ME, MYSELF, AND ID: TOWARDS USABLE, PRIVACY-PRESERVING, FRAUD-RESISTANT DIGITAL IDENTITY SERVICES FOR SMARTPHONE USERS

by
Michael Lutaaya

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of
the requirements for the degree of

MASTER OF COMPUTER SCIENCE

School of Computer Science

at

CARLETON UNIVERSITY

Ottawa, Ontario
August, 2019

Abstract

Between easy-to-counterfeit physical IDs, the poor adoption of some digital alternatives, and the weak security of others, the need for better identity solutions becomes apparent. As such, this thesis explores how digital identity services can be designed in a usable, secure, and privacy-preserving manner.

Our first user study examines the discoverability of participants' personally identifiable information (PII) to determine what information is suitable for identity proofing. We found that many different types of PII could be obtained by an imposter and that individuals significantly underestimate the risk of said information being available.

Our second user study explores the usability of a prototype digital identity service for smartphone users. We identify and categorize users' perceptions, opinions, and concerns over using this type of service. Our findings highlight the need to support users in developing accurate mental models to avoid usability issues.

Acknowledgements

I would like to extend my deepest gratitude to my supervisor, Sonia Chiasson. Your wisdom and unwavering support have been instrumental in both the completion of this thesis and my growth as a researcher from my first year of undergraduate studies to now. Thank you for being so generous with your time and for believing in your students.

I also wish to thank the members of my defence committee: Elizabeth Stobert, Kasia Muldner, and David Mould. I am very appreciative of the feedback you have shared with me and this thesis was made stronger as a result.

Many thanks to members of the CHORUS Lab, current and past. I am lucky to have been surrounded by such knowledgeable and supportive people. I would also like to recognize the hard work of Khadija and Paulina, two research assistants from the lab who helped with data collection and data analysis.

Thank you to all who participated in our studies and thanks also to Bluink Ltd. for collaborating with us.

To my incredible friends, thank you for your patience and encouragement along the way.

Finally, thank you to my parents and my sister Gerri. This thesis would not have been possible without your support and encouragement.

Table of Contents

List of Tables

List of Figures

Glossary

authentication A process through which an entity proves that it is who or what it claims to be.

certificate authority A trusted party that issues certificates indicating that a public key belongs to a particular entity.

chain of trust An attribute of digital certificates whereby validating a certificate's signature requires trusting the CA. That CA's certificate might be signed by another CA, thus requiring any relying party to move up the chain until the root CA's certificate. The root CA's certificate is implicitly trusted.

ciphertext The unintelligible output produced by an encryption algorithm.

cryptographic key A string of random bytes that can be used to transform a message from plaintext into ciphertext, or vice versa.

cryptography A field of study concerned with techniques for communicating securely such that information can not be read by unauthorized parties.

decryption A process where a cryptographic key is used to transform ciphertext into plaintext.

digital certificate A document that indicates to whom a public key belongs.

digital identity A digital representation of the attributes, preferences, and traits that define who or what an entity.

digital identity service A service that allows individuals to establish a digital identity, authenticate with relying parties by providing their digital identity, and have the entity operating the service vouch for the authenticity of their digital identity.

digital signature A message that has been encrypted with the sender's private key so that recipients can use the corresponding public key to decrypt the message and be reasonably assured that the message came from the owner of the key pair.

encryption A process where a cryptographic key is used to transform plaintext into ciphertext.

federated identity A mechanism that allows identity information to be used across otherwise separate systems. A common use case for federated identity is single sign-on where a relying party, such as Yelp, can allow users to sign in via an identity provider, such as Google or Facebook.

hash The fixed-size output of a cryptographic hash function (a function that accepts a message of any size as input). Cryptographic hash functions are designed so that it is infeasible to derive the original input from the function's output and it is improbable that two different inputs would result in the same output.

identity A representation of who or what an entity is, as defined by its identity attributes.

identity attribute An attribute, preference, or trait that describes who or what an entity is (e.g., a car's vehicle identification number or VIN, a person's preference for sci-fi novels over non-fiction books, a computer's date of manufacture).

identity claim An assertion about an entity (e.g., this person was born on June 7, 1958).

identity proofing A process where information about an individual is collected, validated, and verified to establish the applicant's identity with a certain level of confidence.

identity provider An entity who, when presented with credentials, can attest to their validity and whether the party presenting the credentials is genuinely associated with the credentials.

identity transaction An exchange or interaction with another entity (e.g., a business, a hospital) where an individual is required to prove their identity.

identity verification service A service that evaluates whether an individual is who they claim to be. This process often require individuals to provide government-issued documents or other types of documents.

issuing authority An entity that issues digital identities.

key pair Two keys—a public key and a private key—whereby messages encrypted with one key can be decrypted with the other key.

knowledge-based authentication An authentication method where individuals are asked to provide information about themselves which can be verified with sources that maintain data about the individual.

mental model An individual's beliefs for how something works.

National Institute of Standards and Technology A US government agency whose responsibilities include developing scientific and technological standards.

nonce A random **n**umber that is used only **once**. Nonces are typically used in cryptographic communications to prevent replay attacks whereby an adversary eavesdrops on a communication and repeats it to give the appearance of a legitimate communication.

personally identifiable information Information that can distinguish a specific individual from others.

plaintext The human- or machine-readable output produced by a decryption algorithm.

point of sale A location where customer transactions are completed.

private key A cryptographic key that must not be shared publicly. Private keys are used to sign messages and to decrypt messages that were encrypted with the key-pair owner's public key.

public key A cryptographic key that is safe to share publicly. Encrypting messages with a public key ensures that the message can only be read by the key-pair owner (i.e., by decrypting the message with the corresponding private key). Further, the public key can be used to verify messages that were signed with the corresponding private key to prove that the message was sent by the key-pair owner.

public key cryptography Cryptographic techniques where the sender and the recipient each have their own key pair (i.e., public key and private key).

public key infrastructure The facilities and mechanisms used to manage digital certificates.

QR code A two-dimensional barcode made up of black and white squares that can be read by a camera. Data, such as text, URLs, and other types of information, can be embedded in the squares.

relying party An entity that is dependent on assertions made by the issuing authority regarding about an individual's claims.

salt A randomly-generated value that is concatenated to the input passed to a cryptographic hash function. Password databases often assign a unique salt to each user and append the salt to the user's password before hashing and storing the password. This ensures that even if two users has the same password, the values stored in the database would be unique, thus preventing certain types of attacks (e.g., dictionary attacks).

social media service A website or app that allows users to share content or to connect with other users. For the purposes of our research, we excluded websites and apps that only have ephemeral content (e.g., Snapchat) or that only have messaging functionality (e.g., WhatsApp, Facebook Messenger).

symmetric key cryptography Cryptographic techniques where the sender and the recipient share a key that is used for both encryption and decryption.

verification A process for assessing the truth of a particular claim or assertion.

Acronyms

CA certificate authority

KBA knowledge-based authentication

NIST National Institute of Standards and Technology

PII personally identifiable information

PKI public key infrastructure

POS point of sale

Chapter 1

Introduction

1.1 Motivation

The *Universal Declaration of Human Rights* states that "everyone has the right to recognition everywhere as a person before the law" [125]. In practice, however, over 1.1 billion people are living without official identification [122]; a reality that effectively denies them access to essential services like health care and limits their ability to participate in today's economy.

There are still fundamental issues even for those with official identification. The authenticity of physical identity documents can no longer be guaranteed given that counterfeits are accessible and relatively easy to produce [9, 53]. Additionally, physical identity documents lack safeguards for preventing access to personally identifiable information that is irrelevant for a given **transaction**.

And while some countries have adopted digital identity systems (e.g., [4, 38, 64, 103, 110], these systems can have adverse consequences when implemented poorly. For example, a recent software vulnerability put more than 760,000 Estonian citizens at risk of becoming victims of identity theft [55].

Ultimately, the above challenges highlight the need for an ID system that is both secure, maintains privacy, and does not compromise on usability. We believe that **digital identity services** have the potential to meet these goals. Although many definitions are available, we start with **NIST**'s interpretation [61]: "**digital identity** is the unique representation of a subject engaged in an online transaction. A digital identity is always unique in the context of a digital service, but does not necessarily need to uniquely identify the subject in all contexts. In other words, accessing a digital service may not mean that the subject's real-life identity is known."

Canada does not yet have a national digital identity service, and while most agree that one is needed (e.g., [28, 45, 59]), no one approach has been selected yet.

1.2 Research Goal

In this thesis, we aim to bring a greater understanding to how digital identity services can be designed to be usable, secure, and privacy-preserving through the lens of four research questions:

1. To what degree are certain types of **personally identifiable information** (**PII**) discoverable through **social media services** and search engines?
2. How do an individual's behaviours, both online and offline, relate to the ease with which an imposter could obtain their PII?
3. How do users perceive digital identity services?
4. What are users' expectations for digital identity services?

1.3 Contribution

We identify two primary contributions of this work:

- Our first study indicated that many different types of PII could be obtained by an imposter and that individuals significantly underestimate the risk of said information being available. As a result, we recommend that identity proofing processes rely on other types of information and suggest initiatives that can help users both recognize the risk associated with their behaviours and support them in changing their behaviours.
- From our second study, we identify and categorize users' perceptions, opinions, and concerns over using a digital identity service for their own privacy and security. We also demonstrate the problems that occur when a digital identity app fails to convey the system image accurately. Accordingly, we made recommendations to help users develop accurate mental models and to ensure users understand how the service works. This study adds to the sparse literature on the usability of digital identity services.

1.4 Relationship with Bluink

The work presented in this thesis was conducted as joint research between our research group at Carleton University and Bluink Ltd. [19], an Ottawa-based technology company specializing in identity management.

We conceived the design of both studies, with consultation from Bluink. Bluink handled all development of the eID-Me software and provided us with versions of the software for our evaluations. We worked together on aspects of the user interaction, including the app's registration process for which we proposed designs.

After analyzing the data collected for each study, we provided Bluink with reports of our research results. Bluink used the outcomes of the information discoverability study to inform the implementation of the identity proofing algorithm used in their prototype, which participants interacted with in the usability study. In addition, for the usability study, we provided Bluink with recommendations detailing how the app could be modified to address the issues identified through our research.

1.5 Thesis Outline

This thesis is arranged as follows:

- In Chapter 2, we provide a review of relevant research literature relating to identity proofing, personally identifiable information, and authentication. We also identify gaps in the literature that we intend to fill with our work.
- In Chapter 3, we present the findings from our information discoverability study, where we investigated the discoverability of participants' personally identifiable information.
- In Chapter 4, we present the findings from our usability evaluation, where participants interacted with a smartphone-based digital identity solution.
- In Chapter 5, we interpret the findings of both studies in the context of our research questions, set forth recommendations for the digital identity services, discuss the opportunities for future research, and indicate the limitations of our work.

Chapter 2

Background

In this chapter, we introduce essential concepts relating to identity and public key infrastructure, and provide an overview of the relevant research literature concerning identity proofing, personally identifiable information, and authentication. We conclude the chapter by summarizing the gaps we identified in the literature that this thesis intends to fill.

2.1 General Concepts

In this section, we define terminology relating to identity, usability, and public key infrastructure.

2.1.1 Identity

Identity can be defined generally as the attributes, preferences, and traits that define who or what an entity is (e.g., a car's vehicle identification number or VIN, a person's preference for sci-fi novels over non-fiction books, a computer's date of manufacture) [5, 27, 131]. Hereafter, we refer to these attributes, preferences, and traits collectively as ***identity attributes***.

Certain types of identity attributes are considered **personally identifiable information**, or PII. PII is a term used to describe information that can distinguish a specific individual from other people [43]. In one of its publications [82], the National Institute of Standards and Technology (NIST) states that PII includes:

(1) any identity attribute(s) that can be used to distinguish an individual from others (e.g., full name, Social Insurance Number, home address), and

(2) any information that "is linked or linkable to an individual" (e.g., IP address, medical records, gender, postal code).

Identities are commonly used in **transactions**. Transactions are exchanges where entities are required to provide *credentials*—for example, physical proofs like a driver's license or digital proofs like cryptographic keys—to support claims that a specific identity belongs to them [11, 117, 131]. These credentials are typically issued by an ***identity provider***, such as a government agency or an online service provider like Google, who can attest to the validity of the credentials and their association with a specific identity [61].

Related to digital identities, which are used to represent *entities*, the term eID refers to electronic mechanisms that *people* can use to prove their identity to a **relying party** [30].

Blue, Condell, and Lunney [18] offer a recent comprehensive introduction to the concepts surrounding identity, authentication of identity, and digital identity. In addition to providing definitions for key terminology, this overview acknowledges the relative ease with which counterfeit physical documents can be produced but concedes that a superior digital alternative has yet to emerge.

In our research, we focus on this transactional use of identity, specifically when conducted with digital credentials managed by a smartphone app.

2.1.2 Public Key Infrastructure

Cryptography is a field of study concerned with techniques for communicating securely such that information can not be read by unauthorized parties [87].

Two essential cryptography concepts are ***encryption*** and ***decryption*** [84]. Encryption is a process where the sender uses a ***cryptographic key*** to transform their message from its original, readable format (***plaintext***) into an unintelligible format (***ciphertext***). Conversely, decryption is a process where the recipient can use a key to transform ciphertext back into plaintext.

The two main types of cryptography are ***symmetric key cryptography*** and ***public key cryptography*** (or asymmetric cryptography) [85]. With symmetric key cryptography, the sender and the recipient share a key that is used for both encryption and decryption.

With public key cryptography, the sender and the recipient each have their own

key pair [85]. A key pair consists of a ***public key*** and a ***private key***. The public key can be distributed widely while the private key must be kept secret.

When a message is encrypted with a key pair's public key, it can only be decrypted using the corresponding private key [85]. Conversely, when a message is encrypted with a key pair's private key, it can only be decrypted with the corresponding public key. As such, to send an encrypted message using public-key authentication, the sender encrypts their message with the recipient's public key. This ensures that it can not be read by anyone else.

Similarly, senders can encrypt messages with their private key to create a ***digital signature*** [85]. Recipients can decrypt the message using the sender's public key to validate the signature and be reasonably assured that the message was from the sender (assuming that the sender protected their private key).

Public key cryptography is also fundamental to ***digital certificates***. A digital certificate is a document that indicates to whom a public key belongs [68]. A digital certificate is issued by a ***certificate authority (CA)***, a trusted party capable of asserting that the public key belongs to the entity in question. The facilities and mechanisms used to manage digital certificates are collectively known as a ***public key infrastructure (PKI)***.

A standard X.509 certificate [36] includes details about the certificate owner and the certificate authority, information about the certificate's validity, the digital signature of the certificate authority, and other information.

Relying parties can validate the certificate's signature to be assured of its authenticity, however, this requires the CA's certificate to be trusted. The CA's certificate might be signed by another CA, thus creating a ***chain of trust*** until a root CA that is implicitly trusted is found [69]. Relying parties are any entity that must depend on assertions made about the identity of another entity.

PKI is often used to generate digital identity certificates for individuals once sufficient identity attributes have been **verified** and the credential provider is reasonably confident in the individual's identity. These certificates form the basis of a digital identity infrastructure.

2.2 Identity Proofing

Identity proofing is a process where information about an individual is collected, validated, and verified to establish the applicant's identity with a certain level of confidence [11, 24, 61].

In an identity proofing process, the main parties involved are:

- the applicant (the individual who is claiming a particular identity), and
- the credential service provider (the party responsible for asserting whether the identity in fact belongs to the applicant).

As seen in Figure 2.1, identity proofing processes typically consist of three main steps [11, 60]:

1. **Resolution**. During the resolution step, the applicant gives the credential service provider the requested identity attributes (e.g., full name, birth date, home address) and other proofs of identity (e.g., photographs of identity documents). This step serves to uniquely identify the applicant.

2. **Validation**. During the validation step, the credential service provider confirms the authenticity of the information provided in the resolution step. This can involve consulting existing databases (e.g., existing government records, databases maintained by credit bureaus or consumer reporting agencies like Equifax [49]) or checking for signs of tampering (e.g., information embedded in barcodes that does not match the information on the document).

3. **Verification**. During the verification step, the credential service provider attempts to confirm that the identity being claimed by the applicant is associated with a real person. This can be achieved through measures such as requiring applicants to take a photo and checking that the photo matches information provided in the resolution step.

Identity proofing processes can be designed to allow applicants to complete the process in person, remotely, or a combination of the two [88]. With in-person identity

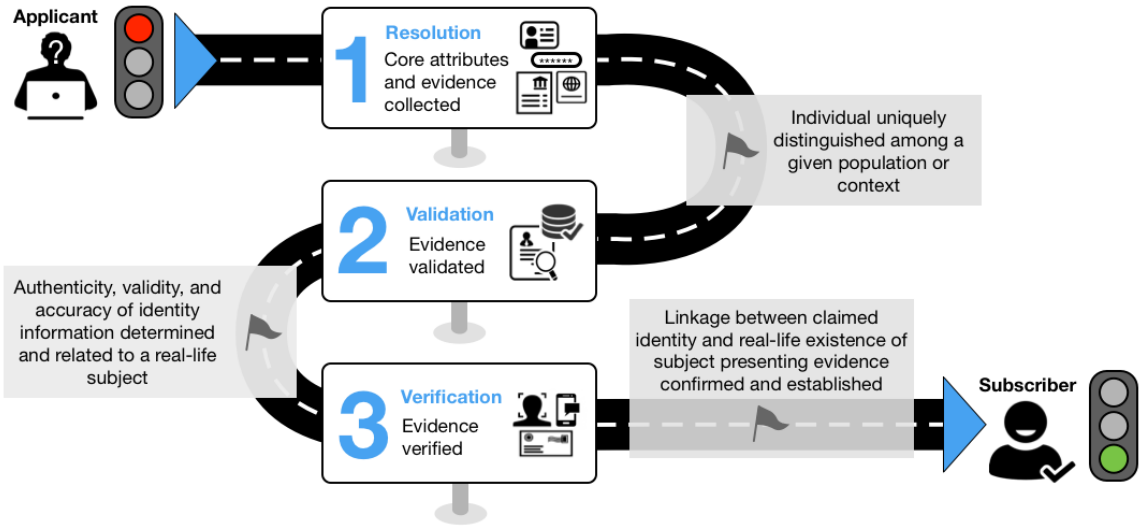


Figure 2.1: Typical steps in an identity proofing process. *Reprinted from the National Institute of Standards and Technology, U.S. Department of Commerce [60].*

proofing, applicants present proofs of identity to someone affiliated with the credential service provider for verification. While the ability to physically verify proofs may be desirable, in-person identity proofing can be expensive to implement (e.g., cost of training staff to recognize counterfeit documents or fraudulent behaviours, cost of deploying any specialized equipment) and inconvenient for applicants (e.g., travel time).

On the other hand, with remote identify proofing, proofs of identity can be provided over the Internet instead of requiring the applicant to be physically present. A downside of this option is that depending on the implementation, an attacker could pass off someone else's information as their own with relatively little effort (e.g., no need to create convincing identity documents).

Accordingly, some researchers have investigated how the weaknesses of remote identity proofing can be addressed. For example, Lewison and Corella developed a *rich credential* that allows three-factor identification (i.e., a private key on their device, a password, and biometrics) of applicants even if their credentials have never been registered with the credential service provider [76]. To accomplish this, the credential service provider is given additional metadata that can be used for verification (i.e., a biometric token and a **salted hash** of the password) in lieu of receiving the

user's password or biometrics. Augot et al. conducted similar investigations into this *zero-knowledge* [84] approach of identity proofing with blockchain technologies [12].

2.2.1 Real-World Implementations

GOV.UK Verify [57] allows residents of the United Kingdom to access online services provided by the government (e.g., filing tax returns, driver's license applications and renewals) once they have completed an identity proofing process with a third-party company, such as Experian [50] or Barclays [15].

Prior to the launch of GOV.UK Verify, the government conducted a survey to understand the likelihood of various demographic groups having the means to complete the required identity proofing [34]. For instance, the survey results indicated that those 75 years of age or older and those between the ages of 16 and 24 would be least likely to have the means. It was suggested that the younger cohort would be less likely to have the necessary identity documents and less likely to have a financial history that companies could use to verify their identity. The older cohort was less likely to have the required technology (i.e., smartphones and tablets) and less likely to have recent activity on their credit file that companies performing the identity proofing could use as evidence.

The survey also found correlation between an individual's occupation and having the means to complete identity proofing. For instance, those in "professional and managerial occupations" were more likely to have evidence accepted by the companies performing the identity proofing than the unemployed, students, and those in "non-classified occupations."

These results highlight the challenges associated with verifying the identities of individuals who may not have enough records with organizations such as the government or financial institutions.

One possible workaround that is currently being investigated is the use of online activity in identity proofing solutions. For instance, exploratory research on the use of online activity for GOV.UK Verify's identity proofing [128] yielded two key findings: (1) users were willing to allow access to their accounts for identity verification purposes, and (2) analyzing online activity can be effective for identity verification. Similarly, in

2019, the United States Department of State [10] announced that its visa applications would begin requiring applicants to provide "social media usernames, previous email addresses and phone numbers" for identity proofing processes. While these approaches may help identity proofing become more accessible, they also raise new questions about the implications for users' privacy.

2.2.2 Literature Gaps

While we identified research papers focusing on technical implementations of identity proofing and examples of identity proofing in use, there is limited understanding on the usability of these solutions and their impact on users' privacy.

2.3 Personally Identifiable Information

The effectiveness of any identity proofing solution will be severely compromised if the information that it requests from users can be easily obtained or spoofed. As such, a key aspect of understanding the potential for identity proofing solutions involves understanding confidentiality issues involving PII. In this section, we explore this topic from several different perspectives.

2.3.1 Metrics

When PII is made available to unauthorized parties, either intentionally or accidentally, this can have adverse effects for individuals and organizations. However, the magnitude of these effects can vary greatly depending on the types of PII involved and the context [99].

To help organizations estimate this potential for harm, NIST established criteria for assessing the *PII confidentiality impact level* of a given instance of PII [82]. The PII confidentiality impact level describes the extent to which the individual owning the PII and to which the organization maintaining the PII would experience adverse effects if the PII was "inappropriately accessed, used, or disclosed." The three impact levels (low, medium, and high) are described in Table 2.1.

Table 2.1: Definitions for PII confidentiality impact levels [61]

Impact Level	Condensed Definition
Low — "Limited adverse effect" on the organization or the individual	• minor financial losses • minor harm to individuals • organization's ability to achieve its mission is diminished but the organization can still accomplish its primary functions with reduced effectiveness • minor damage to the organization's assets
Medium — "Serious adverse effect" on the organization or the individual	• significant financial losses • significant harm to individuals but no loss of life and no life-threatening injuries • organization's ability to achieve its mission is significantly diminished but the organization can still accomplish its primary functions with significantly reduced effectiveness • significant damage to the organization's assets
High — "Severe or catastrophic adverse effect" on the organization or the individual	• major financial losses • severe or catastrophic harm leading to loss of life or life-threatening injuries • organization's ability to achieve its mission is severely diminished and the organization can no longer accomplish one or more of its primary functions • major damage to the organization's assets

NIST also identified six factors, listed in Table 2.2, that it recommends organizations take into account when assigning confidentiality impact levels [82]:

Table 2.2: Additional factors to consider when assessing PII confidentiality impact levels

Factor	**Example**
Identifiability	Typically, Social Insurance Numbers can directly identify individuals whereas less specific attributes, such as an individual's city of residence, often can not.
Quantity of PII	A data breach involving 10 million records might cause more harm than a breach of 10 records (e.g., the organization might have to compensate each affected individual).
Data Field Sensitivity	In many contexts, the unauthorized disclosure of a Social Insurance Number might cause more harm than a disclosure of an email address.
Context of Use	The names and home addresses of an intelligence agency's undercover agents likely have a higher confidentiality impact level than the names and home addresses of a newspaper's subscribers.
Obligation to Protect Confidentiality	An organization operating in a jurisdiction with strict privacy laws might face stronger penalties for an unauthorized disclosure than an organization operating elsewhere.
Access to and Location of PII	A data set that is accessible to many people or can be accessed remotely might be at greater risk of being compromised.

While these metrics may be used in practice (e.g., [42, 51]), the research literature surrounding their use is sparse, making it difficult to assess their impact or the implications arising from their use.

2.3.2 Prior Work

Much of the literature concentrates on how de-identified publicly or semi-publicly available data can ultimately compromise users' privacy. De-identification is a process where identifying information is removed from a data set to minimize the chances of any single person being identified by their data [98]. Conversely, re-identification involves attempting to discover the identity of an individual in the data set by matching their anonymous data to information from other non-anonymized data sets [98].

The literature suggests that individuals can often be re-identified with relatively few identity attributes. After analyzing data from the 1990 United States Census, Sweeney [118] concluded that 87% of the U.S. population could likely be uniquely identified with just three identity attributes: their ZIP code, their gender, and their date of birth. An analysis by Golle [54] using more recent census data supported these findings, although only 63% of the population was likely to be directly identified with the same three attributes.

Others have more concretely demonstrated how feasible it is for individuals to be re-identified. Sweeney's work [118] illustrated this by linking anonymized hospital discharge data from various U.S. hospitals to voter registration lists, which would ultimately reveal information like an individual's diagnosis, any procedures they underwent, and medications they were taking. Journalists [14] were able to re-identify anonymized search queries that AOL had published for researchers to examine. After identifying patterns in the publicly-available Death Master File (a database from the U.S. government containing the Social Security numbers of the deceased) [126], Acquisti and Gross [3] were able to predict most digits in a person's Social Security number if they knew the person's date of birth and place of birth.

Munson et al. [86] conducted a survey with U.S. participants to investigate awareness and attitudes regarding the availability of public records online—specifically, records of donations made to political campaigns and records of real estate transactions. A major outcome of the survey was that there was no difference in the level of awareness between those who were represented in the records (i.e., those involved in real estate transactions and those who made political donations) and those who were not. This finding led the authors to suggest that organizations could improve

how data practices are communicated but cautioned that handling this poorly could discourage people from participating in activities that will ultimately end up being publicized.

Irani et al. [71] analyzed the social footprints (i.e., online presence across multiple social networks) of over 8,000 users to quantify the degree to which select identity attributes can be discovered for a particular user. The study revealed that the "leakage" of identity attributes generally worsens as the number of social network profiles for a given user increases.

Focusing more on criminal behaviours, Copes and Vieraitis [37] conducted interviews with 59 incarcerated identity thieves to learn more about their strategies for carrying out these crimes. A major finding of these interviews was that the thieves' most frequently used methods for acquiring other people's information included purchasing the information (e.g., from employees at various businesses and government agencies, from the black market), taking the information from trashcans and mailboxes, and taking it from friends, family and acquaintances (sometimes with their permission).

Together, this literature provides valuable insight into how the confidentiality of PII can be compromised through public records and through more calculated attempts. However, there appears to be a limited amount of literature examining vulnerabilities in certain mediums, such as the physical spaces users occupy in their daily lives and social networks.

2.3.3 Legislation

As of 2019, 132 countries have enacted privacy laws [63]. In Canada, the *Personal Information Protection and Electronic Documents Act (PIPEDA)* stipulates the rights that individuals have with respect to their personal information and also governs how organizations must conduct themselves when making use of personal information [43].

One of the more impactful pieces of recent legislation is the European Union's *General Data Protection Regulation (GDPR)* [119], which went into effect in 2018. The GDPR requires companies that collect personal information to allow users to determine what information is being maintained about them, how the company will

be using that information, and to be able to request that their information be deleted. Companies that fail to comply can face fines of up to 20 million EUR (approximately 30 million CAD) or four percent of their annual revenue, whichever is higher.

Some have described the GDPR as having a "Brussels effect" [23]—a phenomenon where legislation enacted in the European Union is able to shape laws and markets abroad [40, 107]. The GDPR had international impact because it was applicable not only to companies operating in the EU but to companies *serving* EU residents. This resulted in many companies around the world updating their privacy policies to be GDPR-compliant [32] and building controls for users to access their data [47, 70, 113].

The GDPR has also prompted other jurisdictions to enact similar legislation, such as the *California Consumer Privacy Act (CCPA)* [26], the *Deceptive Experiences To Online Users Reduction (DETOUR) Act* in the US [1], or Brazil's *General Data Protection Act* [39].

While predicting how these legal changes will impact digital identity services is difficult, the current and pending legislation may pressure such services towards offering basic privacy protections and transparency to their users as a baseline.

2.3.4 Literature Gaps

While we identified literature on the subject of the confidentiality of PII, most of this research was limited to the digital world. To construct a more comprehensive understanding of the confidentiality issues involving PII, more aspects of the analog world should be studied (e.g., people's interpersonal relationships, habits with personal belongings).

2.4 Authentication

Authentication is a process through which users prove that they are who they claim to be [35, 84]. Users can accomplish this by presenting proofs known as *authentication factors*. Typically, there are three types of authentication factors [6]:

- **knowledge factors** (or something the user knows), such as a password;
- **possession factors** (or something the user has), such as an identity card; and

- **inherence factors** (or something the user is), such as a fingerprint.

On their own, each of these factors is easier to compromise than if multiple factors were required [120] so it is often advised that multi-factor authentication be implemented. Multi-factor authentication requires users to provide two or more factors as proof. For example, many email providers implement two-factor authentication by allowing users to provide their password as a knowledge factor and a temporary code sent to their mobile phone as their possession factor.

Many of the same principles established for authentication can be applied to identity proofing. In particular, the use of multiple factors of different types is a valuable approach as proof of identity.

2.4.1 Digital Identity

Over the last couple of decades, many countries have considered adopting eIDs; however, the switch from paper-based systems to an electronic one is considered a fundamental transition in how citizens and governments interact [4] and can have socio-technical impact across many aspects of daily life. Further, the adoption of eID can also have significant privacy and security implications. We provide an overview of some of the early implementations of these systems in which some of these challenges arose.

In 1999, Finland became the first country to issue eIDs, however, these were non-mandatory [103]. Finland's eID could be used as a travel document, to access government services, and to sign documents electronically. Finland's implementation of eID did not see widespread adoption; ten years after its initial launch, only 10% of citizens had acquired one. Citizens felt that the eID was not necessary and there were also usability and compatibility issues, particularly with using it for web-based services. We learn two lessons here relating to adoption. First, adoption will be slow unless mandatory, and second, compatibility is difficult but critical to large-scale adoption.

The usability issues with Finland's eIDs further serve as a reminder that for users, digital identity and its management is not a goal in itself [44]. According to Dhamija and Dusseault [44], its usability needs to be seamless, secure, and private so that users can focus on their primary tasks. Besides general usability, this may mean limiting

user options in some cases to minimize chances of over-disclosing information. As an example, users might inadvertently select all options when asked which data to share if it is too confusing to figure out the consequences of not sharing. Digital identity services may also introduce confusion because now users need to figure out who to contact to resolve issues—the organization that issued their card or the organization requesting their credentials?

In the early 2000s, Austria began deploying its Citizen Card (CC) [4], a technology-agnostic implementation of eID that is compatible with a wide range of devices such as cell phones, smart cards, and USB tokens. CC is primarily used for accessing government services and the most popular format of the technology is the smart card. Distinct from other implementations, CC uses a *sector-specific personal identifier (ssPIN)* for each sector (e.g., health sector, tax sector, education sector). ssPINs can only be used to identify a user within the originating sector, thus preventing cross-identification. For example, one's identity with a health institution cannot be linked to their use of the identity for tax filing. This privacy-preserving feature has been important in promoting adoption among a population that is generally privacy-conscious.

Aspects of Austria's solution relate to Hansen et al.'s [65] assessment of 'partial identity'—the idea that in most transactions, only a subset of your identity attributes are necessary for the exchange. Managing how these partial identities are used can be unwieldy given that an individual's digital footprint tends to continuously grow and rarely shrinks. For instance, even upon death, one's Social Insurance Number and associated data might be persisted to facilitate the payment of benefits to a dependent. The authors further discuss issues relating to partial identity management throughout a user's lifespan, including privacy and security concerns, and suggest mechanisms for mitigating these issues, such as packaging privacy policies with user data and logging any misuse automatically.

In 2006, Belgium [38] was the first European country to make electronic identity cards (eIDs) mandatory for its citizens. The cards were plastic smart cards with built-in microchips for storing the cardholder's information (e.g., name, photograph, date of birth, card expiration date) which was also printed on the card itself. The

microchip could also be used for generating keys and validating digital signatures. Notably, these cards lacked features for preserving the cardholder's privacy.

Estonia's eID smart card is also mandatory for its citizens [110]. The card can be used to obtain access to services like banking, voting, and health insurance. Like other cards, its microchip includes users' information but also stores a key pair used to sign information shared with a relying party. However, in 2017, a software vulnerability made it possible for attackers to derive the card's private key from the corresponding public key [16, 55]. This highlights a concern held by many users: how safe is their data and to what extent might it be compromised if breaches occur?

Sweden [64] took a different approach, using a 'market solution' where it defined the legislative framework, but then allowed different providers (mainly banks) to issue eIDs. These could all be used to access government services. The eIDs were made available in two formats—a 'soft' ID in the form of a file on the user's computer, and a 'hard' ID that was embedded into a physical card. Each eID included the user's personal identity number, which is used across all government services. The approach enabled the government to set out the framework without the large infrastructure investment.

Ultimately, digital identities continue to be an active area for innovation as different countries work towards solutions that meet particular needs. This is evidenced by recent patent filings, covering different technical aspects (e.g., [2, 74]) and continued interest from governments around the world [46, 58, 124].

2.4.2 Federated Identity

Federated identity is a mechanism that allows identity information to be used across otherwise separate systems [62, 106]. With federated identity, a user's identity information is stored with an identity provider. When a relying party requires authentication, the user authenticates with the identity provider instead of the relying party.

Rountree [106] explains that trust is an essential component of federated identity. Relying parties never receive users' credentials; only whether the user authenticated successfully and, in some cases, their identity information. As such, relying parties

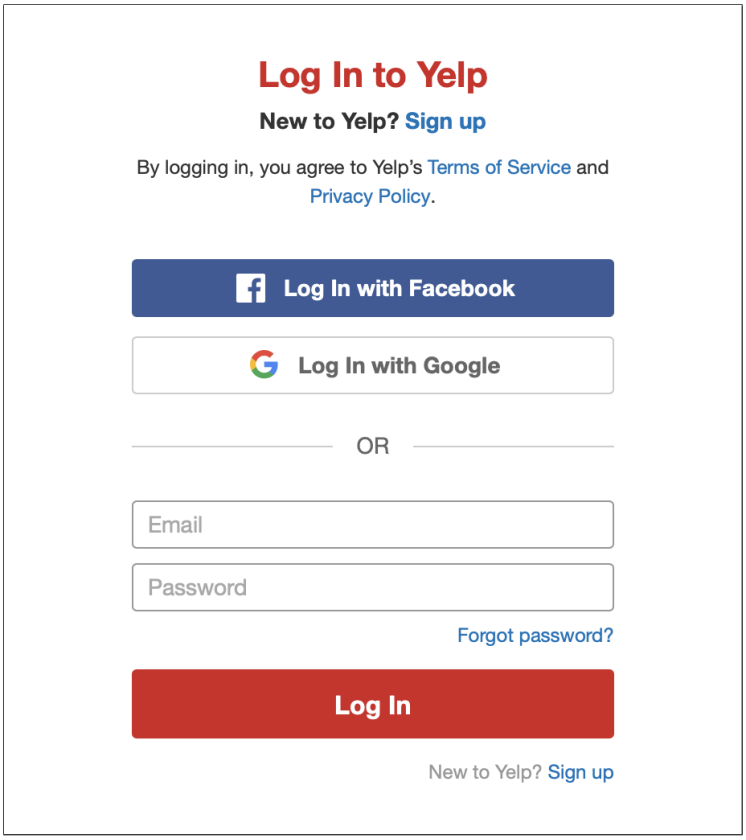


Figure 2.2: Login form for Yelp with buttons for authenticating via Facebook or Google.

must be confident in the abilities of the identity provider to authenticate users and to attest to the validity of users' identity information. Conversely, the identity provider must be confident that the relying party will not act maliciously with any user information they receive.

A common use case for federated identity is single sign-on where a relying party can support signing in via an identity provider. Figure 2.2 shows Yelp's implementation of single sign-on that allows users to sign in with their Facebook or Google credentials.

Industry standards, such as OpenID [121], OAuth [93], and SAML [92], are commonly used to implement federated identity.

In essence, many eID solutions are federated identities since they enable users to access multiple services while only providing an agreed-upon subset of the user's identity attributes. In some instances, only the attestation is provided rather than specific attributes. For example, the system might attest that the user is old enough

to purchase alcohol without disclosing the user's birth date to the merchant.

2.4.3 Literature Gaps

The literature shows continued advancements in the functionality of digital identity and federated identity solutions, but we found few works examined how they fare with users. This presents an opportunity to analyze these technologies from a usability perspective.

2.5 Summary

While the literature we identified covers subjects like identity proofing, PII, and authentication, there is still room for additional research to provide insight into the design of smartphone-based digital identity services. Specifically, much of the literature surrounding identity proofing and authentication focused on implementation and did not analyze these solutions from a usability or privacy perspective. Similarly, the PII-related literature was not comprehensive in that key aspects of the analog world were infrequently discussed. Through studies on the discoverability of PII (both online and offline) and a user study of a digital identity service, this thesis aims to make novel contributions by addressing these gaps.

Chapter 3

Information Discoverability Study

With digital identities, **issuing authorities** often depend on identity proofing mechanisms to be assured that an individual is who they claim to be. These identity proofing mechanisms typically use an authentication method known as **knowledge-based authentication** (**KBA**), where applicants are asked to provide information about themselves which can be verified with sources that maintain data about the individual [33]. A major issue with this type of identity proofing, especially in an age where personally identifiable information is increasingly available online, is ensuring that the challenges can be answered by the applicant while still being difficult for an imposter. To explore this problem, we conducted a study with twenty-eight participants to answer the following research questions:

RQ1.1 To what degree are certain types of PII discoverable through social media services and search engines?

RQ1.2 How do an individual's behaviours, both online and offline, relate to the ease with which an imposter could obtain their PII?

We use the findings of this study to make recommendations on the types of PII that are best suited for identity proofing.

3.1 Methodology

In this section, we describe how we recruited participants, the equipment we used to conduct the study, and the procedures we followed during each session. The study received ethics clearance from Carleton University's Research Ethics Board-B (CUREB-B) on April 16, 2018 (Project #108764).

3.1.1 Recruitment

To recruit participants, we advertised the study by placing posters (see Appendix A.1) across the Carleton University campus with our contact information. We also shared information about the study through social media groups, such as a Facebook page for announcing on-campus HCI user studies.

When prospective participants reached out to us, we sent them an informed consent form, information on how to sign up for available time slots, and asked them to indicate their three most used social media websites and apps. This allowed us to familiarize ourselves with the interfaces and privacy settings for any social media services we had not previously encountered before their session. Participants were also asked to bring either the login information for these accounts or to bring their own device that was already signed into these accounts.

We required all participants to be at least eighteen years of age, fluent in spoken English, and active users of at least one social media service.

3.1.2 Procedure

Sessions were either conducted on campus at Carleton's Human Oriented Research in Usable Security (CHORUS) Lab or off campus at mutually convenient locations. All sessions were conducted in environments that offered participants sufficient privacy and confidentiality (e.g., experiment rooms at CHORUS Lab, quiet areas of public libraries).

We paid participants CAD$25 in cash and an additional CAD$10 for eligible parking expenses. The interview portion of each session was audio recorded if the participant gave consent. Sessions were approximately 90 minutes in length and included the following activities:

1. **Consent form.** Participants were asked to read and sign the consent form (see Appendix A.2), which outlined what the study would entail, described the implications and possible risks associated with participating, and gave participants the ability to choose whether their interview would be audio recorded. All participants received a copy of the consent form via email in advance of their

session to give them ample time to review it.

2. **Demographic questionnaire.** Participants completed a demographic questionnaire (see Appendix A.3) focusing on demographic information, social media usage (including questions about their three most-used accounts with social media services), and their attitudes and behaviours with respect to security and privacy.

3. **Interview.** Through a semi-structured interview (see Appendix A.4), participants were asked about their practices for keeping personal belongings and electronic devices safe, their experiences with data breaches, and their attitudes towards technologies like mobile wallets and digital identities.

4. **Social media exercise.** Starting with their most used social media account, participants were asked to look up generic information about their account, such as their friend count and privacy settings, and to browse their account in search of specific types of PII. Participants completed these activities while signed into their account so that we could see any information associated with their account, including any content that had privacy restrictions.

   The generic information was recorded by the researcher using a paper form (see Appendix A.5). For the PII, we provided participants with a paper checklist of categories of PII to look for (see Appendix A.6). To protect their privacy, participants only recorded the presence or absence of PII (as opposed to recording the PII itself) and remained in full control of the device being used to browse the account. The researcher made sure to observe the presence of the PII on the account and took note of contextual information to aid in classifying the results later on. These notes did not include participants' actual PII.

   For each account declared in the demographic questionnaire, participants were given about ten minutes to complete this activity, however, time constraints did not always allow us to conduct this analysis for all of the accounts declared by a participant.

5. **Search results exercise.** Participants performed searches with Google in

order to find their PII online using queries provided by the researcher (see Appendix A.7). When search results appeared to be about the participant, they opened the search result to verify whether the page contained any of their PII. As before, participants used a paper checklist (see Appendix A.8) to record the presence or absence of their PII and the researcher took note of the context in which the PII was found.

6. **Clear browser data.** On the computer provided to participants, the researcher cleared the browser's data to ensure that no history, cookies, site data, or information entered into forms was saved.

7. **Debriefing.** At the end of the session, participants were given a debriefing document (see Appendix A.9), which reiterated our study's purpose and included our contact information.

3.1.3 Equipment

All participants completed the demographic questionnaire and the search results exercise on a computer that we provided. For the social media exercise, participants were able to use either their own device or our computer.

When participants used our computer, they were set up with a private browsing (incognito) window to ensure that no history, cookies, site data, or information entered into forms was saved. As an extra precaution, we also cleared the browser's data at the end of the session.

3.2 Data Analysis Strategy

3.2.1 Quantitative Data

To communicate the discoverability of the identity attributes that participants were instructed to look for, we classified the results of the social media and search results exercise according to the six terms defined in Table 3.1: *Identified*, *Deducible*, *Not Identified*, *Outdated*, *Fake Data*, or *Not Applicable*. We analyzed this data and responses to the demographic questionnaire using Numbers for Mac [7] and used

summary statistics to highlight key findings. To facilitate an analysis of these findings, we discuss the discoverability of the identity attributes in terms of six categories defined in Table 3.2.

Table 3.1: Terms used to classify results of social media and search result exercises

Classification	**Definition**
Identified	The attribute was located, was correct, and was clearly an attribute of the participant's identity (e.g., the participant's first name was found).
Deducible	The attribute was located and was either incomplete or was missing a clear indication that it was an attribute of the participant's identity (e.g., inferring an individual's birth year from birthday greetings seen on their profile).
Not Identified	The attribute was not located.
Outdated	The attribute was located and was correct at some point in the past but is no longer true or valid (e.g., email addresses that are no longer valid).
Fake Data	The attribute was located but was never true or valid at any point (e.g., incorrect birth dates).
Not Applicable	The attribute did not apply to the participant (e.g., participants without middle names).

Table 3.2: Categories of identity attributes

Category	**Identity Attributes**
Personal Information	First name, middle names, last name, gender, nationality, clearly identifiable photo(s) of the participant, marital status

Table 3.2 (*continued*)

Category	**Identity Attributes**
Birth Information	Place of birth, year of birth, month of birth, day of birth
Address Information	Street (current), street (where lived the longest), city (current), city (where lived the longest), province (current), province (where lived the longest), postal code (current), postal code (where lived the longest)
Contact Information	Email address, phone number (home), phone number (mobile)
Occupational Information	Employer or school, job title
Vehicle Information	License plate number, car make, car model, car year

3.2.2 Qualitative Data

In terms of qualitative data, we had captured approximately nine hours of audio recordings from sessions with twenty-seven participants. The researcher then transcribed these audio recordings resulting in 314 pages of lightly edited transcribed materials that we included the collection of qualitative data being examined as part of this analysis (hereafter referred to as the *dataset*). We also included four pages of notes taken by the researcher during two participants' interviews as they did not consent to being audio recorded.

Content Analysis

Using NVivo [100], a qualitative data analysis program from QSR International, the researcher and an undergraduate research assistant experienced with qualitative data analysis applied qualitative content analysis methodology [48] to analyze the dataset.

Content analysis is a systematic research method for making inferences about a dataset (typically textual data) in order to answer research questions [75, 129]. In

most cases, this involves three phases:

- A **preparation phase** whereby researchers read through the dataset to familiarize themselves with the content.
- An **organizing phase** whereby meaningful parts of the data are labeled with *codes* and analyzed to identify patterns or themes. Researchers can take an inductive approach (i.e., themes are derived from the dataset), a deductive approach (i.e., the researcher uses existing theories or knowledge to define themes before starting their analysis), or a mixture of the two.
- A **reporting phase** whereby the results of the analysis are described and an interpretation of the findings is provided.

The researcher and the researcher assistant were each assigned a set of documents to code. The overlap of these sets represented one-third of the dataset (i.e., eight transcripts and one note). These quantities were proportional to the types of documents contained in the full dataset (25 transcripts and 2 notes).

We began the process by taking a deductive approach and developing a categorization matrix consisting of five categories derived from our research questions (see Table 3.3). We then coded any excerpts from the dataset that we deemed meaningful by assigning them to relevant categories from the matrix. Excerpts were assigned to multiple categories if we felt that the excerpt was better described by more than one category. Throughout this initial round of coding, we met to compare our coding decisions. We resolved any disagreements by discussing the rationale behind our coding decisions and either moving excerpts between categories, clarifying the criteria for a given category, or revising a category's definition.

Table 3.3: Categories used for qualitative data analysis

Category	**Description**	**# of coding assignments**
Breaches and violations	Comments about breaches (including physical breaches and data breaches), hacking, or other forms of unauthorized access.	663
Perception of digital technology replacing analog technology	Comments about digital technologies that can act as substitutes for analog technologies (e.g., Google Pay replacing cash or credit cards, a digital identity app replacing a driver's license).	793
Precautions and safety measures	Behaviours that the participant performs or has in place to protect their physical possessions (e.g., wallet, keys) or digital/electronic possessions (e.g., smartphone, online accounts).	549
Risky behaviours and vulnerabilities	Behaviours that the participant performs or has in place that hinder the protection of their physical possessions (e.g., wallet, keys) or digital/electronic possessions (e.g., smartphone, online accounts).	282
Trust or distrust of others	Trust or distrust of other human beings or organizations. Comments about the participant's confidence or lack thereof in other people or in organizations.	236

Next, we used an inductive approach to develop subcategories within each of the five main categories. This involved reading through the relevant excerpts identified in the initial round to establish appropriate subcategories. This process yielded 19

subcategories and 10 tertiary categories. We then performed multiple iterations of applying these newly developed subcategories and tertiary categories to the excerpts, and resolving disagreements until we had obtained sufficient inter-rater reliability. After eight rounds of coding completed over the course of approximately seven days, percent agreement was 99.86% demonstrating that our analysis met the recommendations set forth in relevant works (i.e., 80% agreement or higher) [83].

In our analysis, we deliberately avoided relying on quantitative measures as an indicator of the importance of any excerpts, as literature [41] warns that there is "no simple relationship" between quantitative measures for a given text and its significance.

3.3 Results

3.3.1 Demographic Questionnaire

We collected data from twenty-eight participants but removed one participant's data from our analysis as he had difficulty completing the questionnaire and answering interview questions. Of the remaining twenty-seven participants, twelve were male and fifteen were female. The participants ranged in age from 18 to 35 years with an average age of 23.5 years ($SD = 4.8$ years); two participants did not disclose their age.

The majority of the participants ($n = 16$) reported high school as the highest level of education that they had completed. The remainder had obtained undergraduate degrees ($n = 5$), post-graduate certificates or diplomas ($n = 3$), graduate degrees or professional degrees ($n = 2$), or had completed college ($n = 1$).

Two-thirds of participants ($n = 18$) described themselves as students in the survey while the remainder were employed in various industries.

Figure 3.1 characterizes the security of our participants' smartphones. In terms of device locking, 22 participants were using a PIN or password to unlock their smartphone, 17 participants had enabled fingerprint authentication, and 16 participants had configured their phone to automatically lock after a period of time. Only one participant had jailbroken or rooted their device, a process which gives the user more privileges within the operating system but consequently, removes certain security mechanisms.

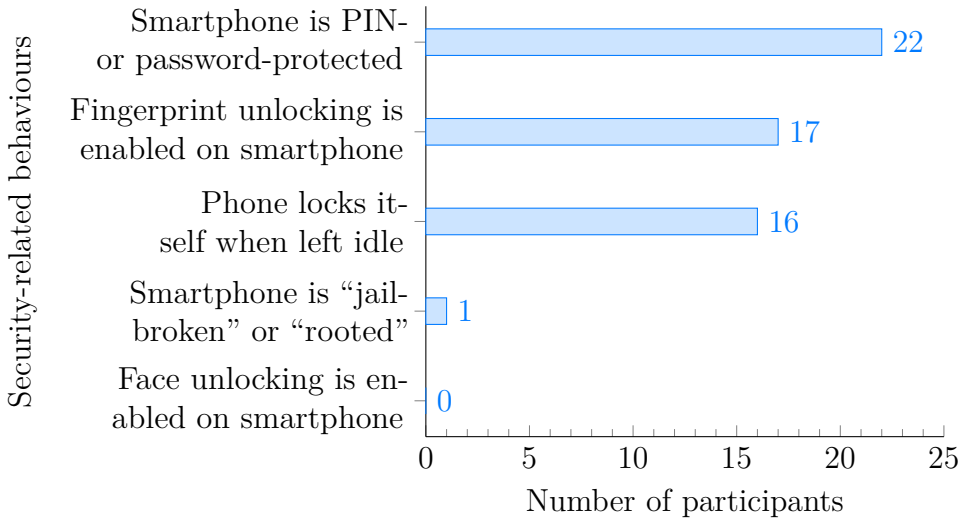


Figure 3.1: Frequency of security-related behaviours on participants' devices. $N = 27$ participants.

As seen in Figure 3.2, when participants were asked to report the social media services that they use most often, Facebook, Instagram, and YouTube were given as answers most frequently.

3.3.2 General Questions

Table 3.4 breaks down by social media service the total number of accounts that were analyzed as part of the social media exercise. These accounts only represent a subset of the accounts disclosed to us by participants in the demographic questionnaire as time constraints did not allow us to analyze all accounts. Together, Facebook and Instagram accounted for 83 percent of the accounts that we analyzed.

Figure 3.3 shows the average number of friends, followers, or subscribers on the accounts that we analyzed. On average, the Facebook accounts had 692.3 friends (SD = 665.0 friends, median = 449.0 friends)[1], the Instagram accounts averaged 329.6 followers (SD = 222.3 followers, median = 302.0 followers) while accounts with other

[1]During the study, we began to notice that the friend count shown by Facebook for a single participant sometimes differed from screen to screen. Online reports of this phenomenon suggest that Facebook includes deactivated accounts in the total friend count displayed on some screens [130], possibly for infrastructure reasons [132]. Once we noticed this pattern, we took note of all friend counts that we observed for a given participant. The results presented are based on each participant's lowest observed friend count.

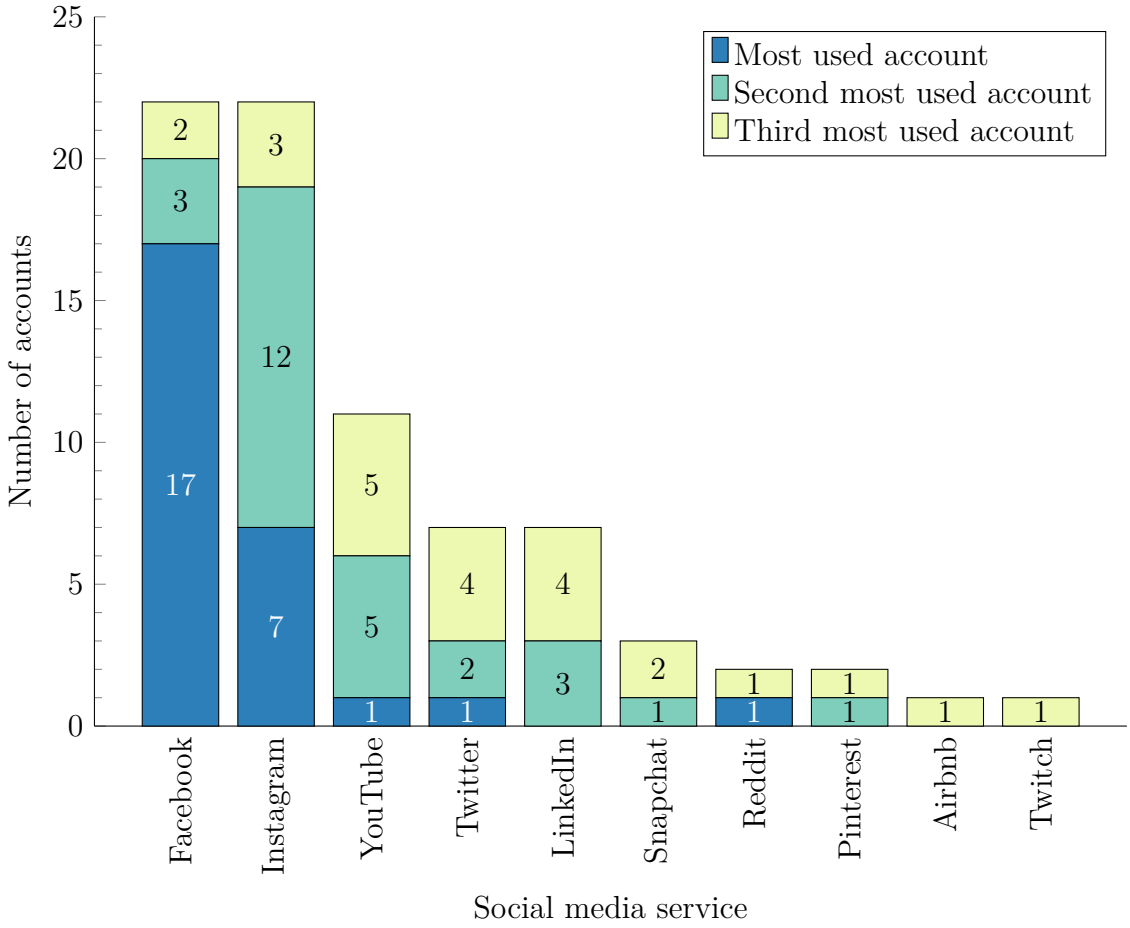


Figure 3.2: Participants' most used social media accounts. $N = 78$ accounts.

Table 3.4: Total number of accounts analyzed, grouped by social media service

Social Media Service	**Number of Accounts Analyzed**
Facebook	19
Instagram	15
LinkedIn	2
Twitter	2
Airbnb	1
Reddit	1
YouTube	1
Total	41

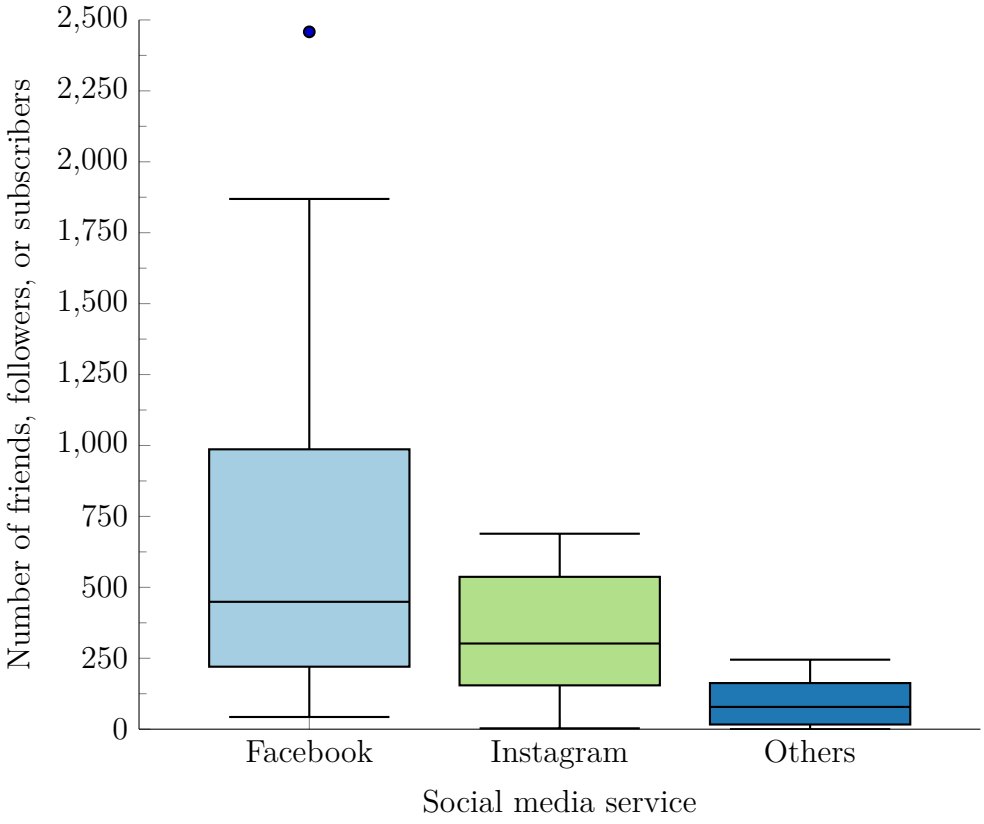


Figure 3.3: Number of friends, followers, or subscribers on accounts analyzed for the social media exercise

social media services averaged only 100.5 followers (SD = 113.0 followers, median = 78.5 followers). Accounts with services that did not have a notion of friendship or followership, such as Reddit and Airbnb, were excluded from these statistics.

Figure 3.4 shows the degree to which the contents of participants' accounts were visible to others. Of the Facebook accounts, the majority (13 out of 18 accounts) had configured their future posts to be publicly viewable while the remainder had different degrees of visibility. Of the Instagram accounts, 8 out of 14 accounts restricted their content to being viewed by their followers whereas the content of the remaining accounts were viewable by the public. Finally, for the accounts with other social media services, the majority (5 out of 7 accounts) made their content viewable to the public and two accounts only made their content viewable to other registered users of the service.

Many social media services allow users to upload their contacts in order to find the

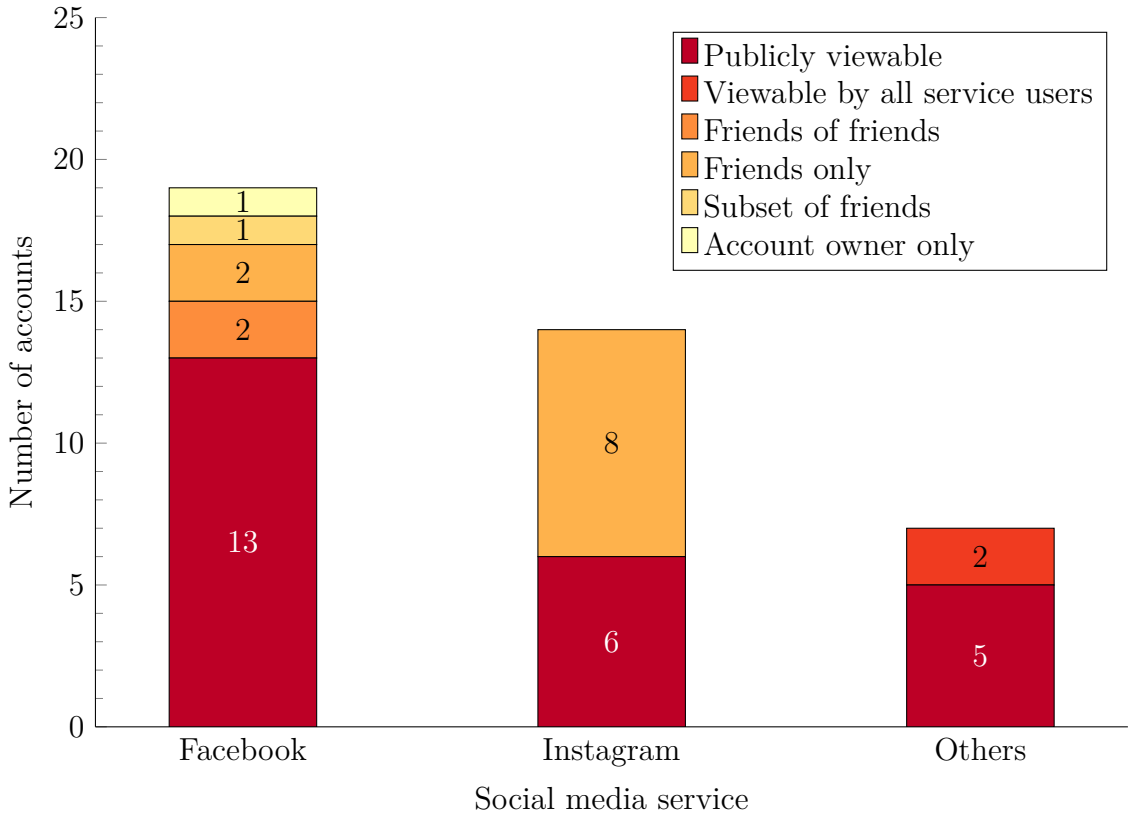


Figure 3.4: Restrictions on content visibility for accounts analyzed for social media exercise

accounts of other users they might know. Figure 3.5a shows the degree to which our participants' accounts were configured to allow other users to look them up by email address. On Facebook, eleven accounts were using the most restrictive option (*Friends*) while eight accounts allowed all users to find them by email. On Instagram, users cannot restrict who can look them up by email since email addresses are mandatory at registration, thus it would appear that all Instagram users can be found by another user who has their email address. Finally, of the accounts on other social media services that supported lookup by email, two had restricted lookup to *Nobody*, one had restricted it to *Friends of Friends* while another had allowed *All Users*.

Figure 3.5b shows the degree to which participants' accounts were configured to allow other users to look them up by phone number. Ten Facebook accounts allowed *Everyone* while nine accounts had restricted this capability to *Friends*. Notably, upon seeing that their account was configured to allow *Everyone* to look them up by their phone number, P8 immediately restricted this setting. As before, Instagram users cannot restrict who can look them up by their phone number and, as such, all users could be looked up by the phone number associated with their account. For the other accounts with social media services that supported this type of lookup, three allowed *Nobody* and one allowed *Friends of Friends*.

3.3.3 Analysis of Social Media Accounts

Figure 3.6 shows the discoverability of participants' PII on their social media accounts. Across all participants, the most easily *identifiable* category of identity attributes was *Personal Information* ($n = 27$ participants). Within this category the most easily identifiable identity attributes were first names ($n = 26$ participants), clearly identifiable photos ($n = 25$ participants), and gender ($n = 20$ participants). The most easily *deducible* category was *Birth Information* with place of birth being the most commonly deduced attribute within this category ($n = 10$ participants).

Vehicle Information was classified most frequently as *not identifiable* ($n = 18$ participants) and was only identified once through a participant who had a picture of their car on their Facebook account. However, it is worth nothing that this category was *not applicable* for eight participants who were not drivers.

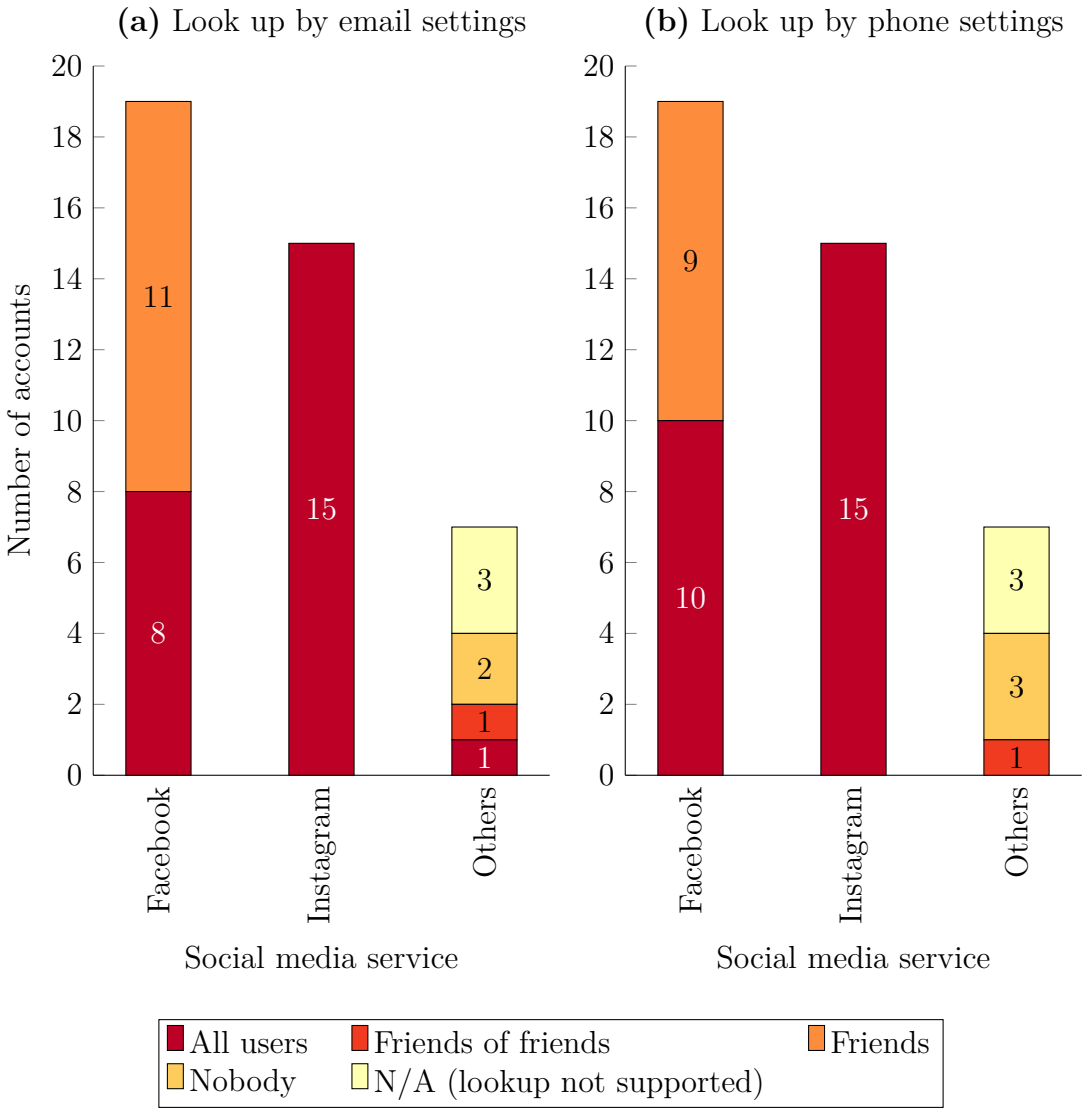


Figure 3.5: Lookup settings for participants' accounts.

Contact Information was the category that was most frequently *outdated* (n = 2 participants). This included two instances of outdated email addresses and one instance of an outdated mobile phone number.

3.3.4 Analysis of Search Results

Figure 3.7 shows the discoverability of participants' PII through Google searches including one participant who did not complete the search exercise due to time constraints. Similar to the results of the social media exercise, the most easily *identifiable* category of identity attributes was *Personal Information* with first name (n = 25 participants), last name (n = 24 participants), and clearly identifiable photos (n = 17 participants) being the top three most identified attributes. *Birth Information* was the most easily *deducible* category with year of birth (n = 3 participants) and place of birth (n = 2 participants) being this category's most easily deduced attributes.

Birth Information also tied with vehicle information for being classified most frequently as *not identifiable* (n = 18 participants). For birth information, day of birth (n = 24 participants) and month of birth (n = 24 participants) were the attributes that were most frequently classified as *not identifiable*. For *Vehicle Information*, all four attributes were classified as *not identifiable* 18 times.

Occupational Information was *outdated* most frequently (n = 2 participants) with participants' employer or school being outdated most frequently. Finally, the information we discovered was rarely faked. In fact, *Birth Information* was the only category classified as *fake data* (n = 1 participant) and even this category's most frequently faked attribute was year of birth (n = 1 participant) which was only faked once.

3.3.5 Interview

We highlight findings from our qualitative data analysis according to the five main categories we established. Table 3.5 provides the complete structure of our categorization matrix.

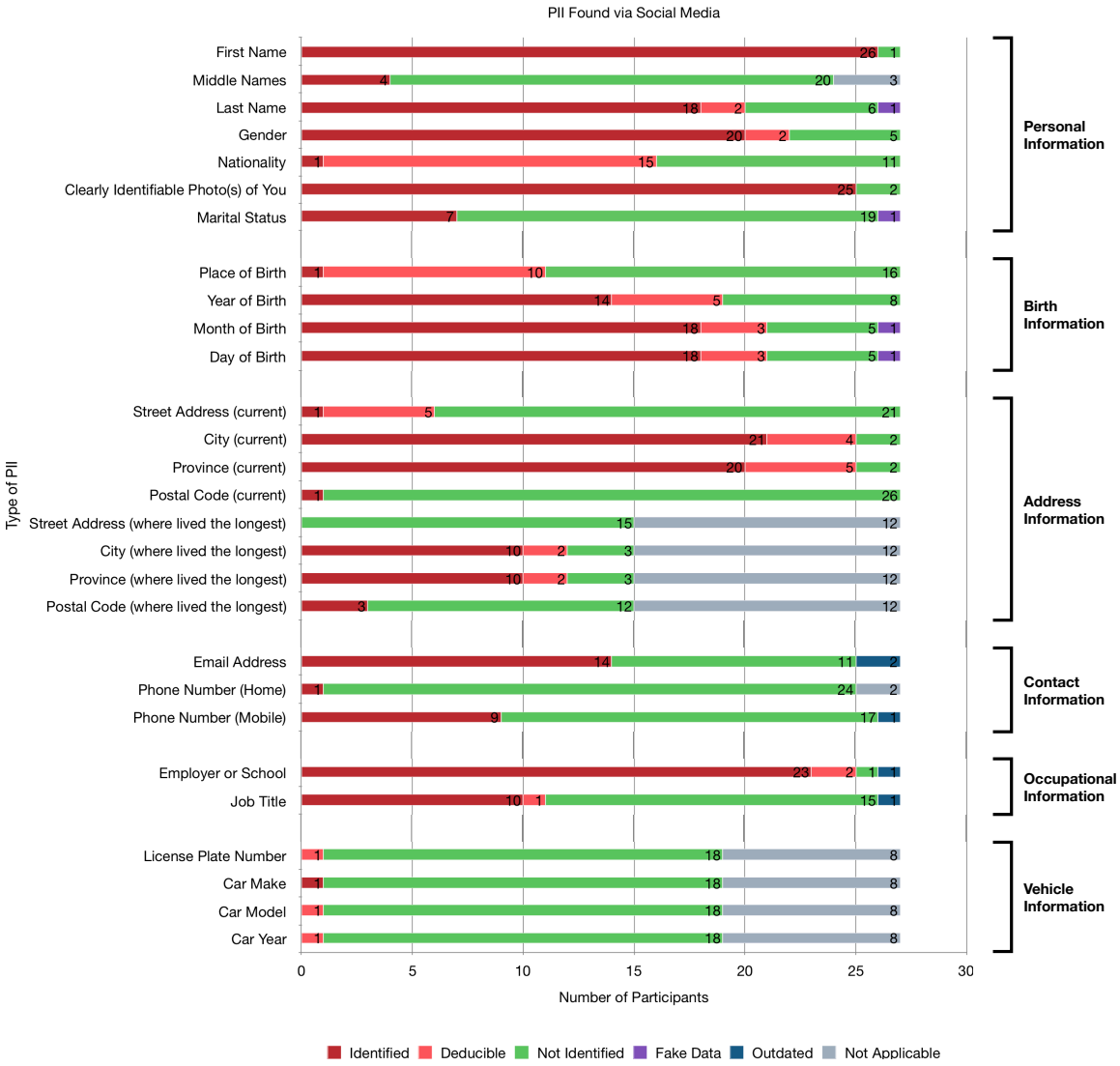


Figure 3.6: Discoverability of different types of PII via social media. $N = 27$ participants.

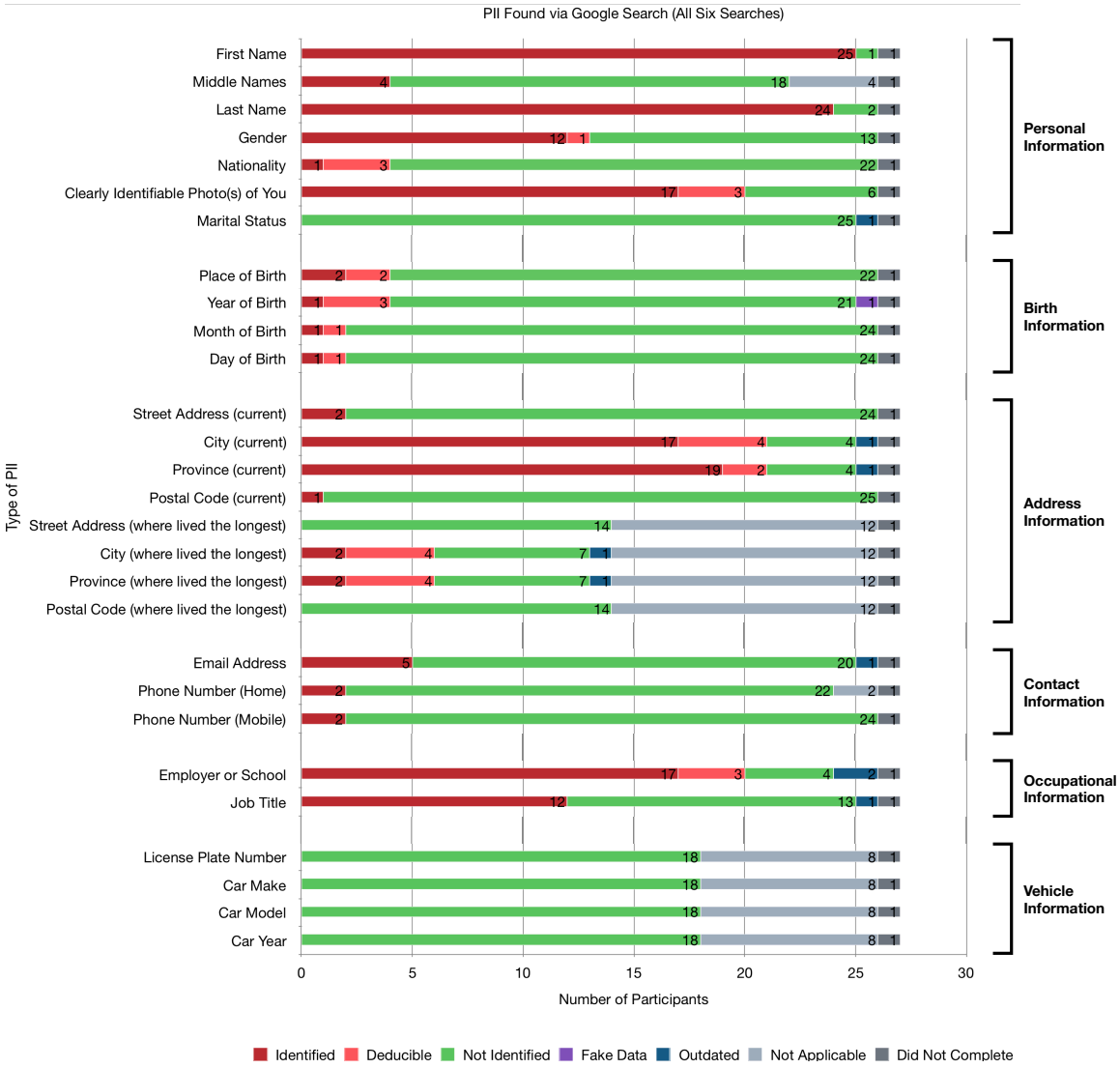


Figure 3.7: Discoverability of different types of PII via search. $N = 27$ participants (including one participant who did not complete the search exercise).

Table 3.5: Categories and subcategories from qualitative data analysis

Category	**Subcategories**
Breaches and Violations	• Digital Breaches ◦ Identity Theft ◦ Unauthorized Modifications ◦ Unauthorized Use ◦ Unauthorized Access of Personal Data • Perception of Risk • Personal Experiences • Physical Breaches
Perception of Digital Technology Replacing Analog Technology	• Factors Influencing Adoption ◦ Availability ◦ Convenience ◦ Government Knows All ◦ Influence of Individuals and Society ◦ Interest ◦ Security and Privacy • Preferences • Technical Knowledge
Precautions and Safety Measures	• Devices • Digital • Physical Environments and Possessions
Risky Behaviours and Vulnerabilities	• Devices • Digital • Physical Environments and Possessions • Self-Awareness

Table 3.5 (*continued*)

Category	Subcategories
Trust or Distrust of Others	• Government • Law Enforcement • People You're Obliged to Trust • Personal Relationships • Private-Sector Companies

Breaches and Violations

Two important issues that emerged from this category were participants being unable to determine exactly how a breach was executed and dismissing the idea of themselves being targeted by an attacker. These findings suggest a need for greater transparency on all activity involving an individual's assets and a need to convey the dangers of assuming that one is immune to attacks or that the effects of a breach can be managed and resolved easily.

- **Digital Breaches**

    - *Identity theft.* When discussing digital technologies, multiple participants expressed concern about identity theft. For example, P23 supposed a digital identity app could enable someone to obtain a credit card on her behalf: "*They could just be like, 'Hey, I want a credit card,' and have information [from the app] on their phone to prove that. And there'd be no way to even know, I think, because how would you know that someone got a credit card in your name?*"

      Some participants, such as P20, felt the potential consequences of identity theft would be severe: "*If someone breaches that [an online database supporting a digital identity app], . . . that's your whole life and then you're kinda screwed.*" Other participants had a more optimistic outlook, like P1: "*Let's say you get hacked and you get your bank account stolen. It's gonna get solved. It'll be an inconvenience in the beginning. . . . But in the end you'll get your money back.*"

- *Unauthorized access of personal data.* Participants were concerned about a broad range of personal information that an attacker could access through their devices including email messages, messages in messaging apps, photos, banking information, browser history, and files containing personal information, such as tax return forms. Some participants, like P10, kept photographs of "*personal data, like [his] passport, [and his] SIN number*" on their devices but were concerned about attackers potentially accessing this information.
- *Unauthorized modifications.* Participants expressed concern over various aspects of their digital lives being tampered with by an attacker. For instance, P28 speculated that "*[if] someone had an enhanced driver's license and they had a criminal record, someone could hack their driver's license so they had no criminal record and could travel to the States.*"
- *Unauthorized use.* We identified some recurring concerns regarding how an attacker might misuse participants' digital information. These concerns included purchases being made with cards stored in a mobile wallet or with website and apps that remember payment information (e.g., Amazon, PayPal), undesirable posts being made on social media, undesirable messages being sent, and the browser's password autofill feature being used to access various accounts. P6 had a unique concern and feared undesirable searches being made by an attacker: "*[It would worry me if an attacker] went to Google and looked up how to make a bomb . . . And it shows that, and the FBI catches it.*"

- **Perception of Risk** Our participants had varied outlooks on the levels of risk associated with their behaviours. For instance, P1 "*didn't view [himself] as that much of a target*" because he was "*not a millionaire.*" P14 shared this opinion explaining that celebrities and politicians are likelier targets because of their societal influence. Others, like P3, were more leery: "*It's sensitive information [the information stored in a wallet app] and it's never one hundred percent safe. . . . And there's always the potential that it could . . . fall into the wrong hands.*"

- **Personal Experiences** Multiple participants discussed their own experiences with breaches. These anecdotes included loss of access to email accounts, money being

withdrawn from bank accounts, and undesired postings or use of a social media account. Within some of these anecdotes, we identified a trend of participants being unable to identify the origin of the breach. For instance, P19 discovered that their Facebook account had been used by someone unauthorized but he "*couldn't trace it back to the actual person that did it.*" Similarly, P17 theorized malware led to his bank account being breached but ultimately, he was "*not a hundred percent sure on how that [the breach] occurred.*"

- **Physical Breaches** Some of the physical environments and possessions that participants feared being breached included their mobile phones being taken, items from their homes being taken, and letters or packages being taken from their mailboxes. Participants sometimes indicated they were doubtful that such attacks would be successful, such as P8's passport being "*kind of out of reach*" for a burglar and P10's colleagues being unlikely perpetrators because they each "*[belonged] to a well-behaved family.*"

Perception of Digital Technologies Replacing Analog Technologies

In discussing mobile wallet technologies and digital identity services, we learned about how participants perceive these technologies. Our findings suggest that the security- and privacy-related aspects of these services are cause for worry for many participants.

- **Factors Influencing Adoption** We identified six key factors that affected participants' willingness to adopt either a mobile wallet technology or a smartphone-based digital identity service.

    - *Availability.* Some participants, such as P01, explained that if mobile wallet services were supported everywhere then "*[he] would just use Apple Pay*" instead of traditional forms of payment like cash, credit cards, and debit cards. Others reported being unable to use certain technologies because their phone was incompatible.
    - *Government knows all.* Multiple participants seemed unconcerned with providing their personal information to a government-operated digital identity service because, in the words of P24, "*they [the government] already have it*

all anyways." Others viewed the transition towards digital by governments as a slippery slope towards constant surveillance.

- *Influence of individuals and society.* Several participants felt more comfortable using a popular technology rather than an emerging one. For instance, P22 was more likely to use a technology if she had "*seen people using it and not heard [about] any problems with [it] or any leak [sic] of information.*" Participants also explained they would look to media reports and the opinions of people they know when considering a technology.
- *Convenience.* Multiple participants viewed these technologies as convenient and spoke of the perceived benefits (e.g., fewer items to carry, no more forgetting documents since they are stored on the phone). However, some comments suggested convenience may not always be desirable. For example, P11's bank supported sending money via iMessage but she felt this was too easy and insecure.
- *Appeal.* These technologies appealed to participants for a variety of reasons; some were interested in specific use cases (e.g., travel documents, rewards programs) while others were interested because they were generally early adopters of technology. Others, such as P25, showed disinterest in these technologies: "*I grew up in the generation where we had cards for things and that's what I'm used to so that's what I stick to.*"
- *Security and privacy.* Participants' comments on the privacy- and security-related aspects of these technologies were broad, covering topics such as their avoidance of these technologies for fear of the security and privacy risks, how to prevent attacks, and how to respond to attacks. Several participants, such as P25, expressed concern for the exploitation of wireless technologies: "*I read in the news as well that some people are able to just scan somebody's phone and take all their information.*"

- **Preferences** Multiple participants preferred conducting transactions with traditional payment methods (e.g., cash, debit cards, credit cards) and with analog versions of their identity documents. Participants cited a wide range of reasons for

not using the digital equivalent, including a general preference for analog formats, a desire to keep personal information offline, aversion towards online technologies, and an inability to understand how the technology would benefit them.

Among those who were interested in newer technologies, we found cases where participants were only amenable to using these technologies under certain circumstances. For instance, P10 gave an example of when he will make an exception and pay with a smartphone-based technology: "*Usually, I pay by credit or debit card but Starbucks are giving some additional schemes and discounts when we pay from the application.*"

- **Technical Knowledge** We found great variation among our participants in their level of understanding of these technologies. While some participants demonstrated accurate **mental models** for these technologies and awareness of the associated risks, others spoke with a lack of conviction. For instance, P24 spoke about the possibility of an attacker intercepting transmissions from her smartphone when using a mobile wallet but quickly downplayed her statements saying, "*I'm not exactly sure how it works to be honest.*"

  P25 also had security concerns regarding these technologies but was unsure if they were valid given she did not have personal experiences with breaches: "*I don't know how realistic it is 'cause I've never had it actually happen to me, personally, where someone just scanned my phone or took information, broke into any of my social media accounts. I've never had it happen to me, personally. But I've just heard so much about it [through reading and in the media] and I know it's such a big market. I know the probability seems high so it causes anxiety.*"

Precautions and Safety Measures

Participants engaged in a variety of behaviours that may have aided in protecting them across various aspects of their lives. We recognized these behaviours as falling into one of three subcategories: protection of their devices, protection of their digital information, and protection of their physical environment and possessions.

- **Devices** Participants made use of passwords, passcodes, biometrics, and pattern

locks to protect their devices. With respect to knowledge-based authentication schemes, many participants believed their secret would be difficult to guess, however, we did not substantiate these claims.

With phones in particular, some participants spoke of extra measures they take to keep their devices safe, such as keeping the device within reach. Additionally, P13 used an older smartphone and explained he "*kinda sorta purposefully [keeps that] phone . . . [to deter] people from wanting to steal [his] possessions.*"

- **Digital** We identified two major patterns in current behaviours that may help participants to protect their digital information. The first was being required to authenticate prior to performing certain activities in apps or when accessing particular apps or files. P22 applied this strategy more aggressively than other participants by being signed out of most apps and "*[having] to sign in pretty much every time for most of the things I care about—at least, my emails and anything that has more personalized information.*" The second was deliberately keeping certain types of information off of their device. For example, both P13 and P20 said they do not keep financial information, such as credit card information, on their phone.

- **Physical Environments and Possessions** We discovered a wide range of behaviours that may aid participants in protecting their physical environments and possessions. For mailboxes, we found that many participants either had mail slots that resulted in their mail being delivered into their residence or mailboxes that required keys.

  We also found that participants often made a concentrated effort to keep their personal belongings, such as keys, identity documents, or wallets, close to them or stored them in locations with restricted access, such as locked storage. Participants also used more obscure locations, such as suitcases and drawers for clothing, in home environments.

  For some participants, safeguarding their identity documents was a collaborative effort with family members or other members of their household. For example, P8 coordinated with her partner: "*My partner got a fire-safe box and his routine is to put all his documents there, so I just threw mine in there as well.*"

Finally, some participants described the factors influencing their decision to whether a document should be carried on their person or stored away. These factors included whether they anticipated needing the document and how challenging it would be to replace it. In P10's case, for example, one reason he did not carry his birth certificate was because replacing it would entail "*[going] back to [his] country*" while P1 pointed to the complex application process for passports as a reason for keeping his at home.

Risky Behaviours and Vulnerabilities

Similar to the *Precautions and Safety Measures* category, participants' behaviours clustered around: their devices, their digital information, their physical environments and possessions, and an additional category involving their self-awareness of the potential ramifications of their behaviour.

- **Devices** Most behaviours that we identified as potential hindrances to the protection of participants' devices were related to authentication. For instance, several participants explained that they were deliberately not using any authentication schemes on their device. P10 made this choice because she viewed entering a password as a "*wastage of time*," P20 had "*nothing to hide,*" while P28 was "*on a lot of medication and [she] sometimes [mistyped] things.*" Other participants had weak passwords, such as single-character passwords and passwords based on personal information. Some participants also mentioned that others knew how to unlock their device.

- **Digital** Multiple participants mentioned their use of their browser's autofill functionality as an aspect of their digital life that may help an attacker fulfill their goals. Participants said this functionality would enable access to various accounts, including accounts with financial institutions, email providers, and credit card information. Participants also indicated that they sometimes remain signed in to certain apps or websites, such as Facebook, Amazon, and Instagram, a practice which could allow an attacker to begin using the account without authenticating.

- **Physical Environments and Possessions** Multiple participants reported leaving

various items (e.g., keys, wallets, purses) unattended in a variety of contexts (e.g., amongst colleagues, in vehicles, in coffee shops while using the washroom). Some participants defended these actions by mentioning their intention to return quickly or their trust of the people in the environment. P6's tendencies were distinct from other participants as he would typically "*leave [his] wallet in [his] car*" and would only "*go and get [the] wallet from the car*" when he needed money.

Despite recommendations from the government against this practice [112], multiple participants reported carrying their SIN card on their person or in their wallet. P28 justified this with her living situation which prevented her from storing these documents safely while P25 explained she "*[didn't] have it [her SIN number] memorized.*"

- **Self-Awareness** Some participants demonstrated awareness of the risky behaviours they engaged in. P6, for example, described his practice of keeping his wallet in his car as "*not very wise*" while P12 said she "*should put a lock on [her] phone*" after reflecting on the data an attacker could access. It is worth noting, however, that in some cases, participants did not feel that a change in their behaviour was warranted, such as P10 who doubted his colleagues posed a threat when he left his belongings unattended in a laboratory.

Trust or Distrust of Others

On the topic of trust relationships, we found there was apparent distrust when discussing organizations, whereas individuals were often given the benefit of the doubt. This way of thinking could prove to be problematic if users relax their security practices around those with whom they have established a rapport.

- **Government** In speaking about digital identity services, some participants, such as P12, felt that if the government operated the service, she could "*trust them a bit more with . . . personal information.*" P12 continued by saying, "*The government's goal is not turning a profit on its citizens. . . . Like, they're not about to sell your information. I think a breach of your security is a huge concern for them as well. They would be really concerned about someone stealing your Social Insurance Number*

just like you [an individual] would be." Others were more conscious of the type of power such a service could afford the government. P6 said, "*When you are dealing with the state, . . . the balance of power is not beneficial to you, right? They have more power. So if I show them I have a phone, they might just confiscate it, right?*"

- **Law Enforcement** P6 expressed concern for police confiscating his device: "*If I'm getting pulled over on the road [and I can text my information to the officer], I'd be happy to do that but I don't want them to be around my phone and they can just grab it.*"

- **People You're Obliged to Trust** Participants made mention of a variety of people they are obligated to trust who might have privileged access to their belongings such as mail carriers, landlords, and doorkeepers in residential settings, and custodians in professional settings.

- **Personal Relationships** In terms of physical belongings, many participants relegated responsibility or trusted others, especially family members, with access to their mail and the safekeeping of their identity documents. Trust was also extended to colleagues in the workplace or at school who were often assumed not to have ulterior motives.

  Some participants intentionally shared device passwords or passcodes with others or added other people's biometrics to the device to facilitate their use of the device. As before, participants did not suspect that these individuals would act maliciously.

- **Private-Sector Companies** We identified some conflicting opinions when it came to the trustworthiness of private-sector companies. For instance, P6, who earlier discussed the imbalanced power relationships between governments and individuals, described interactions between businesses and customers as an "*equal power balance.*" P12, however, wondered whether companies' prioritization of profitability could be harmful for users' security and privacy. In speaking of mobile wallet services, P10 asked, "*Why should I use a third party [such as Apple or Google] to access my credit and debit cards?*" citing concern for "*the privacy of [his] banking details.*"

3.4 Discussion

3.4.1 PII Discoverability

Through this study's tasks, we learned that certain categories of personally identifiable information—specifically, *Personal Information*, *Address Information*, and *Occupational Information*—have relatively high discoverability on both social media and search engines.

We also found that more categories of information were discoverable on social media than search engines—in particular, *Birth Information* and *Contact Information*. The discoverability of these categories of PII may be a function of certain social media services offering additional functionality when this information is provided (e.g., allowing other users to find you by your phone number, reminding other users about your birthday).

The implications of these findings on the types of PII best suited for identity proofing may differ according to which threat models are being considered. For instance, if the attacker is assumed to be a stranger without access to their target's social media, categories such as *Birth Information* and *Contact Information* may be suitable proofs. However, if the attacker can access their target's social media, alternative proofs should be considered.

The reality is that a non-trivial amount of identity theft is committed by non-strangers [102] and as such, we believe that the PII that can be depended on for identity proofing is quite limited. Accordingly, we recommend that identity proofing mechanisms incorporate the following:

(1) *Use dynamic KBA instead of static KBA.* Instead of relying on static KBA, such as the user's birth year, we believe that identity proofing should rely on dynamic KBA. With dynamic KBA, issuing authorities can use a variety of sources to generate identity proofing challenges that are based on the individual's recent activity (e.g., line 200 from your tax return, the amount of your last withdrawal) rather than depending on identifiers that never change or change infrequently. Theoretically, this element of dynamism could make it more difficult for an attacker to anticipate what PII will be requested and more challenging to locate

the correct information.

Nonetheless, this approach is not foolproof as the sources where the PII is originating from can be prone to data breaches. It also has usability drawbacks because it requires the user to find the requested information, which may not be readily on-hand.

(2) *Use multiple types of proof.* Given the weakness associated with relying solely on knowledge factors, we suggest relying on multiple factors. Examples could include the user photographing themselves, providing their documents, accessing sensors in their smartphone, or performing phone number verification. By relying on multiple factors, greater assurances can be made that an imposter is not attempting to obtain an identity.

3.4.2 Impact of User Behaviour on Information Discoverability

Through our qualitative analysis, we found that personal security and privacy came as an afterthought for some participants, thus putting them at greater risk of becoming victims of identity theft.

As an example, we found instances of participants deliberately choosing convenience over their own security and privacy, such as by not having a password on their device to enable quicker access or by carrying a SIN card instead of memorizing their number. We also found cases of participants questioning why an attacker would target them, believing that more valuable information and assets could be obtained from high-profile individuals like politicians or celebrities.

These outlooks are concerning because they suggest that users may not have accurate assessments of the realities of identity theft, such as identity theft often being a crime of opportunity rather than a meticulously crafted scheme to attack the elite [102]. As such, we recommend addressing this issue by educating the general public on how identity theft can be carried out, providing them with strategies for preventing identity theft, and advising them on what to do as a victim.

However, relying solely on education results in new responsibilities—such as staying informed on the latest threats and prevention strategies, parsing through

public discourse on identity theft, and adapting their existing knowledge as new technologies emerge—being thrusted upon users which they may not be able to handle appropriately.

For this reason, we believe that systems designers (i.e., those responsible for the systems being targeted by attackers) must be held accountable. These designers play an important role in influencing users' security and privacy decisions through how they architect systems and this influence can be leveraged to support users in making security- and privacy-preserving decisions. For instance, the Government of Canada no longer issues wallet-sized SIN cards but instead provides Social Insurance Numbers on letter-sized paper to discourage individuals from using their SIN as an identity document and to reduce opportunities for theft [116]. Similarly, by integrating biometric authentication into their devices, smartphone manufacturers relieve users of the burden of memorizing their passwords [21]. Going forward, we recommend that designers take an active role and consider how their works could harm users' security and privacy.

3.5 Limitations

Participants had limited time to complete the social media and search exercises preventing us from conducting an exhaustive, comprehensive review of their social media accounts and search results. It is possible that with additional time, more identity attributes would have been discovered. Future research could allocate more time towards a more extensive investigation of participants' accounts and search results.

Our sample population skewed young and consisted of mostly students who were active social media users. We recognize that social media use varies by demographic group and that certain services may be more popular with specific groups [115]. We acknowledge this may have affected the results we observed during the interviews, the social media tasks, and search tasks.

We also conducted some of our recruitment on Facebook which may have affected the distribution of social media services that we observed from our participants. However, we believe that the other recruitment methods that we used helped to

mitigate against this limitation (e.g., email, posters).

Finally, in the weeks prior to our first sessions, information about Facebook's Cambridge Analytica scandal was publicized [25] and in response, Facebook displayed a banner in the News Feed that linked to a page for reviewing apps and websites connected to a user's account [127]. These events may have led participants to be more concerned about their privacy than they otherwise would have.

3.6 Conclusion

After conducting a study evaluating how users' behaviours affect their personal security and privacy, and examining the availability of their personally identifiable information with social media services and search engines, we recommend that:

(1) identity proofing mechanisms incorporate dynamic KBA and multi-factor authentication to detect imposters, and;

(2) to reduce the proliferation of PII that is publicly available, resources be allocated to help the general public develop accurate mental models for identity theft and to encourage designers to be conscious of their influence on users' security and privacy.

Chapter 4

Usability Study

To learn more about users' perceptions and expectations of digital identity services, we conducted a usability study where participants evaluated a prototype digital identity service. Through this study, we collected data on participants' experiences with the prototype as well as their attitudes towards digital identity services in general. In this chapter, we present the findings of this study in order to address the following research questions:

RQ2.1 How do users perceive digital identity services?

RQ2.2 What are users' expectations for digital identity services?

4.1 eID-Me

eID-Me [20] is a prototype digital identity app developed by Bluink that allows Ontario residents to use their smartphone to authenticate both in person and online. First, users prove who they are via an identity proofing process that involves submitting PII to the app. Once this information has been verified, the user's smartphone is issued a digital identity which can be used to authenticate in person with participating organizations and online with participating websites. Appendix B.4 provides an architectural overview of eID-Me.

The digital identity contains the user's **identity claims** and is stored securely on their phone as a digital certificate. As part of preparing the certificate during registration, the user's device generates a key pair. The device sends the public key to the eID-Me Registration Authority (RA)—a trusted entity that issues the certificates—and uses the private key, which never leaves the device, to sign all identity transactions. Before issuing the certificate to the user's device, the RA embeds salted and hashed versions of the user's identity claims, embeds the public key from their

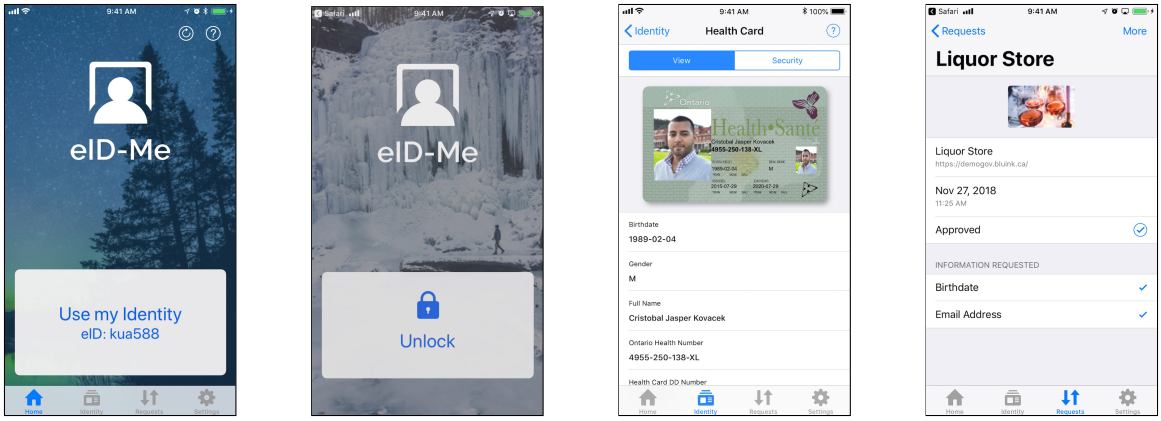


Figure 4.1: **eID-Me overview (from left to right)** – (a) On the Home tab, the app displays the Use my Identity button for in-person authentication and the user's unique identifier for online authentication (eID). (b) eID-Me can be configured to require authentication each time the app is launched. (c) Interface for reviewing identity information linked to digital identity. (d) Interface for reviewing recent transactions.

device, and, finally, signs the certificate. These steps provide security assurances to both the user and any relying parties:

- Salting and hashing the identity claims ensures that personal information is not stored in plain text and makes brute force attacks more difficult especially for claims that have smaller guessing spaces (e.g., a user's date of birth).
- Since all identity transactions are signed with the user's private key, relying parties can verify the signature upon receiving the identity claims and be assured that the information originated from the user's device.
- By verifying the RA's signature on the certificate, relying parties can be assured of the integrity of the information contained in the certificate.

The app also has measures aimed at protecting users:

- upon launch, the app requires the user to authenticate if their device has biometric authentication enabled or a passcode (see Figure 4.1b);
- it stores the user's digital identity on their device instead of in remote storage, enabling users to maintain control over their data;

- it requires user's explicit approval before any PII is shared with relying parties (see Figure 4.3b);
- it allows users to omit sharing identity attributes the relying party has specified as optional (see Figure 4.3b); and
- it displays recent transactions to make it easier for users to notice any suspicious activity (see Figure 4.1d).

4.1.1 Registration

After receiving a registration code via email, users complete an identity proofing process by providing an assortment of PII to the app including location information from their smartphone, a selfie, and photographs of their government-issued identity documents. Figure 4.2 gives an overview of the registration process while Appendix B.5.1 provides more technical details. The app then sends this PII to the eID-Me Registration Service to calculate a strength of identity proofing score (SIP) that reflects the estimated authenticity of the PII.

Factors that influence the SIP score include how closely the PII matches with government databases and third-party **identity verification services**, whether the selfie appears to have been taken by a live person, whether the user appears to have completed registration at home, and whether there is evidence of the submitted images being spoofed.

If the user's SIP score is high enough, the Registration Service issues a digital identity to the user. If not, users can try to improve their score by submitting additional documents to the app or can complete the process in person by visiting a service centre.

4.1.2 In-Person Authentication

Figure 4.3 shows the process for in-person authentication with eID-Me while technical details are provided in Appendix B.5.2

To authenticate in person, users open the app and tap the *Use My Identity* button. The app then displays a **QR code** containing a **nonce** for the user to present to the

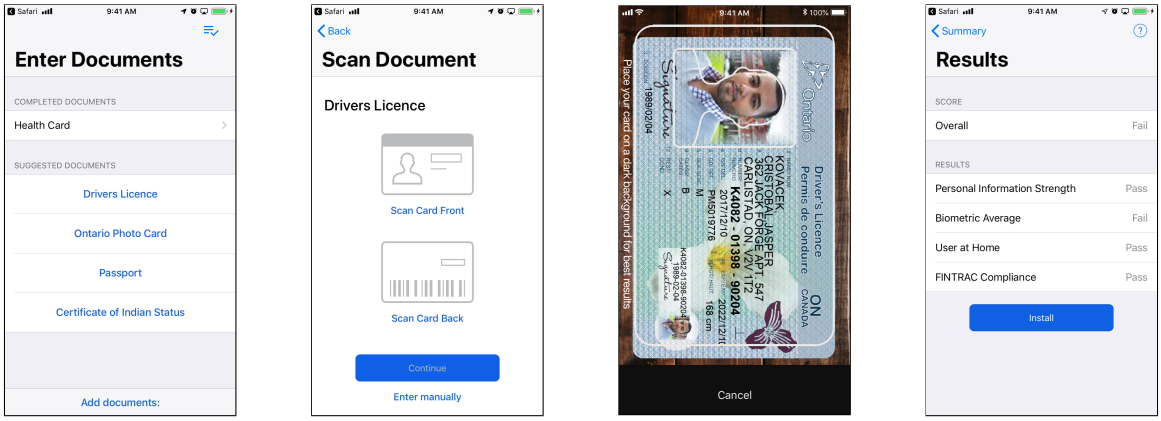


Figure 4.2: **Screenshots of eID-Me's registration process (from left to right)** – (a) To prove their identity, users can provide several different types of identity documents to the app. (b) Users can scan their identity documents with their smartphone's camera or can enter the information manually. (c) When scanning their identity documents, users centre their documents within the outline displayed on screen. (d) Upon submitting their identity documents, users are informed of their strength of identity proofing (SIP) score.

relying party's scanner. Once scanned, the Bluetooth device attached to the relying party's **point of sale** (**POS**) terminal includes the nonce in its advertising data so that the app knows to connect with this device. Once the app and the Bluetooth device are connected, the POS terminal can send its request for identity information securely to the user's device over Bluetooth with no need for an Internet connection. The request appears on the user's device and can either be approved or declined. The entire process typically takes approximately ten seconds.

4.1.3 Online Authentication

eID-Me implements several federated identity protocols, such as OpenID Connect [108], allowing users to authenticate with supported websites using their digital identity instead of a password. Figure 4.4 shows the process for online authentication with eID-Me while Appendix B.5.3 provides more technical details.

To authenticate online, the user visits a supported website and clicks on the federated login button for authenticating with eID-Me. The user's browser is then redirected to a form asking for their eID, a unique identifier assigned during registration. Upon submitting the form, the web page displays a four-digit confirmation code and

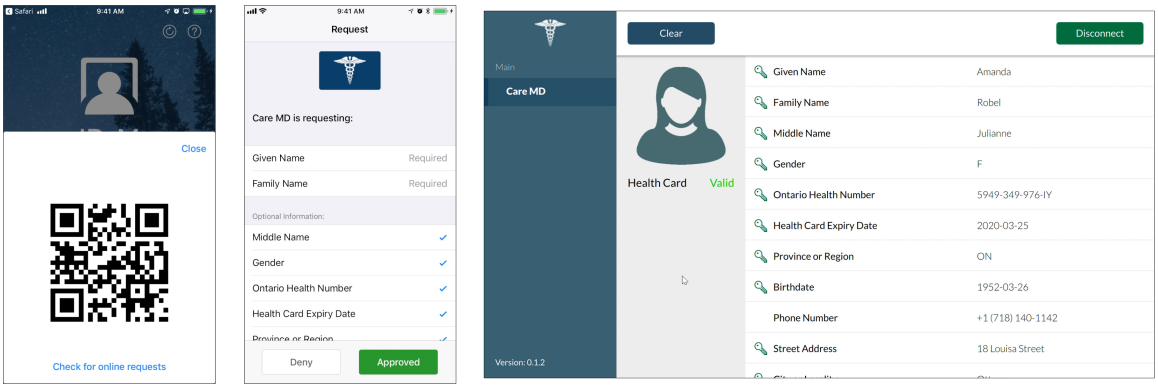


Figure 4.3: **Screenshots of eID-Me's in-person authentication process (from left to right)** – (a) To begin an in-person transaction, users present the QR code to the relying party's scanner. (b) The app displays the relying party's request and buttons for approving or denying the request. (c) If approved, the relying party receives the identity information.

instructs the user to open eID-Me on their phone.

Similar to the in-person authentication process, the app shows what information the relying party is requesting but also requires the user to select the four-digit confirmation code that matches what their browser is displaying. Upon approving the request, the relying party receives the requested information.

4.2 Methodology

In this section, we describe the study's methodology including our recruitment methods, the equipment we used, and the procedures we followed. The study received ethics clearance from Carleton University's Research Ethics Board-B (CUREB-B) on August 31, 2018 (Project #109170).

4.2.1 Recruitment

To find participants, recruitment posters (see Appendix B.1) were put up across the Carleton University campus and were circulated through the Ottawa Public Library's poster distribution program [95]. Information was also posted in various online communities, such as a Facebook page for announcing HCI user studies at Carleton, university mailing lists, and online bulletin boards for neighbourhoods close

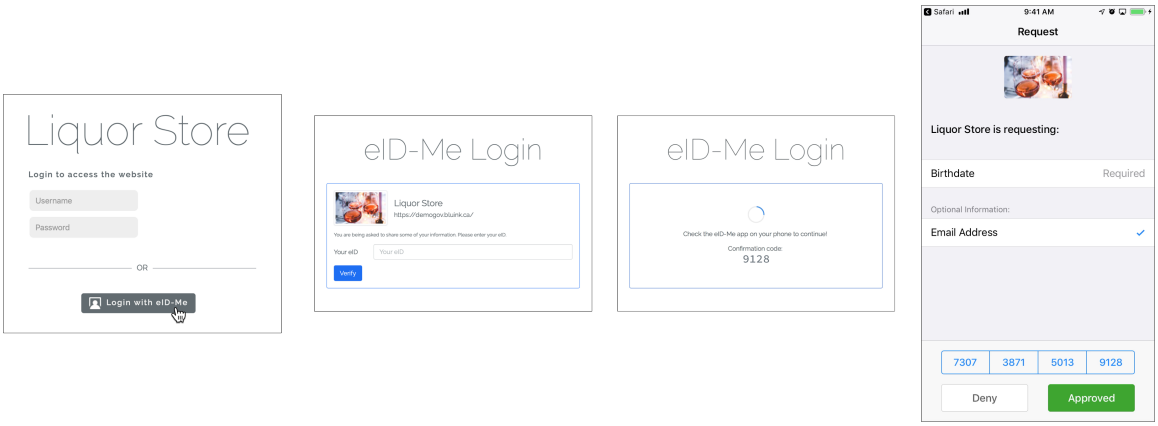


Figure 4.4: **Screenshots of eID-Me's online authentication process (from left to right)** – (a) The user visits the relying party's login page and clicks the button for federated login via eID-Me. (b) The user enters their eID and submits the form for federated login. (c) The web page updates with a confirmation code and instructions to open the eID-Me app. (d) The app displays the relying party's request, buttons for selecting the confirmation code, and buttons for approving or denying the request.

to campus.

When prospective participants contacted us to express their interest in the study, we sent them the study's consent form (see Appendix B.3.1) for informational purposes. This ensured that participants had advance knowledge of the possible risks of the study and had ample time to review the form.

Participants were also asked to complete an online screening questionnaire (see Appendix B.2) hosted on Qualtrics [101], a survey management platform. To minimize sampling bias, we used the results of the screening questionnaire to invite a broad range of respondents to participate in the study.

The screening questionnaire also asked participants whether or not they consented to an employee from Bluink attending their session. This question was only used to help us coordinate schedules with Bluink employees and did not influence who was invited to participate.

In total, 22 individuals (11 female, 10 male, 1 unspecified) participated in the study. The participants ranged in age from 19 years to 59 years ($\bar{x} = 32$ years, $SD = 12$ years). All participants but one ($n = 21$) had completed post-secondary education or were in the process of earning accreditation from a post-secondary institution.

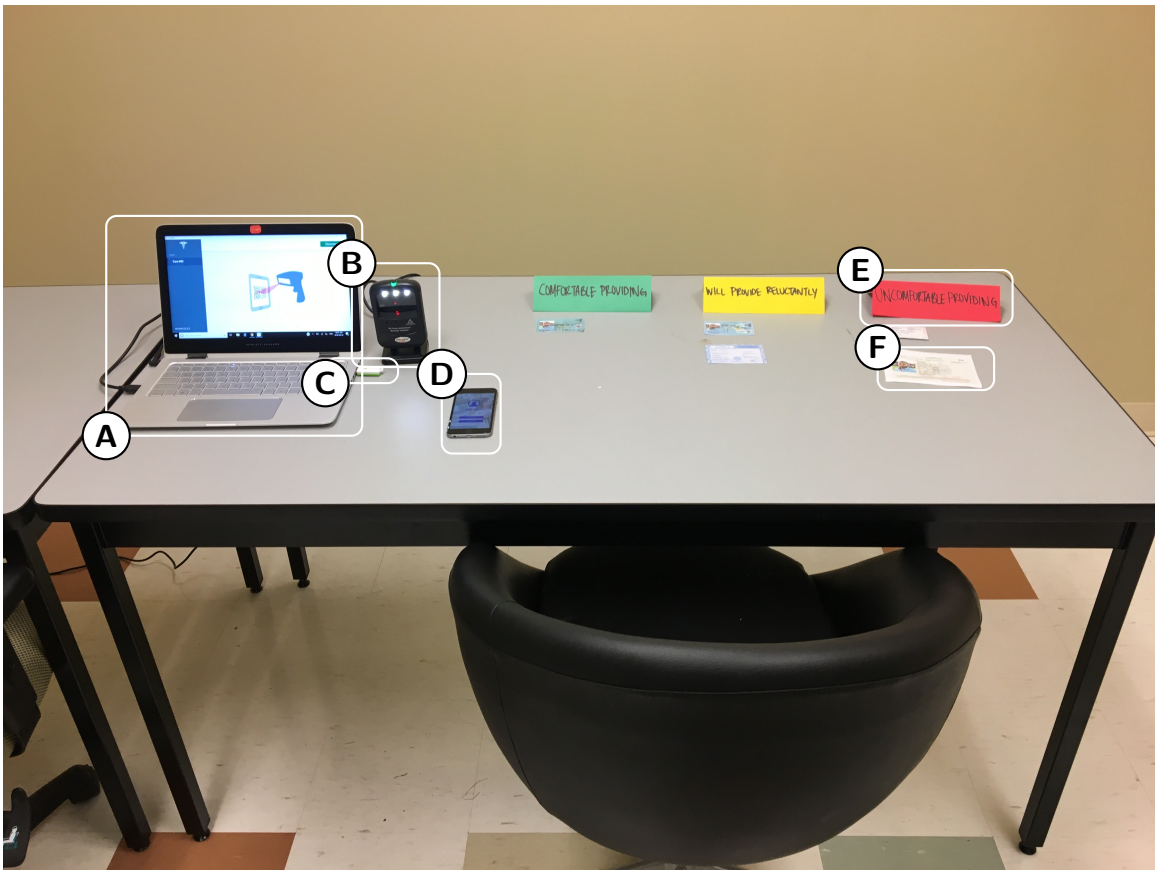


Figure 4.5: **Equipment overview** – (A) A Windows laptop, (B) a two-dimensional barcode scanner, (C) a USB Bluetooth device, (D) an iPhone 6, (E) one of three placards used for sorting identity documents, and (F) one of the mock identity documents provided to participants.

4.2.2 Equipment

Questionnaires

Participants used Google Chrome on the Windows laptop to complete both the pre-test and post-test questionnaires. Both questionnaires were hosted on Qualtrics.

Identity Documents

To complete the sorting exercise and the registration process, participants were provided with mock identity documents including a birth certificate, a Canadian passport, an Ontario driver's license, an Ontario health card, and a Social Insurance Number (SIN) card. The mock documents were roughly the same size as their real

counterparts to help make them convincing and some were laminated or double-sided. Images of the documents have been included in Appendix B.3.3.

For the sorting exercise, placards for each of the three categories (*comfortable providing*, *will provide reluctantly*, and *uncomfortable providing*) were set up in front of the participants.

Tasks Involving the eID-Me App

Participants were provided with an iPhone 6 running iOS 11.4.1, the latest version of iOS available when the study began. Before giving the smartphone to the participant, we installed a development version of eID-Me onto the device.

When participants were asked to prove their identity in the in-person scenario, the researcher operated a Windows laptop running proof-of-concept medical practice software developed by Bluink. The laptop was connected to a USB Bluetooth device for facilitating communications between the smartphone and medical practice software, and a two-dimensional barcode scanner for scanning the participant's QR code.

When participants were asked to prove their identity in the online scenario, they used Google Chrome on the Windows laptop to authenticate with a proof-of-concept website for purchasing alcohol.

4.2.3 Procedure

Sessions were conducted in one of two locations at Carleton University: the Human Computer Interaction (HCI) Building and Carleton's Human Oriented Research in Usable Security (CHORUS) Lab. These locations offered rooms that allowed participants to complete the study with minimal distraction.

Some of the sessions in the HCI Building were observed by members of Bluink's design team from behind a one-way mirror. Bluink employees did not interact directly with participants. Instead, their questions were shared with the researcher discreetly via a messaging service. The researcher would only relay the question to the participant if it was felt that doing so would not compromise the integrity of the data being collected.

An undergraduate research assistant with prior experience running user studies

in usable security helped the principal investigator conduct some of the sessions. Participants were each paid CAD$25 in cash and an additional CAD$10 for eligible parking expenses. Sessions were audio recorded if consent was provided. Each session was approximately ninety minutes in length and consisted of the following activities:

1. **Consent form.** Participants were asked to read and sign the consent form (see Appendix B.3.1). In addition to explaining the possible risks and benefits associated with participating, the form allowed participants to decline audio recording for their session, and to decline observation of their session by a Bluink employee, regardless of how they answered in the screening questionnaire.

2. **Instructions.** To provide participants with enough context, the researcher explained the study's goals, briefly described the concept of digital identity, and gave an overview of the eID-Me app to the participant using a pamphlet (see Appendix B.3.2).

3. **Pre-test questionnaire.** Participants completed a pre-test questionnaire (see Appendix B.3.4) hosted on Qualtrics using a computer that we provided.

4. **Tasks.**

   (a) **Retrieve registration code.** Participants launched the eID-Me app on the smartphone. After granting eID-Me the permissions that it requested, participants entered an email address (provided by the researcher) in order to be sent a registration code. Upon receipt of an email containing the registration code, participants could either tap a link in the message to forward the code to the eID-Me app or copy and paste the code into the eID-Me app.

   (b) **Sort identity documents.** The researcher placed mock identity documents in front of the participant and asked them to sort the documents into three categories: *comfortable providing*, *will provide reluctantly*, and *uncomfortable providing*. Participants were asked to act as if the documents belonged to them and displayed their personal information. They were

also asked to explain the reasoning behind the categories chosen for each document.

(c) **Take a selfie.** For privacy reasons, participants used the app's selfie capture interface to photograph a static image of the individual on the identity documents placed in front of them instead of photographing themselves.

(d) **Register using identity documents.** For each document categorized as *comfortable providing* or *will provide reluctantly*, participants input each document into the app by either using the app's scanning interface or by typing in the document's details manually. After inputting the documents, participants submitted the information for verification and were issued a digital identity.

(e) **Use eID-Me identity: in-person authentication.** Participants used eID-Me to verify their identity as if they were checking-in for an appointment at a hospital. The researcher used prototype medical practice software on the laptop to verify the user's identity.

(f) **Use eID-Me identity: online authentication.** Participants navigated to a proof-of-concept website for purchasing alcohol and were instructed to use their digital identity to log in.

5. **Delete eID-Me.** After transferring log files from the app to the researcher's computer, the app was deleted from the smartphone.

6. **Post-test questionnaire.** Participants completed a post-test questionnaire (see Appendix B.3.5) hosted on Qualtrics using a computer that we provided.

7. **Post-test interview.** In a semi-structured interview, the researcher asked participants questions (see Appendix B.3.6) relating to digital identity in general, their willingness to use certain documents as proofs of identity, and their experience using the eID-Me app.

4.3 Data Analysis Strategy

4.3.1 Quantitative Data

Participants were asked to complete two questionnaires (pre-test and post-test) and to indicate their comfort with providing certain identity documents by sorting each document into one of three categories. We analyzed this data with Numbers for Mac [7] and used summary statistics to describe participants' responses.

4.3.2 Qualitative Data

After conducting our sessions, we had obtained a total of close to twenty-seven hours of audio recordings from twenty participants.[1] These twenty recordings were transcribed by the principal investigator and an undergraduate research assistant, yielding 361 pages of lightly edited transcribed materials. Eight recordings were transcribed in full and twelve recordings were transcribed selectively (i.e., the transcriber listened to the full audio recording but only transcribed noteworthy moments). Additionally, for the remaining two sessions where audio recordings were not available, we included written notes made by the researcher during the session (11 pages). Hereafter, we refer to this collection of documents as our *dataset*.

Content Analysis

To analyse our dataset, the principal investigator and an undergraduate research assistant with a computer science background applied qualitative content analysis methodology.

In applying this methodology, we used the individual questions posed to our participants as the unit of analysis such that the question and the participant's answer comprised a single unit (hereafter referred to as a *data item*). In addition, we determined which documents each researcher would code by splitting the documents into two partially overlapping subsets. The overlap between these subsets represented 3 full transcripts, 4 selective transcripts, and 1 note, or about 36% of the documents.

[1] Our dataset consisted of only twenty recordings instead of twenty-two because one participant declined being recorded while another participant's recording was destroyed as a result of a technical failure.

Table 4.1: Categories used for qualitative data analysis

Category	**Description**
Attribution of Responsibility	Data items in which participants discuss who should be responsible for the operation of a digital identity service, the implications of this role, and who should be accountable for any negative effects resulting from the service being in operation.
Benefits	Data items in which participants express their views on the benefits of digital identity service.
Drawbacks and Concerns	Data items in which participants describe any perceived downsides of a digital identity service or any apprehensiveness they feel about using a digital identity service.
Mental Model Issues	Data items in which participants demonstrate a mismatch between their mental models and the designer's model.
Usability Concerns and User Requirements	Data items in which participants refer to elements of the app that made completing tasks difficult, inefficient, or unpleasant. This category also includes data items where participants describe their expectations of how the service should function.
Uses for Digital Identity	Data items where participants describe the contexts in which they would use a digital identity.

These quantities were proportional to the distribution of document types in the full dataset.

Next, we began our analysis by using a deductive approach and developing a categorization matrix for sorting the data items. As shown in Table 4.1, the matrix included six categories which were related to our research questions. Using this matrix, we reviewed the dataset and coded data items by assigning them to relevant categories. We assigned a data item to multiple categories if it was best described by more than one category.

Once coding was finished, we met to review our results and, following the same process used to analyze data from the information discoverability study, we resolved any disagreements by moving items between categories, clarifying a category's definition, or revising the matrix.

To provide more structure to our analysis, we then developed subcategories within

each of the six categories using an inductive approach. We did this by reading through each category's data items and, together, developed a list of thirty-three subcategories across the six main categories. Using these subcategories, we went through multiple iterations of coding the data items and resolving disagreements until we achieved sufficient inter-rater reliability. After four rounds completed over the course of about nine days, we reached 80% agreement, which met the recommendations set forth in relevant literature [83]. Any remaining differences in the researchers' coding decisions were combined such that the individual decisions made by each researcher were preserved.

As before, we avoided using quantitative measures to serve as indicators for the importance of any particular data items.

Task Completion

We reviewed each transcript to measure our participants' ability to complete the study's tasks. Since this measure required a level of detail only afforded to us by our transcripts, we excluded data from the two participants for whom transcripts were unavailable.

For each task, we considered a participant's attempt as a *success* (i.e., the participant received no assistance from the researcher), a *partial success* (i.e., the participant received some assistance from the researcher or was only able to complete some of the task), or a *failure* (i.e., the participant failed to complete any aspect of the task). We also distinguished between attempts where the participant encountered a *technical issue* outside of their control. Table 4.2 lists the criteria we used to classify our participants' attempts.

4.4 Results

4.4.1 Task Completion

Table 4.3 shows task completion data for the six tasks assigned to our participants.

Table 4.2: Task completion criteria

Task	**Success Criteria**
Registration Code	• Entered the provided email address, • Found a new email message from eID-Me with the registration code, and • Forwarded the registration code to the app by tapping the link in the email or copying and pasting the registration code.
Selfie	• Took a picture of the person on the identity document.
Add Documents	• Took pictures of the identity documents or manually typed in the information from the identity documents.
Set Up Digital ID	• Pressed "submit" to send the documents to the server and receive their score, and • Pressed "install" to set up their identity on the smartphone.
In-Person Authentication	• Located the QR code in the app, • Scanned the QR code, and • Approved the hospital's request in the app.
Online Authentication	• Found their eID (unique identifier), • Typed their eID into the website, and • Approved the website's request in the app.

Registration Code Seventy percent of participants completed this task without assistance. Issues with this task stemmed from matters beyond the participants' control (i.e., network connectivity, issues with email delivery).

Selfie Ninety percent of participants completed this task without assistance while the remaining ten percent required assistance with switching the app to use the rear camera instead of the front camera since we asked participants to photograph a provided picture instead of themselves. Since this step would not be required in the real world, it is possible that the rate of success would have otherwise been higher for this task.

Add Documents While 75% of participants successfully provided identity documents to the app, the remainder required assistance when they encountered issues when providing their documents (e.g., photographing the wrong document, the app's scanning interface failing to detect the document).

Set Up Digital Identity Completion data was omitted for this task because at the time of our study, eID-Me's scoring mechanisms were under development such that participants received a failing SIP score. We informed participants that this scoring mechanism was under development and encouraged them to complete the task in spite of the failing score they received.

In-Person Authentication Only 40% of participants completed the in-person authentication task without help. Those who received assistance struggled with finding how to begin a transaction or experienced technical difficulties when the app attempted to communicate with the POS terminal.

Online Authentication Virtually all participants (95%) successfully completed the online authentication task without assistance. It is worth noting that this task was preceded by the in-person authentication task which has a similar process for approving requests and requires participants to tap the "Use my Identity" button, which shows the unique identifier (eID) needed to authenticate online. As such, this

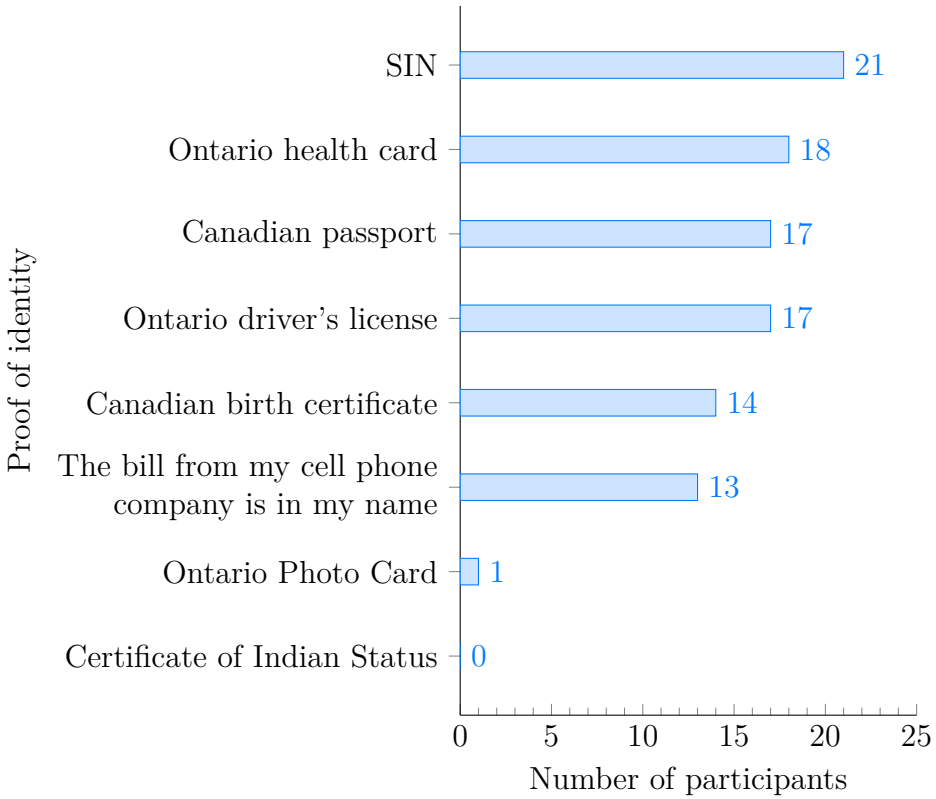


Figure 4.6: Proofs of identity owned by participants. $N = 22$ participants.

task's performance may have been inflated due to a priming effect, or may suggest that users only need to be walked through the process once before understanding it.

4.4.2 Document Ownership

In a pre-test questionnaire, participants were asked to disclose the proofs of identity that they owned. Figure 4.6 shows that the Social Insurance Number (SIN), the Ontario health card, Canadian passport, and Ontario driver's license were the most common. At the other extreme, none of our participants owned a Certificate of Indian Status and only one individual owned an Ontario Photo Card. These results suggest that it is practical for an identity proofing process to require applicants to submit multiple proofs of identity provided that a variety of proofs are accepted and that these proofs are obtainable by a wide range of demographic groups.

Table 4.3: Task completion per participant. S = Success, E = Assistance provided with recovering from an error made by the participant, T = Assistance provided with recovering from a technical issue outside of the participant's control, U = Assistance provided with locating the correct user interface control.

Participant ID	Registration Code	Selfie	Add Documents	Set Up Digital ID*	In-Person	Online
P1	S	U	S	-	T	S
P2	T	S	S	-	T	S
P3	S	S	S	-	T	S
P4	S	S	E	-	S	S
P5	S	S	S	-	S	S
P6	T	S	U	-	U	S
P7	T	S	S	-	S	U
P8	S	S	S	-	U	S
P9	S	S	S	-	U	S
P10	T	S	S	-	S	S
P11	S	U	E	-	U	S
P12	S	S	U	-	U	S
P13	T	S	S	-	S	S
P14	Data unavailable†					
P15	S	S	E	-	U	S
P16	S	S	S	-	S	S
P17	S	S	S	-	U	S
P18	Data unavailable†					
P19	S	S	S	-	S	S
P20	S	S	S	-	S	S
P21	S	S	S	-	U/T	S
P22	T	S	S	-	U	S
# of successes	14	18	15	-	8	19
# of partial successes	6	2	5	-	12	1

* Completion data was omitted for the "Set Up Digital ID" task because at the time of our study, eID-Me's scoring mechanisms were under development such that participants received a failing SIP score. We informed participants that this score was under development and encouraged them to complete the task in spite of the outcome.

† Transcripts for P14 and P18 were unavailable

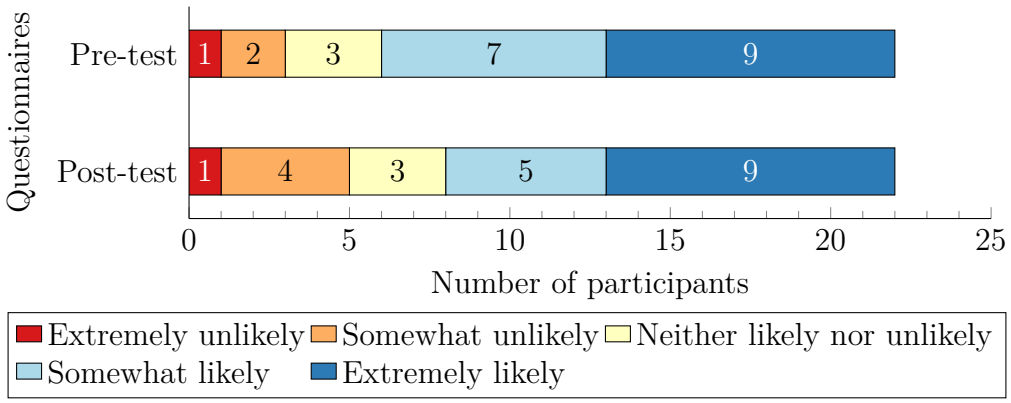


Figure 4.7: Participants' likelihood of using a government-approved app instead of traditional identification. $N = 22$ participants.

4.4.3 Likelihood of Use

In both the pre-test and post-test questionnaires, we asked participants how likely they would be to use a government-approved smartphone app as their proof of identity instead of traditional identity documents. Figure 4.7 shows that the results skewed positive for both questionnaires, although opinions became slightly more negative after participants used the app (pre-test: 72% extremely or somewhat likely, 14% neutral, 14% extremely or somewhat unlikely; post-test: 63% extremely or somewhat likely, 14% neutral, 23% extremely or somewhat unlikely).

Both questionnaires also asked participants how likely they would be to use digital identities in various scenarios (see Figure 4.8). With the exception of law enforcement, which performed noticeably worse in the post-test questionnaire, participants' opinions were fairly similar between the pre- and post-test questionnaires for each scenario. However, where there were differences, they mostly skewed towards more negative in the post-test.

4.4.4 Document Ranking

Figure 4.9 presents the results of the document sorting exercise and shows that participants had high levels of comfort providing certain types of documents or information (e.g., phone numbers, driver's license, photo card) and strong opposition to providing others (e.g., Social Insurance Number).

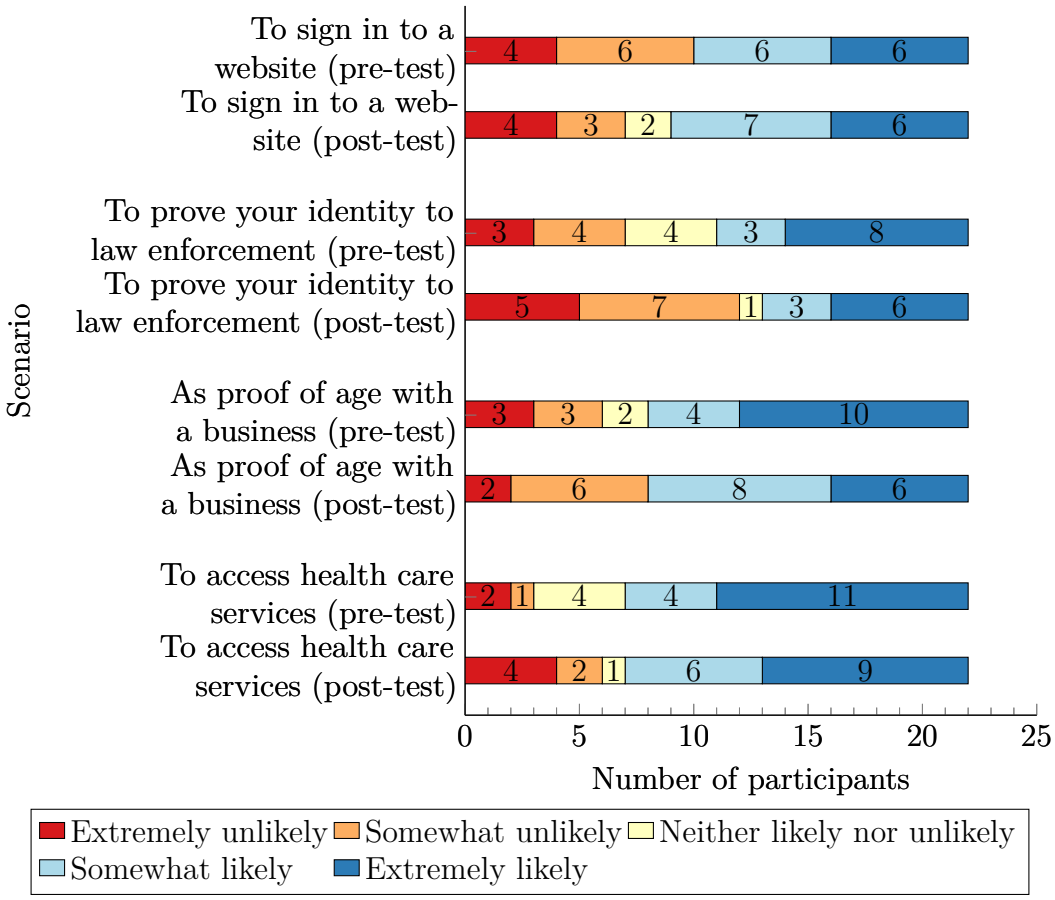


Figure 4.8: Participants' likelihood of using digital identity in various scenarios. $N = 22$ participants.

From speaking with participants about their rankings, we learned that participants' comfort with providing information generally depended on their views of how the information could be misused. For example, many participants were opposed to providing a Social Insurance Number because it can be used to commit fraud (e.g., redirect government benefits or tax refunds).

4.4.5 Qualitative Analysis

We present the results of our qualitative analysis by highlighting notable findings from across the six categories relating to our research questions. Table 4.4 provides an overview of the categories and subcategories we generated through our qualitative analysis.

Table 4.4: Categories and subcategories from qualitative data analysis

Categories	Subcategories
Attribution of Responsibility	Government
	For-profit companies
	Personal responsibility
Benefits	Speed and efficiency
	Redundancy
	Less to carry
	Security
	Resource efficiency
	Convenience (miscellaneous)
Drawbacks and Concerns	Hacking and identity theft
	Tracking and monitoring
	Selling of personal information
	Reliance on mobile phones and network connectivity
	Use with law enforcement
	Giving away too much personal information
Mental Model Issues	Access to information
	Information storage
	Information security
	Limited knowledge of particular subjects
	QR code scanning
Usability Concerns and User Requirements	Requests for additional security features
	Requests for additional functionality
	Requests for clearer communication
	Operational requirements
	Technical support requirements
	Technical difficulties encountered
	Misunderstandings
	Visual appeal
Uses for Digital Identity	Interactions with the government
	Health care
	Purchases
	Travel
	Interactions with law enforcement

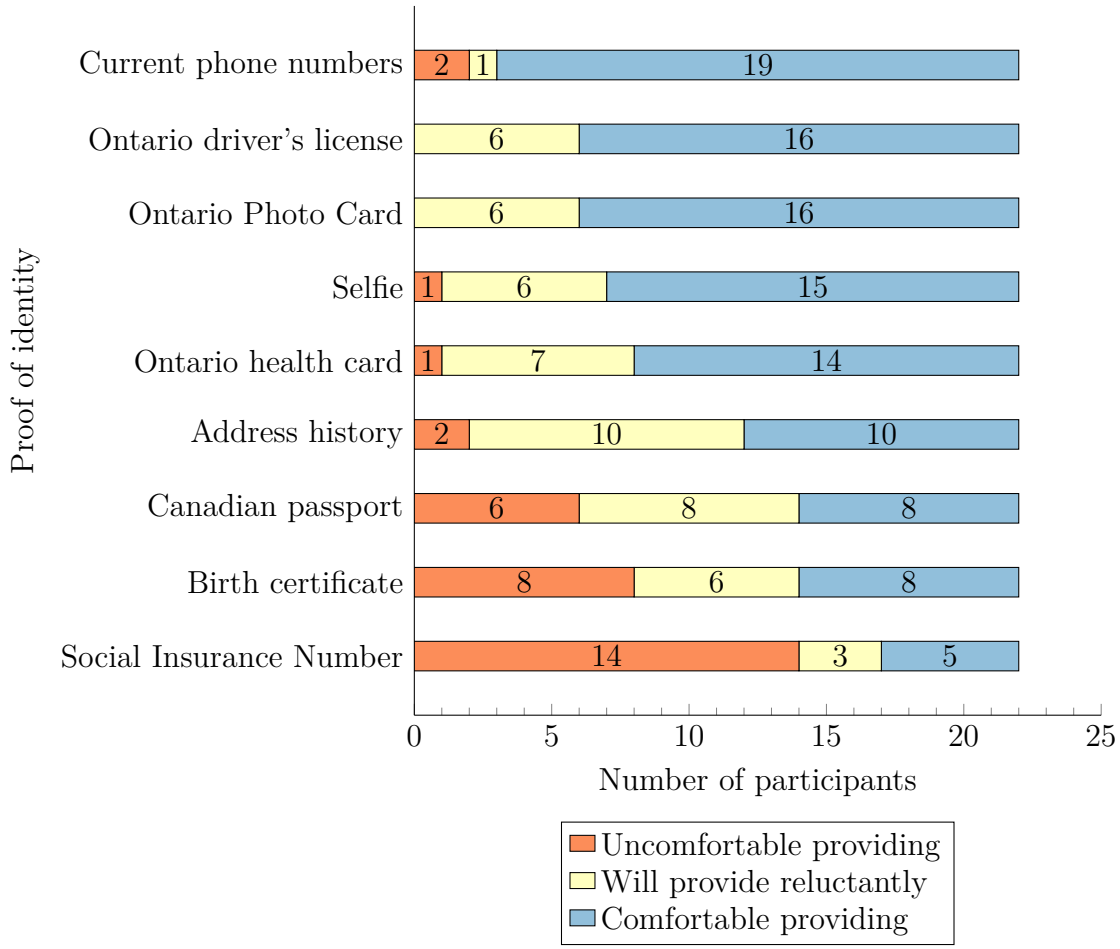


Figure 4.9: Participants' willingness to provide select types of documents and information to eID-Me. $N = 22$ participants.

Attribution of Responsibility

In discussing who they felt should be responsible for a digital identity service, most participants mentioned the government, given its role as an issuer of identity documents. Other participants felt corporations were well-positioned to take on this role but even fewer mentioned taking personal responsibility. Regardless of which entity would end up taking responsibility for the deployment of a real digital identity service, it is clear that users already have expectations for who should occupy this role and how this party should conduct themselves. As such, digital identity services should manage user expectations by clearly communicating each party's assumed liabilities.

Government. Like many other participants, P7 presumed the government's backing of a digital identity service would imply certain security-related activities and preventative measures: "*the government is screening everything and they're taking the responsibility.*" For P6, the government's endorsement would be persuasive in getting them to use the service since this participant "*already [trusted] the government when . . . [giving] them . . . documents*" to receive health care. Others were skeptical of government-backed digital identity systems, such as P22 who pointed to issues with Phoenix, a payroll system for federal employees: "*How much can we trust the government to do [technical projects] properly?*"

For-profit companies. Some participants were more amenable to using a digital identity system operated by a for-profit company. For instance, P4 felt that companies are incentivized to "*guarantee that everything is secured and that the customer is happy*" whereas with the government, "*you feel like they care a little less.*" P8 felt companies have clear financial motives whereas the government would be elusive with its motives: "*They [the government] pretend they don't get anything out of [data collection]. . . . They're just liars.*"

Personal responsibility. Multiple participants felt they should take on at least some level of responsibility for their use of a digital identity service. P6 suggested a need to be prudent when using a digital identity by saying that "*it's [the user's] responsibility to make sure someone doesn't steal [their] phone or make off with [their] information.*" Similarly, P22 believed that although businesses should face repercussions for data breaches, individuals "*need to be careful what [they] share*" since owning a digital identity could entail being liable for any misuse.

Benefits

The benefits identified by participants were mostly related to their information being in a digital format rather than a traditional printed format. Surprisingly, there was little mention of the app's privacy-preserving features possibly indicating a need to adjust how the service is framed.

Speed and efficiency. Several participants felt that a digital identity could be faster than traditional identity documents, such as P21 who said it would be quicker than "*searching [her] bag and bringing out the ID*" and P7 who said users could "*just scan once and go*" instead of filling out lengthy forms at hospitals, for example.

Others pointed to the reduction in manual work as a key benefit. P4 observed that first-time visitors to a website normally "*have to register [and] put in all their information again...*" while digital identities could be "*much faster*" and "*[save] a lot of time*" by automatically transferring the information to the relying party. Similarly, P15 liked that she had "*less typing*" to do and "*[fewer] passwords to remember*" with a digital identity.

Redundancy. Multiple participants thought of a digital identity as a backup of their physical identity documents. P3 explained how if he had forgotten his wallet, it would be advantageous to "*have [a driver's license] in [his] phone*" since it is common for people to "*carry [their] phone all the time.*" While P13 still preferred using his physical identity documents, he expressed interest in setting up the app and "*[using] it once in a while if [he] didn't have [his] wallet.*"

Less to carry. Several participants explained that digital identities would eliminate their concerns with needing to carry physical identity documents. P4 said that a digital identity "*saves space in your pockets.*" P15 explained that a digital identity could alleviate her "*[paranoia] as to whether [she] had a certain form of ID*" since mobile phones are "*the one thing [people] don't leave the house without.*"

Drawbacks and Concerns

Participants' concerns and views on the downsides of digital identity services were primarily rooted in how their data would be handled and the risks associated with providing personal information to the app. This finding signals the importance of addressing these concerns by having protections in place for users' personal information and communicating these measures to users.

Hacking and identity theft. Some participants feared identity theft, believing that determined attackers would be able to surmount any protections that are in place. P9 said, "*Somebody's going to become me whether I try to prevent it or not, or whether I hide under a rock, or never own a phone. . . .*" According to P16, eID-Me would be just as susceptible as any other system: "*One way or another, it [the information] can get stolen. My credit card information is another database that can get stolen and hacked. Anything can get hacked at the end of the day.*"

Tracking and monitoring. Participants were concerned about whether their use of the service and any personal information they provide would be tracked. For instance, P19 expressed weariness over private-sector companies operating a digital identity service and wondered if her use of the service could be used against her: "*. . . I guess media has conditioned me to think, 'Oh, what are they [private-sector companies] doing with this information? What are they gonna do with it later? Could this become Big Brother? Could this be used to discriminate for jobs?'*"

In another example, P16 asked if using his digital identity at a liquor store would mean that "*the government [would know] how often [he goes] to the liquor store.*" Finally, in response to the app requiring access to the user's location, P11 said, "*Nobody needs to know where I am unless I tell them where I am. . . . I don't need to be spied on.*"

Selling of personal information. For P20, a major concern was his personal information "*[being] put into the wrong hands, . . . [such as] marketers, advertisement agencies, or just someone that might have malicious intent.*" Similarly, P16 felt that companies were not incentivized to put users first: "*. . . companies are accountable to shareholders, board [sic] of directors, and other people who are more in it for profits than they are for providing a national service to the people.*"

Reliance on mobile phone and network connectivity. P11 felt that using a digital identity was too risky, saying, "*Lose your phone, lose your identity, lose everything.*" P22 explained that digital identities should not replace traditional documents "*because of the problems with devices where they break down or run out of*

power or lose connectivity."

P15 said the unreliability of his "*older model*" phone would encourage him to "*always make sure [he] has [his] paper documents . . . in case the app didn't work. . . .*" P3 wondered if the app had to be "*connected to Wi-Fi*" while providing identity information to a relying party: "*I would switch off data when I'm not using it [the app] 'cause I have a limited amount of data. . . .*"

Use with law enforcement. Multiple participants voiced concerns with using a digital identity with law enforcement, especially in the early phases. P3 explained that if an officer asked for her driver's license, she would be "*paranoid*": "*I don't know if this officer knows about this [the existence of digital identity]. I don't know if she's fine with me providing a digital copy.*" Meanwhile, P15 feared her use of a digital identity could be life-threatening:

> *[From the officer's perspective,] I could be pulling out a gun. I could be pulling out a variety of things. . . . Just the police in general make any person nervous, I feel, once they've pulled up to your vehicle. And doing anything outside of tradition in that area might make your very life feel like it's in jeopardy.*

Mental Model Issues

Many participants had inaccurate mental models for how eID-Me works. Some participants were not accustomed to the technology used by the service and thus lacked background knowledge that may have helped them complete tasks more easily. Others tried drawing upon prior experiences that seemed applicable on the surface but, ultimately, translated poorly and led these participants astray. While accounting for gaps in users' understanding may be challenging, identifying and focusing on the common issues may make eliminating these gaps more feasible.

Access to information. P4 believed that Apple could obtain access to a digital identity and its contents though iCloud Backup saying, "*as soon as the information's on the cloud, they [Apple] have access at all times.*" In discussing their avoidance

of OS updates, P15 said, "*Every single time that you do an update, more of your personal information just gets accessible to—not everybody but, like, law enforcement and such, which is good to a certain extent.*"

Information storage. At least four participants incorrectly believed that their identity information was embedded in the QR code when in reality, the QR code was random, contained no personally identifiable information, and was only used to connect the smartphone to the POS terminal. P19 thought that "*all the identifiers and information . . . [were] represented in that one unique barcode.*" Meanwhile, P4 speculated that relying parties would "*only [scan] the information that they need and not any additional information even though all the information is in your barcode.*"

Information security. Despite mechanisms existing for verifying users' identity claims, some participants believed that it would be possible to be issued an identity containing fraudulent information. P4 imagined a scenario where a fourteen-year-old could pretend to be nineteen because identity claims are "*not verified by anyone.*" P15 did not know what confirming the legitimacy of a digital identity entailed and wondered whether the absence of physical security features like holograms might make it "*more difficult for bartenders and such to determine if [the ID is] real.*"

QR code scanning. Multiple participants struggled with completing the in-person authentication task. P15 first tried scanning an image of their health card found within the app instead of the QR code. Similarly, P6 tried transferring knowledge from their use of traditional identity documents and erroneously scanned an image of a document: "*I found it was confusing . . . seeing the driver's license [in the app] and thinking I had to scan that because I'm used to a traditional card.*"

Possibly influenced by the in-person authentication task, P11 had difficulty understanding how information could be transmitted between the smartphone and the laptop during the online authentication task:

> *When you present this screen [a login page on a laptop computer], a separate unit to me, I'm like, "How am I going to get it [the QR code] from here to here?". . . I can't hold this [the smartphone] up and scan.*

Usability Concerns and User Requirements

Throughout the study, users mentioned aspects of the app that they felt had poor usability and also described what they expect a digital identity service to offer. These remarks help to identify pitfalls that designers can seek to avoid in future implementations of digital identity services.

Requests for additional security features. Some participants expressed interest in mechanisms for disabling their digital identity. P20 wanted to "*[be] able to lock [his digital identity] out completely*" upon losing his phone and be able to reactivate his identity on a replacement device.

Multiple participants felt that it was inadequate for the app to support authentication with the device passcode as opposed to the app having its own passcode. P15 explained her qualms with this setup: "*. . . my sister already knows my general passcode then if she wanted to use my ID to get into a club, she looks similar to me. . . .*" P4 explained that a separate passcode would enable him to use a more complex passcode and ultimately, lead to him "*[feeling] more secure.*"

P8 wanted some degree of ephemerality when sharing information with relying parties: "*Once they've printed off my hospital bracelet, I [should be able to] press on my phone 'Disconnect' so that they can no longer see my information. They have it already recorded, they don't need to stare at it anymore.*"

Requests for additional functionality. Participants gave a wide range of feedback on how they would like to use a digital identity. For example, P8 wanted to add documents for people in her care: "*. . . As a mom, if I'm going to do my health card, I'd like to do my children's so it's all together, right? And maybe, perhaps if I'm responsible for other people.*" P4 wanted to add "*health and car insurance information*," while P19 wanted to "*link [a] student card to a wallet*" in order to facilitate transactions, such as printing or food purchases, and add credit card information provided that it was "*guaranteed [to be more] secure than just [being] in the Apple Wallet.*"

Requests for clearer communication. When participants had issues submitting their identity information, many felt the app did not clearly indicate the problem or

how it could be resolved. For instance, when information from a submitted passport was missing, P19 saw "*a strange pop-up... with some app-like development jargon*" which she felt insufficiently explained what was missing.

When entering information manually from identity documents, P6 had difficulty locating certain numbers on the documents and requested that the app use language like "*on the front or the back of the card.*"

Operational requirements. Some participants described the qualities they would expect in the people interacting with their personal information. For example, P4 expected "*competent staff*" but did not specify of what the staff's skill set should consist. P7 spoke with similar ambiguity saying that anyone accessing their personal information should be "*a clean person.*"

Others spoke about their expectations for organizations operating a digital identity service. P7, for instance, explained that the organization's "*security system should be monitored*" but did not specify how the system would work or what kinds of monitoring would be in place. P8 also discussed how any organization responsible for such a service "*would have to be a very secure place*" but did not elaborate beyond this.

Technical support requirements. P17 suggested that contact information should be included for the people responsible for the app so users can ask questions: "*You know when you buy something from IKEA and they tell you, 'Well, don't break it. If you don't know how to do it, call us.' Something like that.*" P8 agreed saying that a toll-free number should be provided to make your identity "*inoperable*" or to "*put a pause on the app.*"

P5 said the app should include resources for identity theft since she would not know what to do if she became a victim of identity theft. She believed that such resources would "*[show] the friendliness of the app*" and reinforce that someone is "*looking out for [the user] in a sense.*"

Technical difficulties encountered. Several participants encountered technical difficulties while using the app. P6, for instance, explained that "*certain documents like the health card didn't scan*" and they had to resort to a manual process for entering

information that "*took forever*" to input. P4 encountered a similar issue where he was only able to scan the front side of a card but still had to manually enter information for the information that was captured.

Other participants, such as P21, experienced technical difficulties during in-person authentication task. Although the scanner gave audible feedback when P21 presented her QR code, the app had actually failed to establish a connection to the POS software. Lacking the appropriate signals, she asked, "*Does it work?*"

Misunderstandings. Like other participants, P15 mistakenly tried scanning a digital copy of an identity document displayed by the app rather than scanning the QR code for pairing. This prompted P15 to state, "*I feel like you just need to... test it out at home before you go to the clinic.*"

Despite directly interacting with it, P21 failed to notice the screen where users can review and select which of their identity attributes will be shared with the relying party: "*I didn't notice that.*" P19 also overlooked this screen saying, "*... maybe it just glazed over me... it wasn't quite obvious to me that you could do that...*".

Uses for Digital Identity

Participants discussed the appeal of certain use cases for digital identity. For instance, some participants were more willing to use their identity with health care institutions knowing that the institution would already have access to their information through other means. Participants also raised questions relating to the practicality of using a digital identity in certain contexts (e.g., Is the infrastructure ready to support use by travellers?, Would using digital identity in emergency situations require entail sacrificing security?). These results suggest that there is value in tailoring a digital identity service for specific use cases and communicating considerations that have been made for these situations to users.

Interactions with the government. P19 stated she would be comfortable using a digital identity to receive government services online, viewing it as an alternative means for relaying the same information that traditional documents provide: "*They*

need that stuff [identity information] anyway. It's either me manually typing in the driver's license or just scanning it anyway."

Health care. Several participants were interested in the use of a digital identity for health care services. P4 elaborated on their interest saying, "*it'd be actually really helpful to have your [health card] number handy if it's in your phone.*" P13 also saw the convenience of using digital identities for health care but added their use would be rooted in the fact that "*you're gonna be in the [health care provider's] system anyways*" possibly signaling that whether the relying party already has the user's PII may be a factor when deciding whether to use the app in a given context.

Multiple participants said a digital identity would be useful in the context of emergency medical services. For instance, P9 imagined being able to identify someone unconscious and learn about their medical history but that there would need to be "*some way they [relying parties] could bypass the [smartphone's and the app's] password.*" P9, however, quickly acknowledged that the existence of such a mechanism would call into question "*why there's a password if anyone can bypass it.*"

Purchases. Multiple participants were open to the concept of using digital identity for purchases. P6, for example, foresaw the benefit of such technology when "*buying marijuana . . . or liquor or things like that where you'd have to verify more.*" Despite having earlier voiced concerns about the app making personal information available too easily, P8 called the online authentication feature "*convenient*" after signing into the liquor store website: "*It's very hypocritical of me but that's the way it is, right? I wanna buy something online or I need them to check my date of birth, well, I'm gonna have to deal with it [my concerns].*"

4.5 Discussion

Through this study, we learned about users' perceptions and expectations of a digital identity service. In this section, we provide insights into the results of this study and how they can be applied more generically to the design of digital identity services.

4.5.1 Disconnect Between Participants' Mental Models and the Designer's Model

Norman [90] defines *conceptual models* as explanations for how something works. He states that designers can help users form accurate *mental models* (conceptual models in their minds) by ensuring that the *system image* (the artefacts that the user interacts with) reflects the *designer's model* (the conceptual model that the designer wishes to convey).

Our qualitative analysis revealed instances where our participants' mental models differed substantially from the designer's model. A major example of this was the failure by some participants to observe the app's privacy-preserving features, such as being able to select which identity attributes are shared with relying parties and identity information not being stored in the cloud. A similar disconnect was also observed in participants' understanding of the information flows that occur when participating in transactions.

It is important to acknowledge that participants' mental models were likely influenced by their past experiences; a phenomenon known as *transfer of learning* [97]. In some cases, these prior experiences helped participants accomplish tasks more easily (*positive transfer*). For instance, multiple participants drew upon their experiences scanning and depositing cheques within mobile banking apps in order to complete the document scanning portion of the registration task.

In other situations, these experiences appeared to hinder a participant's ability to complete a task (*negative transfer*). For instance, some of our participants had experience using digital wallet platforms, such as Apple Wallet [8] or Google Pay [56]. In-person transactions involving these platforms typically require users to tap the image of the card they wish to use and, in some cases, to scan a barcode displayed on the card. During our in-person authentication task, multiple participants tried to authenticate by scanning a barcode from an identity card displayed within the app instead of scanning the QR code. One possible way to counteract this would be to display an eID-Me identity card in the app containing the QR code and asking users to scan this instead.

Participants' conceptual models for the digital identity service may have also

been shaped by their own privacy expectations and public discourse on privacy. For instance, some participants were aware that some technology companies collect user data and share it with third parties for marketing purposes and questioned whether this was the case with eID-Me. Some participants made specific mention of Facebook's data-sharing agreements with Cambridge Analytica [25] which were made public a few months before we started conducting our study. Other participants mentioned data breaches that they had heard about in the news and speculated about the possibility of identity information being stolen from eID-Me.

Ultimately, these results suggest that there are challenges associated with helping users form accurate conceptual models for a digital identity service. Accordingly, we suggest that designers of digital identity services apply the following recommendations:

Communicate the service's data practices prominently. It is natural for users to apply their prior experiences to a new context, however, it is simultaneously important that designers mitigate against the potential for these experiences to become counterproductive for their users. In the case of eID-Me, we observed many users make incorrect conclusions about the app, such as the notion that service operators would be able to track their use of the service. Statements like these are at odds with the app's actual behaviour and, when left unaddressed, can result in users assuming that the app is more privacy adverse than it actually is. This suggests a need to proactively counter users' misconceptions early in their use of the service.

To do this, designers of digital identity services should clearly indicate what is happening with personal information so that users understand what information is being collected, who can access it, and where it is being stored. While these types of statements are typically reserved for privacy policies, these policies have limited effect because they are often ignored by users [94]. Further, many websites and apps employ privacy policies as *dark patterns* [22] to purposefully hide undesirable aspects of their data practices within confusing legalese.

Digital identity services should emphasize, not conceal, their data practices, especially in the case of a privacy-preserving app such as eID-Me. Otherwise, users could fail to discover privacy-related features and dismiss the app altogether.

Set expectations early. The term *user onboarding* [80,89] refers to mechanisms for introducing first-time users of an app to its main features and guiding them through any setup processes. During user onboarding, designers typically provide relevant information, such as how the app might benefit the user or instructions on how to perform key tasks, that will help make the experience as straightforward as possible.

User onboarding is especially critical for apps that have complex workflows or use technology in ways unfamiliar to users. In the case of eID-Me, the app's registration process was more complex than conventional registration processes that involve picking a username and password; participants were required to take a selfie and scan particular documents. Outside of the lab environment, this process would be even more complex as users would have to be at home (or at a service centre) to complete the process and would need to locate the necessary identity documents.

The app also made use of technologies in ways that participants might not be accustomed to. One participant thought QR codes were meant to be scanned by customers who wanted additional information about a product as opposed to how eID-Me users must present a QR code to a scanner in order to start a transaction. In another instance, a participant failed to recognize that they could review the information being requested and, for optional attributes, select what the relying party would receive.

Also, our quantitative results showed that users' willingness to use the app generally decreased after completing the study's tasks. One possible explanation for this phenomenon is that the participants whose opinions became more negative had expectations of how the app would work and that these expectations were not met by the end of the evaluation.

Ultimately, since digital identity services make use of concepts unfamiliar to typical users and users are likely to have preconceived notions of how the service works, we recommend providing information upon launch to help users develop a more accurate understanding. Doing so can help users be successful in their use of the app and prevent negative outcomes that could occur from misuse, such as inadvertent sharing of personal information.

Focus on familiar concepts and actionable information. Some participants found it confusing to have a score assigned to the personally identifiable information they provided. While the lack of understanding likely stems from the technical format in which the scoring information was presented, it is worth considering the value of exposing this level of detail.

The concept of scoring might be serviceable as an internal representation of the authenticity of a user's personally identifiable information but for users, this information is of little value without additional context. We recommend that instead of communicating scoring information to users, designers should provide the user with actionable information such as telling them to provide more documents or to visit a service centre to finish registration.

4.5.2 Participants' Concerns Vary in Their Ability to Be Addressed

Our qualitative and quantitative results showed that users have wide-ranging concerns. For some of these concerns, devising solutions is rather straightforward. As an example, concerns about the implications of sharing personal information can be addressed by calling attention to the app's data practices and highlighting some of its privacy-preserving features. Similarly, concerns about the service operator's motives could be alleviated by being transparent about who the service operators are and how the service is financed.

For other concerns, conceiving solutions may seem infeasible at first. For example, in our examination of participants' concerns, we discovered an apparent paradox: participants wanted the service operator to implement mechanisms that would safeguard their digital identities from identity theft but also expressed concern about the use of their digital identity being tracked.

While solutions for concerns like these may seem unattainable, best practices suggest otherwise. Cavoukian's Privacy by Design (PbD) framework [31] claims that systems can avoid making such trade-offs by accounting for privacy from the onset. With an awareness of what users value, our findings present an opportunity to address these concerns while also maintaining their privacy.

4.6 Limitations

The following limitations may have impacted participants' behaviour and, ultimately, the ecological validity of this study:

- We provided participants with sample identity documents to complete the registration process instead of asking them to use their own identity documents. Participants were asked to treat these documents as if they were their own and contained their personal information. However, given that these documents did not contain their personal information, participants may have reported different levels of comfort with providing specific types of information than if their own documents were in front of them. Similarly, participants' feedback for the in-person and online authentication tasks may have been different if their own identity information was being shared with relying parties.

- All sessions were conducted in a laboratory environment. In the real world, participants would be required to complete registration at their home addresses, and would complete any in-person authentication at real venues, such as hospitals and retail stores. Naturally, this would have created a considerable lapse of time between registration and use. Participants' opinions and reactions may have differed from those expressed in the lab environment.

- Instead of using their own devices, we provided participants with devices (i.e., a smartphone and a laptop computer) to complete the study's tasks. Although eID-Me allows users to unlock the app via biometric authentication (e.g., facial recognition, fingerprint recognition), these devices had not been configured with participants' biometrics. As such, the app used the device passcode as the fallback authentication scheme. Participants may have expressed different opinions and felt more secure if they had used biometrics instead of passcodes.

Despite these limitations, we still believe this study helped us gain insights into how users perceive digital identity services at a high level rather than identifying issues with the implementation of a specific service. Ultimately, the results provide us

with an understanding of the elements needed for a digital identity service to be both usable and protective of users' privacy.

4.7 Conclusion

After conducting our usability evaluation, we learned that designing a digital identity service to be both usable and privacy-protective requires non-trivial consideration for users' prior experiences, the context in which the identity will be used, and the expectations that users have for the service. Based on our findings, we recommend that designers confront these challenges through upfront communication, familiar metaphors, and the use of best practices to balance privacy and functionality.

Chapter 5

Discussion and Conclusion

5.1 Addressing Our Research Questions

RQ1.1: To what degree are certain types of PII discoverable through social media services and search engines? Our evaluation of participants' online presence found categories of PII that, for the majority of our participants, were identified through social media (i.e., *Personal Information*, *Birth Information*, *Address Information*, *Contact Information*, and *Occupational Information*) and were identified through a search engine (i.e., *Personal Information*, *Address Information*, *Occupational Information*).

Given that a main use case of social media is to maintain existing relationships [78, 114], the discovery that more categories were identified through social media than through search engines may be an indication that people known to a user could be better positioned to carry out an attack. However, given that approximately half of the accounts were public and that, on average, participants had hundreds of friends in their network, the visibility of this data is widespread. These findings are consistent with other literature that reached similar conclusions about the ability for an individual's personal information to be accumulated by reviewing their social media accounts [71]. Additionally, our findings fill a literature gap on the ability for personally identifiable information to be obtained through search engines.

RQ1.2: How do an individual's behaviours, both online and offline, relate to the ease with which an imposter could obtain their PII? Through semi-structured interviews, we learned that participants engage in a wide range of behaviours both online and offline. Some of these behaviours, such as choosing not to carry certain documents on their person or using passwords and biometrics to protect their devices, had the potential to preserve participants' personal security and privacy.

Other behaviours—sometimes deliberate—may have jeopardized their security and privacy (e.g., not using any authentication schemes on their devices, receiving mail through a box without a lock, leaving personal belongings unattended).

Some literature argues that when users reject security advice, the root cause is neither negligence nor unusable software but rather users concluding that the costs of protecting themselves are greater than any resulting benefits [66]. Examples of this surfaced during our study with users describing themselves as unlikely targets of identity theft or explaining that they preferred quick access to their device instead of being slowed down by passwords. These findings present a challenge in coming up with appealing solutions for maintaining users' security and privacy.

RQ2.1: How do users perceive digital identity services? From our usability study, we discovered that our participants mainly understood digital identity services to be convenient alternatives to physical documents but that their interest depended on the circumstances (e.g., hesitant to use with law enforcement). Participants also viewed digital identity services as a potentially risky technology, raising questions about data breaches, identity theft, the implications of drained mobile phone batteries, surveillance, and more. These findings help to address the lack of literature on the usability of digital identity services.

Our participants' perception of digital identity services was frequently inaccurate. In some cases, these misunderstandings stemmed from participants' prior experiences with other technologies, such as presuming that the interface for the eID-Me app would function similarly to the interface for a mobile wallet app or speculating that data provided to the app might be harvested for marketing or tracking purposes given that other apps have engaged in this practice.

In other cases, these misconceptions were seen in areas of the app where users were provided with little guidance (e.g., when scanning a document) or were presented with highly technical information (e.g., when reviewing a submission). Finding a happy medium between these two extremes may prove to be difficult given evidence that users often skip educational materials to try completing their task first rather than first learning about how to complete the task [29].

RQ2.2: What are users' expectations for digital identity services? The usability study also provided valuable insight into users' expectations for digital identity services. The main expectations included clarity surrounding the app's practices when handling users' data, safeguarding users' security and privacy through principled decision-making in terms of how the service is operated and who is hired to operate it, and processes for dealing with a lost or stolen smartphone.

These findings mirror relevant works [17, 105] and ultimately, serve as a reminder that technical security alone is often not enough to foster trust with users; how users perceive a system's security and privacy mechanisms can influence their adoption of the system.

5.2 Recommendations

Based on the results from the two studies we conducted, we provide recommendations for the design and successful deployment of smartphone-based digital identity services.

5.2.1 Strengthen Identity Proofing Mechanisms

Our work highlights that many types of PII are easily accessible by strangers and non-strangers alike. Accordingly, we suggest two preventative measures for managing this reality.

First, when initially registering for an identity, digital identity services should use out-of-wallet questions to verify a user's identity. Out-of-wallet questions, sometimes referred to as dynamic KBA [81], are authentication challenges that are generated on the fly, in real-time using information available from a variety of sources (e.g., credit bureaus, mobile network operators) [33, 77]. These types of questions are intended to be more difficult for an imposter to answer correctly because the necessary information is often harder to obtain than other types of information like the user's full name or date of birth.

As an example, a scheme relying on out-of-wallet questions might consult a user's credit file and ask them to identify the credit provider that issued them a loan two months ago. Another user, who might not have any debt, could be asked about recent activities on their mobile phone.

While this lack of predictability may be ideal for hindering an attacker's efforts, it also presents usability challenges to legitimate applicants. Rather than relying exclusively on this mechanism, it may be more appropriate to use it in combination with other mechanisms when developing an identity proofing solution.

Second, digital identity services should require users to provide multiple factors when they are initially registering for an identity. Relying solely on PII is risky given the possibility of information becoming obtainable through data breaches. By using multiple factors at the time of registration, such as how eID-Me makes use of facial data and the location of the user's device, these services can increase the level of effort and skill needed by an attacker thus impeding their ability to obtain a digital identity fraudulently.

5.2.2 Discourage Self-Harming Behaviours

Through our information discoverability study, we observed participants either downplaying or altogether failing to acknowledge the riskiness of behaviours that could in fact compromise their security and privacy. As such, we believe that efforts should be made to discourage users from engaging in these behaviours. However, one of the challenges involved with effecting this type of change, according to Bada et al. [13], is that simply educating users on the risks and how to protect themselves has limited impact. For changes in behaviour to occur, users need to comprehend the recommendations and also be motivated enough to apply them [13], thus indicating that successful campaigns will be highly tailored for their intended audience.

5.2.3 Provide Assurances for Users' Privacy

Participants were vocal with their distaste for their personal information being misused whether through the sale of information or through government monitoring. As such, we believe it is crucial for digital identity services to be privacy-preserving and to actively emphasize any protections that they have in place.

Today, many apps communicate these protections by using privacy policies, but research has shown that users frequently struggle to comprehend these texts [72], making privacy policies ineffective tools for relaying information. Instead of relying

solely on privacy policies—which are often filled with legalese—Schaub, Balebako, and Cranor [109] advocate for the use of privacy notices. They describe privacy notices as user-centric documents that serve to complement privacy policies and contain "timely, relevant, actionable, and understandable information" that can support users in making informed decisions about their use of a particular service.

Based on our users' feedback, we suggest that services should make use of shorter, more digestable privacy notices before users interact with key features in addition to making a full privacy policy available to users. For example, prior to taking a selfie, a brief privacy notice could be presented to users explaining how facial data will be used. In another instance where the user is authenticating on the web, a separate notice could describe who would have access to their information in that context.

5.2.4 Upfront Communication and Familiar Language

After observing some users encounter difficulties while using the app (e.g., confusion over technical terms, uncertainty surrounding the document scanning process) and in line with best practices [52, 91], we suggest explaining essential concepts about the app upfront. While some users were able to complete tasks independently, it is clear that a sizeable portion would have benefitted from additional information on how to accomplish key activities.

Further, we recommend conveying information to users with concepts that they will find intuitive. Digital identities are relatively new and unfamiliar to many. To minimize confusion, it is important that these concepts be framed using ideas that are already familiar to users [73].

5.3 Future Directions

The results of our studies present compelling opportunities for future research and in this section, we highlight some of these areas.

5.3.1 Methodology and Scope

We conducted both the information discoverability study and the usability study in a lab environment with each session being relatively short. Conducting these studies over longer periods of time and in more realistic environments might bring additional insights. For example, parts of the information discoverability study could be conducted as a diary study where participants would take note of behaviours like the safekeeping of their personal belongings over several days.

It may also be worthwhile to investigate the discoverability of PII on more niche social networks given that we examined fairly mainstream networks with our participants. For example, online support communities, such as PatientsLikeMe [96], encourage users to share their personal experiences in order to connect with others that have similar health conditions. Are these users more likely to publicize their PII?

5.3.2 Employ Gamification for Effective Onboarding

One of the challenges that we identified in our usability study for smartphone-based digital identity services was the need to support users in developing accurate conceptual models for how the app works. We believe this would be particularly difficult because of negative transfer effects and users' tendencies to skip instructional information.

A possible approach to this problem could be to incorporate elements of games into the app to keep users engaged, a concept known as *gamification* [67, 104]. Many video games dedicate their first levels to introducing the game's *mechanics* (i.e., rules) and controls, helping to ensure users are successful as they go on to accomplish other tasks within the game.

As an example, eID-Me could employ gamification to educate users on how the app is expected to be used in real-world scenarios, such as in retail stores. The app could simulate a retail environment and require users to complete a trial use of their identity. This would help to familiarize users with the workflow while also giving them an opportunity to learn about the app in a low-pressure environment as opposed to while others are waiting in line behind them.

5.3.3 Eliminate Barriers for Assessing PII Availability

A key aspect of preventing identity theft is being able to manage the availability of one's PII online. If users are to take an active role in doing this, tools for this process should be improved.

While the passing of privacy legislation, such as the General Data Protection Regulation [119], has resulted in users gaining the ability to learn about the scope of data that businesses have collected about them and to request the deletion of this data, these processes are quite primitive. It is common practice for businesses to generate large ZIP files containing users' data—sometimes represented as complex JSON or CSV files—which users must first download before examining using whichever tools they can find.

This information could be presented in a comprehensible format directly on the business's website so that users do not have to wait for a download and can act directly on a given piece of information.

5.3.4 Increase Transparency for How Mobile Apps Behave

While today's mobile operating systems offer control over the types of information on a user's device that installed apps can access (i.e., permissions), these operating systems do not provide insight into how these apps are using personal data. Further, they do not allow more granular restrictions beyond a binary *access granted* or *access denied*.

Lutaaya [79] developed a proof-of-concept prototype exploring extended capabilities for permissions on mobile operating systems such as substituting users' personal data with sample data and monitoring the frequency and destination of communications sent to remote servers that involve protected resources.

We believe that if mobile operating systems had these additional capabilities, it could remove some of the mystery surrounding apps' behaviour and possibly alleviate users' concerns with using a digital identity service on their smartphone.

5.4 Contributions

In this thesis, our aim was to advance knowledge concerning digital identity services by investigating the discoverability and suitability of users' PII for identity proofing and by conducting a usability evaluation of a smartphone-based implementation of a digital identity service. The primary contributions of this work were the following:

- We conducted a study examining how widely available participants' PII is and how their day-to-day behaviours affect the availability of their PII. Our findings indicated that many different types of PII could be obtained by an imposter and that individuals often underestimate the risk of said information being available. This led us to recommend that identity proofing processes rely on other types of information and to call for initiatives that can help users both recognize the risk associated with their behaviours and support them in changing their behaviours.
- We conducted a usability study to better understand how users perceive digital identity services and their expectations for these services. We observed that users had difficulty completing the study's tasks and many expressed concern over the implications of using a digital identity service for their own privacy and security. In many cases, these task failures and knowledge gaps stemmed from the app failing to convey the system image accurately. Accordingly, we suggested that future implementations of digital identity services use familiar concepts to help users develop accurate mental models and prioritize communicating with user-friendly language to ensure users understand how the service works.

5.5 Conclusion

As more aspects of everyday life shift towards digital technologies, the success of these transitions hinges on the technology being reliable enough to earn the trust of both individuals and organizations. For digital identity services, this is no different. Accordingly, we looked into this issue by conducting two studies on topics that affect a service's ability to be trusted: (1) the discoverability of users' PII and its suitability for identity proofing, and (2) the usability of a smartphone-based implementation of a digital identity service.

Through the information discoverability study, we identified types of PII that are most easily discovered and learned how our participants think about their personal security and privacy. Through the usability study, we learned about users' expectations for how a digital identity service should be operated and factors that impacted their conceptual models for a smartphone-based implementation. Based on our findings, we recommended that: (1) identity proofing processes request other types of information from users, (2) concerted efforts be made to encourage users to change behaviours that may jeopardize their own security and privacy, (3) the privacy implications of digital identity services be communicated to users in a clear, digestable manner, and (4) digital identity services make use of familiar concepts and approachable language to when communicating with users. Finally, we identified opportunities for future research that we felt have the potential to advance knowledge in the identity management space.

Bibliography

[1] 116th Congress. S.1084 – Deceptive Experiences To Online Users Reduction Act, 2019. https://www.congress.gov/bill/116th-congress/senate-bill/1084. [Accessed: 2019-06-23].

[2] William F Abt, Robert W Harris, Asher C Johnson, et al. Deterministic verification of digital identity documents, March 5 2019. US Patent App. 10/225,088.

[3] Alessandro Acquisti and Ralph Gross. I just found 10 million SSNs. In *Black Hat USA 2009*, 2009.

[4] Georg Aichholzer and Stefan Strauß. Electronic identity management in e-government 2.0: Exploring a system innovation exemplified by Austria. *Information Polity*, 15(1, 2):139–152, 2010.

[5] Gergely Alpár, Jaap-Henk Hoepman, and Johanneke Siljee. The identity crisis. security, privacy and usability issues in identity management. *arXiv preprint arXiv:1101.0427*, 2011.

[6] Dimitrios Antoniou and Krzysztof Socha. Authentication methods. Whitepaper, Computer Emergency Response Team (CERT-EU) for the EU institutions, bodies, and agencies, 2016. http://cert.europa.eu/static/WhitePapers/CERT-EU%20SWP-16_003_Authentication%20Methods.pdf. [Accessed: 2019-06-07].

[7] Apple Inc. Numbers - Apple. https://www.apple.com/numbers/. [Accessed: 2019-03-30].

[8] Apple Inc. Use Wallet on your iPhone or iPod touch - Apple Support. https://support.apple.com/en-us/HT204003. [Accessed: 2019-04-23].

[9] Amelia M. Arria, Kimberly M. Caldeira, Kathryn B. Vincent, Brittany A. Bugbee, and Kevin E. O'Grady. False identification use among college students increases the risk for alcohol use disorder: Results of a longitudinal study. *Alcoholism: Clinical and Experimental Research*, 38(3):834–843, 2014.

[10] Associated Press. State Department now requires U.S. visa applicants to share social media accounts, 2019. https://www.cbsnews.com/news/state-department-now-requires-us-visa-applicants-to-share-social-media-accounts-2019-06-01/. [Accessed: 2019-06-26].

[11] Attorney-General's Department. National identity proofing guidelines. Technical report, Commonwealth of Australia, 2016. https://www.homeaffairs.gov.au/

criminal-justice/files/national-identity-proofing-guidelines.pdf. [Accessed: 2019-01-30].

[12] Daniel Augot, Hervé Chabanne, Olivier Clémot, and William George. Transforming face-to-face identity proofing into anonymous digital identity using the Bitcoin blockchain. In *2017 15th Annual Conference on Privacy, Security and Trust (PST 2017)*, pages 25–34, 2017.

[13] Maria Bada, Angela Sasse, and Jason R. C. Nurse. Cyber security awareness campaigns: Why do they fail to change behaviour? In *International Conference on Cyber Security for Sustainable Society*, pages 118–131, 2015.

[14] Michael Barbaro and Tom Zeller Jr. A face is exposed for AOL searcher no. 4417749. *The New York Times*, 2006. https://www.nytimes.com/2006/08/09/technology/09aol.html. [Accessed: 2019-06-24].

[15] Barclays Bank UK PLC. Barclays Identity Service. https://www.barclays.co.uk/identity/. [Accessed: 2019-06-26].

[16] BBC News. Security flaw forces Estonia ID 'lockdown'. *BBC News*, 2017. http://www.bbc.com/news/technology-41858583. [Acccessed: 2019-06-25].

[17] France Belanger, Janine S Hiller, and Wanda J Smith. Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *The Journal of Strategic Information Systems*, 11(3):245–70, 2002.

[18] Juanita Blue, Joan Condell, and Tom Lunney. A review of identity, identification & authentication. *International Journal for Information Security Research*, 8(2):794–804, 2018.

[19] Bluink Ltd. Bluink | Secure your digital identity. https://bluink.ca. [Accessed: 2019-02-04].

[20] Bluink Ltd. eID-Me | Fast and secure mobile identity verification and digital identity | Bluink. https://bluink.ca/eid-me. [Accessed: 2019-04-22].

[21] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy*, pages 553–567, 2012.

[22] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: Privacy dark strategies and privacy dark patterns. In *Proceedings on Privacy Enhancing Technologies*, volume 4, pages 237–254, 2016.

[23] Anu Bradford. The Brussels effect. *Northwestern University Law Review*, 107(1), 2012.

[24] Cabinet Office. Identity proofing and verification of an individual. Good Practice Guide 45 Version 3, Government of the United Kingdom, 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/720963/good_practice_guide_45_identity_proofing_version_3_february_2017.pdf. [Accessed: 2019-01-31].

[25] Carole Cadwalladr and Emma Graham-Harrison. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, March 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. [Accessed: 2019-04-15].

[26] California State Legislature. The California Consumer Privacy Act of 2018, 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375. [Accessed: 2019-06-23].

[27] L. Jean Camp. Digital identity. *IEEE Technology and Society Magazine*, 23(3):34–41, 2004.

[28] Canadian Bankers Association. Canada's digital ID future - a federated approach. https://cba.ca/embracing-digital-id-in-canada. [Accessed: 2019-06-26].

[29] John M. Carroll and Mary Beth Rosson. Paradox of the active user. In John M. Carroll, editor, *Interfacing Thoughts: Cognitive Aspects of Human-Computer Interaction*, chapter 5, pages 80–111. MIT Press, 1987.

[30] Daniel Castro. Explaining international leadership: Electronic identification systems. Technical report, The Information Technology & Innovation Foundation, 2011. http://www.itif.org/files/2011-e-id-report-final.pdf. [Accessed: 2019-06-26].

[31] Ann Cavoukian. The 7 foundational principles: Implementation and mapping of fair information principles. Technical report, Information and Privacy Commissioner of Ontario, 2010. http://www.ontla.on.ca/library/repository/mon/24005/301946.pdf. [Accessed: 2019-03-29].

[32] Brian X. Chen. Getting a flood of G.D.P.R.-related privacy policy updates? Read them. *The New York Times*, 2018. https://www.nytimes.com/2018/05/23/technology/personaltech/what-you-should-look-for-europe-data-law.html. [Accessed: 2019-06-22].

[33] Ye Chen and Divakaran Liginlal. Bayesian networks for knowledge-based authentication. *IEEE Transactions on Knowledge and Data Engineering*, 19(5):695–710, 2007.

[34] Michael Clark and Kyran Dale. Estimating what proportion of the public will be able to use GOV.UK Verify, 2016. https://identityassurance.blog.gov.uk/2016/01/25/estimating-what-proportion-of-the-public-will-be-able-to-use-gov-uk-verify/. [Accessed: 2019-06-26].

[35] Computer Security Divison. Minimum security requirements for federal information and information systems. Federal Information Processing Standards Publications (FIPS PUBS) 200, National Institute of Standards and Technology, 2006. https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.200.pdf. [Accessed: 2019-06-07].

[36] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, and W. Polk. Internet X.509 Public Key Infrastructure certificate and Certificate Revocation List (CRL) profile. RFC 5280, RFC Editor, May 2008. http://www.rfc-editor.org/rfc/rfc5280.txt.

[37] Heith Copes and Lynne Vieraitis. Identity theft: Assessing offenders' strategies and perceptions of risk. Report, Inter-university Consortium for Political and Social Research [distributor], 2007. https://www.ncjrs.gov/pdffiles1/nij/grants/219122.pdf. [Accessed: 2017-12-12].

[38] Danny De Cock, Christopher Wolf, and Bart Preneel. The Belgian electronic identity card (overview). In *Sicherheit*, volume 77, pages 298–301, 2006.

[39] Deloitte LLP. Brazilian General Data Protection Act. https://www2.deloitte.com/br/en/pages/risk/articles/lgpd.html. [Accessed: 2019-06-25].

[40] Deloitte LLP. A new era for privacy: GDPR six months on. Technical report, Deloitte LLP, 2018. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/risk/deloitte-uk-risk-gdpr-six-months-on.pdf. [Accessed: 2019-06-23].

[41] Hilde Van den Bulck. Tools for studying the media. In Chris Newbold, Oliver Boyd-Barrett, and Hilde Van den Bulck, editors, *The Media Book*, chapter 2, pages 55–100. Arnold, 2002.

[42] Department of Homeland Security. The DHS Privacy Office Guide to Implementing Privacy. Technical report, Department of Homeland Security, 2010. https://www.dhs.gov/xlibrary/assets/privacy/dhsprivacyoffice-guidetoimplementingprivacy.pdf. [Accessed: 2019-06-27].

[43] Department of Justice Canada. Personal Information Protection and Electronic Documents Act, 2000. https://laws-lois.justice.gc.ca/eng/acts/P-8.6/FullText.html. [Accessed: 2019-06-19].

[44] Rachna Dhamija and Lisa Dusseault. The seven flaws of identity management: Usability and security challenges. *IEEE Security & Privacy*, 6(2):24–29, 2008.

[45] Digital ID & Authentication Council of Canada. https://diacc.ca/. [Accessed: 2019-06-26].

[46] Digital Transformation Agency, Commonwealth of Australia. Vision 2025, 2018. https://dta-www-drupal-20180130215411153400000001.s3.ap-southeast-2.amazonaws.com/s3fs-public/files/digital-transformation-strategy/digital-transformation-strategy.pdf. [Accessed: 2019-06-26].

[47] Erin Egan and Ashlie Beringer. It's time to make our privacy tools easier to find, 2018. https://newsroom.fb.com/news/2018/03/privacy-shortcuts/. [Accessed: 2019-06-22].

[48] Satu Elo and Helvi Kyngäs. The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1):107–115, 2008.

[49] Equifax Inc. eIDverifier customer ID verification | Equifax. https://www.consumer.equifax.ca/business/eidverifier/. [Accessed: 2019-06-18].

[50] Experian Information Solutions, Inc. GOV.UK verify | identity & fraud | Experian UK. http://www.experian.co.uk/identity-and-fraud/govuk-verify/. [Accessed: 2019-06-26].

[51] Federal Aviation Administration. Protecting personally identifiable information (PII), 2008. https://www.faa.gov/documentLibrary/media/Order/1280.1B_.pdf. [Accessed: 2019-06-27].

[52] Steven M. Furnell, Adila Jusoh, and Dimitris Katsabas. The challenges of understanding and using security: A survey of end-users. *Computers & Security*, 25(1):27–35, 2006.

[53] Angela Gilbert. Novelty fake IDs fly under forgery radar. *CBC News*, 2012. https://www.cbc.ca/news/canada/hamilton/news/novelty-fake-ids-fly-under-forgery-radar-1.1217994. [Accessed: 2019-06-25].

[54] Philippe Golle. Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, WPES '06, pages 77–80, 2006.

[55] Dan Goodin. Millions of high-security crypto keys crippled by newly discovered flaw. *Ars Technica*, 2017. https://arstechnica.com/information-technology/2017/10/crypto-failure-cripples-millions-of-high-security-keys-750k-estonian-ids/. [Accessed: 2019-06-25].

[56] Google LLC. Google Pay (CA) – pay in apps, on the web, and in stores. https://pay.google.com/about/. [Accessed: 2019-04-23].

[57] Government Digital Service. GOV.UK Verify - GOV.UK, 2019. https://www.gov.uk/government/publications/introducing-govuk-verify/introducing-govuk-verify. [Accessed: 2019-06-26].

[58] Government Digital Service, Cabinet Office, and Oliver Dowden. Government technology innovation strategy. Policy paper, Government of the United Kingdom, 2019. https://www.gov.uk/government/publications/the-government-technology-innovation-strategy/the-government-technology-innovation-strategy.

[59] Government of Canada. Canada's trusted digital identity vision. https://www.canada.ca/en/treasury-board-secretariat/corporate/news/canada-trusted-digital-identity-vision.html. [Accessed: 2019-06-26].

[60] Paul A. Grassi, James L. Fenton, Naomi B. Lefkovitz, Jamie M. Danker, Yee-Yin Choong, Kristen K. Green, and Mary F. Theofanos. Digital identity guidelines: Enrollment and identity proofing. NIST Special Publication 800-63A, National Institute of Standards and Technology, 2017. https://doi.org/10.6028/NIST.SP.800-63a. [Accessed: 2019-06-18].

[61] Paul A. Grassi, Michael E. Garcia, and James L. Fenton. Digital identity guidelines. NIST Special Publication 800-63-3, National Institute of Standards and Technology, 2017. https://doi.org/10.6028/NIST.SP.800-63-3. [Accessed: 2019-01-30].

[62] Paul A. Grassi, Justin P. Richer, Sarah K. Squire, James L. Fenton, Ellen M. Nadeau, Naomi B. Lefkovitz, Jamie M. Danker, Yee-Yin Choong, Kristen K. Greene, and Mary F. Theofanos. Digital identity guidlines: Federation and assertions. NIST Special Publication 800-63C, National Institute of Standards and Technology, 2017. https://doi.org/10.6028/NIST.SP.800-63c. [Accessed: 2019-06-11].

[63] Graham Greenleaf. Global data privacy laws 2019: 132 national laws & many bills. *157 Privacy Laws & Business International Report, 14–18*, 2017.

[64] Åke Gŕonlund. Electronic identity management in sweden: governance of a market approach. *identity in the information society*, 3(1):195–211, 2010.

[65] Marit Hansen, Andreas Pfitzmann, and Sandra Steinbrecher. Identity management throughout one's whole life. *Information security technical report*, 13(2):83–94, 2008.

[66] Cormac Herley. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop*, NSPW '09, pages 133–144. ACM, 2009.

[67] Kai Huotari and Juho Hamari. Defining gamification: A service marketing perspective. In *Proceeding of the 16th International Academic MindTrek Conference*, MindTrek '12, pages 17–22, 2012.

[68] IBM. What is a digital certificate. `https://www.ibm.com/support/knowledgecenter/en/SSB23S_1.1.0.12/gtps7/s7what.html`. [Accessed: 2019-06-27].

[69] IBM. Digital certificates, 2019. `https://www.ibm.com/support/knowledgecenter/en/SSMKHH_10.0.0/com.ibm.etools.mft.doc/ac55140_.htm`. [Accessed: 2019-06-27].

[70] Apple Inc. Get a copy of the data associated with your Apple ID account, 2019. `https://support.apple.com/en-us/HT208502`. [Accessed: 2019-06-22].

[71] Danesh Irani, Steve Webb, Kang Li, and Calton Pu. Modeling unintended personal-information leakage from multiple online social networks. *IEEE Internet Computing*, 15(3):13–19, 2011.

[72] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 471–478, 2004.

[73] Anna Kaley. Match between the system and the real world: The 2nd usability heuristic explained, 2018. `https://www.nngroup.com/articles/match-system-real-world/`. [Accessed: 2019-06-14].

[74] Mark Joseph Kapczynski. Digital identity, February 15 2018. US Patent App. 15/662,712.

[75] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology*, chapter 11, pages 221–236. Sage Publications, Inc., second edition, 2004.

[76] Karen Lewison and Francisco Corella. Rich credentials for remote identity proofing. `https://pomcor.com/techreports/RichCredentials.pdf`. [Accessed: 2019-06-25], 2017.

[77] LexisNexis. Financial service identity management in an era of tell-all and technology overload. White paper, LexisNexis, 2012. `http://lexisnexis.com/risk/downloads/idm/financial-services-identity-management.pdf`. [Accessed: 2019-06-21].

[78] Kuan-Yu Lin and Hsi-Peng Lu. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, 27(3):1152 – 1161, 2011.

[79] Michael Lutaaya. Rethinking app permissions on iOS. SIGCHI Conference on Human factors in Computing Systems (CHI) Student Research Competition, 2018. Extended Abstracts and Workshop Papers.

[80] Lukas Mathis. *Designed for Use: Create Usable Interfaces for Applications and the Web*. Pragmatic Bookshelf, second edition, 2016.

[81] Priya Matta and Bhasker Pant. TCpC: a graphical password scheme ensuring authentication for IoT resources. *International Journal of Information Technology*, 2018.

[82] Erika McCallister, Tim Grance, and Karen Scarfone. Guide to protecting the confidentiality of personally identifiable information (PII): Recommendations of the National Institute of Standards and Technology. NIST Special Publication 800-122, National Institute of Standards and Technology, 2010. `https://doi.org/10.6028/NIST.SP.800-122`. [Accessed: 2019-06-19].

[83] Mary L. McHugh. Interrater reliability: the kappa statistic. *Biochemia Medica*, 2(3):276–282, 2012.

[84] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[85] Mozilla and individual contributors. Encryption and decryption, 2019. `https://developer.mozilla.org/en-US/docs/Archive/Security/Encryption_and_Decryption`. [Accessed: 2019-06-27].

[86] Sean A. Munson, Daniel Avrahami, Sunny Consolvo, James Fogarty, Batya Friedman, and Ian Smith. Attitudes toward online availability of US public records. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, dg.o '11, pages 2–9, 2011.

[87] National Institute of Standards and Technology. Guideline for identifying an information system as a national security system. NIST special publication, National Institute of Standards and Technology, 2003. `https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-59.pdf`. [Accessed: 2019-06-27].

[88] National Institute of Standards and Technology. Measuring strength of identity proofing. Discussion draft, Inforation Technology Laboratory, NIST, 2015. `https://www.nist.gov/sites/default/files/nstic-strength-identity-proofing-discussion-draft.pdf`. [Accessed: 2019-06-25].

[89] Theresa Neil. *Mobile Design Pattern Gallery: UI Patterns for Smartphone Apps*. O'Reilly Media, second edition, 2014.

[90] Don Norman. *The Design of Everyday Things*. Basic Books, 2013.

[91] Jason R. C. Nurse, Sadie Creese, Michael Goldsmith, and Koen Lamberts. Guidelines for usable cybersecurity: Past and present. In *2011 Third International Workshop on Cyberspace Safety and Security (CSS)*, pages 21–26, 2011.

[92] OASIS Standard. Profiles for the OASIS Security Assertion Markup Language (SAML) v2.0, 2005. `http://docs.oasis-open.org/security/saml/v2.0/saml-profiles-2.0-os.pdf`. [Accessed: 2019-06-10].

[93] OAuth. OAuth community site. `https://oauth.net`. [Accessed: 2019-06-10].

[94] Jonathan A. Obar and Anne Oeldorf-Hirsch. The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, page To appear, 2018.

[95] Ottawa Public Library. Post public information. `https://biblioottawalibrary.ca/en/library/distributing-public-information`. [Accessed: 2019-01-13].

[96] PatientsLikeMe. Live better, together! | PatientsLikeMe, 2019. `https://www.patientslikeme.com`. [Accessed: 2019-06-05].

[97] David N. Perkins and Gavriel Salomon. Transfer of learning. In *International Encyclopedia of Education*. Pergamon Press, second edition, 1995.

[98] Privacy Analytics Inc. De-identification 101. White paper, Privacy Analytics Inc., 2017. `https://privacy-analytics.com/files/De-identification-101.pdf`. [Accessed: 2019-06-24].

[99] Privacy Office. Handbook for safeguarding sensitive personally identifiable information. Technical report, United States Department of Homeland Security, 2012. `https://www.dhs.gov/sites/default/files/publications/Handbook%20for%20Safeguarding%20Sensitive%20PII_0.pdf`, [Accessed: 2019-06-20].

[100] QSR International. NVivo qualitative data analysis software. `https://www.qsrinternational.com/nvivo`. [Accessed: 2019-05-03].

[101] Qualtrics. The leading research & experience software | Qualtrics. `https://www.qualtrics.com`. [Accessed: 2019-04-22].

[102] Donald J. Rebovich, Kristy Allen, and Jared Platt. The new face of identity theft: An analysis of federal case data for the years 2008 through 2013. Technical report, Center for Identity Management and Information Protection, 2015. `https://www.utica.edu/academic/institutes/cimip/New_Face_of_Identity_Theft.pdf`. [Accessed: 2019-05-31].

[103] Teemu Rissanen. Electronic identity in finland: ID cards vs. bank IDs. *Identity in the Information Society*, 3(1):175–194, 2010.

[104] Karen Robson, Kirk Plangger, Jan H. Kietzmann, Ian McCarthy, and Leyland Pitt. Is it all a game? understanding the principles of gamification. *Business Horizons*, 58(4):411–420, 2015.

[105] Juan Carlos Roca, Juan José García, and Juan José de la Vega. The importance of perceived trust, security and privacy in online trading systems. *Information Management & Computer Security*, 17(2):96–113, 2009.

[106] Derrick Rountree. *Federated Identity Primer*. Syngress, 2012.

[107] Michael L. Rustad and Thomas H. Koenig. Towards a global data privacy standard. *Florida Law Review*, 71, 2018.

[108] Nat Sakimura, John Bradley, Michael B. Jones, Breno de Medeiros, and Chuck Mortimore. Final: OpenID Connect Core 1.0 incorporating errata set 1. Technical report, The OpenID Foundation, 2014. `https://openid.net/specs/openid-connect-core-1_0.html`. [Accessed: 2019-04-22].

[109] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. Designing effective privacy notices and controls. *IEEE Internet Computing*, 21(3):70–77, 2017.

[110] Mark Scott. Estonians embrace life in a digital world. *The New York Times*, 2014. `https://www.nytimes.com/2014/10/09/business/international/estonians-embrace-life-in-a-digital-world.html`. [Accessed: 2019-06-25].

[111] Securefact Transation Services, Inc. Securefact - SIDni^{TM} (Secure Identification Network for Individuals). `https://www.securefact.com/sidni/`. [Accessed: 2019-02-07].

[112] Service Canada. The Social Insurance Number code of practice. Technical report, Service Canada, 2013. `https://www.canada.ca/en/employment-social-development/services/sin/reports/code-of-practice.html`. [Accessed: 2019-05-23].

[113] Slack. Slack's GDPR commitment. `https://slack.com/intl/en-ca/gdpr#`. [Accessed: 2019-06-22].

[114] Aaron Smith. Why Americans use social media. Technical report, Pew Research Center, 2011. `https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2011/Why-Americans-Use-Social-Media.pdf`. [Accessed: 2019-06-02].

[115] Aaron Smith and Monica Anderson. Social media use in 2018. Survey report, Pew Research Center, 2018. https://www.pewinternet.org/wp-content/uploads/sites/9/2018/02/PI_2018.03.01_Social-Media_FINAL.pdf. [Accessed: 2019-05-26].

[116] Standing Committee on Finance. Evidence. Transcript 60, House of Commons of Canada, 2012. https://www.ourcommons.ca/DocumentViewer/en/41-1/FINA/meeting-60/evidence. [Accessed: 2019-05-24]].

[117] Clare Sullivan. *Digital Identity: An Emergent Legal Concept*. The University of Adelaide Press, 2011.

[118] Latanya Sweeney. Simple demographics often identify people uniquely. http://dataprivacylab.org/projects/identifiability/. [Accessed: 2019-06-24], 2000.

[119] The European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119:1–88, 2016.

[120] The MITRE Corporation. CWE-308: Use of single-factor authentication, 2018. https://cwe.mitre.org/data/definitions/308.html. [Accessed: 2019-06-08].

[121] The OpenID Foundation. OpenID Foundation website. https://openid.net. [Accessed: 2019-06-10].

[122] The World Bank. 1.1 billion 'invisible' people without ID are priority for new High Level Advisory Council on Identification for Development, 2017. https://www.worldbank.org/en/news/press-release/2017/10/12/11-billion-invisible-people-without-id-are-priority-for-new-high-level-advisory-council-on-identification-for-development. [Accessed: 2019-06-23].

[123] Trans Union of Canada, Inc. Identity verification | TransUnion Canada. https://ca-pvw.transunion.com/product/identity-verification. [Accessed: 2019-02-07].

[124] Treasury Board of Canada, Secretariat. Digital Operations Strategic Plan: 2018–2022. Monograph, Treasury Board of Canada, Secretariat, 2018. http://publications.gc.ca/pub?id=9.866296&sl=0. [Accessed: 2019-06-26].

[125] United Nations. Universal Declaration of Human Rights, 1948. https://www.un.org/en/universal-declaration-human-rights/index.html. [Accessed: 2019-06-23].

[126] United States Social Security Administration. Requesting SSA's death information, 2006. https://www.ssa.gov/dataexchange/request_dmf.html. [Accessed: 2019-06-24].

[127] Tom Warren. Facebook starts notifying millions of users that their data was 'improperly obtained', 2018. https://www.theverge.com/2018/4/9/17214814/facebook-data-notification-cambridge-analytica. [Accessed: 2019-05-26].

[128] Harry Weber-Brown. Use of online activity as part of the identity verification. Technical report, Open Identity Exchange, 2016. https://oixuk.org/wp-content/uploads/2016/09/Useofonlineactivityaspartoftheidentityverification FINAL-1-1.pdf. [Accessed: 2019-06-26].

[129] Marilyn Domas White and Emily E. Marsh. Content analysis: A flexible methodology. *Library Trends*, 55(1):22–45, 2006.

[130] Alexis Wilke. Getting rid of deactivated Facebook friends | Made to Order Software Corporation. https://www.m2osw.com/delete-deactivated-facebook-friends, 2017. [Accessed: 2019-04-30].

[131] Phillip J. Windley. *Digital Identity*. O'Reilly Media Inc., 2005.

[132] Josh Wolford. Facebook including deactivated accounts in total friend count [updated]. https://www.webpronews.com/facebook-including-deactivated-accounts-in-total-friend-count/, 2012. [Accessed: 2019-04-30].

Appendix A

Materials for Information Discoverability Study

A.1 Recruitment Poster

Carleton UNIVERSITY

Participate in a study on the discoverability of personal information

To participate in this study, you must be:

- ✓ Active on at least one social media account
- ✓ At least 18 years old
- ✓ English-speaking

This is a 90-minute study. You will be asked to examine your social media accounts and search results about yourself for certain types of personal information. You will also participate in an interview on smartphone usage, your handling of personal belongings, and your handling of documents with your personal information.

Participants will be compensated with $25 (CAD) and related parking costs ($10) if applicable.

The ethics protocol for this project has been reviewed and cleared by the Carleton University Research Ethics Board (Study Clearance #108764). If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Please contact the researcher, Michael Lutaaya, for more details on this study at michael.lutaaya@carleton.ca

Michael Lutaaya
michael.lutaaya@carleton.ca

A.2 Consent Form

Carleton UNIVERSITY
Canada's Capital University

Study Clearance #108764

Title: What Types of Personally Identifiable Information Are Most Easily Discovered?

Funding Source: NSERC Discovery Grant

Date of ethics clearance: April 16, 2018

Ethics Clearance for the Collection of Data Expires: April 30, 2019

I, _____, choose to participate in a study on the discoverability of personally identifiable information (PII). This study aims to identify the types of PII that can be discovered most easily by an individual looking to commit identity fraud. **The researcher for this study is Michael Lutaaya in the School of Computer Science.** He is working under the supervision of Dr. Sonia Chiasson in the School of Computer Science.

This study will last 90 minutes and consist of four activities: (1) a demographic questionnaire; (2) an interview covering your attitudes and behaviours regarding the safekeeping of your belongings, smartphone usage, and documents with containing your personal information; (3) an exercise examining the types of your PII that can be found on social media; and (4) an exercise evaluating the types of your PII that can be found by looking yourself up on a search engine.

This study is done in collaboration with Bluink, an Ottawa-based company that specializes in software for identity and access management. We will use the findings to make recommendations on the types of PII that are most robust against being discovered by potential attackers. Bluink may use some of these recommendations to reduce fraud in a digital identity solution that they are developing.

To participate in this study, you must be:

- Active on at least one social media account
- At least 18 years old
- English-speaking

During your session, the researcher will take notes of their observations by typing on a computer. Additionally, with your consent, the interview portion of the study will be audio recorded to help with accurately capturing your comments. The recordings will be encrypted and will be deleted after one year. If you do not want

to be audio-recorded, you can still participate and the researcher will take notes of the interview instead.

All research data will be anonymized and stored on a password-protected computer that is only accessible by the research team. The results of the questionnaire, transcriptions of the interview, data recorded on the checklists during the social media and web search exercises, and the researcher's notes may be shared with Bluink; however, this data will already be anonymized and will include no PII whatsoever. Audio recordings of the interview will not be shared with Bluink.

Data collected during your session will include the anonymous username (e.g., P1, P2) assigned to you when you completed the demographic questionnaire. Any paper documents (e.g., consent forms) will be kept in a locked cabinet at Carleton University that is only accessible by the research team and will be shredded after one year.

Since this project deals with your personal information, there are some potential emotional risks to you if you are uncomfortable with the researcher viewing your personal information. To mitigate this risk, several precautions will be taken:

- The participant will complete a checklist indicating whether certain *types* of personally identifiable information were discovered (e.g., full name, email address). Neither the checklist nor any notes taken by the researcher will include any of your personally identifiable information.
- During the social media and web search exercises, you will operate the lab computer and/or the device that you bring to the session. If you do not feel comfortable following any of the steps given by the researcher, you have the right to decline.
- On the lab computer, we will use the browser's incognito mode to ensure that browser history, cookies, website data, and any information entered into forms is not saved. As an extra precaution, the researcher will clear the browser's data and ask you to witness this to be certain that no traces of your browsing activity are kept on the lab computer.
- During the interview, you have the right decline answering any of the questions.

You have the right to end your participation in the study at any time, for any reason, up until the end of your scheduled session. If you withdraw from the study before it ends, all information you have provided will be immediately destroyed.

As a token of appreciation, you will receive $25 in cash and, associated parking expenses ($10) if applicable. If you withdraw from the study, you will still be compensated the full amount.

Once the project is completed, electronic research data will be kept and potentially used for other research projects on this same topic. Consent forms will be shredded

after one year. Anonymized results may be used in publications or presentations.

If you would like a copy of the finished research project, you are invited to contact the researcher to request an electronic copy which will be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research, CUREB-B Clearance #108764. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Researcher contact information:
Michael Lutaaya
School of Computer Science
Carleton University
Email: michael.lutaaya@carleton.ca

Supervisor contact information:
Dr. Sonia Chiasson
School of Computer Science
Carleton University
Tel: (613) 520-2600 ext. 1656
Email: chiasson@scs.carleton.ca

Do you agree to be audio-recorded during the interview: ___Yes ___No

_____ _____
Signature of participant Date

_____ _____
Signature of researcher Date

This document has been printed on both sides of a single sheet of paper.
Please retain a copy of this document for your records.

A.3 Demographic Questionnaire

PII Discoverability — Demographic Questionnaire

Start of Block: Default Question Block

Q1 What is your gender?

- Male (1)
- Female (2)
- Other (3)
- Prefer not to say (4)

Q2 What is your age?

_____

Q3 What is the highest level of education you have completed?

- ○ Elementary school (1)
- ○ High school (2)
- ○ College (3)
- ○ Technical, trade school, or apprenticeship (4)
- ○ Undergraduate degree (Bachelor's) (5)
- ○ Post-graduate certificate or diploma (6)
- ○ Graduate degree or professional degree (7)
- ○ Other (8) _____
- ○ Prefer not to say (9)

Display This Question:
If 1 = 2

Q4 What is your occupation/field of study?

_____

Q21 What is your occupation?

- ○ Administrative Support (e.g.,secretary, assistant) (1)
- ○ Art, Writing, Journalism (e.g., author, reporter, sculptor) (4)
- ○ Business, Management and Financial (e.g., manager, accountant, banker) (5)
- ○ Education (e.g., teacher, professor) (6)
- ○ Legal (e.g., lawyer, law clerk) (7)
- ○ Medical (e.g., doctor, nurse, dentist) (8)
- ○ Science, Engineering, and IT professional (e.g., researcher, programmer, IT consultant) (9)
- ○ Service (e.g., retail clerk, server) (10)
- ○ Skilled Labor (e.g., electrician, plumber, carpenter) (11)
- ○ Student (specify your program of study) (12) _____
- ○ Unemployed (13)
- ○ Retired (14)
- ○ Other (specify) (15) _____
- ○ Decline to answer (16)

Q5 Which of the following devices do you own? Check all that apply.

- ☐ Laptop computer (1)
- ☐ Desktop computer (2)
- ☐ Tablet computer (e.g., iPad, Kindle Fire) (3)
- ☐ Smartphone (e.g., Android, iPhone, Windows Phone, Blackberry) (4)
- ☐ Smartwatch or fitness tracker (e.g., Fitbit, Apple Watch) (5)
- ☐ A device whose primary purpose is for reading e-books (e.g., Kindle, Nook) (6)
- ☐ Stationary game console (e.g., PlayStation, Xbox) (7)
- ☐ A portable device whose primary purpose is for playing video games (e.g., PlayStation Vita, Nintendo 3DS) (8)
- ☐ Other (9) _____

End of Block: Default Question Block

Start of Block: Block 5

Q23 The term social media refers to websites and apps that enable users to: (1) create and share content, or (2) communicate and interact with each other. For the social media accounts that you post to most frequently, answer the questions below.

Websites or apps that only have short-term content (e.g., Snapchat) or that only have messaging functionality (e.g., WhatsApp, Facebook Messenger) should not be included in your answers.

End of Block: Block 5

Start of Block: Most Used Social Media Account

Q2 What is your most used social media account?

- ○ Facebook (1)
- ○ Pinterest (2)
- ○ Instagram (3)
- ○ LinkedIn (4)
- ○ Twitter (5)
- ○ YouTube (6)
- ○ Tumblr (7)
- ○ Other (8) _____

Q3 What do you use this account for? Check all that apply.

- ☐ To stay in touch with friends and family (1)
- ☐ To follow public figures (e.g., celebrities, athletes, politicians) (2)
- ☐ To connect with other people with similar interests (3)
- ☐ To find romantic partners (4)
- ☐ To find news and information (5)
- ☐ To follow political or social issues (6)
- ☐ For professional networking or job seeking (7)
- ☐ To express myself (8)
- ☐ To relieve boredom or to find entertaining links and videos (9)
- ☐ Other (10) _____

Q4 On what devices do you use this account?

- [ ] Laptop computer (1)
- [ ] Desktop computer (2)
- [ ] Tablet computer (e.g., iPad, Kindle Fire) (3)
- [ ] Smartphone (e.g., Android, iPhone, Windows Phone, Blackberry) (4)
- [ ] Smartwatch or fitness tracker (e.g., Fitbit, Apple Watch) (5)
- [ ] A device whose primary purpose is for reading e-books (e.g., Kindle, Nook) (6)
- [ ] Stationary game console (e.g., PlayStation, Xbox) (7)
- [ ] A portable device whose primary purpose is for playing video games (e.g., PlayStation Vita, Nintendo 3DS) (8)
- [ ] Other (9) _____

Q5 How often do you use this account?

▼ Several times a day (1) ... Every few weeks or less often (4)

End of Block: Most Used Social Media Account

Start of Block: Second-Most Used Social Media Account

Q2 What is your second-most used social media account?

- ○ Facebook (1)
- ○ Pinterest (2)
- ○ Instagram (3)
- ○ LinkedIn (4)
- ○ Twitter (5)
- ○ YouTube (6)
- ○ Tumblr (7)
- ○ Other (8) _____

Q3 What do you use this account for? Check all that apply.

- ☐ To stay in touch with friends and family (1)
- ☐ To follow public figures (e.g., celebrities, athletes, politicians) (2)
- ☐ To connect with other people with similar interests (3)
- ☐ To find romantic partners (4)
- ☐ To find news and information (5)
- ☐ To follow political or social issues (6)
- ☐ For professional networking or job seeking (7)
- ☐ To express myself (8)
- ☐ To relieve boredom or to find entertaining links and videos (9)
- ☐ Other (10) _____

Q4 On what devices do you use this account?

- [ ] Laptop computer (1)
- [ ] Desktop computer (2)
- [ ] Tablet computer (e.g., iPad, Kindle Fire) (3)
- [ ] Smartphone (e.g., Android, iPhone, Windows Phone, Blackberry) (4)
- [ ] Smartwatch or fitness tracker (e.g., Fitbit, Apple Watch) (5)
- [ ] A device whose primary purpose is for reading e-books (e.g., Kindle, Nook) (6)
- [ ] Stationary game console (e.g., PlayStation, Xbox) (7)
- [ ] A portable device whose primary purpose is for playing video games (e.g., PlayStation Vita, Nintendo 3DS) (8)
- [ ] Other (9) _____

Q5 How often do you use this account?

▼ Several times a day (1) ... Every few weeks or less often (4)

End of Block: Second-Most Used Social Media Account

Start of Block: Third-Most Used Social Media Account

Q2 What is your third-most used social media account?

- ○ Facebook (1)
- ○ Pinterest (2)
- ○ Instagram (3)
- ○ LinkedIn (4)
- ○ Twitter (5)
- ○ YouTube (6)
- ○ Tumblr (7)
- ○ Other (8) _____

Q3 What do you use this account for? Check all that apply.

- ☐ To stay in touch with friends and family (1)
- ☐ To follow public figures (e.g., celebrities, athletes, politicians) (2)
- ☐ To connect with other people with similar interests (3)
- ☐ To find romantic partners (4)
- ☐ To find news and information (5)
- ☐ To follow political or social issues (6)
- ☐ For professional networking or job seeking (7)
- ☐ To express myself (8)
- ☐ To relieve boredom or to find entertaining links and videos (9)
- ☐ Other (10) _____

Q4 On what devices do you use this account?

- [ ] Laptop computer (1)
- [ ] Desktop computer (2)
- [ ] Tablet computer (e.g., iPad, Kindle Fire) (3)
- [ ] Smartphone (e.g., Android, iPhone, Windows Phone, Blackberry) (4)
- [ ] Smartwatch or fitness tracker (e.g., Fitbit, Apple Watch) (5)
- [ ] A device whose primary purpose is for reading e-books (e.g., Kindle, Nook) (6)
- [ ] Stationary game console (e.g., PlayStation, Xbox) (7)
- [ ] A portable device whose primary purpose is for playing video games (e.g., PlayStation Vita, Nintendo 3DS) (8)
- [ ] Other (9) _____

Q5 How often do you use this account?

▼ Several times a day (1) ... Every few weeks or less often (4)

End of Block: Third-Most Used Social Media Account

Start of Block: Block 1

Q7 If you have a smartphone, check any of the following that apply:

- ☐ My smartphone has a PIN or is password-protected (1)
- ☐ I have enabled fingerprint unlocking on my smartphone (2)
- ☐ I have enabled face unlocking only my smartphone (3)
- ☐ My smartphone is “jailbroken” or “rooted” (4)
- ☐ If I haven’t been using it for a few minutes, my phone locks itself. (5)

Q8 On a scale of 1 to 5, how concerned are you about your online privacy?

- ○ 1 (Not concerned at all) (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (Extremely concerned) (5)

End of Block: Block 1

A.4 Interview Themes

- **Possessions**
  - What the participant typically carries (e.g., wallet, backpack, briefcase, purse) and the typical contents of that item
  - Practices for safeguarding wallet/purse/briefcase while at work/school
  - Security of the participant's home mailbox and who can access it
- **Important Documents**
  - Who can access the participant's sensitive/personal documents or could obtain the information written on these documents (e.g., birth certificate, passport, health card, driver's license, passport, SIN)
  - Whether the participant destroy documents with personal/sensitive information before discarding them (e.g., medical records, financial documents, bills, envelopes with your address, etc.)
  - Circumstances where they have presented documents to someone else or an organization (e.g., health card, birth certificate, passport, tax records)
- **Technology**
  - What types of authentication the participant uses on their devices (e.g., password on laptop, fingerprints on smartphone)
  - Whether someone else could unlock their smartphone (e.g., this person knows the passcode or has a registered fingerprint on the device)
  - Whether the participant conducts certain activities on their smartphone (e.g., financial management, password management, photo storage)
  - Whether the participant has ever had their credentials or phone stolen and what happened afterwards
  - Whether the participant has used their smartphone as a digital wallet (e.g., Apple Pay, Google Pay)

– If the government released an app where you would enter your personal information in exchange for a digital ID on your smartphone that can used instead of other forms of official ID, how likely would you be to use it? Why?

A.5 General Questions (Social Media)

Appendix F — Session Forms and Interview Themes

General Questions (Social Media)

Most Used Social Media Account

The researcher will write down the participant's anonymous identifier (e.g., P1): _____

The researcher will write down the name of the social media app or website (e.g., "Facebook" or "LinkedIn") here: _____

Please navigate to your account's profile page. If you are not sure how to do this, please let me know and I can assist you.

- How many friends/followers/subscribers do you have? _____

Depending on the app or website being analyzed, some of these questions may be omitted.

- Please navigate to the settings page for your account and find the privacy section. _____
- When did you last look at the privacy settings for this account: within the last 7 days, within the last 30 days, within the last 12 months, more than a year ago, when your account was first created, never, unsure/can't remember, prefer not to disclose? _____
- Who is able to search for your [website or app name] account? _____
- Who is able to view the contents of your profile on [website or app name]? _____
- Could someone find you on [website or app name] by using your email? _____
- Could someone find you on [website or app name] by using your phone number? _____
- Do you have restrictions on what can be seen by particular friends or can all of your friends see the same information? _____
- [website or app name] gives you the ability to grant third-party apps or websites access to some of the information on your account. What apps or websites have

_____

you granted permission to? _____

Second-Most Used Social Media Account

The researcher will write down the participant's anonymous identifier (e.g., P1):
_____

The researcher will write down the name of the social media app or website (e.g., "Facebook" or "LinkedIn") here: _____

Please navigate to your account's profile page. If you are not sure how to do this, please let me know and I can assist you.

- How many friends/followers/subscribers do you have?
  _____

Depending on the app or website being analyzed, some of these questions may be omitted.

- Please navigate to the settings page for your account and find the privacy section. _____
- When did you last look at the privacy settings for this account: within the last 7 days, within the last 30 days, within the last 12 months, more than a year ago, when your account was first created, never, unsure/can't remember, prefer not to disclose? _____
- Who is able to search for your [website or app name] account?
  _____
- Who is able to view the contents of your profile on [website or app name]?
  _____
- Could someone find you on [website or app name] by using your email?
  _____
- Could someone find you on [website or app name] by using your phone number?
  _____
- Do you have restrictions on what can be seen by particular friends or can all of your friends see the same information? _____
- [website or app name] gives you the ability to grant third-party apps or websites access to some of the information on your account. What apps or websites have you granted permission to? _____

Third-Most Used Social Media Account

The researcher will write down the participant's anonymous identifier (e.g., P1): _____

The researcher will write down the name of the social media app or website (e.g., "Facebook" or "LinkedIn") here: _____

Please navigate to your account's profile page. If you are not sure how to do this, please let me know and I can assist you.

- How many friends/followers/subscribers do you have? _____

Depending on the app or website being analyzed, some of these questions may be omitted.

- Please navigate to the settings page for your account and find the privacy section. _____
- When did you last look at the privacy settings for this account: within the last 7 days, within the last 30 days, within the last 12 months, more than a year ago, when your account was first created, never, unsure/can't remember, prefer not to disclose? _____
- Who is able to search for your [website or app name] account? _____
- Who is able to view the contents of your profile on [website or app name]? _____
- Could someone find you on [website or app name] by using your email? _____
- Could someone find you on [website or app name] by using your phone number? _____
- Do you have restrictions on what can be seen by particular friends or can all of your friends see the same information? _____
- [website or app name] gives you the ability to grant third-party apps or websites access to some of the information on your account. What apps or websites have you granted permission to? _____

A.6 Social Media Checklist

PII Checklist for Social Media Exercise			
Participant's Anonymous Identifier			
Personally Identifiable Information	**Most Used Social Media Account**	**Second-Most Used Social Media Account**	**Third-Most Used Social Media Account**
First Name			
Middle Names			
Last Name			
Gender			
Street Address (current)			
City (current)			
Province (current)			
Postal Code (current)			
Is your current address different from the place where you have lived the longest? (If participant replies yes, check for the next four PII elements.)	**Circle one: YES / NO**		
Street Address (where lived the longest)			
City (where lived the longest)			
Province (where lived the longest)			
Postal Code (where lived the longest)			
Email Address			
Phone Number (Home)			
Phone Number (Mobile)			
Place of Birth			

Nationality			
Year of Birth			
Month of Birth			
Day of Birth			
Clearly Identifiable Photo(s) of You			
Marital Status			
Employer or School			
Job Title			
License Plate Number			
Car Make			
Car Model			
Car Year			

A.7 Search Queries

- First and last names (in quotes); full name (in quotes) (e.g., `"jane doe" OR "jane emily doe"`)

- Name; current city; city resided in the longest (e.g., `jane doe ottawa OR toronto`)

- Name; current or most recent employer (e.g., `jane doe air canada`)

- Name; current or most recent school (e.g., `jane doe carleton university`)

- Primary email address (in quotes), typical username (in quotes) (e.g., `"jane.doe@gmail.com" OR "janedoe123"`)

- Your phone number; your longest-held phone number (e.g., `6135551234 OR 4165551234`)

A.8 Search Checklist

PII Checklist for Search Results Exercise						
Participant's Anonymous Identifier						
	Search Term #					
Personally Identifiable Information	**1**	**2**	**3**	**4**	**5**	**6**
How many results are on the first page?						
How many results on the first page are about you?						
First Name						
Middle Names						
Last Name						
Gender						
Street Address (current)						
City (current)						
Province (current)						
Postal Code (current)						
Is your current address different from the place where you have lived the longest? (If participant replies yes, check for the next four PII elements.)	Circle one: YES / NO					
Street Address (where lived the longest)						
City (where lived the longest)						
Province (where lived the longest)						
Postal Code (where lived the longest)						
Email Address						
Phone Number (Home)						
Phone Number (Mobile)						

Place of Birth						
Nationality						
Year of Birth						
Month of Birth						
Day of Birth						
Clearly Identifiable Photo(s) of You						
Marital Status						
Employer or School						
Job Title						
License Plate Number						
Car Make						
Car Model						

A.9 Debriefing

Carleton UNIVERSITY
Canada's Capital University

What are we trying to learn in this research?

This study explores the types of personally identifiable information (PII) that can be obtained by an attacker when they target an individual. The interview you participated in as well as the analysis of your social media accounts and web search results will help us assess the types of personally identifiable information that are obtained most easily.

None of your personal information was recorded in the researcher's notes and the researcher cleared browser data (i.e., browsing history, cookies, caches, and other data) on the lab computer. If you used a mobile device that was provided by the CHORUS Lab, the researcher deleted apps that you used (and therefore, any data associated with those apps). The researcher must have ensured that you witnessed this deletion process.

Why is this important to scientists or the general public?

The results of this study may be used to inform the design of a digital identity solution. The eventual goal is to develop an identity proofing measure that will reject individuals registering for a digital identity with someone else's PII. Identity proofing is an important countermeasure used by many organizations to guard against identity fraud. Especially in interactions requiring high levels of assurance, identity proofing is instrumental in preventing damages to individuals as well as organizations. As digital credentials increasingly replace physical identification documents in everyday scenarios, it is expected that the results of this research will make a positive contribution to identity management and computer security. We will share anonymized results with Bluink, an Ottawa-based company that specializes in software for identity and access

management.

What are our hypotheses and predictions?

We predict that the personally identifiable information of frequent social media users will be discovered most easily. Additionally, we predict that the behaviours and attitudes of participants regarding the safekeeping of their belongings, smartphone usage, and documents with their personal information will depend on demographic factors such as age, occupation, and education.

Where can I learn more?

If you are interested in learning more about how to protect yourself against identity fraud or what steps victims should take, you may consult resources from the Government of Ontario (`https://www.ontario.ca/page/how-avoid-or-recover-identity-theft`) or the Royal Canadian Mounted Police (`http://www.rcmp-grc.gc.ca/scams-fraudes/id-theft-vol-eng.htm`)

What if I have questions later?

If you have any remaining concerns, questions, or comments about the experiment, please feel free to contact
Michael Lutaaya (Principal Investigator), at: `michael.lutaaya@carleton.ca`,
Dr. Sonia Chiasson (Faculty Sponsor), at: `chiasson@scs.carleton.ca` (613-520-2600, ext. 1656).

If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at `ethics@carleton.ca`).

Thank you for participating in this research!

This research was cleared by Carleton University Research Board-B Clearance #108764.

Appendix B

Materials for Usability Study

B.1 Recruitment Poster

Carleton UNIVERSITY

Participants needed to study a smartphone-based digital identity app

To be eligible, you must:

- ✓ be at least 18 years of age,
- ✓ speak English fluently, and
- ✓ own an iPhone or Android smartphone

We need participants for a study entitled "Usability Evaluation of a Smartphone-Based Identity Solution."

The study will take approximately 90 minutes. You will be asked to:

- try a smartphone-based digital identity app,
- complete questionnaires, and
- participate in an interview focusing on identity documents, your experience with using a digital identity, and your willingness to use the app.

You will not be required to use your own identity documents.

Participants will receive $25 and related parking expenses ($10) if applicable.

The ethics protocol for this research has been cleared by the Carleton University Research Ethics Board-B (CUREB-B Clearance #109170). If you have any ethical concerns with the study, please contact CUREB-B at 613-520-2600 ext. 4085 or ethics@carleton.ca

If interested, please email Michael Lutaaya at michael.lutaaya@carleton.ca

Study: Smartphone-based Identity App
michael.lutaaya@carleton.ca

B.2 Screening Questionnaire Materials

B.2.1 Consent Form

Title: Usability Evaluation of a Smartphone-Based Identity Solution
Funding Source: NSERC Discovery Grant
Date of ethics clearance: August 31, 2018
Ethics clearance: CUREB-B Clearance #109170 (expires July 31, 2019)

This is a study on the usability of a smartphone-based identity solution known as eID-Me. This study aims to (1) establish which identity documents individuals have access to and which are best suited for identity verification, (2) evaluate the usability of eID-Me, and (3) collect feedback about the concept of digital identities. **The lead researcher for this study is Michael Lutaaya.** He is working under the supervision of Dr. Sonia Chiasson in the School of Computer Science.

This study is done in collaboration with Bluink, an Ottawa-based company that specializes in software for identity and access management. We will use the findings to suggest changes to Bluink that could improve eID-Me's usability. An employee from Bluink may observe some of the sessions to better understand any usability issues with the app. You will be asked whether you agree to have an observer from Bluink present. You may still participate in the study if you choose no.

The study starts with a 2-minute questionnaire to determine whether you qualify for the in-person session. You will only be contacted about the in-person session if you qualify. As such, your responses to this initial questionnaire are associated with your email address, but not to your data from the session if you participate.

While the risk involved with completing this questionnaire is expected to be minimal, measures are in place to protect your identity. Qualtrics, the company running the online questionnaire, stores collected data on secured servers in a secured data centre based in Canada. Once copied from the survey-hosting server, the data will be deleted from the server. It will then be stored on the researcher's encrypted, password-protected computer. Once the responses have been analyzed to determine those who qualify for the in-person session, the questionnaire data will be destroyed. The data will only be accessible by the researchers directly involved in this project.

You have the right to end your participation in the survey at any time, for any reason, up until you hit the "submit" button. You can withdraw by exiting the survey at any time before completing it. If you withdraw from the study, all information you provided will be immediately destroyed.

If you would like a copy of the finished research project, you are invited to contact the lead researcher (Michael Lutaaya) to request an electronic copy which will be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Researchers:

The researchers involved in this research are all from the School of Computer Science at Carleton University and members of the Carleton Human Oriented Research in Usable Security (CHORUS) Lab.

Michael Lutaaya, Masters student in Computer Science

michael.lutaaya@carleton.ca

Khadija Baig, Undergraduate student in Computer Science
KhadijaBaig@cmail.carleton.ca

Research supervisor contact information:

Sonia Chiasson
Associate Professor
School of Computer Science
Carleton University
Tel: (613) 520-2600 ext. 1656
chiasson@scs.carleton.ca

By clicking the "→" button below, you consent to complete this screening questionnaire as described above.

B.2.2 Questionnaire

1. Email address we should use to contact you to set up a session if selected to participate: [*text field*]

2. What is your gender?

   - Male
   - Female
   - Other
   - Prefer not to disclose

3. What is your age?

   - Under 18
   - 18 – 24

- 25 – 34
- 35 – 44
- 45 – 54
- 55 – 64
- 65 – 74
- 75 – 84
- 85 or older

4. Choose either the level of education for which you are currently enrolled or the highest level of education you have completed.

    - Elementary school
    - High school
    - College
    - Technical, trade school, or apprenticeship
    - Undergraduate degree (Bachelor's)
    - Post-graduate certificate or diploma
    - Graduate degree or professional degree
    - Other: [*text field*]
    - Prefer not to disclose

5. What is your occupation?

    - Student (specify your program of study): [*text field*]
    - Unemployed
    - Retired
    - Administrative Support (e.g., secretary, assistant)
    - Art, Writing, Journalism (e.g., author, reporter, sculptor)
    - Business, Management, and Financial (e.g., manager, accountant, banker)

- Education (e.g., teacher, professor)
- Legal (e.g., lawyer, law clerk)
- Medical (e.g., doctor, nurse, dentist) Science, Engineering, and IT Professional (e.g., researcher, programmer, IT consultant)
- Service (e.g., retail clerk, server)
- Skilled Labour (e.g., electrician, plumber, carpenter)
- Other (specify): [*text field*]
- Prefer not to disclose

6. Do you own a cellphone?

   - Yes
   - No

7. [If answered "Yes" to "Do you own a cellphone?"]: Which of the following best describes the type of cell phone that you have?

   - Android
   - iPhone
   - Windows
   - BlackBerry
   - Symbian
   - Other smartphone
   - Other non-smartphone
   - Not sure
   - Prefer not to disclose

8. If you are selected to participate, do you agree to have an employee from Bluink attend your session in order to observe how you use the app and the issues that you may encounter?

- Yes
- No

B.3 Session Materials

B.3.1 Consent Form

Carleton UNIVERSITY
Canada's Capital University

CUREB-B Clearance #109170

Consent Form (Session)

Title: Usability Evaluation of a Smartphone-Based Identity Solution

Funding Source: NSERC Discovery Grant

Date of ethics clearance: July 31, 2018

Ethics Clearance for the Collection of Data Expires: July 31, 2019

I, _____, choose to participate in a study on the usability of a smartphone-based identity solution known as eID-Me. This study aims to (1) establish which identity documents individuals have access to and which are best suited for identity verification, (2) evaluate the usability of eID-Me, and (3) collect feedback about the concept of digital identities. **The lead researcher for this study is Michael Lutaaya.** He is working under the supervision of Dr. Sonia Chiasson in the School of Computer Science.

This study will last 90 minutes and you will complete four activities: (1) answer a pre-test questionnaire; (2) try the eID-Me app to create and use a digital identity (3) answer a post-test questionnaire; and (4) participate in an interview focusing on your use and opinion of eID-Me and digital identities in general.

This study is done in collaboration with Bluink, a for-profit company based in Ottawa that specializes in software for identity and access management. eID-Me is a product of Bluink and we will use the findings of this study to suggest changes to Bluink that could improve eID-Me's usability.

If you provided consent, an employee from Bluink may sit in on your session to observe how you use the app and to see the usability issues that you discover.

To participate in this study, you must meet the following criteria:
- Be at least 18 years old
- Speak English fluently
- Own an iPhone or Android smartphone

During your session, the app generates a log file for the researcher that describes how you used the app. The researcher will also take notes of their observations by typing on a computer. Additionally, with your consent, the session will be audio-recorded to help with accurately capturing your comments. The recordings will be

This document has been printed on both sides of a single sheet of paper.
Please retain a copy of this document for your records.

encrypted and will be deleted after one year. If you do not want to be audio-recorded, you can still participate and the researcher will type notes instead.

Privacy: Research data will use a pseudonym assigned to you (e.g., P1, P2, P3) but no master list will be kept that would allow your pseudonym to be directly linked to you. All data will be stored on an encrypted, password-protected computer that is only accessible by the researchers directly involved in the study. Any paper documents (e.g., consent forms, compensation receipts) will be kept in a locked cabinet at Carleton University that is only accessible by the research team.

Your data will be stored and protected by Qualtrics on Toronto-based servers but may be disclosed via a court order or data breach. The participant's pseudonym will be used to link together their responses but no IP addresses will be recorded. Once copied from the survey-hosting server, the data will be deleted from the server. It will then be stored on the researcher's encrypted, password-protected computer.

Bluink's servers perform identity verification on the information that you submit through eID-Me, however, the researcher will provide mock identity documents for you to use. These documents do not contain your personal information. As such, none of your personal information is sent to servers and the digital identity that will be created will be based on the information contained on the mock documents.

Withdrawal: You have the right to end your participation in the study at any time, for any reason, up until the end of your scheduled session. If you withdraw from the study before it ends, all information you have provided will be immediately destroyed. Additionally, during any of the questionnaires or interviews, you have the right to decline answering any of the questions.

Compensation: As a token of appreciation, you will receive $25 in cash and, associated parking expenses ($10) if applicable. If you withdraw from the study, you will still be compensated the full amount.

Data use: Once the project is completed, electronic research data will be kept and potentially used for other research projects on this same topic. Consent forms will be shredded after one year. Results may be used in publications or presentations. Any data that is retained or published will use pseudonyms only and will not be attributable to you.
The results of the questionnaire, transcriptions of the session, log files, and the researcher's notes may be shared with Bluink. This data will already be using the pseudonyms described above. Audio recordings of the session will not be shared with Bluink.

Follow up: If you are interested in our future studies on digital identity, you can provide an email address at the end of the session. We will email you information

about these studies when we are looking for participants. Your email address will not be provided to Bluink.

If you would like a copy of the finished research project, you are invited to contact the researcher to request an electronic copy be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research, CUREB-B Clearance #109170. Should you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or by email at ethics@carleton.ca). For all other questions about the study, please contact the researcher.

Researcher contact information:

Michael Lutaaya
Master's student
michael.lutaaya@carleton.ca

Khadija Baig
Undergraduate assistant
KhadijaBaig@cmail.carleton.ca

Supervisor contact information:

Dr. Sonia Chiasson
School of Computer Science
Carleton University
(613) 520-2600 ext. 1656
chiasson@scs.carleton.ca

Do you agree to be audio-recorded during the session: ___Yes ___No

Do you agree to have an observer from Bluink present: ___Yes ___No

I will not disclose or discuss with anyone the proprietary information that I am granted access to or may encounter. This includes, but is not limited to, Bluink's product designs, product plans, software, technology, and know-how process, interactive features, interfaces, inventions, and other intellectual property. I also acknowledge that I may be quoted (no images, audio, or video) in Bluink's marketing materials and I will be credited anonymously with an identifier such as "User Testing Participant." _____ INITIALS

_____ _____
Signature of participant Date

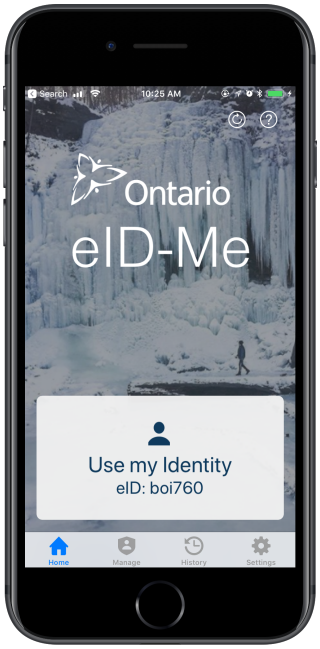
_____ _____
Signature of researcher Date

This document has been printed on both sides of a single sheet of paper.
Please retain a copy of this document for your records.

B.3.2 Pamphlet

How to Use eID-Me

Ontario digital identity on your smartphone



① Get Your Registration Code

Check your email for your registration code.

② Add Your Identity Info

Verify your identity by adding documents like a driver's license or a passport.

③ Use Your Smartphone Instead of Identity Cards

Show your digital identity's QR code in person. Approve or deny requests for your personal information and choose what gets revealed.

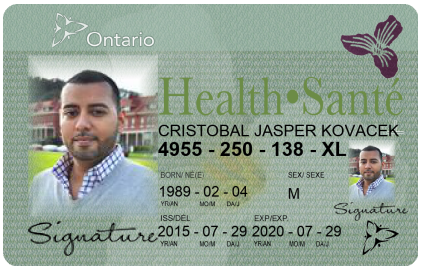
④ Sign in to Online Services with Your Smartphone

Enter your eID (username) on a supported website. Approve the sign-in from your smartphone. No passwords needed.

B.3.3 Identity Documents Provided to Participants

(a) Ontario driver's license (front)

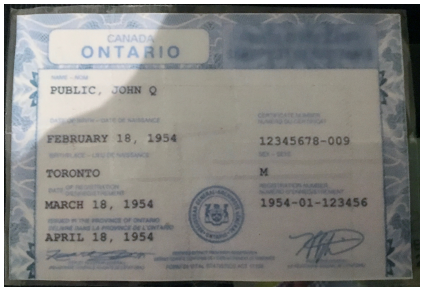
(b) Ontario driver's license (back)



(c) Ontario health card (front)

(d) Ontario health card (back)

(e) Social Insurance Number (SIN) card



(f) Ontario birth certificate

(g) Canadian passport

Figure B.1: Sample identity documents provided to participants

B.3.4 Pre-Test Questionnaire

1. Please enter your Participant ID: [*text field*]
2. What is your gender?
    - Male
    - Female
    - Other
    - Prefer not to disclose
3. What is your age?: [*text field*]
4. Choose either the level of education for which you are currently enrolled or the highest level of education you have completed.
    - Elementary school
    - High school
    - College
    - Technical, trade school, or apprenticeship
    - Undergraduate degree (Bachelor's)
    - Post-graduate certificate or diploma
    - Graduate degree or professional degree
    - Other: [*text field*]
    - Prefer not to disclose
5. What is your occupation?
    - Administrative Support (e.g., secretary, assistant)
    - Art, Writing, Journalism (e.g., author, reporter, sculptor)
    - Business, Management, and Financial (e.g., manager, accountant, banker)
    - Education (e.g., teacher, professor)

- Legal (e.g., lawyer, law clerk)
- Medical (e.g., doctor, nurse, dentist) Science, Engineering, and IT Professional (e.g., researcher, programmer, IT consultant)
- Service (e.g., retail clerk, server)
- Skilled Labour (e.g., electrician, plumber, carpenter)
- Student (specify your program of study): [*text field*]
- Unemployed
- Retired
- Other (specify): [*text field*]
- Prefer not to disclose

6. When you are awake, what percentage of the time is your phone within reach and operable?: [*text field*]

7. Which of the following best describes the type of cell phone that you have?

   - Android
   - iPhone
   - Not sure

8. Which of the following methods have you used to make a mobile payment with your smartphone? Check all that apply.

   □ Holding my smartphone near a payment reader to pay for a good or service (e.g., Apple Pay)

   □ Using my smartphone's camera to scan a barcode in order to pay for a good or service

   □ Displaying a barcode on my smartphone's screen for scanning in order to pay for a good or service

   □ Using the card information stored on my smartphone to pay for a good or service inside of an app

- □ Using the card information stored on my smartphone to pay for a good or service on a website
- □ Completing a person-to-person payment through a mobile app (e.g., PayPal, Scotiabank app, e-transfer)
- □ Other
- □ Never used any

9. [If did **not** answer "Never used any" for "Which of the following methods have you used to make a mobile payment with your smartphone?"]: How often do you use your smartphone for mobile payments?
   - Every day
   - Several times a week
   - Once a week
   - 1–3 times a month
   - Less than once a month
10. [If answered "Never used any" for "Which of the following methods have you used to make a mobile payment with your smartphone?"]: Why don't you use your smartphone for mobile payments? Check all that apply.
   - I prefer using cash
   - I prefer using my physical debit or credit card
   - I didn't know about this functionality
   - My smartphone doesn't support this
   - My bank doesn't support this
   - I'm concerned about security / privacy / phone being stolen.
   - Other
11. To prevent fraud, eID-Me will rely on personal information that can be verified with the government or third parties. Read each of the statements below and check all that apply.

- □ I own a valid Ontario driver's license.
- □ I own a valid Ontario health card (OHIP).
- □ I own a valid Canadian passport.
- □ I own a valid Ontario Photo Card. (An Ontario Photo Card is government-issued identification for Ontario residents who don't have a driver's license.)
- □ I have a valid Social Insurance Number (SIN).
- □ I have a birth certificate from a Canadian province or territory.
- □ I have a valid Certificate of Indian Status.
- □ The bill from my cell phone company (e.g., Rogers, Bell, Telus) is in my name.
- □ The address on my cell phone bill is different than the address on my driver's license.

12. Which of the following best describes how often you are asked for government-issued ID (e.g., driver's license, Ontario Photo Card, passport, etc.)?

    - Daily
    - A few times per week
    - A few times per month
    - A few times a year
    - Rarely or never

13. How likely are you to use a digital identity in the following situations?

What is digital identity?
A *digital identity* is the electronic equivalent of traditional identity documents such as driver's licenses or passports.

Digital identities can be used in person (e.g., proving age to buy alcohol) or online (e.g., signing in to a website instead of entering a username and password).

	Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
To sign in to a website	○	○	○	○	○
To prove your identity to law enforcement	○	○	○	○	○
To prove your identity to a business requiring proof of age (e.g., entering a bar or casino, buying alcohol)	○	○	○	○	○
To access health care services	○	○	○	○	○

14. How likely are you to use a government-approved smartphone app as your ID instead of traditional ID cards?

    - Extremely likely
    - Somewhat likely
    - Neither likely nor unlikely
    - Somewhat unlikely
    - Extremely unlikely

15. Have you ever used password management software on your smartphone? (e.g., LastPass, 1Password, iCloud Keychain, storing passwords with Google Chrome)

    - Yes

- No

B.3.5 Post-Test Questionnaire

1. Please enter your Participant ID: [*text field*]

2. After having used the app, how likely are you to use a government-approved smartphone app as your ID instead of traditional ID cards?

   - Extremely likely
   - Somewhat likely
   - Neither likely nor unlikely
   - Somewhat unlikely
   - Extremely unlikely

3. After having used the app, how likely are you to use a digital identity in the following situations?

   What is digital identity?
   A *digital identity* is the electronic equivalent of traditional identity documents such as driver's licenses or passports.

   Digital identities can be used in person (e.g., proving age to buy alcohol) or online (e.g., signing in to a website instead of entering a username and password).

	Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
To sign in to a website	○	○	○	○	○

	Extremely unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Extremely likely
To prove your identity to law enforcement	○	○	○	○	○
To prove your identity to a business requiring proof of age (e.g., entering a bar or casino, buying alcohol)	○	○	○	○	○
To access health care services	○	○	○	○	○

4. For each statement, please select the option that best describes how you feel about the statement.

	Very difficult	Difficult	Neither difficult nor easy	Easy	very easy
The language used in the app was _____ to understand.	○	○	○	○	○
I think that the app was _____ to use.	○	○	○	○	○
I think that the app was _____ to navigate.	○	○	○	○	○
I think using the app to sign in to a website was _____.	○	○	○	○	○
I think using the app to prove my identity in person (e.g., at a hospital) was _____.	○	○	○	○	○

5. For each statement, please select the option that best describes how you feel about the statement.

	Very unappealing	Unappealing	Neither unappealing nor appealing	Appealing	Very appealing
Visually, the app was _____.	○	○	○	○	○

B.3.6 Interview Questions

Smartphone Questions

- Several times a year, updates to Android/iOS (the operating system that's running on your smartphone) are made available. Note that this is different from updates to the apps on your phone. How often do you update your phone's operating system? As soon as you see that an update is available? When it's convenient for you? Never?
- Do you have your smartphone with you? Can you check what version of Android/iOS is installed on your smartphone?

Digital ID Questions

- In general, what are your thoughts about using a form of digital ID instead of traditional ID cards?
- What do you see as the advantage of using digital IDs?
- Do you have any concerns about using digital IDs?
- Are there instances where you would be particularly comfortable or uncomfortable in using digital IDs?

- Who should be responsible for managing a digital ID system?
- From your understanding, how does a digital ID system work? Who has access to your information? When? Who is determining whether you can access a service?

App-Specific Questions

General

- Is there anything that you would add or change about the app?
- Did you find anything about the app confusing?
- Are there other things that you would like this app to do?
- Is there anything specific that you liked about the app?
- Is there anything specific that you did not like about the app?
- If this app was backed by the Ontario government, how likely would you be to use it?
- If this app was developed a for-profit company, how likely would you be to use it?

Registration

- Now that you have used the app, would you have picked different documents to use for registration? Why?
- Are there documents that weren't listed in the app that you would have liked to use during registration?
- What did you think of the process for submitting the identity documents into the app? Did you find anything about the scanning process difficult?

Usage

- What did you think about the using your identity to sign in to the website? Would you ever use this feature?
- What did you think about using your identity and the QR code to prove your age? Would you ever use this feature?
- When you were using your identity to sign in to the website and to prove your age, you were able to see what information was being requested and when some of the information was optional, you could choose whether or not to share that information. What were your thoughts on how this worked?

B.4 eID-Me: Architectural Overview

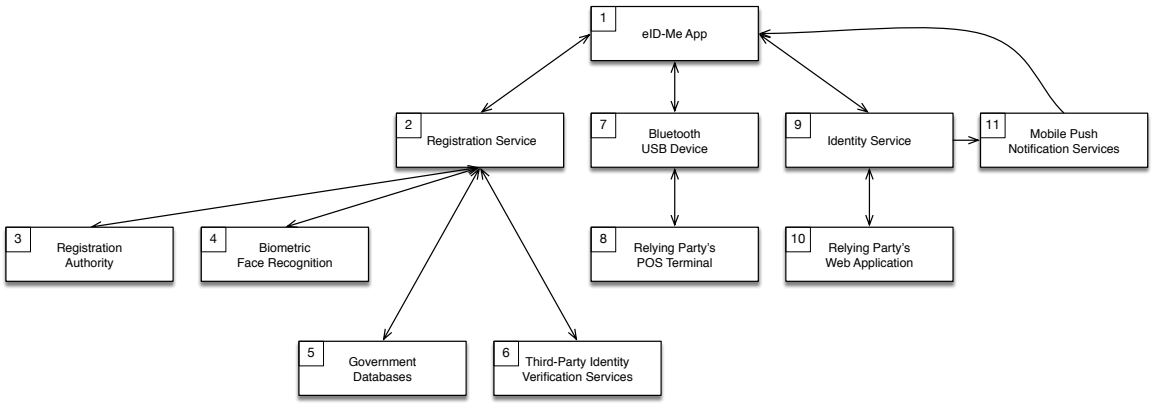


Figure B.2: Architecture diagram for eID-Me. Adapted from internal documents provided by Bluink.

Table B.5: Descriptions of system components for eID-Me. Adapted from internal documents provided by Bluink.

#	Component	Description
1	eID-Me App	Smartphone app running on the user's device.
2	Registration Service	Server operated by Bluink that receives proofs of identity from the user and, upon verification, sends the certificate to the user's device. This server also corroborates submitted identity claims with government databases and identify verification services to compute a strength of identity proofing score.
3	Registration Authority	Server operated by Bluink dedicated to the signing and issuing of digital certificates. The Registration Service communicates with this server to request a digital certificate with the user's identity claims.

Table B.5 (*continued*)

#	**Component**	**Description**
4	Biometric Face Recognition	Server operated by Bluink dedicated to analyzing submitted images of faces for the purpose of verifying the identity of the person in the image. The Registration Service communicates with this server to obtain information about the credibility of submitted images.
5	Government Databases	Records maintained by the government departments (e.g., Ministry of Transportation of Ontario) that can be used to corroborate submitted identity claims. The Registration Service consults these databases as part of computing the strength of identity proofing score.
6	Third-Party Identity Verification Services	A service offered by firms from the private sector (e.g., Securefact [111], TransUnion [123]) whereby an individual's identity claims are verified against a variety of data sources, such as credit bureaus, records with mobile network operators, and active fraud alerts.
7	Bluetooth USB Device	A USB accessory that facilitates communications between the user's device smartphone and POS terminal.
8	Relying Party's POS Terminal	System used by the relying party to request identity attributes from the user.

Table B.5 (*continued*)

#	**Component**	**Description**
9	Identity Service	Server operated by Bluink that enables websites to rely on the eID-Me service for authentication (identity provider). The relying party's web application communicates with this server to request specific identity attributes from the user with eID-Me acting as the intermediary.
10	Relying Party's Web Application	Web application operated by the relying party that communicates with the Identity Service to request identity attributes from the user.
11	Mobile Push Notification Services	Services offered by mobile platform operators that enable mobile app developers to send push notifications to their users (e.g., Apple Push Notification Service, Google's Firebase Cloud Messaging).

B.5 eID-Me: Sequence Diagrams

B.5.1 Registration

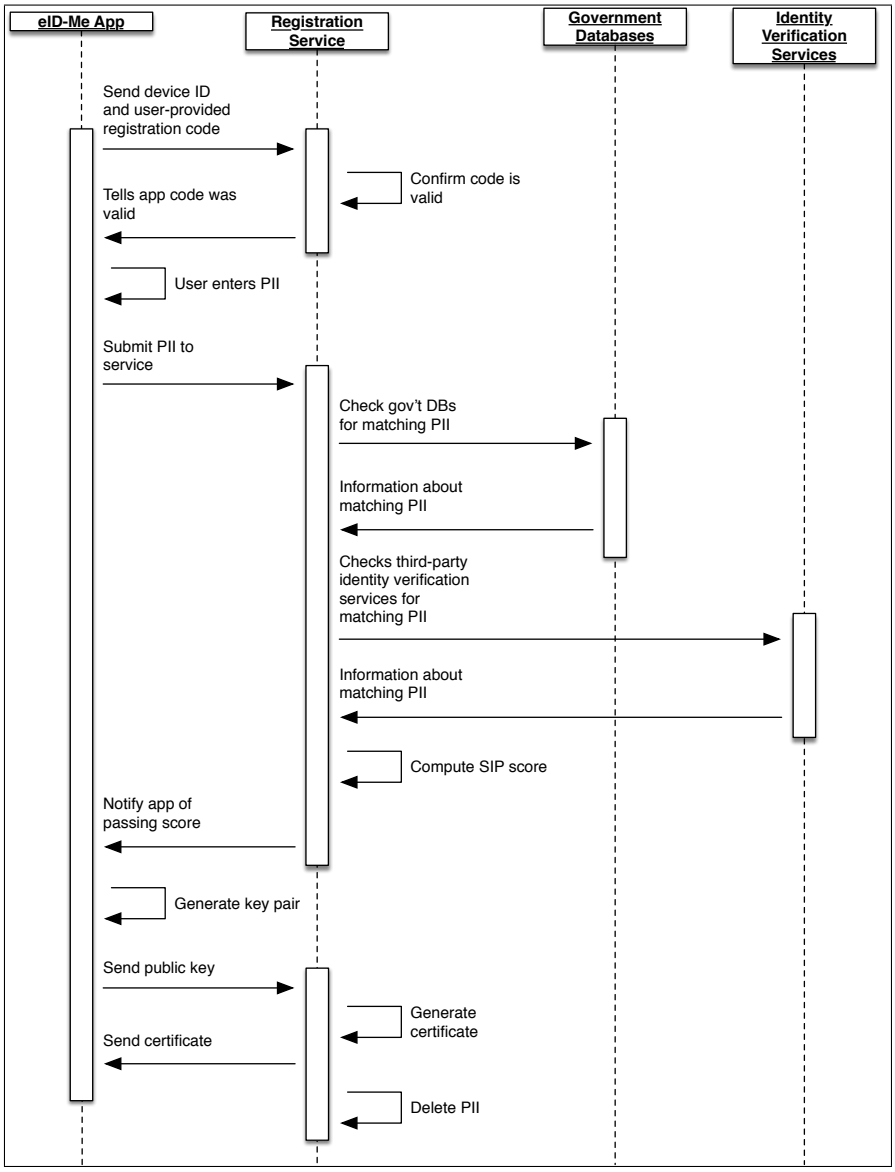


Figure B.3: Sequence diagram for registration. Adapted from internal documents provided by Bluink.

B.5.2 In-Person Authentication

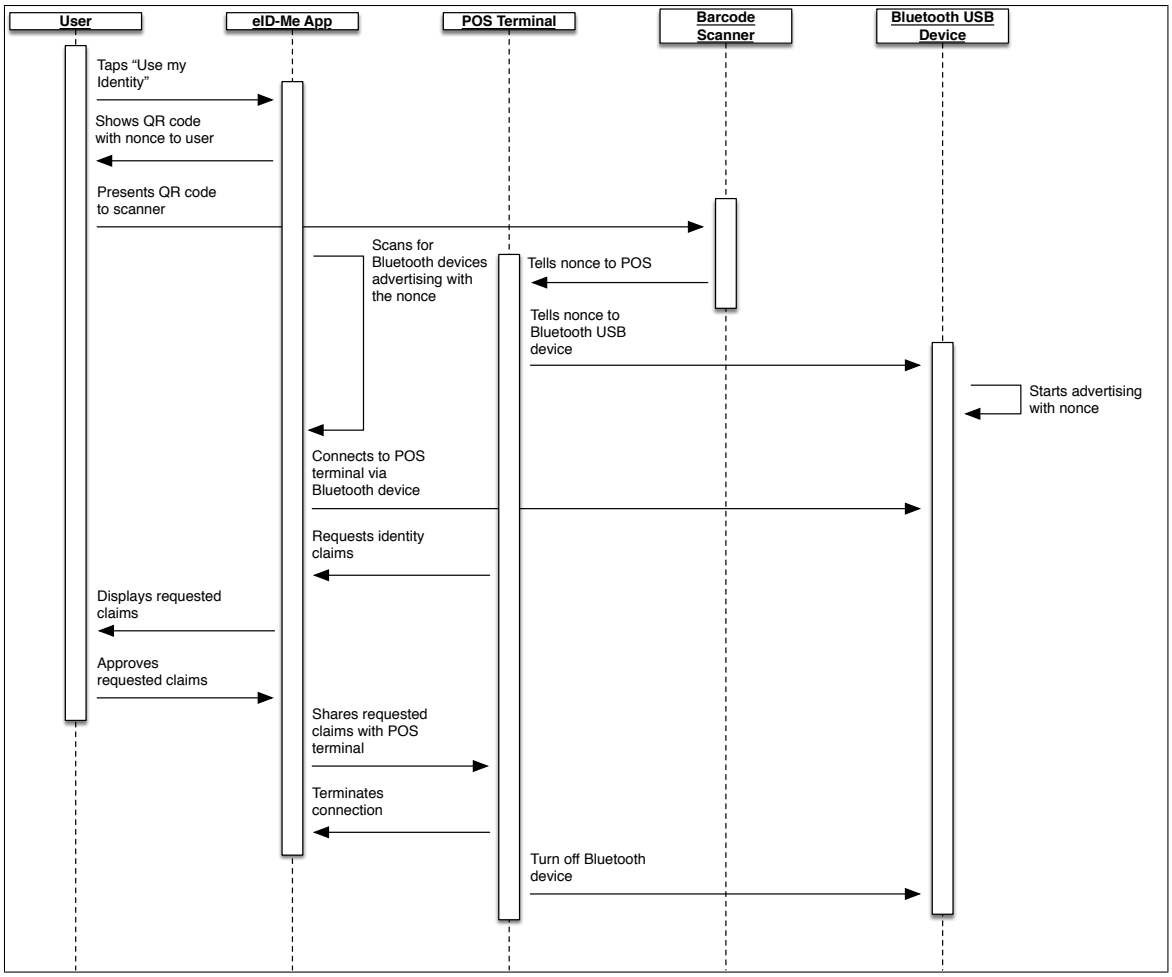


Figure B.4: Sequence diagram for in-person authentication. Adapted from internal documents provided by Bluink.

B.5.3 Online Authentication

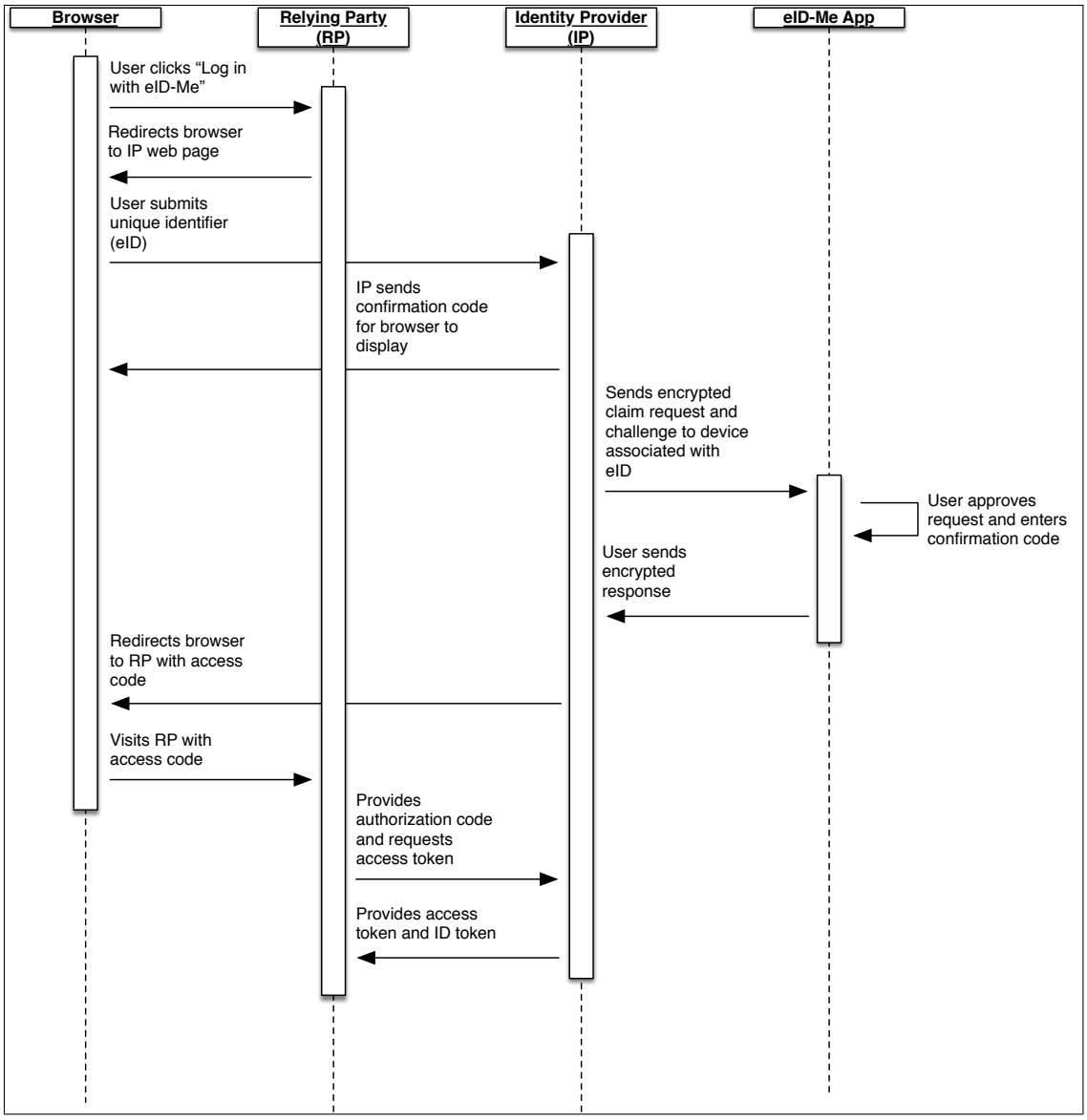


Figure B.5: Sequence diagram for online authentication. Adapted from internal documents provided by Bluink.