verification is taking place. A subsequent study could then evaluate the appropriateness of chosen methods.

5.3. Outlook

In this last section, we take a look at upcoming technological developments and discuss how lessons learned in this thesis can be applied to future designs. With continuous technological advancements, new types of threats to individual privacy appear and new security challenges need to be solved. We start by discussing a couple of emerging applications, continue with security and privacy threats they pose, and finally argue that designers need to take approaches similar to those presented in this thesis in future system.

The deployment of smart energy meters in homes promises to help inform users in real-time about the energy consumption of individ-

ual devices and advise users on options to save energy [110]. Fine-grained information on consumer energy-usage patterns could help energy providers load balance their network and incentivize consumers to lower energy consumption at peak hours through price variation. Furthermore, a number of other parties could use the (perhaps annonymized) energy-usage information to improve their services. For example, device manufacturers of home appliances could offer free maintenance services in exchange for usage data they can use to improve their products and make them more fault-tolerant. Users might volunteer to give Greenpeace access to their household consumption data to help verify if home appliances are really as energy efficient as their manufacturers claim. Finally, advertisement companies could recommend new products and services based on what devices users already have in their homes and how these devices are being used.

An automated, smart home which preheats dinner as the tenant approaches home, pre-orders the grocery list from favorite supermarkets when the fridge is empty, and knows exactly to what temperature to set the washing machine for a given laundry load, seems the dream of any tenant.

A number of sensors attached to the body or devices that users wear could in the future provide very accurate information on current sport activities and monitor the health status of patients. Storing measured data in the cloud allows users to share their achievements with friends and benefit from real-time advice from their doctors. For example, a cloud service that processes the sensor data could alert the hospital when a patient is about to have a heart attack.

In a future, digitized world, security and privacy threats will abound, many of which will have to be managed by non-technical users. Ideally, it will be up to consumers to opt-in to such services and decide how much data about themselves they want to reveal in exchange for different benefits. Energy consumption information is highly sensitive, as it could be used to infer location, daily schedules, eating habits, hobbies and interests. For example, Lisovich and Wicker [104] showed that, even with unsophisticated hardware and algorithms, one can infer based on the energy consumption in the home sleeping and eating habits, as well as shower and bathroom usage of the inhabitants. Users might want to make sure that, if the data leaves their home, it will not end up being sold to their employees, who want to check presence

at home during a sick leave days. Health insurance companies could abuse the data to secretly monitor sleeping and eating habits, and then increase charges for consumers labeled to have a high risk of getting sick.

As the physical world becomes virtually connected in an *Internet of Things* in which "physical items are no longer disconnected from the virtual world, but can be controlled remotely and can act as physical access points to Internet services" [109], security vulnerabilities and misconfigurations pose a much higher threat than today. Sensor data from body area networks could also be used to infer stress level in certain situations, emotional attachment to one's partner, and could even act as lie detectors in business meetings. Body area network might get infected by a new virus when users fail to promptly update the firmware to the latest version. Hackers might break into home appliances and pose risks to physical safety.

Nowadays, many consumers have loyalty cards with several shops or supermarkets and enjoy the occasional discounts and benefits the stores provide them in exchange of their shopping records. Most people do not consider information on buying habits of mundane articles such as food purchases to be privacy sensitive. It is very unintuitive that, based on this data only, supermarkets could figure out, before her father does, that a 16 year old girl is pregnant. However, that was the recent case with the Target merchandise store [73]. By analyzing buying habits on tens of products through data-mining techniques, Target could identify pregnant women and even estimate their due dates. Target then advertised baby products before any competitor at specific pregnancy stages; revenues increased by $23 billion in 8 years, from $44 billion to $67 billion [46].

Buying customer data from several sources and aggregating it for data mining and user profiling has become a common practice both in business [46] and national defense sectors [140]. Although generally unknown to end-users and almost never in the media attention, a number of companies currently make their business by selling consumer data [131]. For example, at Emarketing Solutions [49] one can buy 100,000 customer records (including personal data such as name, email address, date of birth, address and phone number) for $99; 5 million consumer records cost less than $500. Purchased data comes with time stamp, IP address and website source of where the data was collected,

and is for "unrestricted usage" [50]. Further options allow the buyer to filter records by city, hobbies, and occupation, thus creating targeted user lists.

Imagine a world in which—besides buying habits, education, personal information, and maybe documents you share with your friends online—all the information about your private life, social activities, health, and habits is secretly traded without your consent by advertisers, insurance companies and governments. Without proper legislation and developments of privacy-friendly technological solutions, this is what might happen in a future digitized world when technologies like smart energy meters and body area networks are widely deployed. Politicians might buy private data to get inside the minds of voters by creating psychological profiles they can use to target voters. Thieves might buy data to find out when people are away and when to break into their homes. Understanding the significance and threat of all these systems and data is a paramount task. If it currently takes users 200 hours per year to read the privacy policies of the websites they visit [113], it might take them a full-time job to understand and review the data collection mechanisms in their home, and to appropriately set desired privacy levels for their data.

If the ways people interact through technology are likely to change in time, so are users' privacy needs and expectations. In future systems, security and privacy mechanisms that let users specify their preferences on the level of detail and recipients of their data, should fit users' mental models and be easy to use. System designers should conduct studies similar to the one we presented in Chapter 2 to elicit users' privacy needs, and to the one we presented in Chapter 4 to elicit users' mental models related to the security of different device interactions. The design of such usable systems is a challenging task, especially given that users currently struggle with less complex systems that deal with less sensitive data, such as mechanisms for protecting personal data in the cloud which we investigated in Chapter 3.

Finally, as computers come out of the lab and take the form of embedded sensors or speaking glasses, technology should become more socially-aware. In the future, people will need to interact with more and more devices and smart, embedded sensors. Secure device pairing protocols and key distribution among devices with limited input and output capabilities will become even more needed. Interactions with

embedded devices might happen in the comfort of one's home or in public. In designing future, ubiquitous systems, the lessons learned during the study presented in Chapter 4 of this thesis should be kept in mind.

Appendices

A. Cloud Privacy Study Scripts

In the following we present the interview questions and the online questionnaire used in the user study presented in Chapter 2.

A.1. Interview Study Script

A.1.1. Introduction

Note to self: Give a small description, introduction about the scope of the study, will be told to the participant when starting the interview. We will refer to people and organizations whom you have not explicitly given permission to see your files as "unauthorized parties", and to online websites where you put your files like Dropbox, Google or Facebook as "online site" or "online storage."

A.1.2. Current Practices

1. Please draw a diagram showing what kind of files you store on your own computers (laptops, desktops) and what you store on online services such as Google Docs, Dropbox, Facebook, Flickr, Picasa Web.
2. What data do you upload online? Since when? Why? (e.g., to share with friends, for backup, to be able to access it from other computers, etc)
3. What data do you store in more than one location?
4. What documents do you still keep only on your computer and why?

A.1.3. Where would you store the following documents?

1. Where do you store/would you store...
    a) Financial files such as bank transactions, income, or your tax documents.
    b) ID documents such as copy of your passport or copies of passports of your family members, scanned visa application forms, in case somebody steals your documents while you travel.
    c) Your password list or bank login information and credit card number so you can log in from anywhere.
    d) Health history so that your doctors can access your entire profile fast when you go to a new hospital.

A.1.4. Expectation of Privacy

Physical location:

1. When you write a Word document, where is it stored?
2. How about your email attachments or document in Google Docs, where are they stored?
3. How many copies of your online data are out there?
4. In which country? Does it matter to you?
5. Would you be willing to pay extra to have the guarantee that your data is stored in a specific country, like in Switzerland? For which documents?

Data protection:

1. How do you think your online data is being protected?
2. Have you heard of data encryption?
3. Do you think your data is safer on an online storage than on your computer? Why?
4. When do you think the risk is higher of somebody obtaining unauthorized access to your files: when stored locally or online?
5. Would you like to be able to request higher protection levels for more sensitive data? By what means? Would you pay?

Unauthorized access:

1. Who else, accept for you, might be able to see the private data you store online [pick an example: in your Dropbox, Gmail inbox]?
2. How easy would it be for ...
   a) Hackers
   b) Employee of your online storage provider
   c) Your government
   d) US government
3. How likely do you think it is, the above entities would access some of your online stored documents intentionally or maliciously?
4. Do you think that any of these parties have already accessed your documents?
5. Do you think you would be informed, if an unauthorized party/person accessed your data?
6. Do you think you should be informed?

Third-parties:

1. Imagine that instead of storing your data on their own servers, your storage provider (e.g., Google/Dropbox) hired another company to store your data, on their servers
   a) How concerned would you be if this happened?
   b) Do you think this might currently be the case?
   c) How likely do you think this is to happen?
   d) How upset would you be if they did?
   e) Do you think you would be informed? Should you be informed? Through which means?
   f) How upset would you be if you were not informed?
2. Have you heard of Terms of Service and Privacy Policies? Did you read them? What do you think they say?
3. Are the Terms of Services and Privacy Policies legal contracts, enforceable in court?
4. If your data is stored in another country (e.g. the U.S.), do you think that Swiss/Indian or U.S. regulations apply?
5. For Dropbox users: Which one of the following four statements do you think apply to your contract? ... Discussion.
   a) Dropbox may sell, transfer or otherwise share some or all of its assets, including your
   b) Personal Information, in connection with a merger, acquisition, reorganization or sale of assets or in the event of bankruptcy.
   c) Dropbox will not sell, transfer nor otherwise share any of your Personal Information to another party.
   d) Dropbox may only sell, transfer or otherwise share some or all of its assets, including your Personal Information, in the event of bankruptcy.
   e) Dropbox may sell or otherwise share some or all of its assets, including your Personal Information, in connection with a sale of assets.
6. Do you think you are allowed to store third-party data (music, videos, photos, text etc.) on your Google Docs/Dropbox account?

Data Integrity:

1. Imagine that you are accessing your online documents and notice that somebody modified or deleted some of your data (e.g., emails you know you sent now contain a different text).
   a) If this happened, what would you do?
   b) How likely do you think this is to happen?
   c) Whom would you suspect to have modified your data?
   d) Do you think anybody has the right to modify or delete your data?

2. Which of the following statements do you think is in the Google privacy policies document?
    a) Google reserves the right (but shall have no obligation) to pre-screen, review, flag, filter, modify, refuse or remove any or all Content from any Service.
    b) Google reserves the right (but shall have no obligation) to pre-screen, flag, filter, refuse or remove any or all Content from any Service.
    c) Google reserves the right (but shall have no obligation) to pre-screen, review, flag, filter, modify, refuse or remove Content from any Service, except for personal documents in user accounts.
3. The first option is the correct one. Do you think there might be a reason for this policy?

Controls:

Imagine that you could set a lock on your data (set a flag) before you upload it, or when it is already uploaded. If you set this lock then nobody can modify your data.

1. Do you think such a technology could be possible?
2. If your provider offered it to you, would you use it?
3. How much would you be willing to pay for this feature?

Guaranteed Deletion of Data:

1. Can your data still be recovered after you delete it from your computer?
2. How about after you deleted from your email account?
3. Which one of the following statement is correct?
    a) You may permanently delete any files you create in Google Docs. Residual copies of your files will be deleted within 24 hours.
    b) Because of the way we maintain this service, residual copies of your files reside on several active servers and offline backup systems. We therefore do not guarantee permanent deletion of files you create in Google Docs.
    c) You may permanently delete any files you create in Google Docs. Once you do, all copies of your files will be deleted from all of our servers.
    d) You may permanently delete any files you create in Google Docs. Because of the way we maintain this service, residual copies of your files may take up to 30 days to be deleted from our active servers and may remain in our offline backup systems for up to an additional 60 days.

Lock-out/Data Migration:

Imagine that in the future you will decide to abandon Dropbox/Gmail/Yahoo and move to a new system that is gaining popularity. Perhaps these companies are going bankrupt.

1. What data would you save?

2. Do you know how to get your data out of the system easily?

Account Disabling:

Imagine that tomorrow when you are trying to access your Google/Dropbox account you are being informed that your account has been disabled and you may no longer log in.

1. How likely do you think this is to happen?
2. Do you think your provider has the right to disable your account?
3. What would be the worse/irepleaceable thing/data to lose?
4. Whom would you turn/complain to?

Liability in case of failure:

Imagine that your storage provider lost some of your data, perhaps an administrator accidently deleted it or there was a server crash.

1. What do you think your rights are in such a case? What actions would you take?
2. What if you paid for the service? Does it change your rights?
3. Do you think you would have to file a lawsuit?
4. Or a complaint with a privacy protection authority? In which country?

Government, surveillance and coercion:

1. Do you think the Swiss/Indian police or government can access the data you store online?
2. Would they need a court order?
3. Would you be informed if this happened? Should you be informed?

Coercion (US vs. Swiss, local vs online):

1. Could you be forced by the Swiss/Indian police to give your Gmail/Yahoo password? How about the password of your laptop? Would they need a court oder?
2. How about the US police or government? When could they access your data? Would they need a court order?
3. Do you think the technology exists for you and a friend to communicate electronically and exchange data without any other party being able to decrypt or see your communication?
4. Is such a technology possible? Why not?

The right to privacy:

1. Do you think YOU should be able to protect the privacy of your data and communications, whether stored locally or online?
2. Do you think EVERYBODY should?
3. Do you think TERRORISTS should?

Regulation:

1. Do you think somebody is responsible to check that your online storage provider does not sell your data and that they apply appropriate data protection levels?
2. What data protection laws do you think apply to the personal data you store online? (e.g., Swiss, internationals, EU, US?)
3. If your data is stored in another country, e.g. the U.S., do you think that you will have the same rights & privileges as U.S. citizens or do you think special rules apply to you because you are located in Switzerland?
4. Would you like to be able to insure the data you store online, in a similar way that you insure your car or the assets in your home? If something bad would happen and you would lose your data you would be reimbursed by your insurance company.

A.2. Online Survey Questionnaire

1. **Where do you consider your private data to be safer: on your computer or stored online (for instance as email attachment)? Order the following arguments from 1 to 6, according to their relevance for you, where 1 is the one you most agree with.**
   On my computer, because I can look after it and physically protect my data, whereas online I cannot see where it is actually stored or who has access to it.
   On my computer, because I can disconnect it from the Internet, whereas online it is always exposed to hackers.
   On my computer, because hackers target big companies. They would need to identify my computer first, and they don't know where I am.
   Online, because my computer might crash or somebody might steal it and then I would lose all my data, but if I put it online I can always access it.
   Online, because big companies have more security experts and can guarantee better protection than what I could do for my laptop.
   Online, because on those servers there are many documents, from many users. Nobody would have the time to look at mine.
   Select the correct option
2. **When you delete a file stored on the Internet or an email in your Webmail account, what do you think happens?**

   The file gets permanently deleted just as when I would delete it from my computer.
   Copies will still be kept for security reasons, in case they are ever needed in criminal investigations.
   Some copies might still exist, but only for a few weeks, until the company manages

to delete all of them.
I don't know.
Other (please specify)

3. **Google began in January 1996 as a research project by Larry Page and Sergey Brin. Its initial public offering took place on August 19, 2004. In which year did the initial public offering of Google take place?**

   1996 1998 2004 2006 2011

4. **You want to open a new account with a company that provides storage space for personal documents on a server on the Internet. You have come across these two companies. Which one do you choose and why?**

   Company A: Offers the service for free, but their privacy policy says that they may sell, transfer or share your personal information and documents to another company.
   Company B: Asks you to pay $20 per year. Their privacy policy says that they will not sell, transfer nor share any of your personal information to another companies.
   Company A, because it is free.
   Company A, because I don't have sensitive data anyway.
   Company A, because I can never be sure what they do with my data anyway.
   Company B, because I value my privacy.
   Company B, if the price was lower.
   Company B, if I am sure they are trustworthy.
   Other (please specify)

5. **Does your Webmail provider have the right to see or modify the documents you have as attachments in your email account?**
   They don't have the right to look at nor modify any of my documents.
   They can see them, but not modify them, because these are my documents and they belong to me, even if I store them there.
   They have the right to see and modify my documents only in criminal or terrorists cases.
   They have the right to see and modify any of the documents I store.
   I don't know.
   Other (please specify)

6. **Does your Webmail provider have the right to disable your account?**
   Yes, at any time, without advanced notice and without explanation.
   Yes, but only with advanced notice and a valid reason.
   Only if I am using it for criminal purposes.
   No.
   I don't know.
   Other (please specify)

7. **If your Webmail provider lost some of the data you store with them, what would your rights be?**

They should pay me for the damages, regardless whether it was a paid for or free service. We had a contract.
If it is a free service, I have no rights, but if I paid for it, they would have to pay me for the damages.
I have no rights even if it is a paid-for service. There are no guarantees.
My data is lost anyway. I wouldn't care about money. An apology would be enough.
I don't know.
Other (please specify)

8. **How much do you agree with each of the following statements?**
   Mark on the likert scale(Strongly agree, somewhat agree, somewhat disagree, strongly disagree, N/A)

   I try to keep local backups of every important document I store on the Internet.
   I try not to store important, sensitive documents on the Internet, and instead keep them offline, on my personal computers.
   Most businesses handle the personal information they collect about customers in a proper and confidential way.
   If people put their private data on the Internet and it gets hacked, it is their own fault. They should know that nothing is really safe on the Internet
   There is no such thing as consumer protection service or police on the Internet whom I could turn to, if I felt that my rights were violated.
   If the government had access to every document users store on the Internet, that would be a major violation of individual privacy.
   Consumers have lost all control over how personal information is collected, circulated and used by companies.
   Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.
   It is good if the government monitors every Internet communication and all user accounts. National security comes first.

9. **How important do you consider the data you store online (data that is private, not for public access on the Internet)?**
   Mark it on a likert-scale one to seven:
   Very Important to Only spam or things I can live without

10. **What is your age?** 18-24 25-34 35-44 45-54 55+

11. **What is your gender?** Male Female

12. **What is your nationality?**

13. **What country do you live in?**

14. **What is the highest education degree that you completed?**

    High school Bachelor Masters PhD Other (please specify)

15. **How would you rate your computer skills?**

Novice Intermediate Proficient Expert Comment

16. **What Webmail accounts do you use?**

Yahoo Mail Gmail Hotmail AOL Other (please specify)

17. **Which is the main Webmail account you use?**

18. **Do you use any of the following systems for storing your documents online?**

Dropbox FolderShare GoogleDocs Other (please specify)

19. **Three survey participants will be randomly selected to win a USD 100 Amazon vouchers. If you want to take part in the lucky draw, please specify an email address or phone number where we could contact you. All data collected during this survey will be anonymized and aggregated. Your answers are treated confidentially and used for research purposes only. We will not use your contact information for any other purposes but to contact you to collect**

B. XPath-based Steganography Rules

```
<!-- GMAIL -->
- <page id="email">
    <url>https://mail.google.com/*</url>
  - <block id="mailMessage">
      <region>//div[@class='gs']</region>
      <sender>.//div[@class='iw']/span[@class='gD']/@email</sender>
    - <data type="text">
        .//*[contains(@class,'ii gt')]//div[not(contains(@class,'gmail_quote'))]/text()
      </data>
    </block>
  - <block id="chatMessage">
      <region>//*[contains(@class,'aec')]</region>
      <sender> .//@title</sender>
      <data type="text">.//div[@class='kl']</data>
    </block>
  </page>
```

Figure B.1.: XPath-based webpage parsing rules for Gmail

```
- <pages>
    <!-- TWITTER -->
  - <page id="tweet">
      <url>twitter.com/*</url>
    - <block id="twitt">
        <region>//div[@class='tweet-content']</region>
        <sender>.//div[@class='twitt-user-name']</sender>
        <data type="text">.//*[contains(@class,'twitt-text')]</data>
      </block>
    </page>
  </pages>
```

Figure B.2.: XPath-based webpage parsing rules for Twitter

```
- <pages>
    <!-- FACEBOOK -->
  - <page id="messages">
      <url>www.facebook.com/messages/</url>
    - <block id="message">
        <region>//*[contains(@class,'MessagingMessage')]</region>
        <sender>.//strong/a</sender>
        <data type="text">.//*[contains(@class,'uiListItem')]</data>
      </block>
    </page>
  - <page id="wall">
      <url>www.facebook.com/*</url>
    - <block id="wallPost">
        <region>//div[@class='mainWrapper']</region>
        <sender>.//div[contains(@class, 'actorName')]/a</sender>
        <data type="text">.//span[@class='messageBody']</data>
      </block>
    - <block id="messageComment">
        <region>//div[contains(@class, 'commentContent')]</region>
        <sender>.//div[contains(@class, 'actorName')]/a</sender>
        <data type="text">.//span[@class='commentBody']</data>
      </block>
    - <block id="tickerFeedMessage">
        <region>//*[@class='tickerFeedMessage']</region>
        <sender>.//*[contains(@class, 'actorName')]</sender>
        <data type="text">.//span[@class='messageBody']</data>
      </block>
    - <block id="chatMessage">
        <region>//*[contains(@class,'fbChatMessageGroup')]</region>
        <sender>.//*[contains(@class,'actorName')]/a</sender>
      - <data type="text">
          .//*[contains(@class,'messages')]//div[contains(@class, 'fbChatMessage')]
        </data>
      </block>
    </page>
  - <page id="album">
      <url>www.facebook.com/photo.php*</url>
    - <block id="image">
        <region>/html</region>
      - <sender>
          //*[contains(@id, 'fbPhotoSnowboxAuthorName') or contains(@id, 'fbPhotoPageAuthorName')]/a
        </sender>
      - <data type="img">
          //img[contains(@class, 'spotlight') or contains(@class, 'fbPhotoImage')]/@src
        </data>
      </block>
    </page>
  </pages>
```

Figure B.3.: XPath-based webpage parsing rules for Facebook

C. Device Pairing Study Script

C.1. Allgemeines

In dieser Studie, werden Sie vier Methoden kennenlernen, um drahtlose Verbindungen zwischen Geräten zu erstellen.

Sie könnten zum Beispiel solche Methoden benutzen, um beim gemeinsamen Mittagessen im Restaurant einige Fotos Ihrer Familie von Ihrem Handy auf das Handy eines guten Freundes zu übertragen.

Das Problem beim Verwenden solcher drahtlosen Verbindungen ist, die zwei *richtigen* Geräte zu verbinden. In einem beliebten Lokal können schnell mehrere Dutzend Handys auf engstem Raum zusammen kommen und es ist leicht möglich, die Familienfotos aus versehen auf ein Handy am Nachbartisch zu senden. Die Methoden die Sie hier verwenden werden, dienen dazu sicherzustellen, dass Ihre Verbindungen tatsächlich zwischen den von Ihnen gewählten Handys erstellt wird.

Unser Ziel ist es herauszufinden, welche der hier vorgestellten Methoden Sie bevorzugen. Dazu werde ich Sie nun gleich bitten, drei kleine Aufgaben zu erfüllen, in denen zwei Geräte in einer Reihe von verschiedenen Alltagssituationen verbunden werden sollen.

Bevor wir beginnen, werde ich gerne ein Paar weitere Punkte erklären:

1. Ihre Teilnahme in dieser Benutzerstudie ist freiwillig. Sie können die Studie jederzeit abbrechen, ohne uns Gründe zu nennen.

2. Wenn Sie die Studie vollenden, werden Sie mit 20 CHF belohnt.

3. Sie können mich jederzeit unterbrechen, wenn Sie Fragen haben.

4. Whärend der Studie können wir jederzeit eine Pause einlegen. Geben Sie mir einfach bescheid.

5. Ein ganz wichtiger Punkt vorweg: Es geht in dieser Studie nicht darum, *Sie* und ihre Fähigkeiten zu testen. Wenn etwas nicht funktioniert oder Sie bei einer Aufgabe Schwierigkeiten haben, liegt das an den offensichtlich noch nicht ausgereiften Methoden. Diese Probleme aufzudecken ist das Ziel unserer Studie.

6. Am Ende der Aufgaben werde ich Ihnen Fragen über Ihre Erfahrungen stellen. Bitte sagen Sie uns offen Ihre Meinung. Dies ist uns sehr wichtig!

Bitte lesen Sie die Einverständniserklärung und unterschreiben Sie sie.

Jetzt würde ich Sie bitten den Fragebogen auszufüllen.

C.2. Studieneinführung

Die Studie wird wie folgt laufen:

- Sie werden diese zwei Mobiltelefone benutzen. Die Geräte sollen drahtlos miteinander verbunden werden.
- Bitte verwenden Sie die Mobiltelefone nur wie ich es Ihnen sage. (Drcken Sie nur die Taste die ich Ihnen zeige.)
- Wenn etwas nicht funktioniert, ist es nicht Ihre Schuld. Geben Sie mir bitte einfach das Gerät, so dass ich das Problem beheben kann.
- Die Studie wird ungefähr 60 Minuten dauern.

Haben Sie Fragen, bevor wir beginnen?

C.3. Die Methoden

Sie werden jetzt die vier Methoden kennenlernen. Um eine drahtlose Verbindung zu erstellen, werden Sie die zwei Mobiltelefonen benutzen. Dieses Mobiltelefon hier werde ich von jetzt an IHR GERÄT nennen. Sie müssen IHR GERÄT mit DEM ANDEREN GERÄT verbinden.

Die Tasten, die für Sie wichtig sind, sind der AUSWAHL-Taste und die vier PFEIL-Tasten.

Bitte gehen Sie wie folgt vor:

Die erste Methode heisst....

C.3.1. Die Wähl das Gerät Methode

Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. IHR GERÄT beginnt, nach anderen Geräten in Ihrer Umgebung zu suchen. IHR GERÄT zeigt eine Gerätliste an. Drücken Sie die NACH OBEN und UNTEN Pfeile an IHREM GERÄT und wählen Sie den Namen des ANDEREN GERÄTES, der auf dem Bildschirm des ANDEREN GERÄTES steht. Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

C.3.2. Die Drück den Knopf Methode

Drücken Sie die AUSWAHL-Taste an IHREM GERÄT, woraufhin ein Einführungstext angezeigt wird. IHR GERÄT wird von 3 auf 1 herunterzählen und anschliessend drei Mal folgenden Ablauf erwarten: Sobald IHR GERÄT beginnt zu vibrieren, drücken Sie bitte die AUSWAHL-Taste am ANDEREN GERÄT. Es ist wichtig, dass Sie schnell reagieren, sonst wird die Verbindung fehlschlagen. Warten Sie auf die nächsten Vibrationen und wiederholen Sie das Ganze.

Um zu beginnen, drücken Sie jetzt bitte noch mal die AUSWAHL-Taste an IHREM GERÄT. Durch Ihr Knopfdrücken, synchronisieren sich die zwei Geräte und stellen eine drahtlose Verbindung her. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

C.3.3. Die Mach ein Foto Methode

Drücken Sie die AUSWAHL Taste an IHREM GERÄT. DAS ANDERE GERÄT wird einen Barcode anzeigen. Gleichzeitig, startet IHR GERÄT die Kamera. Zielen Sie mit der Kamera IHRES GERÄTES so, dass Sie den Barcode auf der Anzeige IHRES GERÄTES sehen können. (Spielen Sie mit dem Abstand.) Die Kamera wird dabei selbständig fokussieren. Drücken Sie die AUSWAHL-Taste an IHREM GERÄT, um ein Foto zu machen. (Versuchen Sie noch mal.) IHR GERÄT verarbeitet das Foto, welches eine Nachricht enthält, und stellt die Verbindung mit DEM ANDEREN GERÄT her. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

C.3.4. Die Hör zu Methode

Wählen Sie die Methode und drücken Sie den AUSWAHL-Taste an IHREM GERÄT. Halten Sie dabei Ihr Gerät in die nähe des anderen Gerätes. DAS ANDERE GERÄT spielt eine drei Sekunden lange Melodie, auf welche Sie nicht weiter achten müssen. IHR GERÄT nimmt die Melodie auf und dekodiert die enthaltene Nachricht. IHR GERÄT verarbeitet diese und stellt eine Verbindung mit DEM ANDEREN GERÄT her. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

Nun würde ich Sie bitten, noch mal alle 4 Methoden selbstständig durchzuführen.

Haben Sie Fragen?

C.4. Einführung zu den Aufgaben

Ich werde Ihnen drei hypothetische Situationen vorstellen und bitte Sie so zu tun, als ob Sie wirklich in diesen Situation wären.

Bitte wählen Sie für jede der Situationen eine der vier vorgestellten Methoden und stellen Sie eine drahtlose Verbindung zwischen beiden Geräten her. Wählen Sie die Methode, die Sie auch im wirklichen Leben nutzen würden.

[For each task, print pictures and show them while you explain the scenario.]

C.4.1. Aufgabe 1: Finanzbericht drucken

Sagen wir, dass Sie für eine Beratungsfirma arbeiten. Sie sind am Flughafen und werden bald in Ihr Flugzeug steigen. Sie werden nach London fliegen, um Ihren Kunden zu besuchen. Sie haben den vertraulichen Finanzbericht Ihres Kunden auf Ihrem Mobiltelefon gespeichert. Im Wartebereich befindet sich ein Drucker. Verbinden Sie Ihr Mobiltelefon mit dem Drucker, so dass Sie den Finanzbericht drahtlos an den Drucker senden können.

Stellen Sie sich vor, dass das andere Gerät das Display des Druckers darstellt.

Interview

Vielen Dank dass Sie die Aufgabe erledigt haben. Jetzt werde ich Ihnen ein paar Fragen über Ihre Erfahrung stellen. Ausgewählte Methode: PK, VIC, HAPADEP oder BEDA

1. Warum haben Sie diese Methode gewählt? (über welche andere Kriterien haben Sie nachgedacht?)
2. Würden Sie eine andere Methode benutzen, wenn Sie zum Beispiel einen weiteren Bericht drucken wollen?
3. Hätten Sie eine andere Methode gewählt, wenn Sie allein in Ihrem Büro anstatt einem öffentlichen Ort wären?
4. Welche Methode wäre Ihre zweite Wahl?

C.4.2. Aufgabe 2: Mit Handy bezahlen

In London werden Sie auch einen guten Freund besuchen. Sie wollen ihm eine Flasche Whisky im Duty-Free kaufen, bevor Sie ins Flugzeug steigen. Sie hören, dass das Boarding Ihres Fluges soeben begonnen hat. Verbinden Sie Ihr Mobiltelefon mit der Kasse, um die Flasche zu bezahlen.

Stellen Sie sich vor, dass das andere Gerät das Display der Kasse darstellt.

Interview

1. Warum haben Sie diese Methode gewählt?
2. Würden Sie eine andere Methode nutzen, wenn Sie zum Beispiel eine neue Zahlung durchführen wollen?
3. Hätten Sie eine andere Methode gewählt, wenn Sie nicht in Eile gewesen wären Ihren Flug zu erwischen?
4. (Welche Methode würden Sie wählen, wenn Sie nachts an einer Tankstelle zahlen würden? Es gibt keine anderen Kunden die nach Ihnen warten.)

C.4.3. Aufgabe 3: Visitenkarte austauschen

Sie sind jetzt in London bei Ihrem Kunden. Auf einer Veranstaltung lernen Sie den CEO einer anderen Firma kennen, der an einer Geschäftsbeziehung interessiert ist. Daher wollen Sie Ihre elektronischen Visitenkarten austauschen. Benutzen Sie Ihr Mobiltelefon und stellen Sie eine drahtlose Verbindung mit seinem Handy her.

Interview

1. Warum haben Sie diese Methode gewählt?
2. Würden Sie eine andere Methode nutzen, wenn Sie zum Beispiel Visitenkarten mit jemand anders austauschen wollten?
3. Hätten Sie eine andere Methode gewählt, wenn Sie statt mit dem CEO mit Ihrem Freund Adressen austauschen wollten?
4. Welche Methode hätten Sie gewählt, wenn Sie mit dem CEO alleine in seinem Büro gewesen wären?

5. Was ist Ihr genereller Eindruck von den vier Methoden?
6. Haben Sie sich in den drei Situationen Sorgen bezüglich der Sicherheit der Verbindung gemacht?
7. Wenn ja, welche und wie hat das Ihre Auswahl beeinflusst?
8. Besitzen Sie ein Laptop? [Sie haben gesagt dass Sie ein Laptop besitzen.] Haben Sie sich jemals beim Verbinden von Geräten Sorgen bezüglich der Sicherheit der Verbindung gemacht?
9. Haben Sie noch etwas anzumerken?

C.5. Sicherheitsstandard Anpassen

So wie wir die Methoden bisher genutzt haben, gab es keine Sicherheitsgarantien. Jeder mit geeigneten Werkzeugen könnte Ihre Kommunikation abhören und womöglich verändern. Unbekannte könnten so z.B. Ihren Finanzbericht, Ihr Kreditkartendaten oder Ihre Visitenkarte sehen, und diese potentiell whärend der übertragung sogar verändern.

Nun werde ich Ihnen zeigen, wie Sie das Mass an Sicherheit bestimmen können.

Allgemein, gibt es drei Sicherheitsstufe: Keine Sicherheit (genau wie früher), mittlere Sicherheit und hohe Sicherheit. Gehen wir nun die Methoden noch einmal durch. Bitte folgen Sie die folgenden Schritte, um eine Gerätverbindung zu erstellen.

Die Wähl das Gerät Methode

Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. IHR GERÄT zeigt einen Bildschirm an, in dem Sie den gewünschten Sicherheitsstufe auswählen können. Drücken Sie die Pfeiltasten LINKS und RECHTS, um zwischen den drei verfügbaren Sicherheitsstufen zu wählen: unsicher (nur das Gerät auswählen), sicher (Gerät auswählen und 6-stellige PIN eintippen) und sehr sicher (Gerät auswählen und 9-stellige PIN eintippen). Wählen Sie mittlere Sicherheit. Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. IHR GERÄT beginnt, nach anderen Geräten in Ihrer Umgebung zu suchen. IHR GERÄT zeigt eine Gerätliste an. Drücken Sie die NACH OBEN und UNTEN Pfeile an IHREM GERÄT und wählen Sie den Namen des ANDEREN GERÄTES. Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. [IHR GERÄT zeigt eine 6- oder 9-stellige Nummer. Tippen Sie diese am ANDEREN GERÄT ein.] Sie können die letzte Ziffer durch die C-Taste DES ANDEREN GERÄTES löschen. Wenn Sie fertig sind, drücken Sie die AUSWAHL-Taste am ANDEREN GERÄTES.] Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

Die Drück den Knopf Methode

Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. IHR GERÄT zeigt einen Bildschirm an, in dem Sie die gewünschte Sicherheitsstufe auswählen können. Drücken Sie die Pfeiltasten LINKS und RECHTS, um zwischen den drei verfügbaren Sicherheitsstufen zu wählen: unsicher (3-maliges Knopfdrücken), sicher (6-maliges Knopfdrücken) und sehr sicher (9-maliges Knopfdrücken). Wählen Sie mittlere Sicherheit. Drücken Sie die AUSWAHL-Taste an IHREM GERÄT. IHR GERÄT zeigt einen Einführungstext an. IHR

GERÄT wird von 3 auf 1 herunterzählen und anschliessend drei Mal folgenden Ablauf erwarten: Sobald IHR GERÄT beginnt zu vibrieren, drücken Sie bitte die AUSWAHL-Taste am ANDEREN GERÄT. Es ist wichtig, dass Sie schnell reagieren, sonst wird die Verbindung fehlschlagen. Warten Sie auf die nächsten Vibrationen und wiederholen Sie das Ganze. Um zu beginnen, drücken Sie jetzt bitte noch mal die AUSWAHL-Taste an IHREM GERÄT. Durch Ihr Knopfdrücken, synchronisieren sich die zwei Geräte und stellen eine drahtlose Verbindung her. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

Die Mach ein Foto Methode

Drücken Sie die AUSWAHL Taste an IHREM GERÄT. IHR GERÄT zeigt einen Bildschirm an, in dem Sie die gewünschte Sicherheitsstufe auswählen können. Drücken Sie die Pfeiltasten LINKS und RECHTS, um zwischen den drei verfügbaren Sicherheitsstufen zu wählen: unsicher (ein Foto aufnehmen), sicher (2 Fotos aufnehmen) und sehr sicher (3 Fotos aufnehmen). Wählen Sie mittlere Sicherheit. Um zu beginnen, drücken Sie jetzt bitte noch mal die AUSWAHL-Taste an IHREM GERÄT.

DAS ANDERE GERÄT wird einen Barcode anzeigen. Gleichzeitig, startet IHR GERÄT die Kamera. Zielen Sie mit der Kamera IHRES GERÄTES so, dass sie den Barcode auf der Anzeige IHRES GERÄTES sehen können. Drücken sie den AUSWAHL Knopf an IHREM GERÄT, um ein Foto zu machen. (Spielen Sie mit dem Abstand.) [Drücken sie die AUSWAHL-Taste an IHREM GERÄT, um ein weiteres Foto zu machen.] IHR GERÄT verarbeitet das Foto, welches eine Nachricht enthält, und stellt die Verbindung mit DEM ANDEREN GERÄT her. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

Die Hör zu Methode

Wählen Sie die Methode und drücken Sie den AUSWAHL-Taste an IHREM GERÄT. IHR GERÄT zeigt einen Bildschirm an, in dem Sie die gewünschte Sicherheitsstufe auswählen können. Drücken Sie die Pfeiltasten LINKS und RECHTS, um zwischen den drei verfügbaren Sicherheitsstufen zu wählen: unsicher (3-sekündige Melodie), sicher (6-sekündige Melodie) und sehr sicher (9-sekündige Melodie). Wählen Sie mittlere Sicherheit. Um zu beginnen, drücken Sie jetzt bitte noch mal die AUSWAHL-Taste an IHREM GERÄT.

DAS ANDERE GERÄT spielt eine drei, sechs oder neun Sekunden lange Melodie, auf welche Sie nicht weiter achten müssen. IHR GERÄT nimmt die Melodie auf und dekodiert die enthaltene Nachricht. IHR GERÄT verarbeitet diese und stellt eine Verbindung mit DEM ANDEREN GERÄT her. Sie haben jetzt die Verbindung zwischen den beiden Geräten erfolgreich hergestellt.

Nun würde ich Sie bitten, noch mal alle 4 Methoden selbstständig durchzuführen, diesmal mit mittlerer Sicherheitsstufe.

C.5.1. Aufgaben - mit Sicherheit

Bitte denken Sie noch mal an die drei Aufgaben und wählen Sie die geeignete Methode und den gewünschten Sicherheitsstufe.

[Run through the methods and re-read the tasks]

Aufgabe 1 (Finanzbericht drucken): Interview

1. Warum haben Sie diese Methode und Sicherheitsstufe gewählt? (über welche Kriterien haben Sie nachgedacht?)
2. Würden Sie auch eine der andere Methoden nutzen, zum Beispiel wenn Sie einen weiteren Bericht drucken wollten?
3. Auf einer Skala von 1 bis 7 (wobei 1 der niedrigste und 7 der höchste Wert ist), wie sicherheitskritisch schätzen Sie den Finanzbericht ein? Warum?
4. Auf einer Skala von 1 bis 7 (wobei 1 der niedrigste und 7 der höchste Wert ist), wie besorgt sind Sie, dass jemanden Ihre drahtlos gesendeten Daten einsehen könnte? Warum?
5. Welche Methode und Sicherheitsstufe hätten Sie gewählt, wenn Sie zum Beispiel Ihre eigene Steuererklärung gedruckt hätten?
6. Sortieren Sie von 1 (am relevantesten) bis 5 (am wenigsten relevant) die folgenden Kriterien beim Drucken:
   - Einfache Nutzung
   - Schnelligkeit
   - Sicherheit
   - Professionelles Design/Erscheinungsbild
   - Spass
   - Andere ?
7. Hätten Sie eine andere Methode und/oder Sicherheitsstufe gewählt, wenn Sie allein in Ihrem Büro wären und nicht an einem öffentlichen Ort?
8. Hätten Sie eine andere Methode und/oder Sicherheitsstufe gewählt, wenn Sie zu Hause anstatt an einem öffentlichen Ort gewesen wären?

Aufgabe 2 (Mit Handy Zahlen): Interview

1. Warum haben Sie diese Methode und Sicherheitsstufe gewählt?
2. Wenn Sie ein weiteres Geschenk vor Ihrer Rückfahrt kaufen würden, welche Methode würden Sie nutzen?
3. Auf einer Skala von 1 bis 7 (wobei 1 der niedrigste und 7 der höchste Wert ist), wie sicherheitskritisch schätzen Sie die Kreditkarteinformation ein? Warum?
4. Finden Sie die Zahlung sicherheitskritischer als den Finanzbericht? Warum? Warum nicht?
5. Sagen wir dass Sie statt der teuren Whiskyflasche jetzt eine Schachtel Zigaretten kaufen. Welche Methode und Sicherheitsstufe benutzen Sie?
6. Auf einer Skala von 1 bis 7 (wobei 1 der niedrigste und 7 der höchste Wert ist), wie sicherheitskritisch schätzen Sie die Zahlung ein? Warum?

7. Auf einer Skala von 1 bis (wobei 1 der niedrigste und 7 der höchste Wert ist), wie besorgt sind Sie, dass jemand Ihre Kreditkarteinformation, die Sie drahtlos gesendet haben, abfangen könnte, wenn Sie keine Sicherheit benutzt hätten? Und mit Sicherheit?
8. Sortieren Sie folgenden Kriterien bei der Zahlung nach ihrer Relevanz für Sie (1 am relevantesten, bis 5 am wenigsten relevant):
   - Einfache Nutzung
   - Schnelligkeit
   - Sicherheit
   - Professionelles Design/Erscheinungsbild
   - Spass
   - Andere ?
9. Hätten Sie eine andere Methode oder Sicherheit gewählt, wenn Sie nicht in Eile gewesen wären Ihren Flug zu erwischen?
10. Welche Methode und Sicherheitsstufe würden Sie wählen, wenn Sie an einer Tankstelle zahlen würden? Es gibt keine anderen Kunden die nach Ihnen warten.

Aufgabe 3 (Visitenkarten Austauschen): Interview

1. Warum haben Sie diese Methode und Sicherheitsstuffe gewählt?
2. Auf einer Skala von 1 bis 7 (wobei 1 der niedrigste und 7 der höchste Wert ist), wie sicherheitskritisch schützen Sie die Informationen auf Ihren Visitenkarte ein? Warum?
3. Sehen Sie ein höheres oder niedrigeres Risiko als in der vorherigen Aufgabe, in der Sie zahlten?
4. Wäre es Ihnen unangenehm oder gar peinlich beim Austausch von Visitenkarten eine andere Methode zu verwenden?
5. Welche Methode und Sicherheitsstufe hätten Sie gewählt, wenn Sie zum Beispiel Ihre eigene Steuererklärung gedruckt hätten?
6. Sortieren folgenden Kriterien beim Austausch der Visitkarten nach ihrer Relevanz für Sie, beginnend mit dem wichtigsten:
   - Einfache Nutzung
   - Schnelligkeit
   - Sicherheit
   - Professionelles Design/Erscheinungsbild
   - Spass
   - Andere ?
7. Welche Methode und Sicherheitsstufe hätten Sie gewählt, wenn Sie mit dem CEO alleine in seinem Büro gewesen wären?

8. Welche Methode und Sicherheitsstufe hätten Sie gewählt, wenn Sie mit dem CEO in einem Cafe gewesen wären?

C.5.2. Debriefing

Das war alles! Vielen Dank für Ihre Teilnahme.

Haben Sie noch weitere Bemerkungen?

Gibt es etwas, was Ihnen besonders gefallen oder auch besonders missfallen hat, eine der Methoden? Fanden Sie etwas besonders praktisch oder unpraktisch?

Noch Mal vielen Dank und einen schönen Tag.

Bibliography

[1] 2peer. http://2peer.com. Accessed on August 23, 2012.

[2] Alessandro Acquisti and Ralph Gross. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Proceedings of the 6th Workshop on Privacy Enhancing Technologies (PET 2006)*, pages 36–58, Cambridge, UK, June 2006.

[3] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42:40–46, 1999.

[4] Robert Adelmann, Marc Langheinrich, and Christian Floerkemeier. Toolkit for bar code recognition and resolving on camera phones—Jump starting the Internet of Things. In *GI Jahrestagung (2)*, pages 366–373, 2006.

[5] Lilian Adkinson-Orellana, Daniel A. Rodriguez-Silva, Francisco J. Gonzalez-Castano, and David Gonzalez-Martinez. Sharing secure documents in the cloud - a secure layer for Google Docs. In *Proceedings of the International Conference on Cloud Computing and Services Science*, pages 439–444, Noordwijkerhout, The Netherlands, May 2011.

[6] Chloe Albanesius. Google: Software bug caused Gmail deletions. PCMAG News, March 1, 2011. Available at: http://www.pcmag.com/article2/0,2817,2381168,00.asp. Accessed on August 23, 2012.

[7] Janna Q. Anderson and Lee Rainie. The 2011 cloud trends and best practices report. Available at: http://www.hosting.com/resources/ebooks/2011-cloud-computing-trends-report. Accessed on August 23, 2012.

[8] Janna Q. Anderson and Lee Rainie. The future of cloud computing. Pew Research Center, June 2010. Available at: http://pewinternet.org/Reports/2010/

The-future-of-cloud-computing.aspx. Accessed on August 23, 2012.

[9] Annie I. Anton, Julia B. Earp, Qingfeng He, William Stufflebeam, Davide Bolchini, and Carlos Jensen. Financial privacy policies and the need for standardization. *IEEE Security and Privacy*, 2(2):36–45, March 2004.

[10] Annie I. Antón, Julia B. Earp, and Jessica D. Young. How Internet users' privacy concerns have evolved since 2002. *IEEE Security and Privacy*, 8(1):21–27, January 2010.

[11] Ching Man Au Yeung, Ilaria Liccardi, Kanghao Lu, Oshani Seneviratne, and Tim Berners-Lee. Decentralization: The future of online social networking. In *Proceedings of the W3C Workshop on the Future of Social Networking*, Barcelona, Spain, January 2009.

[12] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. Persona: an online social network with user-defined privacy. *ACM SIGCOMM Computer Communication Review*, 39(4):135–146, 2009.

[13] Dirk Balfanz, Glenn Durfee, Rebecca E. Grinter, D. K. Smetters, and Paul Stewart. Network-in-a-box: how to set up a secure wireless network in under a minute. In *Proceedings of the 13th USENIX Security Symposium*, pages 207–221, San Diego, CA, August 2004.

[14] Dirk Balfanz, Glenn Durfee, and Diana Smetters. Making the impossible easy: Usable PKI. In *Security and Usability: Designing Secure Systems that People Can Use*, pages 319–334. O'Reilly, 2005.

[15] Adam Barth, Dan Boneh, and Brent Waters. Privacy in encrypted content distribution using private broadcast encryption. In *Proceedings of the 10th International Conference on Financial Cryptography and Data Security*, pages 52–64, Anguilla, British West Indies, March 2006.

[16] Lujo Bauer, Lorrie F. Cranor, Michael K. Reiter, and Kami Vaniea. Lessons learned from the deployment of a smartphone-based access-control system. In *Proceedings the Symposium on*

Usable Privacy and Security (SOUPS 2007), pages 64–75, Pittsburgh, PA, July 2007.

[17] Filipe Beato, Markulf Kohlweiss, and Karel Wouters. Enforcing access control in social networks. In *Proceedings of the 9th Privacy Enhancing Technologies Symposium (HotPETs 2009)*, Seattle, WA, August 2009.

[18] Filipe Beato, Markulf Kohlweiss, and Karel Wouters. Scramble! your social network data. In *Proceedings of 11th Privacy Enhancing Technologies Symposium*, Waterloo, Canada, July 2011.

[19] Steven Bellman, Eric J. Johnsonb, Stephen J. Kobrinc, and Gerald L. Lohse. International differences in information privacy concerns: A global survey of consumers. *The Information Society*, 20:313–324, 5 November 2004.

[20] Fabrício Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgílio Almeida. Detecting spammers on Twitter. In *Proceedings of the 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS 2010)*, Redmond, WA, July 2010.

[21] Tamas Besenyei, Adam Mate Foldes, Gabor Gyorgy Gulyas, and Sandor Imre. StegoWeb: Towards the ideal private web content publishing tool. In *Proceedings of the 5th Int. Conference on Emerging Security Information, Systems and Technologies (SECUREWARE 2011)*, French Riviera, France, August 2011.

[22] Davic Blei and John Lafferty. A correlated topic model of science. *Annals of Applied Statistics*, 1:17–35, 2007.

[23] David M. Blei. Probabilistic topic models. *Communications of the ACM*, 55:77–84, 2012.

[24] Bluetooth SIG. Bluetooth Special Interest Group. Simple Pairing Whitepaper (Revision V10r00), 2006.

[25] Danah Boyd and Alice Marwick. Social steganography: Privacy in networked publics. In *International Communication Association*, Boston, MA, May 2011.

[26] Bump. `http://bu.mp/`. Accessed on September 3, 2012.

[27] Jon Callas, Lutz Donnerhacke, Hal Finney, David Shaw, and Rod-

ney Thayer. OpenPGP Message Format. RFC 4880 (Proposed Standard). Available at: `http://www.ietf.org/rfc/rfc4880.txt`, November 2007. Updated by RFC 5581.

[28] Rajarshi Chakraborty, Srilakshmi Ramireddy, Santanam T. Raghu, and H. Raghav Rao. The information assurance practices of cloud computing vendors. *IT Professional*, 12:29–37, 2010.

[29] J. Chang and D. Blei. Hierarchical relational models for document networks. *Annals of Applied Statistics*, 4(1):124–150, 2010.

[30] Richard Chow, Philippe Golle, Markus Jakobsson, Elaine Shi, Jessica Staddon, Ryusuke Masuoka, and Jesus Molina. Controlling data in the cloud: outsourcing computation without outsourcing control. In *Proceedings of the 2009 ACM workshop on Cloud computing security (CCSW 2009)*, Chicago, IL, USA, November 2009.

[31] Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016. CISCO White Paper. February 14, 2012. Available at: `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf`. Accessed on September 3, 2012.

[32] Efthymios Constantinides, Maria del Carmen Alarcón del Amo, and Carlota Lorenzo Romero. Profiles of social networking sites users in the netherlands. In *Proceedings of the 18th Annual High Technology Small Firms Conference (HTSF 2010)*, Enschede, The Netherlands, May 2010.

[33] Mauro Conti, Arbnor Hasani, and Bruno Crispo. Virtual private social networks. In *Proceedings of the First ACM Conference on Data and Application Security and Privacy (ACM CODASPY 2011)*, San Antonio, TX, 2011.

[34] Emiliano De Cristofaro, Claudio Soriente, Gene Tsudik, and Andrew Williams. Hummingbird: Privacy at the time of Twitter. In *Proceedings of the IEEE Symposium on Security and Usability*, San Francisco, CA, May 2012.

[35] Leucio A Cutillo, Refik Molva, and Thorsten Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *IEEE Communications Magazine*, 47(12):94–101, 2009.

[36] Gabriele D'Angelo, Fabio Vitali, and Stefano Zacchiroli. Content cloaking: Preserving privacy with Google Docs and other Web applications. In *Proceedings of the 25th ACM Symposium on Applied Computing (SAC 2010)*, Sierre, Switzerland, 2010.

[37] Data Security Council of India (DSCI). Data protection challenges in cloud computing. December 2010. Available at: `http://www.dsci.in/node/539`. Accessed on September 3, 2012.

[38] Deloitte. Raising the bar. 2011 TMT global security study—key findings. Available at: `http://www.deloitte.com/view/en_YE/ye/industries/technology-media-telecommunications/38bd76226faf3310VgnVCM2000001b56f00aRCRD.htm`. Accessed on September 3, 2012.

[39] Diaspora. `https://joindiaspora.com/`. Accessed on September 3, 2012.

[40] How many users are in the diaspora network? Data as of August 23, 2012. Available at: `https://diasp.eu/stats.html`. Accessed on September 3, 2012.

[41] Tamara Dinev, Paul Hart, and Michael R. Mullen. Internet privacy concerns and beliefs about government surveillance - an empirical investigation. *Journal of Strategic Information Systems*, 17(3):214–233, 2008.

[42] DocCloak. `http://www.gwebs.com/doccloak.html`. Accessed on September 3, 2012.

[43] Saar Drimer and Steven J. Murdoch. Keep your enemies close: distance bounding against smartcard relay attacks. In *Proceedings of 16th USENIX Security Symposium*, Boston, MA, August 2007.

[44] Dropbox. `https://www.dropbox.com`. Accessed on September 3, 2012.

[45] Where does Dropbox store everyone's data? Available at: `https://www.dropbox.com/help/7/en`. Accessed om September 12, 2012.

[46] Charles Duhigg. How companies learn your secrets. The New York Times, February 16, 2012. Available at: `http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html`.

[47] Jim Dwyer. Four nerds and a cry to arms against Facebook. May 11, 2010. Available at: http://www.nytimes.com/2010/05/12/nyregion/12about.html. Accessed on September 3, 2012.

[48] Carl M. Ellison. Establishing identity without certification authorities. In *Proceedings of the 6th USENIX Security Symposium*, San Jose, CA, July 1996.

[49] Emarketing Solutions. http://www.americaint.com/. Accessed on September 3, 2012.

[50] Emarketing Solutions. UK Email List. http://www.americaint.com/worldwideemaillists/uk-consumer-email-list.html. Accessed on September 3, 2012.

[51] Richard Esguerra. Google CEO Eric Schmidt dismisses the importance of privacy. Electronic Fronteer Foundation, 10 December 2009. Available at: https://www.eff.org/deeplinks/2009/12/google-ceo-eric-schmidt-dismisses-privacy. Accessed on September 3, 2012.

[52] Facebook and your privacy: Who sees the data you share on the biggest social network? Consumer Reports Magazine, June 2012. Available at: http://www.consumerreports.org/cro/magazine/2012/06/facebook-your-privacy/index.htm. Accessed on September 3, 2012.

[53] Facebook Newsroom—Key Facts. Available at: http://newsroom.fb.com/content/default.aspx?NewsAreaId=22. Accessed on September 3, 2012.

[54] The Swiss Federal Data Protection Act. Available at: http://www.edoeb.admin.ch/org/00828/index.html. Accessed on September 3, 2012.

[55] FireGPG. Available at: http://getfiregpg.org. Accessed on September 3, 2012.

[56] Christoph Gaffga. DCT-watermark: Robust watermarks for color JPEG in Java. Available at: https://code.google.com/p/dct-watermark/. Accessed on September 3, 2012.

[57] Simson Garfinkel. *PGP: Pretty Good Privacy*. Oreilly Series. O'Reilly Media, Incorporated, 1994.

[58] Hofstede Geert. Geert hofstede cultural dimensions. Available at: http://www.geert-hofstede.com. Accessed on September 3, 2012.

[59] Hofstede Geert. *Cultural and Organizations, Software of the Mind: Intercultural Cooperation and its importance for survival.* McGraw-Hil, 1991.

[60] Christian Gehrmann and Kaisa Nyberg. Enhancements to Bluetooth baseband security. In *Proceedings of the Nordic Workshop on Secure IT-Systems (NordSec 2001)*, Copenhagen, Denmark, November 2001.

[61] Christian Gehrmann and Kaisa Nyberg. Manual authentication for wireless devices. *RSA Cryptobytes*, 7:2004, 2004.

[62] Robert Gellman. Privacy in the clouds: Risks to privacy and confidentiality from cloud computing. World Privacy Forum, 23 February 2009. Available at: http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf. Accessed on September 3, 2012.

[63] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, August 1986.

[64] Michael T. Goodrich, Michael Sirivianos, John Solis, Gene Tsudik, and Ersin Uzun. Loud and clear: Human-verifiable authentication based on audio. In *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS 2006)*, Lisbon, Portugal, July 2006.

[65] Google 'may pull out of China after Gmail cyberattack'. BBC News, January 13, 2010. Available at: http://news.bbc.co.uk/2/hi/8455712.stm. Accessed on Sptember 3, 2012.

[66] A new approach to China. Google Official Blog, January 13, 1010. Available at: http://googleblog.blogspot.com/2010/01/new-approach-to-china.html. Accessed on September 3, 2012.

[67] Google Docs. https://docs.google.com. Accessed on September 3, 2012.

[68] Bernd Grobauer, Tobias Walloschek, and Elmar Stöcker. Un-

derstanding cloud-computing vulnerabilities. *IEEE Security & Privacy*, 9:50–57, March/April 2011.

[69] Grant Gross. Cloud computing may draw government action. PCWorld, September 12, 2008. Available at: `http://pcworld.about.com/od/businesscenter/Cloud-Computing-May-Draw-Gover.htm`. Accessed on September 3, 2012.

[70] Oliver Günther and Sarah Spiekermann. RFID and the perception of control: the consumer's view. *Communications of the ACM*, 48(9):73–76, 2005.

[71] Peter Gutmann. Plug-and-play PKI: a PKI your mother can use. In *Proceedings of the 12th conference on USENIX Security Symposium*, Washington, DC, August 2003.

[72] Cormac Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the New Security Paradigms Workshop (NSPW 2009)*, pages 133–144, Oxford, UK, September 2009.

[73] Kashmir Hill. How Target figured out a teen girl was pregnant before her father did. Forbes, 16 February 2012. Available at: `http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/`. Accessed on September 3, 2012.

[74] Chris J. Hoofnagle, Jennifer King, Su Li, and Joseph Turow. How different are young adults from older adults when it comes to information privacy attitudes and policies? *SSRN eLibrary*, 14 April 2010. Available at: `http://ssrn.com/paper=1589864`. Accessed on September 3, 2012.

[75] Chris Jay Hoofnagle and Jennifer King. Research report: What Californians understand about privacy online. *SSRN eLibrary*, 3 September 2008. Available at: `http://ssrn.com/abstract=1133075`. Accessed on September 3, 2012.

[76] John B. Horrigan. Use of cloud computing applications and services. Pew Research Center, September 12, 2008. Available at: `http://www.pewinternet.org/Reports/2008/Use-of-Cloud-Computing-Applications-and-Services.aspx`. Accessed on September 3, 2012.

[77] Wenjin Hu, Tao Yang, and Jeanna N. Matthews. The good, the bad and the ugly of consumer cloud storage. *ACM SIGOPS Operating Systems Review*, 44(3):110–115, 2010.

[78] Jessica Hunter. Identity theft—don't let the dancing pigs fool you. Available at: `http://www.identitytheftfixes.com/dancing_pigs_and_identity_theft.html`. Accessed on September 3, 2012.

[79] IBM'er argues: Clouds more secure than your data center. Gabriel Consulting Group, February 27, 2011. Available at: `http://www.gabrielconsultinggroup.com/gcg-news-and-views/20-general-blog/258-ibmer-argues-clouds-more-secure-than-your-data-center.html`. Accessed on September 3, 2012.

[80] Internet growth 2000-2005. Internet World Stats, 2005. Available at: `http://www.internetworldstats.com/pr/edi008.htm`. Accessed on September 3, 2012.

[81] Internet World Stats. `http://www.internetworldstats.com/`. Accessed on September 3, 2012.

[82] Iulia Ion, Ponnurangam Kumaraguru Marc Langheinrich, and Srdjan Capkun. Influence of user perception, security needs, and social factors on device pairing method choices. In *Proceedings of Symposium on Usable Privacy and Security (SOUPS 2010)*, Redmond, WA, July 2010.

[83] Iulia Ion, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Capkun. Home is safer than the cloud! Privacy concerns for consumer cloud storage. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS 2011)*, Pittsburgh, PA, July 2011.

[84] Sonia Jahid, Shirin Nilizadeh, Prateek Mittal, Nikita Borisov, and Apu Kapadia. DECENT: a decentralized architecture for enforcing privacy in online social networks. In *Proceedings of the 4th IEEE Workshop on Security and Social Networking (SESOC 2012)*, Lugano, Switzerland, March 2012.

[85] Wayne Jansen and Timothy Grance. Guidelines on security and privacy in public cloud computing. NIST Special Publication 800-144, December 2011. Available at: `http://csrc.nist.gov/`

publications/nistpubs/800-144/SP800-144.pdf. Accessed on September 3, 2012.

[86] Andrew Joint, Edwin Baker, and Edward Eccles. Hey, you, get off of that cloud? *Computer Law & Security Review*, 25(3):270–274, 2009.

[87] Harvey Jones and Jos Hiram Soltren. Facebook: Threats to privacy, 2005.

[88] Ronald Kainda, Ivan Flechais, and A. W. Roscoe. Usability and security of out-of-band channels in secure device pairing protocols. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS 2009)*, Mountain View, CA, July 2009.

[89] Apu Kapadia. A case (study) for usability in secure email communication. *Security Privacy, IEEE*, 5(2):80–84, March-April 2007.

[90] Colleen Kehoe, Jim Pitkow, Kate Sutton, Gaurav Aggarwal, and Juan D. Rogers. GVU's tenth WWW user survey results. Georgia Institute of Technology, 14 May 1999. Available at: http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/.

[91] Mohammed Nyamathulla Khan. Does India have a data protection law? *Legal Service India*, November 2009. Available at: http://www.legalserviceindia.com/article/l406-Does-India-have-a-Data-Protection-law.html. Accessed on September 12, 2012.

[92] Erika Kinetz. Google, Skype targeted in India security crackdown. The Huffington Post. February 9, 2011. Available at: http://www.huffingtonpost.com/2010/09/02/google-skype-targeted-in-_n_703198.html. Accessed on September 3, 2012.

[93] Alfred Kobsa, Rahim Sonawalla, Gene Tsudik, Ersin Uzun, and Yang Wang 0005. Serial hook-ups: a comparative usability study of secure device pairing methods. In *Symposium on Usable Privacy and Security (SOUPS 2009)*, Mountain View, CA, July 2009.

[94] Hugo Krawczyk. SIGMA: The 'SIGn-and-MAc' approach to authenticated Diffie-Hellman and its use in the IKE-protocols. In

Proceedings of the 23rd Annual International Cryptology Conference (CRYPTO 2003), Santa Barbara, CA, August 2003.

[95] Balachander Krishnamurthy and Craig E. Wills. On the leakage of personally identifiable information via online social networks. *ACM SIGCOMM Computer Communication Review*, 40(1):112–117, January 2010.

[96] Arun Kumar, Nitesh Saxena, Gene Tsudik, and Ersin Uzun. Caveat emptor: A comparative study of secure device pairing methods. In *Proceedings of the 7th IEEE International Conference on Pervasive Computing and Communications (PerCom 2009)*, Galveston, Texas, March 2009.

[97] Arun Kumar, Nitesh Saxena, Gene Tsudik, and Ersin Uzun. A comparative study of secure device pairing methods. *Pervasive and Mobile Computing*, 5(6):734–749, 2009.

[98] Arun Kumar, Nitesh Saxena, and Ersin Uzun. Alice meets Bob: A comparative usability study of wireless device pairing methods for a "two-user" setting. *CoRR*, abs/0907.4743, 2009. Available at: `http://arxiv.org/abs/0907.4743`. Accessed on September 3, 2012.

[99] Ponnurangam Kumaraguru and Lorrie F. Cranor. Privacy in India: Attitudes and awareness. In *Proceedings of the Workshop on Privacy Enhancing Technologies (PET 2005)*, Dubrovnik (Cavtat), Croatia, June 2005.

[100] Ponnurangam Kumaraguru and Lorrie F. Cranor. Privacy indexes: A survey of Westin's studies. Technical Report CMU-ISRI-05-138, Carnegie Mellon University, 2005.

[101] Cynthia Kuo, Adrian Perrig, and Jesse Walker. Designing an evaluation method for security user interfaces: lessons from studying secure wireless network configuration. *Interactions*, 13(3):28–31, 2006.

[102] Sven Laur and Kaisa Nyberg. Efficient mutual data authentication using manually authenticated strings. In *Proceedings of the 5th International Conference on Cryptology and Network Security (CANS 2006)*, pages 90–107, Suzhou, Shanghai, December 2006.

[103] Benoit Libert, Kenneth G. Paterson, and Elizabeth A. Quaglia. Anonymous broadcast encryption: Adaptive security and efficient constructions in the standard model. In *Proceedings of Public Key Cryptography Conference (PKC 2012)*, pages 206–224, Darmstadt, Germany, May 2012.

[104] Mikhail Lisovich and Stephen Wicker. Privacy concerns in upcoming residential and commercial demand-response systems. In *Proceedings of the Clemson University Power Systems Conference*. Clemson, SC, March 2008. Available at: `http://www.truststc.org/pubs/332.html`. Accessed on September 3, 2012.

[105] Matthew M. Lucas and Nikita Borisov. FlyByNight: mitigating the privacy risks of social networking. In *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society (WPES 2008)*, Alexandria, VA, October 2008.

[106] Wanying Luo, Qi Xie, and Urs Hengartner. FaceCloak: An architecture for user privacy on social networking sites. In *Proceedings of the International Conference on Computational Science and Engineering*, Vancouver, Canada, August 2009.

[107] MailCloak. Available at: `http://www.gwebs.com/mailcloak.html`. Accessed on September 3, 2012.

[108] MALLET. `http://mallet.cs.umass.edu/`. Accessed on August 31, 2012.

[109] Friedemann Mattern and Christian Floerkemeier. *From the Internet of Computers to the Internet of Things*, volume 6462 of *LNCS*, pages 242–259. Springer, 2010.

[110] Friedemann Mattern, Thorsten Staake, and Markus Weiss. ICT for green – How computers can help us to conserve energy. In *Proceedings of the e-Energy Conference 2010*, pages 1–10, Passau, Germany, April 2010.

[111] Andrew McCallum, Xuerui Wang, and Andrés Corrada-Emmanuel. Topic and role discovery in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research (JAIR)*, 30:249–272, 2007.

[112] Jonathan M. McCune, Adrian Perrig, and Michael K. Reiter. Seeing-is-believing: Using camera phones for human-verifiable au-

thentication. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 110–124, Oakland, CA, May 2005.

[113] Aleecia M. McDonald and Lorrie F. Cranor. The cost of reading privacy policies. *ACM Transactions on Computer-Human Interaction*, 0389(3):1–22, 2008.

[114] Aleecia M. McDonald and Lorrie F. Cranor. Americans' attitudes about Internet behavioral advertising practices. In *Proceedings of the 9th ACM Workshop on Privacy in the Electronic Society (WPES 2010)*, Chicago, IL, October 2010.

[115] Mike Melanson. Facebook wants to be your one true login. ReadWriteWeb, February 10, 2010. Available at: `http://www.readwriteweb.com/archives/facebook_wants_to_be_your_one_true_login.php`. Accessed on September 3, 2012.

[116] Mike Melanson. How Google failed its users and gave birth to an Internet meme. ReadWriteWeb, 11 February 2010. Available at: `http://www.readwriteweb.com/archives/how_google_failed_internet_meme.php`. Accessed on September 3, 2012.

[117] Henry Michael. *International Privacy, Publicity and Personality Laws*. Reed Elsevier, 2001. pages 233-250.

[118] MTurk. `https://www.mturk.com/`. Accessed on August 31, 2012.

[119] Ellen Nakashima. Feeling betrayed, Facebook users force site to honor their privacy. Washington Post, November 30, 2007. Available at: `http://www.washingtonpost.com/wp-dyn/content/article/2007/11/29/AR2007112902503.html`. Accessed on September 3, 2012.

[120] National Intelligence Grid (NATGRID). February 13, 2010. Available at: `http://currentaffairs.gktoday.in/2010/02/national-intelligence-grid-natgrid.html`. Accessed om September 3, 2012.

[121] Anthony Nicholson, Ian Smith, Jeff Hughes, and Brian Noble. LoKey: Leveraging the SMS network in decentralized, end-to-end trust establishment. In Kenneth Fishkin, Bernt Schiele, Paddy

Nixon, and Aaron Quigley, editors, *Pervasive Computing*, volume 3968 of *Lecture Notes in Computer Science*, pages 202–219. Springer Berlin/Heidelberg, 2006.

[122] Anthony J. Nicholson, Junghee Han, David Watson, and Brian D. Noble. Exploiting mobility for key establishment. In *Proceedings of the 7th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2006)*, pages 61–68, Semiahmoo Resort, WA, August 2006.

[123] Yossef Oren and Avishai Wool. Perfect privacy for webmail with secret sharing. February 2009. Available at: `http://www.eng.tau.ac.il/~yos/spemail/OrenWool-SPEmail.pdf`. Accessed on September 3, 2012.

[124] Andreas Pashalidis, Nikos Mavrogiannopoulos, Xavier Ferrer, and Benat Bermejo Olaizola. For human eyes only—security and usability evaluation. In *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2012)*, Raleigh, NC, October 2012.

[125] Sameer Patil, Alfred Kobsa, Ajita John, and Doree Seligmann. Comparing privacy attitudes of knowledge workers in the U.S. and India. In *Proceedings of the 3rd International Conference on Intercultural Collaboration (ICIC 2010)*, pages 141–150, Copenhagen, Denmark, August 2010.

[126] Wayne A. Pauley. Cloud provider transparency – an empirical evaluation. *IEEE Security & Privacy*, 8(6):32–39, November–December 2010.

[127] Chris Potter and Grant Waterfall. Information security breaches survey. PWC Technical Report, April 2012. Available at: `http://www.pwc.co.uk/en_UK/uk/assets/pdf/olpapp/uk-information-security-breaches-survey-technical-report.pdf`. Accessed on September 3, 2012.

[128] Cara Pring. 100 more social media statistics for 2012. Web Blog "the social skinny". February 13, 2012. `http://thesocialskinny.com/100-more-social-media-statistics-for-2012/`. Accessed on September 6, 2012.

[129] Umar Rashid and Aaron J. Quigley. Interaction techniques for binding smartphones: A desirability evaluation. In *Proceedings*

of the First Conference on Human Centered Design (HCD 2009), pages 120–128, San Diego, CA, July 2009.

[130] Jun Rekimoto. SyncTap: synchronous user operation for spontaneous network connection. *Personal Ubiquitous Computing*, 8(2):126–134, 2004.

[131] John Rendleman. Customer data means money. InformationWeek, August 20, 2001. Available at: `http://www.informationweek.com/news/6506304`. Accessed on September 3, 2012.

[132] Christopher Riederer, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, and Pablo Rodriguez. For sale : your data: by : you. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, HotNets-X, pages 13:1–13:6, New York, NY, USA, 2011. ACM.

[133] David Rosenthal. New data protection act in Switzerland: more transparency, additional costs. *Privacy Laws & Business International Newsletter*, 2006(3), December 2007.

[134] Esther Schindler. Cloud development survey. Evans Data Corporation, Strategic Reports, July 2010. Available at: `http://www.evansdata.com/reports/view\Release.php?reportID=27`.

[135] Scramble! `http://sourceforge.net/projects/scramble-it/`. Accessed on August 31, 2012.

[136] SecreTwit. Available at: `http://code.google.com/p/secretwit/`. Accessed on September 3, 2012.

[137] Tali Sharot. The optimism bias. *Current Biology*, 21(23):941–945, December 2012.

[138] Kim B. Sheehan. Toward a typology of Internet users and online privacy concerns. *The Information Society*, 18(1):21–32, 2002.

[139] Erik Sherman. Privacy policies are great—for PhDs. 4 September 2008. Available at: `http://industry.bnet.com/technology/1000391/privacy-policies-are-great-for-phds/`.

[140] Joshua L. Simmons. Buying you: The government's use of fourth-parties to launder data about the people. *Columbia Business Law Review*, 2009(3):950, 2009.

[141] Ryan Singel. Information Technology Act 2000. Department of Information Technology. Available at: http://www.mit.gov.in/content/it-act-2000-dpl-cyber-laws.

[142] Ryan Singel. Twitter's response to WikiLeaks subpoena should be the industry standard. Wired, January 10, 2011. Available at: http://www.wired.com/threatlevel/2011/01/twitter/. Accessed on September 3, 2012.

[143] Smith v. Maryland. 442 U.S. 735 (1979): Available at: http://laws.findlaw.com/us/442/735.html. Accessed on September 3, 2012.

[144] Christopher Soghoian. Caught in the cloud: Privacy, encryption, and government back doors in the Web 2.0 era. *Journal on Telecommunications & High Technology Law 359*, 8(2), 2010.

[145] Claudio Soriente, Gene Tsudik, and Ersin Uzun. BEDA: Button-enabled device pairing. In *Proceedings of the First International Workshop on Security for Spontaneous Interaction (IWSSI 2007)*, September 2007.

[146] Claudio Soriente, Gene Tsudik, and Ersin Uzun. HAPADEP: Human asisted pure audio device pairing. In *Proceedings of the International Information Security Conference (ISC 2008)*, Taipei, Taiwan, September 2008.

[147] Sarah Spiekermann. RFID and privacy: what consumers really want and fear. *Personal and Ubiquitous Computing*, 13(6):423–434, 2009.

[148] Polly Sprenger. Sun on privacy: 'get over it'. Wired, January 26, 1999. Available at: http://www.wired.com/politics/law/news/1999/01/17538. Accessed on September 3, 2012.

[149] Emily Steel and Jessica E. Vascellaro. Facebook, MySpace confront privacy loophole. The Wall Street Journal, May 21, 2010. Available at: http://online.wsj.com/article/SB10001424052748704513104575256701215465596.html. Accessed on September 3, 2012.

[150] Strict origin policy. http://kb.mozillazine.org/Security.fileuri.strict_origin_policy. Accessed on August 31, 2012.

[151] Jani Suomalainen, Jukka Valkonen, and N. Asokan. Security as-

sociations in personal networks: A comparative analysis. In *Proceedings the 4th European Workshop on Security and Privacy in Ad hoc and Sensor Networks (ESAS 2007)*, pages 43–57, Cambridge, UK, July 2007.

[152] SurveyMonkey. `http://www.surveymonkey.com`. Accessed on September 3, 2012.

[153] Dan Svantesson and Roger Clarke. Privacy and consumer risks in cloud computing. *Computer Law & Security Review*, 26(4):391–397, July 2010.

[154] Amin Tootoonchian, Kiran Kumar Gollu, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. Lockr: social access control for Web 2.0. In *Proceedings of the First Workshop on Online Social Networks (WOSN 2008)*, Seattle, WA, August 2008.

[155] Tor. `http://www.torproject.org`. Accessed on August 31, 2012.

[156] Touring Hub. `http://testing.turinghub.com/`. Accessed on August 31, 2012.

[157] Janice Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. In *Proceedings of the 6th Workshop on the Economics of Information Security (WEIS 2007)*, Pittsburgh, PA, June 2007.

[158] Twitter turns six. Twitter Blog, March 21, 2012. Available at: `http://blog.twitter.com/2012/03/twitter-turns-six.html`. Accessed on September 3, 2012.

[159] Unique Identification Authority India (UID). `http://uidai.gov.in/`. Accessed om September 3, 2012.

[160] US-CERT. Tips. Available at: `https://www.us-cert.gov/cas/tips/`. Accessed on September 3, 2012.

[161] Ersin Uzun, Kristiina Karvonen, and Nadarajah Asokan. Usability analysis of secure pairing methods. In *Proceedings of the Usable Security Workshop (USEC 2007)*, Scarborough, Trinidad/-Tobago, February 2007.

[162] Jukka Valkonen, Aleksi Toivonen, and Kristiina Karvonen. Us-

ability testing for secure device pairing in home networks. In *Proceedings of the First International Workshop on Security for Spontaneous Interaction (IWSSI 2007)*, September 2007.

[163] Jessica E. Vascellaro. Google discloses privacy glitch. WJS Blogs, March 8, 2009. Available at: `http://blogs.wsj.com/digits/2009/03/08/1214/`. Accessed on September 3, 2012.

[164] Mario Čagalj, Srdjan Čapkun, and Jean-Pierre Hubaux. Key agreement in peer-to-peer wireless networks. *Proceedings of the IEEE Special Issue on Cryptography and Security*, 94(2):467–478, February 2006.

[165] Cong Wang, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou. Secure ranked keyword search over encrypted cloud data. In *Proceedings of the 30th IEEE International Conference on Distributed Computing Systems (ICDCS 2010)*, pages 253–262, Genoa, Italy, June 2010.

[166] Yang Wang and Greg Mori. A discriminative latent model of image region and object tag correspondence. In *Proceedings of the 24th Annual Conference on Neural Information Processing Systems*, Canada, 2010.

[167] Zoe A. Y. Weinberg. Google settles Buzz lawsuit. The Harvard Crimson, September 7, 2010. Available at: `http://www.thecrimson.com/article/2010/9/7/google-mason-privacy-settlement/`. Accessed on September 3, 2012.

[168] Mark Weiser. The computer for the 21st century. In Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg, editors, *Human-computer interaction*, pages 933–940. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995.

[169] Alan F. Westin. *Privacy and Freedom.* Atheneum, New York, first edition, 1967.

[170] Doug Whiting, Russ Housley, and Niels Ferguson. Counter with CBC-MAC (CCM). RFC 3610 (Informational), September 2003. Available at: `http://www.ietf.org/rfc/rfc3610.txt`. Accessed on September 3, 2012.

[171] Alma Whitten and Doug J. Tygar. Why Johnny can't encrypt: a usability evaluation of PGP 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium*, Washington, D.C., August 1999.

[172] Wi-Fi Alliance announces groundbreaking specification to support direct Wi-Fi connections between devices. October 14, 2009. Available at: `http://www.wi-fi.org/media/press-releases/wi-fi-alliance%C2%AE-announces-groundbreaking-specification-support-peer-peer-wi-fi`. Accessed on September 3, 2012.

[173] XPath. `http://www.w3schools.com/xpath/`. Accessed on September 3, 2012.

[174] Chao M. Zhang and Vern Paxson. Detecting and analyzing automated activity on Twitter. In *Proceedings of the 12th International Conference on Passive and Active Measurement (PAM 2011)*, pages 102–111, Atlanta, GA, 2011.

[175] Mary Ellen Zurko. User-centered security: Stepping up to the grand challenge. In *Proceedings of the Annual Computer Security Applications Conference (ACSAC 2005)*, pages 187–202, Tucson, AZ, December 2005.

Curriculum Vitae

IULIA ION

Education

Sept 2005 – Feb 2007	Master of Science in Computer Science, International University in Germany, Bruchsal, Germany
Sept 2002 – Aug 2005	Bachelor of Science in Information Technology, International University in Germany, Bruchsal, Germany

Work Experience

Feb 2007 – Feb 2008	Junior Security Researcher, CREATE-NET Research Center, Trento, Italy

Internships

June 2011 – Sept 2011	Google Inc., Mountain View, CA
June 2009 – Sept 2009	Carnegie Mellon University, Pittsburgh, PA
June 2011 – Sept 2011	CREATE-NET Research Center, Trento, Italy
April 2005 – Aug 2005	University of Cambridge, UK
Jan 2005 – April 2005	Deutsche Börse Systems, Frankfurt, Germany
Sept 2004 – Dec 2004	Université de Provence, Marseille, France