Toward Understanding and Improving the User Experience with Smartphone Physical Security

by

Masoud Mehrabi Koushki

B.Sc., University of Isfahan, 2011

M.Sc., Sharif University of Technology, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

The University of British Columbia
(Vancouver)

September 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Toward Understanding and Improving the User Experience with Smartphone Physical Security

submitted by **Masoud Mehrabi Koushki** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Electrical and Computer Engineering**.

Examining Committee:

Konstantin Beznosov, Professor, Electrical and Computer Engineering, UBC
Supervisor

Julia Rubin, Associate Professor, Electrical and Computer Engineering, UBC
Supervisory Committee Member

Karthik Pattabiraman, Professor, Electrical and Computer Engineering, UBC
University Examiner

Hasan Chavoshoglu, Associate Professor, Sauder School of Business, UBC
University Examiner

Additional Supervisory Committee Members:

Sidney Fels, Professor, Electrical and Computer Engineering, UBC
Supervisory Committee Member

Abstract

The incumbent physical security system on smartphones is known to dissatisfy users. It comprises explicit authentication (e.g., passcode), which imposes high time and cognitive overhead, and all-or-nothing authorization, which limits flexibility. Consequently, an estimated 20% of users have decided to forgo physical security entirely.

In response, alternative solutions have been proposed by researchers. These include implicit authentication (IA) solutions, which harnesses behavioural data for user identification, and finer-grain (e.g., app-level) authorization solutions, which are more accurate. However, several important aspects of these alternatives are understudied.

Firstly, it is unclear how widely users would adopt IA, and whether they can understand its semantics well enough to avoid dangerous security errors when using it. Secondly, it is unknown how well can the proposed authorization schemes balance usability with security. These unknowns bring into question whether the alternatives can, in fact, improve the user experience (UX) or, conversely, disservice users by providing a false sense of security.

This dissertation contributes insights from several studies that aim at bridging these knowledge gaps. Regarding IA, we took Smart Lock (SL)—currently the most-widely-available solution—as a case. We conducted cognitive walkthroughs, think-aloud sessions, and online surveys to understand how users perceive and understand SL. Regarding authorization, we conducted a longitudinal diary study to obtain a detailed view on users' needs and how well existing solutions meet them.

Results show that SL is not widely adopted, which correlates to its perceived lack of usefulness and security. Regarding semantics, we found users often confused about IA's capabilities and the nature of the data it harnesses. To avoid these issues, we provide UX design recommendations for better communication of the value and intricacies of IA.

Regarding authorization, we found app-level schemes to outperform other solutions; hence we argue for wider deployment of them. However, we also found that users' needs vary significantly based on individual preferences and the functionality being protected; hence we argue for adaptable granularity in authorization.

Overall, our studies demonstrate the inadequacy of the incumbent system, show how current deployment of alternatives potentially disserves users, and provide recommendations for improved deployment in the future.

Lay Summary

The current phone unlocking systems, such as PIN, are known to take too much time and effort to use, leading 20% of users to forgo enabling them entirely. Implicit Authentication (IA) has been proposed as an alternative, which uses behavioural traits (e.g., gait) for unlocking. However, it is not known how well users perceive IA or understand its workings, and also if other phone systems need to change in accordance with IA, to improve security. This dissertation makes first steps towards addressing these gaps by finding that a perceived lack of utility and security deters most users from IA; the capabilities of IA cause frequent confusions for users; and the authorization system on phones need to change as well, as it interrupts the users unnecessarily too often.

Preface

This research was the product of a fruitful collaboration between myself (the author of the dissertation) and the following people: Konstantin Beznosov (supervisor), Julia Rubin (member of the supervisory committee), Yue Huang, and Borke Obada-Obieh from the University of British Columbia, and Jun Ho Huh from Samsung Research, Republic of Korea. The work presented herein consists of research studies that have either been published or are being prepared to be submitted to peer-reviewed international conferences.

The mixed-method study on Android users' perception of Smart Lock, which is presented in Chapter 2 and partly discussed in Chapter 6, is based on the following publication:

> M. Mehrabi Koushki, B. Obada-Obieh, J. Ho Huh, and K. Beznosov. "Is Implicit Authentication on Smartphones Really Popular? On Android Users' Perception of Smart Lock for Android." In *Proceedings of the 22nd ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*, Oldenburg, Germany, 2020, Pages 1-17.

I was responsible for designing and conducting the cognitive walkthrough, think-aloud and survey studies, with Borke Obada-Obieh auditing and taking field notes during the in-person study sessions. I was also responsible for performing all quantitative data analysis, with Borke Obada-Obieh and I performing the qualitative data analysis jointly to reduce personal bias. I was also responsible for writing the paper. Jun Ho Huh, Konstantin Beznosov, and also Borke Obada-Obieh provided feedback on the design of the study, and the manuscript drafts. They also participated in discussions of the results. Prior to conducting the studies, I obtained approval from the Behavioural Research Ethics Board (BREB) at UBC (Ethics ID H18-01370).

The above study showed, notably, that the users' (mis)understanding of the Smart Lock semantics could contribute significantly to their hesitation to adopt the technology. Hence, I conducted a follow up mixed-method study to investigate in more detail the users' understanding of Smart Lock semantics. This study, presented in Chapter 3 and partly discussed in Chapter 6, is based on the following publication:

> M. Mehrabi Koushki, B. Obada-Obieh, J. Ho Huh, and K. Beznosov. "On Smartphone Users' Difficulty with Understanding Implicit Authentication." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI)*, Online Virtual Conference, 2021, Pages 1-14.

For this project, I was responsible for the design and administration of the survey study, as well as analysis of the quantitative data. A re-analysis of the qualitative data from the previous cognitive walkthrough and think-aloud sessions was done jointly by me and Borke Obada-Obieh, to reduce personal bias. I was also responsible for writing the paper. Jun Ho Huh and Konstantin Beznosov provided feedback on the design of the study and discussion of the results. Prior to conducting the study, I obtained approval from the Behavioural Research Ethics Board (BREB) at UBC (Ethics ID H18-01370).

From the results of the above two studies, I realized that there is a need to study the access control systems of smartphones as well, to investigate how they cloud affect the user experience with smartphone physical security overall. As such, I next conducted a longitudinal diary study to investigate users' access control needs, and how well existing solutions can meet them. This study, presented in Chapter 4 and partly discussed in Chapter 6, is based on the following publication:

> M. Mehrabi Koushki, Y. Huang, J. Rubin, and K. Beznosov. "Neither Access nor Control: A Longitudinal Investigation of the Efficacy of User Access-Control Solutions on Smartphones." In *Proceedings of the 31st USENIX Security Symposium*, Boston, MA, USA, 2022.

I was responsible for designing and conducting the diary study, including developing the necessary software (a custom Android app and a web server). Also, all quantitative and statistical data analysis was done by me, whereas Yue Huang and I performed the qualitative analysis jointly to reduce personal bias. I was also responsible for writing the paper. Julia Rubin and Konstantin Beznosov provided feedback on the design of the study, the data analysis, and the paper writing. Prior to conducting the study, I obtained approval from the Behavioural Research Ethics Board (BREB) at UBC (Ethics ID H20-03155).

Lastly, based on the results of the above longitudinal study, I realized that there is currently no practical solution for fine-grained task-based access control on smartphones, which was found in our diary study to be a necessity for meeting the needs of certain users. Hence, as the last project, I implemented such solution and estimated its efficacy using the data collected in the diary study. This endeavour, presented in Chapter 5 and partly discussed in Chapter 6 is based on the following manuscript (currently being prepared to be submitted to an international peer-reviewed conference):

> M. Mehrabi Koushki, J. Rubin, and K. Beznosov. "GUS: A Gradual Access Control System to Complement Implicit Authentication on Smartphones."

I was responsible for the design of the system and the implementation of the prototype. All data analysis and the writing of the manuscript was done by me. Julia Rubin and Konstantin Beznosov provided feedback on the design of the study, data analysis, and the manuscript.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my PhD supervisor, Prof. Konstantin Beznosov, for guiding me through this journey, by providing insightful feedback on the design and execution of all research presented in this dissertation. I am also grateful for the life lessons he taught me, and the great memories he created for me, all of which I will cherish forever.

I would also like to thank Prof. Julia Rubin, for all the detailed feedback she patiently provided me. Her guidance contributed significantly to elevating the scientific rigor of this work.

I thank all my colleagues and collaborators at LERSSE and Samsung for their support throughout this journey. Thank you: Jun Ho Huh, Artemij Voskobojnikov, Yue Huang, Borke Obada-Obieh, and Azadeh Mokhberi.

Lastly, I would like to thank Samsung Electronics and Scotiabank for supporting parts of this research, through providing generous grants to UBC.

Dedication

To my kind, caring, wonderful and beautiful wife Zhila, for supporting me in countless ways throughout this journey. This would not be possible without you.

To my parents, Ali and Zahra for always believing in me and inspiring me to be ever ambitious.

To my brother Naser, his lovely wife Maryam, and their soon-to-be-born baby girl.

To my in-laws, Hasan, Fatemeh, Saeed, Saeedeh, Elham, Ali, Taha, Iylin, Iylar and Mahdi.

Chapter 1

Introduction

Strong security is a necessity for smartphones nowadays. The continued improvements in phone capabilities have enabled their use for a wide range of applications, leading to users storing and accessing highly sensitive data and services on the devices [75]. Phones, for example, are now regularly used to access government services, such as tax filings, or store private data, such as health information [33, 75]. This has increased the cost for users of any unauthorized access to their phones [75, 76], which, in turn, has highlighted the importance of apt security protections.

In specific cases where potential attackers have physical access to a phone, solutions dubbed “physical security” are employed to prevent any unauthorized access. Generally, such solutions consist of two components: (1) the authentication subsystem, which confirms the user’s identity; and (2) the access-control (i.e., authorization) subsystem, which ensures that the authenticated user can access only allowed functionalities [55].

Currently, phones have an authentication system designed around the something-you-know-or-you-are model. This model entails a user either providing a passcode or a valid biometric to gain access to the phone (e.g., a fingerprint, a face scan) [14, 74, 95]. Simultaneously, the phones’ access-control system is based on the all-or-nothing screen locking model; the phone is either fully unlocked (where the user can utilize all of its functionalities) or fully locked (where virtually all functionality is unavailable), without any in-between state [54, 78].

While simple in implementation, these current designs are known to often dissatisfy users. On the authentication side, users find the design too cumbersome. Studies have shown that users unlock their phones more than 30 times a day on average [74, 95], and each unlocking attempt usually takes two to three seconds to perform [74]. Cumulatively, users might spend up to 80% of their phone usage time with unlocking alone [74]. This excessive and often unnecessary effort increases the cognitive and physical overhead of using phones, which creates an incentive for users to weaken or even disable authentication altogether [54, 98].

Moreover, the secret-based unlocking schemes (e.g., a PIN or password) are known to have usability [81] and memorability [56] issues. Biometric schemes are shown to impose negative externalities, like the awkwardness of holding the device in front of one's face [29], as well as reliability and security problems [29]. Consequently, studies estimate that between 10%–35% of phone users do not use any user authentication mechanism on their devices, posing a significant security risk to their sensitive data [33, 51, 52, 82].

Similarly, the all-or-nothing access-control system is shown to be increasingly insufficient for the needs of modern phone users in two key areas: (1) sensitivity to task security requirements [26, 33, 62] and (2) support for phone sharing [54, 60].

Firstly, even though tasks performed on smartphones have different levels of sensitivity [26, 33, 54], the access-control system treats all of them the same. For example, whether a user wants to simply read a book or, conversely, perform a financially sensitive operation (e.g., online shopping), the phone requires the same level of unlocking [6]. Studies suggest that users perceive this model of protection as sometimes unnecessary (e.g., for reading the book) and sometimes insufficient (e.g., for banking) and are dissatisfied with it [60, 62, 98].

Secondly, the all-or-nothing system also lacks support for phone sharing, even though the practice is known to be prevalent [58]. Phone owners share their devices with others for a variety of reasons, such as financial necessity or technical help [50, 54, 60, 77]. Yet current phones are designed to be single user.[1] This has resulted in the insecure practice of primary users sharing their Passcodes with others or adding the biometrics of secondary users to their phones [77], which limits the ability of

[1] This is evident from Apple's iOS, which does not allow creating more than one user account on a phone [6]. While Android does have multi-user support, it is only fully enabled on tablets [34].

primary users to control what secondary users can do/access on their phones. This, in turn, leads to unauthorized access and user dissatisfaction [50, 60, 75–77, 85].

In summarily, it is evident the current authentication and authorization systems suffer major inadequacies. Naturally, therefore, researchers have proposed alternative solutions to address these issues. Regarding authentication, one of the main proposed solutions is implicit authentication (IA). IA proposes the use of behavioural biometrics (e.g., gait patterns) and contextual data (e.g., location) to identify users continuously and unobtrusively [27, 36, 37, 62]. Since this type of authentication does not require explicit action from the users for each phone access, it can alleviate the unlocking burden to a great extent [63]. Phone manufacturers have also tried to make authentication more convenient, through commercial deployments of IA solutions that automatically unlock phones at certain locations, or when close to certain Bluetooth devices [41].

Regarding access control, several alternatives have also been proposed and/or implemented. For example, to facilitate ad hoc phone sharing, both Android and iOS have solutions to lock an app on the screen and not allow switching between apps without the passcode [7, 43]. More recently, phones also allow users to hide certain apps from their launchers, in order to prevent unauthorized access to them (or even hide their existence all together), when the phone is being used by a secondary user. Researchers have also proposed ideas for finer-grain access-control solutions [72] for deliberate sharing of mobile devices.

As of now, however, several important technical and human aspects of these proposed alternatives have remained understudied. In the case of IA, for example, we do not know how (at least in a rather generalizable way) users perceive it (i.e., their adoption intentions for IA). We also do not know whether users understand IA semantics (i.e., how/when it authenticates users) well enough to use it without making dangerous security errors. Regarding access control, we do not know how well the proposed solutions can meet the needs of users (e.g., not causing too many unnecessary interruptions to the users) or what potential trade-offs they come with (e.g., the required upfront configuration effort). Lastly, there is a lack of technical solutions as well; currently, no practical approach exists on smartphones for implementing access control to a more granular degree than at app level.

All of these unknowns make it difficult to judge where we currently stand on research and practice in smartphone physical security. We do not know whether IA is indeed the way forward with authentication. Could it even do users a disservice by providing them with a false sense of security? We also do not know quantitatively how real the need for a new access-control solution is or, if there is indeed a need for such a solution, which model (e.g., in terms of granularity) future systems should base their designs on.

In this dissertation, we aim to take a first step toward addressing these knowledge gaps. We provide the first insight into how smartphone users perceive a commercially available IA scheme, Google Smart Lock (SL) for Android [40], and gauge factors that might correlate with users' adoption intentions for it. We also investigate to what degree users understand how SL works and whether that understanding is sufficient for them to avoid dangerous security errors. We also evaluate how various access-control solutions fare in the number of user interruptions they unnecessarily force or inappropriately miss. Lastly, we propose a solution for finer-grain task-based access control and implement a prototype of it on Android 10.

We should note that our position in this research is not that every user should adopt IA. Naturally, as we will discuss in Chapter 2, some users might be perfectly happy with their current authentication methods (e.g., a fingerprint). Therefore, we aim to understand how IA technology can be designed to better support a user's needs when dissatisfied with the current solutions (and hence has likely left their phone without security protection). We also aim to understand how the semantics of this technology can be better conveyed to users to avoid providing them with a false sense of security. Also, the access-control solution we propose in Chapter 5 is not intended to be the ultimate solution that rectifies all issues with the incumbent system. Instead, we only focus on addressing the level of granularity with smartphone user access control. Much further research is needed to fully address the issues with the current system discussed above.

1.1 Problem Statement

The overarching problem this research seeks to address is **the lack of understanding of the user experience with smartphone physical security solutions, both incumbent and proposed**. This lack of understanding is with regards to efficacy and user preference, two important human aspects of the user experience. Regarding authentication, we focus on IA as it is the most heavily researched solution (see Chapter 2). Regarding authorization, we aim to investigate all available solutions.

We break down this overarching lack of understanding into the following four subproblems: (1) users' perception of IA, (2) users' understanding of IA semantics, (3) access-control solutions' efficacy, and (4) task-based access-control.

1.1.1 Lack of Understanding of Smartphone Users' Perception of IA

An extensive body of research exists that explores the technical feasibility of IA [37, 63, 68]. For instance, Frank et al. [37] showed how a machine learning classifier could continuously authenticate users based on the way they interact with the touchscreen of a smartphone. However, we observed a lack of understanding in the literature regarding how widely and why users might actually be interested in adopting this technology. A few existing studies suggest users might be interested in adopting this technology [26, 64]. However, the findings of these studies have limited generalizability, as they involve mostly lab studies or studies with low- to medium-fidelity prototypes. For example, Khan et al. [64] estimated that 63% of users might be interested in using IA. However, in their study they only used low-fidelity prototypes that mostly resembled a "perfect" IA solution. The prototype did not use real behavioural biometrics. As such, it did not suffer from real world issues, such as occasional lack of authentication data (e.g., in the case of a gait-pattern-based scheme, the user might stop walking). Consequently, it is difficult from existing study results to understand the extent to which factors like reliability might correlate with a user's intention to adopt IA.

Our goal, therefore, is to explore IA perception in a more ecologically valid manner than the existing literature. To this end, we refrain from using prototypes and instead investigate how Android users perceive a real world IA solution called "Smart Lock" (SL), which has been available on more than 500 million Android devices since 2015 [40]. As of the time of writing SL is the only widely available IA solution on smartphones, which makes it a good candidate for the community to learn from its successes/failures.

Through a mixed-method study, we estimated the percentage of users actually using SL and their reasons for adopting/rejecting this technology. We used this insight to provide design recommendations for future systems to be more appealing to users. We do not claim our results to be generalizable beyond SL. This study is presented in Chapter 2.

1.1.2 Lack of Understanding of Users' Understanding of IA Semantics

One of the main findings of our IA-perception study described in 1.1.1 was that a perceived lack of security is a major reason for users to reject SL. Subsequently, we found this perception often stems from an incorrect understanding of SL semantics. However, we found no prior research investigating how users misunderstand IA semantics and how prevalent such misunderstandings are.

Gaining this insight is important because misunderstanding IA can result in unintentionally leaving a smartphone unlocked, undermining the core purpose of IA. Take, for example, an IA solution that unlocks the user's phone at certain "trusted" locations (e.g., home). If the user is unaware of the location detection's accuracy (which is an inherent limitation of the GPS-based solutions), they might leave their phone unsupervised at a semi-public location (e.g., a neighbour's place) presuming that it would be locked even though it might not be. This can give opportunistic attackers (e.g., social insiders) a chance to snoop on the user's data. While this lack of accuracy on IA solution's part might be due to technological limitations that could be rectified in the future, other limitations might arise. Hence, proper communication to the user about the existence of such limitations (of which there might be many others in the future) could help mitigate such misunderstandings, and, therefore, merit further research. Our study aims to investigate what sort of communication is necessary in situations like this.

For this second subproblem, we again focused on SL as a case and investigated how Android users understand its semantics. We identified points that can cause common confusion for users; we leveraged this insight to recommend the type of information future IA UX designs should clearly communicate. The mixed-method study we conducted in this regard is presented in Chapter 3.

1.1.3 Lack of Understanding of Access-Control Solutions' Efficacy

As we discussed before, the all-or-nothing access control is already known to fail at meeting the needs of users. However, our IA-perception study uncovered a new issue—users blame access-control shortcomings on IA. For example, some users were concerned that location-based unlocking could lead to social insiders accessing their private information, (e.g., spouses if the phone were always unlocked at home). While this is certainly an authentication deficiency, as the phone unrealistically assumes anyone at a trusted location to be an authorized user, the all-or-nothing access control is also partially to blame. A more granular authorization system could prevent this issue from occurring by only allowing access to "non-sensitive" apps when authentication is conducted contextually (e.g., through location). Such issues make it clear that authentication and authorization deficiencies should be studied together to create apt physical security (i.e., one cannot set to improving one while completely ignoring the other).

A first step toward accomplishing this goal is to have a clear picture of the status quo of research and practice. Yet we found that currently there is a lack of understanding in the literature about how well the all-or-nothing system and its alternatives meet the needs of users (e.g., the number of phone activities they fail to protect due to insufficient granularity). We found this was mainly due to a lack of detailed data on users' authorization needs.

Thus, the goal for our third subproblem was to solicit a comprehensive set of authorization needs from a representative sample of smartphone users and use this data to gauge the efficacy of different access-control solutions in a quantitative manner. Based on the results, we provide insight into which type of solution might perform the best in different scenarios (e.g., in phone sharing settings). To this end, we conducted a longitudinal diary study, which is presented in Chapter 4.

1.1.4 Lack of Understanding of Customizable Task-Based Access-Control Solutions

One of the main findings of our longitudinal diary study was that while app-level granularity for access control could strike the best balance between usability and security for most users, finer-grain task-level systems are needed for users with more complex preferences (e.g., they want only parts of an app to be available to others). Such system can be used, for example, by a parent to allow their children to play a game without them being able to purchases items online. Yet currently we found that there is no practical solution proposed in the literature to perform this type of access control. By "practical" we mean solutions that provide a machine-understandable way of distinguishing between users' tasks.

We should note that many designs other than task-based ones could follow from our longitudinal study. We decided to focus on task-based design as it would benefit IA deployment and access control at the same time. In short, deploying IA in the real world is known to be tricky. Sporadic availability of behavioural data forces IA to trigger frequent re-authentication prompts, which lock out the user mid-session and ask them to authenticate explicitly using a PIN or biometrics. If this happens too often, it can annoy users [2].

Task-based authorization can alleviate this problem by only showing prompts when a user is performing a sensitive task. As a result, for our final subproblem we propose a design for such a task-based access-control system. We call the system the "Gradual Unlocking System (GUS)". In Chapter 5, we describe the design of GUS, our implementation of a GUS prototype on Android 10, and our use of the data from our longitudinal diary study to evaluate its efficacy.

As mentioned before, GUS is not aimed at or positioned to be the solution to all problems with all-or-nothing. It serves a specific purpose to improve IA deployment and authorization granularity. It is out of the scope of this research (and is up to future work) to investigate how GUS can further evolve to address other issues as well, such as the need for phone sharing.

1.2 Research Summary

Overall, our research found that **the current state of smartphone physical security is far from optimal**. The incumbent solutions fail to meet the needs of users in most cases, and the proposed alternatives do not seem to improve the situation significantly. To explain further, in the following, we give the methodological details of the research projects we conducted to address each subproblem we described above.

1.2.1 Smartphone Users' Perception of IA

To investigate how Android users' perceive SL, we first conducted a cognitive-walkthrough-with-users (CWU) [46, 70, 73] study, comprising two cognitive walkthrough sessions with 10 HCI (Human Computer Interaction)-proficient participants and 16 individual think-aloud sessions with ordinary smartphone users. We were interested in understanding how the participants perceive SL after interacting with its UX to perform certain tasks (e.g., enabling some features of it). We then conducted an online survey with 343 Android users (recruited through Amazon Mechanical Turk (MTurk)) to verify the results of the CWU study, using a near-representative sample of the US smartphone user population. Subsequently, we formulated our findings into an extension of the Technology Acceptance Model (TAM) [28] that aims at reasoning about SL adoption.

Main Findings

- Despite its five year history, Smart Lock is not a widely adopted technology. Only around 13% of SL-capable participants in our survey reported that they were using SL. (An SL-capable participant is someone who knows about SL and has a phone that supports it.)

- Lack of availability is not a major barrier to SL adoption for Android users. As part of the Google Play Services package, Smart Lock has been deployed on hundreds of millions of Android devices over the past five years. More than 91% of our survey participants were using phones that supported SL.

- Lack of awareness is not a major barrier to SL adoption either. Nearly 60% of our survey participants had some knowledge about SL before participating in our study.

- Perceived lack of utility is one of the main factors that could deter potential users from adopting SL. The majority of our SL-novice participants (those who have not used SL before) indicated they did not see enough value in adopting SL and thought fingerprint unlocking was more convenient. Our results are the first to offer insight into such a comparative user perception between IA and traditional unlocking.

- Perceived lack of security is another deterrent for SL use. The majority of SL-novice participants were unwilling to use it because they thought adopting SL could allow unauthorized access to their phones. Our results are the first to show the statistical significance of this perceived-insecurity-on-adoption rate in the context of IA.

1.2.2 Users' Understanding of IA Semantics

We started by reanalyzing the data from our CWU study with a focus on identifying aspects of SL that caused confusion for the participants (i.e., any case in which they answered questions about SL semantics incorrectly). We then conducted a new online survey with 331 participants (recruited through MTurk) to evaluate the prevalence of SL misunderstandings. Based on the results, we generated design recommendations for IA UX on smartphones.

Main Findings

- SL misunderstandings were found to be prevalent. Most of the CWU participants (96%) had difficulty understanding the semantics of at least one SL method. Also, about 80% of the survey participants incorrectly answered at least one question about SL semantics.

- Misunderstandings, however, were not uniformly distributed across SL features. SL features that used contextual (e.g., location) or behavioural (e.g., body movement) data caused confusion more frequently.

- Four particular aspects of SL were found to cause the observed confusion: (1) its capabilities, (2) its modalities (what data it uses for authentication), (3) the inter-operation between its IA and EA (Explicit Authentication) features, and (4) its range parameters.

- Depth of smartphone adoption was found to be a significant antecedent of correct SL understanding; age, computer literacy, and security proficiency were not.

1.2.3 Efficacy of Access-Control Solutions

We conducted a longitudinal diary study. We asked a near-representative sample (N=55, recruited through MTurk) of the US smartphone user population to install our custom Android app on their personal phones for 30 days. They used it to report the tasks (i.e., fine-grain actions) they performed or shared on their phones and their access-control preference for each task. We also gathered detailed reports of the contextual factors (e.g., location, time, and identity of the sharee) surrounding each phone sharing event. Using this data, we estimated how frequently different access-control solutions would fail to meet the needs of users by either forcing unnecessary authentications (i.e., false negatives) or by failing to prevent unauthorized users from accessing a task (i.e., false positives). We also estimated the effort needed for the user to switch to each model by counting the number of configuration operations they would need to perform. Lastly, we evaluated how consistent the contextual factors of phone sharing are across the reported events. Based on the results, we discuss the shortcomings of current systems and suggest research directions for the future.

Main Findings

- A great diversity of functionality and a high complexity of access-control needs were observed in the reported data. The participants had performed 19 distinct categories of tasks, with nearly 55% of tasks perceived as shareable.

- The all-or-nothing solution was found to be suboptimal in meeting users' access-control needs. We observed that if no authentication is enabled on the phone (i.e., the all scenario), all-or-nothing exposes 90.3% of the user's tasks to unauthorized users. Conversely, if authentication is enabled (i.e., the nothing scenario), up to 21.2% of unlocking attempts are unnecessary.

- App-level solutions could potentially reduce the number of exposed tasks to 3.5% while also reducing unnecessary re-authentications to 11.3%.

- Task-level solutions could further reduce unnecessary authentications to 1.7%, albeit with a 15% increase in required configuration effort.

- Overall, app-level solutions were found to strike the best balance between security/privacy and usability for most users, based on the results above. Hence we argue for wider deployment of such solutions.

- Amongst solutions that aim to provide phone sharing support, we found profile switching to be suboptimal despite its ubiquity. It would increase the required configuration effort noticeably (a nearly eightfold increase when compared to the all-or-nothing model), as it would require specific profile-app mapping.

- Session-based phone sharing solutions (e.g., those that limit access per session as opposed to per resource) were found to provide the best balance between security and usability compared to all-or-nothing and other alternatives. They showed a 20% reduction in unnecessary interruptions, with a 1% trade-off in increased configuration effort. Hence we believe greater focus on this type of access control can be a promising avenue for future research.

- Phone sharing was found to happen in the same context (e.g., with the same people, at the same location, or for the same set of tasks) up to 75% of the time, showing high promise for context-based access control in multi-user scenarios. Our results are the first to provide such quantitative evidence.

1.2.4 Gradual Unlocking for Task-Based Access Control

To provide a practical way of performing task-based access control, we designed and implemented a solution that allows the user to specify different security requirements for each Android Activity (application screens; see Chapter 5 for more details). This solution (GUS) considers the level of authentication confidence when permitting or declining users to launch Android Activities. (“Authentication confidence” is a measure provided by an IA system to indicate the degree to which the current user’s behaviour matches that of the registered user.) We estimated the efficacy of our solution using the data from our longitudinal diary study.

Main Findings

- Android Activities were found to somewhat aptly represent tasks. Roughly 50% of user-defined tasks could unequivocally be represented using Android Activities, and 25% more could at least be partially represented.
- GUS was able to reduce unnecessary re-authentications noticeably (by 96%), albeit with a security trade-off (a 1% increase in missed re-authentications).
- Implementing a prototype of GUS on Android is feasible by modifying the WindowManager service of the operating system.

1.3 Main Contributions

In this section, we outline the main contributions of our research and discuss how they could impact future research in smartphone physical security.

1.3.1 Shedding Light on What Attracts/Deters Users from IA

Prior to our study, existing literature on users' perceptions of IA [26, 63] were of limited ecological validity (i.e., not wholly reflective of the users' real world behaviour), as they mostly involved role-playing. One main contribution of our study, therefore, is providing the first empirical evidence on smartphone users' perceptions of commercial IA (SL) in a realistic setting. Its results suggest SL is not widely adopted because it is perceived to be insecure and/or inconvenient. This sheds lights on the necessity for future research on communicating these aspects to users effectively through apt UX design.

Another major contribution of this research is proposing and evaluating a new technology acceptance model (which we dub "SL-TAM") for reasoning about SL/IA adoption. The model unveils three distinct factors correlated with SL adoption: perceived usefulness, perceived ease of use, and perceived security and privacy. This insight can inform the design of future IA schemes, as it sheds light on users' requirements/expectations for such systems.

1.3.2 Insights into How IA Semantics Should Be Communicated to Users

To the best of our knowledge, no research had been conducted on smartphone users' understanding of IA semantics prior to our study. Thus, a main contribution of this research is providing the first insight into how and with what prevalence smartphone users' make mistakes when comprehending IA. It uncovers specific aspects of SL that can be difficult to understand for the average user (those without a technical computer and security knowledge background). As a result, new avenues for future research on IA UX design are opened to find ways of communicating these aspects more clearly to users.

Another major contribution of this work is being the first to investigate how a combination of IA with EA is understood by smartphone users. We provide insights into the pitfalls of this combination (e.g., how putting all of them under the same name could lead to misunderstandings about the fundamental differences between IA and EA) and suggest precautionary measures that should be taken when using it.

This work also provides evidence that IA misunderstandings have no particular antecedents (e.g., age, security proficiency) other than depth of smartphone adoption. This demonstrates that the difficulty with comprehending IA is rather universal, which can inform assumptions about potential users' background knowledge during the design of future IA schemes.

1.3.3 Characterizing the Status Quo in Access Control

As mentioned before, while multiple access-control solutions were proposed prior to this research, they were not compared empirically. Our literature review showed this was mainly due to a lack of detailed data on users' access-control needs. Thus, a first contribution of our work is to provide such data, giving a detailed view of the tasks smartphone users perform on their phones and their sharing preferences for each task. This data can be used to evaluate the efficacy of any current or future access-control solution.

Another major contribution of this work is that by using the collection of tasks, it quantitatively evaluates how the all-or-nothing model and its alternatives meet the access-control needs of users. We show how severely most current solutions fail and discuss where future research needs to focus in order to propose better solutions.

This work also sheds light on the contextuality of phone sharing. It demonstrates that, given the consistency of contextual factors, incorporating them into access-control decisions could help with a better balance between users' needs for privacy and usability. We provide ideas on how this can be accomplished.

1.3.4 Proposing a New Access-Control Solution

This work contributes the first practical task-based access-control solution for smartphones by tying users' ability to launch Android Activities to the confidence of authentication. This solution can improve the usability of IA by reducing the number of unnecessary re-authentication prompts.

As a minor contribution, we also make available the solution's source code[2] to facilitate future research on IA deployment techniques. The code can help researchers experiment with different UX designs for re-authentication prompts or different modalities for user identification, without having to deal with complex operating-system-level code.

We should note our solution comes with a trade-off in the form of increased user effort for upfront configuration. However, this also opens a new avenue for future research: investigating methods to predict users' preferences (e.g., using matching learning techniques).

[2] https://github.com/mehrabik/GUS

Chapter 2

Investigating Smartphone Users' Perception of Implicit Authentication

This chapter presents the results of a mix-method study to investigate Android users' perception of Smart Lock (SL) and factors that correlate with their intentions to adopt it. It is a first step to inform the broader investigation of smartphone users' perception of IA in future research. First, in section 2.1 we provide an overview of the existing literature on IA, a brief introduction to Smart Lock for Android, and related papers on technology adoption. Second, in section 2.2, we present the methodology of our study. In section 2.3, we present our results and discuss their implications. In section 2.4, based on the results of our study, we propose an extension of Technology Acceptance Model (TAM), which we call SL-TAM, to reason about SL adoption. Section 2.5 discusses the limitations of our study, and, finally, section 2.6 concludes the chapter.

2.1 Background

2.1.1 Implicit Authentication

Smartphone users' attitudes toward unlocking their devices have been well-studied, and research shows that users perceive unlocking to be a burden. This perception seems to be justified, as it has been estimated that 80% of short phone sessions is spent unlocking the phone [74].

Implicit authentication (IA)[1] is a promising solution to alleviate this unlocking burden [27, 63]. This is because IA does not require explicit action from the user. There have been numerous studies examining the feasibility of various modalities for IA on smartphones. For instance, Frank et al. [37] demonstrated how a machine learning classifier could continuously authenticate users, based on the way that they interact with the touchscreen of a smartphone. Other proposed modalities (contextual or behavioural data) for IA include gait patterns [30], body movements [94], biomedical signals [84], and app usage [38].

When it comes to smartphone users' perceptions of IA, however, the existing literature is much more limited. Khan et al. [64] conducted a two-part study, consisting of lab-based experiments and a three-day field study where 37 participants used IA on their own smartphone. However, for experimental control, both parts used a low-fidelity IA prototype that simulated re-authentication events manually (that is, no actual behavioural biometric was used. Rather, the scheme prompted the user for re-authentication as pre-specified intervals). They reported that 81% of their participants were satisfied with the level of security that IA provided. They found that 63% of participants were interested in using IA, but 30% were not sure whether they would use it, and 7% did not want to use it. Similar work was conducted by Crawford and Renaud [26]. In a lab study, they asked 30 participants to complete a series of tasks on a smartphone that was protected with a pseudo-IA scheme (again, one that does not actually use behavioural biometrics). They found that 90% of their participants indicated that they would consider using IA on their mobile devices, should it become available.

[1] We define IA as any solution that can authenticate the user without asking for explicit authentication input (e.g., a PIN or a fingerprint).

As discussed, the above studies were either conducted in lab settings or used low- or medium-fidelity prototypes, which limits their ecological validity (this refers to the extent to which the experimental conditions of the study are representative of real tasks being done by real users in their natural environment [65]). This methodological choice is understandable given that the goal of these studies was to gauge the users' level of interest in IA technology in general, as compared to the incumbent explicit authentication methods. At the same time, however, since these studies involve role-playing (i.e., performing a set of predefined tasks with pre-provided data), it is difficult to judge whether users would behave the same way in their interest towards IA when their actual sensitive data is at risk (see Sotirakopolus et al. [109] on the penitential negative effects of role-playing on usable security lab studies).

Our studies aim to address this limitations by investigating how users have acted in response to the opportunity of adopting a real-world IA scheme on their personal phone. The difference between our study and prior work, therefore, goes beyond just methodological differences. Fundamentally, we are not trying to extrapolate the future of IA adoption based on current estimates. Rather, we study users' historical behaviour (the participants have already decided whether they would adopt/reject SL before even participating in our study), in order to find lessons learned from the successes/failures of SL. This insight could help better design the next iteration of the technology.

2.1.2 Smart Lock for Android

SL is the first, and currently only, widely deployed IA scheme used on smartphones. Other commercial IA solutions for smartphones also exist, such as the UnifyID [114] or Kryptowire [67], but due to their centralized design, these schemes are not suited for smartphone unlocking. SL was first introduced during the keynote of the 2014 Google I/O conference [40]. In its essence, SL is designed to reduce the number of times users have to unlock their phones by automatically unlocking the phone (or at least keeping it unlocked) when the surrounding environment is deemed "secure." SL offers the following 5 different methods (three IA and two EA ones) of unlocking the phone, each of which can be enabled separately:

- **On-Body Detection (BODY)** is an IA method that operates based on behavioral traits. It keeps the user's phone unlocked while it is "on-person" (i.e., in movement, like running) by detecting the user's body movements and gait patterns. BODY cannot automatically unlock the phone, but will lock it if no movement is detected.
- **Trusted Places (PLACE)** is a contextual[2] IA method that uses GPS signals to automatically unlock the phone at specific locations (e.g., home). It will automatically lock the phone when the device leaves the trusted location.
- **Trusted Devices (DEVICE)** is a contextual IA method that automatically unlocks the user's phone when a designated Bluetooth device is connected to it. It will also automatically lock the phone when the device is disconnected.
- **Trusted Face (FACE)** is an EA method. It allows the user to scan their face to manually unlock the phone. It is not capable of automatically locking or unlocking the phone.
- **Voice Match (VOICE)** is another EA method. It allows the user to say"Ok Google" to manually unlock the phone with their voice. It is not capable of automatically locking or unlocking the phone.

The EA methods, however, are outside the scope of this study, as we are focused on users' perception of IA. We will consider these methods in our next study, presented in Chapter 3.

SL is considered by Google to be an important part of Android operating system, as it is actively advertised on Android devices. For example, whenever a new Bluetooth device is paired with an SL-capable phone, a notification is shown encouraging the user to add the device as a trusted for Smart Lock. Other manufacturers may opt to either not include SL on their phones, remove some SL methods (e.g., Samsung phones do not provide *FACE*), add other methods (e.g., Wear Recognition on Huawei smartphones), or provide SL as is.

To the best of our knowledge, there are no previous studies that analyzed users' perception of SL.

[2] Meaning it operates based on contextual data rather than behavioral data.

2.1.3 Technology Acceptance

Why people accept or reject new technologies and the factors that affect their decisions have been heavily researched for decades. Numerous theories, such as the technology acceptance model (TAM) [28], diffusion of innovations theory [99], unified theory of acceptance, usage of technology [116], have been put forward to try to explain users' behavior with technology adoption. Among these theories, TAM has gained a lot of traction and has been examined, expanded, and applied to various domains, such as OpenID [111] and online shopping [117].

In TAM, proposed by Davis et al. [28], two main factors are shown to affect users' attitudes toward adopting a new technology: *perceived usefulness* and *perceived ease of use*. *Perceived usefulness* is defined as the degree to which an individual believes that a particular system would enhance his or her performance [22]. *Perceived ease of use* is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental effort" [22].

TAM is frequently used as a tool to explain users' acceptance or rejection of new technologies [22]. To expand TAM's applicability to different domains, numerous extensions to it have been proposed. For example, Venkatesh and Davis [115] proposed TAM2, which included an additional set of variables (e.g., relevance to the user's task, and subjective norm) that could influence users' technology adoption decisions. Another noteworthy extension was proposed by Vijayasarathy [117], who used TAM to predict consumer intentions for using online shopping. They introduced several new variables, such as privacy and security, that can potentially affect users' adoption intentions.

In this work, we use TAM to give structure to our findings. This allows us to ground our findings in a theoretical framework that can more cohesively explain the users' intention to adopt SL. We discuss this matter further in Section 2.4.1.

2.2 Methodology

Our methodology was designed to answer the following research questions:

1. **RQ1**: How widely is SL adopted by Android users?

2. **RQ2**: What factors attract or deter potential users from adopting SL?

2.2.1 Qualitative Studies

Since there were no previous studies to inform our investigation, we started by performing an exploratory qualitative study, to gain preliminary insight into how smartphone users would perceive SL. We opted to conduct a cognitive walkthrough with users (CWU) study [46, 70, 73]. The study comprised of cognitive walkthrough (CW) sessions with participants proficient in HCI, and think-aloud sessions with regular smartphone users (not proficient in HCI or usability).

To explain briefly, cognitive walkthroughs involve a group of HCI proficient participants going over a documented series of actions that an ordinary user would has to perform to accomplish a task with the user interface (UI) under evaluation [124] (e.g., in the case of SL, enabling one its methods, such as BODY. The complete set of tasks we used in our study is provided in the Appendix). After performing each action, the participants would engage in a group discussion as to whether the UI is clear enough to allow an ordinary user to easily identify what they need to do next, to achieve the task. If not, they would suggest ways of improving the UI, to increase its learnability. Think-aloud sessions, conversely, are one-on-one sessions where an ordinary user (not an HCI proficient) would perform the actions necessary to achieve a given task, while thinking out loud about how they figure out what the next action they take is [73]. It would usually be performed to verify the results of the cognitive walkthrough. We discuss the details of the CWU design further in the next Chapter (specifically, Section 3.3).

We preferred CWU over other alternative methodologies (e.g., semi-structured interviews or quantitative user satisfaction studies) due to its task-oriented nature and its focus on the learnability of the UI. The focus on learnability, in particular, was important to us, as we believed that most potential users find out about the semantics of SL UI through exploratory learning, which is shown to be a primary method for users to discover new features on smartphones [113]. Further, we chose to use the CWU method because it addresses the shortcomings of the traditional cognitive walkthrough (e.g., lack of real user involvement) by adding think-aloud sessions with users. We combined the results of these two studies to provide more breadth and depth in usability evaluation [73].

We further refined our study design by introducing a new set of questions to the

CWU method. We asked the participants about how they thought each SL method locked or unlocked the phone, what they thought SL was good for, whether they would consider adopting any SL methods, and the reasons for their decisions. All materials that we developed for our CWU study (e.g., persona definition, task lists, etc.) are presented in Section A.1 in the Appendix.

To test the methodology, we recruited 9 participants through word-of-mouth advertising and conducted two pilot cognitive walkthroughs (five participants in one session and two in the other) and two pilot think-aloud sessions. Pilot studies showed that think-aloud participants had less opportunity to explain their SL adoption attitudes than did the CW participants.[3] Therefore, we decided to conduct a semi-structured interview with each think-aloud participant to allow them to explain their SL adoption attitudes in their own words. We included this change in the consent form, specifying that the study investigators might ask participants clarification questions about their answers in the reporting form.

For the final study, we used both online (Facebook, Craigslist, and mailing lists) and offline (word-of-mouth) channels to recruit participants for our CW and think-aloud sessions. Overall, we recruited 26 participants: 10 for CW and 16 for think-alouds. We compensated the cognitive walkthrough participants with CAD 20 cash and refreshments, and the think-aloud participants with CAD 30 cash only. We offered refreshments to CW participants to reduce fatigue, as each CW session lasted for more than 2 hours (compared to 40-minute-long think-aloud sessions). We conducted all CW and think-aloud sessions in person between September 2018 and February 2019. We performed data collection and analysis concurrently and continued until we reached theoretical saturation (no new codes emerged as a result of the last two data collection sessions).

To analyze the data, we transcribed the audio recordings from all sessions, then anonymized and analyzed the transcripts. While we conducted CW sessions before recruiting for and conducting think-aloud sessions, the collected data were analyzed collectively. The average duration of Cognitive walkthrough sessions was 2 hours. Think-aloud sessions lasted for 40 minutes, on average. Overall, we analyzed the transcripts of approximately 14 hours of audio recordings. We

[3] In the initial design, think-aloud participants submitted written answers to questions about their adoption attitudes, as part of the handout.

chose Thematic analysis [15, 48] as our analysis method for these two datasets. To increase the validity of the results, all of the transcripts were coded by two researchers, who used an agreed-upon shared codebook. We calculated the inter-coder reliability and found that it was satisfactory (80%).

2.2.2 Quantitative Study

Based on the findings of our qualitative studies, we developed a series of hypotheses about **RQ1** and **RQ2** and then conducted a confirmatory online survey to evaluate the hypotheses and answer the research questions.

In our survey, participants were first asked a series of demographic questions, followed by questions about their smartphone usage habits and the screen unlocking methods that they had enabled on their phones. Afterward, participants watched an introductory video about SL, prepared by us, which we used to remind participants about what SL was.[4] We carefully crafted the video to use the exact words that are already used in existing SL set-up UI.[5]

After the participants watched the video, we asked them questions about how familiar they were with SL before our study, their experiences of using SL, and their attitudes toward adopting it. Based on the answers to these questions, participants were asked to rank, in order of importance to them, a list of potential reasons for their attitudes toward adopting SL. The list of potential reasons was informed by the findings of our qualitative studies. Participants could also choose not to rank a reason, or to add a new reason and rank it.

Afterward, participants were asked to rank the common smartphone unlocking methods (fingerprint, face unlock, and PIN/password) against all SL methods, based on how convenient they thought each method was. The same procedure was repeated for the perception of security, and the speed of the unlocking methods (i.e., we asked participants to rank the methods in order of how secure and how fast they thought these methods were). The complete list of survey questions is provided in Section A.2 in the Appendix.

[4] Based on our experience with our qualitative study participants, we believed that participants might not necessarily remember what SL was, or that SL might be named differently on their phones.

[5] The video is publicly available on YouTube through this link: https://www.youtube.com/watch?v=N-pC6-kWW0c.

Finally, to calculate an SL adoption rate, we needed to know whether each participant's phone supported SL. To do so, we asked our survey participants to enter the model number of their phone so we could determine if SL was supported on their device. Since we anticipated that some non-tech-savvy participants might not know the model number of their device, we suggested them to use their main mobile phone to visit the study webpage and to enter their assigned participant ID. This webpage reflected the name of the phone model back to the participant. This was done by examining the user agent field of the HTTP header of their requests. Our web server did not store any user data.

To test our design, 7 HCI researchers reviewed our survey questions and provided feedback. We also conducted a pilot study on MTurk with 10 participants. Accordingly, we made minor adjustments to the wording of some questions.

We conducted the main survey on MTurk in September 2019. We chose MTurk as its participants samples are shown to provide meaningful results for the area of usable security [51]. Our advertisement message mentioned a study about Smart Lock for Android, but stated that Turkers (MTurk workers) did not need prior experience with Smart Lock to participate in the study. The survey was only visible to Turkers living in North America[6] and had an approval rating higher than 90%. Non-Android users were excluded from the study. It took the participants 17 minutes on average to complete the survey, and each participant was compensated with USD 4. Overall we received 407 responses to our survey, but eliminated 64 responses due to either inconsistencies in their answers (39), using IP addresses outside North America (10), or being flagged as duplicates or bots by our survey platform (15).

To analyze the participants' reasons for adopting/rejecting SL, two researchers cooperatively labelled each reason given by them with one of the five broad categories we identified in our CWU study. These categories include Utility, Security, Privacy, Reliability, and Other. Without assuming any particular distribution for our data, we used chi-squared tests of association and binomial logistic regression to analyze how our hypothesized factors correlated with our outcome variables. To analyze our rank-order data (e.g., the average ranking the participants gave to each

[6] We opted to only include North American participants in our survey because otherwise it would be difficult for us to account for cultural differences and their potential effects on the results.

category of reasons), we used Friedman and Durbin-Conover tests. All pair-wise comparisons were corrected using Bonferroni.

To address *RQ1*, we reported on the adoption rate of each SL method, based on the results of the survey study. We defined adoption rate as the ratio of the number of participants who reported using a particular SL method to the number of participants who could have been using that method. We referred to the latter as SL-capable participants. There is a distinction between participants who knew SL was available on their phones and those who knew about SL but did not know it was on their phones. In calculating the SL adoption rate, we used the former.[7]

To get a better understanding of the participants' overall attitudes toward SL, we also reported on the rejection rate of each SL method in our survey study. We defined the rejection rate as the ratio of participants who decided to not use an SL method to SL-capable participants. Participants who rejected the SL method had either experimented with the method but decided not to use it, or used the method for a while but then stopped using it. A high rejection rate would likely indicate the prevalence of important adoption barriers (deal-breakers) for potential users.

Gaining an estimate of these rates are important because while a high adoption rate could indicate a high degree of acceptance and adoption of the technology, a low adoption rate would likely indicate the existence of external barriers to adoption of the technology, or the existence of other factors that deterred potential users from adopting SL, which could merit further research.

Moreover, we also reported on the interest rate for each SL method in our survey study. We defined the interest rate as the ratio of SL-capable participants who were undecided about the SL method but indicated their willingness to adopt it to the total number of undecided participants. Undecided participants were those who had neither rejected nor accepted SL. Since interested participants have not had any real experience with SL, a low interest rate among them could likely indicate the existence of external factors (e.g., a perceived lack of added convenience when compared to biometric-based phone unlock methods) that deterred them from

[7] We excluded participants who were not using an SL-enabled phone, or who reported not having prior knowledge of SL before the study, from the SL adoption/rejection rate calculations. This was because we believed that the adoption attitude data from such participants (without awareness) would be of low ecological validity.

Table 2.1: Participant demographics for our CWU and survey studies.

		CWU Study	Survey Study
		($N = 26$)	($N = 343$)
Parameter	**Property**	**% (#) of participants**	**% (#) of participants**
Gender	Female	53.8 (14)	58.3 (200)
	Male	46.2 (12)	41.4 (142)
	Other	0.0 (0)	0.3 (1)
Age	19-24	38.5 (10)	5.8 (20)
	25-34	57.7 (15)	46.6 (160)
	35-44	3.8 (1)	28.9 (99)
	45-54	0.0 (0)	14.3 (49)
	55-64	0.0 (0)	3.8 (13)
	65-74	0.0 (0)	0.3 (1)
	75-84	0.0 (0)	0.3 (1)
Education	Less than high school	0.0 (0)	0.3 (1)
	High school	7.7 (2)	33.2 (114)
	University (bachelor's)	38.7 (10)	58.3 (200)
	Master's or PhD	53.4 (14)	8.2 (28)

using SL. This can be a promising avenue for future research to improve adoption.

Finally, to address **RQ2**, we report on how our survey participants ranked the reasons for their interest in and adoption of each SL method. Based on the findings of the CWU study, we placed these reasons into 5 categories: "utility," "security," "privacy," "reliability," and "other reasons."

We should note that UBC Behavioural Research Ethics Board (BREB) approved the data collection and analysis of all of our studies before any data collection took place.

2.3 Results

2.3.1 Demographics

Our CWU sample was fairly diverse in terms of participant age, gender, familiarity with SL, and occupation. Detailed demographics of our CWU participants are presented in Table 2.1 (occupation was omitted due to the high diversity of responses). Also, while all CW participants were graduate students, only about one-third of the participants in the think-aloud sessions were students. As for their familiarity with SL, 50% of the participants had never heard of SL; 20% had heard of it but had never used it; 16% previously experimented with it but had never used it; 7% had previously used it regularly but had stopped since then; and 7% were using SL regularly at the time of the study.

Our survey sample was also fairly diverse in terms of gender (58% males and 41% females), and age (min = 19, max = 75, mean = 36, median = 34). In terms of education, 33% of survey participants had high school degrees, 58% had college education, and 9% had master's or PhD degrees. Participants were also relatively diverse in terms of occupation, with computer and mathematical occupations having the highest frequency (17%).

To gauge the representativeness of our survey sample, we compared the distribution of certain demographic factors (age, gender, and education) among our participants to the smartphone population in the US reported by the Pew Research Center [91]. Chi-squared tests revealed no significant differences between distributions of the two samples, in terms of age (p = 0.21), gender (p = 0.19), or education levels (p = 0.16). Hence, we believe that our sample is fairly representative of the target population. However, as seen in Table 2.1, age distribution was somewhat skewed, even if the difference was not statistically significant. We discuss sample limitations further in Section 2.5.

We should also note that, in alignment with our research goals stated in Chapter 1, we focused on the use of SL on personal devices in this research. Hence, we did not focus on diversity in terms of context of use (e.g., corporate phone users). As for personal use of SL, we aimed for diversity in our sample in terms of age, gender, and education (similar to other studies [33, 95]). We only chose those

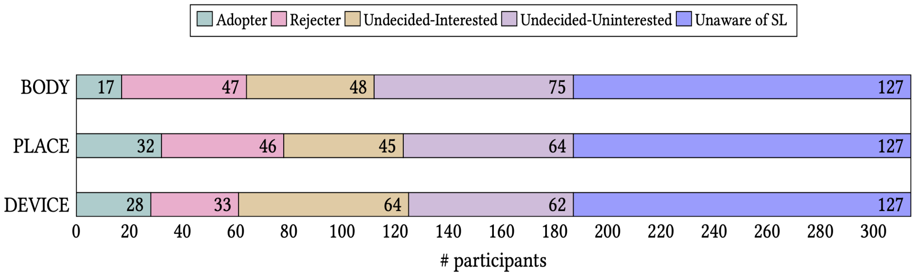


Figure 2.1: Experience with and attitudes toward adopting each SL method, for survey participants who had SL-capable phones and indicated having prior knowledge of SL before the study.

Table 2.2: Adoption, rejection, and interest rates for different SL methods in our survey study.

SL Method	**Adoption Rate (%)**	**Rejection Rate (%)**	**Interest Rate (%)**
BODY	9.1	25.1	39.0
PLACE	17.1	24.6	41.3
DEVICE	15.0	17.6	50.8
Average	13.7	22.4	43.7

factors they have been shown to be correlated with users' phone authentication behaviour [33, 74, 95]. Hence, we anticipated that any differences in the need for IA would manifest itself in a sample that is diverse along those axis. For example, since a difference in age has entailed different usage pattern with traditional phone unlocking solutions [95], we anticipated that recruiting people from different age ranges would uncover differences in IA need as well. We did not specifically aim for diversity in more detailed factors (e.g., co-habitation composition).

2.3.2 Adoption, Rejection, and Interest Rates

Overall, we found that 91.3% (N = 313) of our survey participants were using phones that supported SL. Among those participants, 60% (N = 184) reported having had prior knowledge of SL before participating in our study (as depicted in Figure 2.1), making them SL-capable participants.

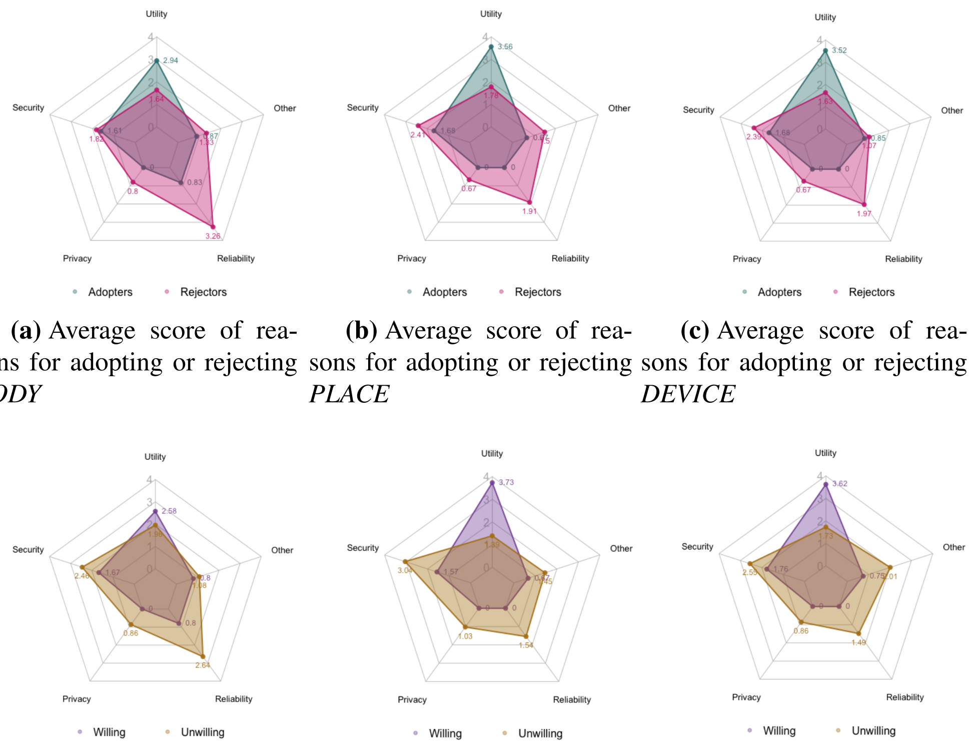
Figure 2.1 shows the number of survey participants who reported that they had either: (1) been unaware of, (2) adopted, (3) rejected, (4) stated interest in, or (5) stated no interest in, each SL method. Based on this data, Table 2.2 presents the adoption, rejection, and interest rates of each SL method. As indicated, the adoption rate for all SL methods is less than 20% and the rejection rate is unanimously higher than the adoption rate. Also, the interest rate is higher than either the adoption or the rejection rate for all SL methods. These findings suggest the following:

- SL is not a widely adopted technology, with an average adoption rate of 13.7%. In comparison, the PIN, fingerprint, and pattern phone unlocking methods had adoption rates of 68%, 47%, and 26% respectively[8], among the SL-capable participants.
- SL had a relatively high rejection rate, with an average of 22.5%. This likely indicates existence of deal-breakers that deter users from adopting it.
- The interest rate in SL among SL-capable participants was 44%. Our observed interest rate is lower than that of the existing IA perception studies, where more than 60% of their participants indicated a willingness to adopt IA [26, 64]. We suspect that this is due to the low ecological validity of such studies, where the participants only experimented with a pseudo-IA scheme that did not suffer from reliability issues, and where the participants did not have to put in the effort to set up the solution on their phones.

Hence, to answer **RQ1**, it seems that with an adoption rate of only 13.7%, SL is not a widely adopted technology. This finding reiterates the importance of investigating **RQ2**, which can help the research community gain insight into possible real-world barriers to IA adoption.

To answer **RQ2**, Figure 2.2 depicts the average rank that the adopters/rejecters or the interested/uninterested participants of each SL method assigned to each category of reasons to justify their decisions. Note that the correlation is different between Adopters/Rejecters. For example, as seen in sub-Figure a, lack of reliability was a major reason for SL rejecters to reject SL, but not a significant reason

[8] The numbers do not add up to 100% because it is possible for one user to have enabled more than one unlocking method.



(a) Average score of reasons for adopting or rejecting *BODY*

(b) Average score of reasons for adopting or rejecting *PLACE*

(c) Average score of reasons for adopting or rejecting *DEVICE*

(d) Average score of reasons for being (un-)interested in *BODY*

(e) Average score of reasons for being (un-)interested in *PLACE*

(f) Average score of reasons for being (un-)interested in *DEVICE*

Figure 2.2: The average importance score that participants assigned to each category of reasons for adopting, rejecting, or being (un-)interested in adopting SL methods.

for Adopters to adopt SL.

In sections 2.3.3 to 2.3.7, we examine in detail the role of each category as a potential barrier to the adoption of SL.

2.3.3 Security

Perceived lack of security has been shown to be a major deterrent for adoption of new technologies. For example, this has been the case with Apple Pay [57] and face unlock [29]. In the case of SL, we observed a perceived lack of security (i.e., the possibility of someone gaining access to the phone without the owner's authorization) to be a major adoption barrier as well.

First, we found evidence of this, first in our qualitative study. Some of the CWU participants justified their unwillingness to use SL by explicitly citing lack of security as the reason. For example, when asked whether he would adopt *BODY*, one of our think-aloud participants (P-TA-15, male, student) said: *"I think [in the case of] the on-body detection [the answer] is no. Even when you're not at home, like you're at the mall or something like that. It's still possible [for someone to unlock the phone]."*

Our survey study confirmed this association further. Among our undecided SL-capable participants, those who indicated an unwillingness to adopt *BODY*, *PLACE*, or *DEVICE* ranked insecurity as one of the top two highest-ranking reason (alongside reliability, as depicted in Figures 2.2d, e, and f) as to why they were unwilling to adopt the SL methods. They ranked it significantly higher than any other category of reasons ($p < 0.05$, Durbin-Conover test). This seems to be a valid concern, as those of our participants who reported having stopped using these methods also ranked insecurity as one of the top three highest-ranking reasons (as depicted in 2.2a, b, and c) for their decision to abandon the feature, again ranking it higher than any other remaining category of reasons ($p < 0.05$, Friedman test).

As further evidence in support of our hypothesis, when we asked participants to rank SL methods against PIN and fingerprint phone unlocking methods, we found that participants who indicated unwillingness to adopt SL ranked fingerprint and PIN as significantly more secure than SL ($p < 0.05$, Durbin-Conover test). We found these results to be in contrast to Khan et al. [64], who reported that 81% of their participants were more satisfied with the level of security that their pseudo-IA scheme provided.

As an illustrative case of such concern, most of our think-aloud participants expressed that SL might allow their family members or co-workers to access their phones without their permission. One participant (P-TA-13, male, student) explained: *"I don't think it [DEVICE] is something that I would use at all, because even if it's at my home and a friend is over or something, they can just unlock my phone. It'll be easy for them to do it. It could be even ... my brother!"* Considering that snooping on mobile devices is known to be both frequent [75] and antagonizing [76], this reasoning appears to be justified.

2.3.4 Privacy

Studies have shown that privacy is a major concern for smartphone users when it comes to authentication [59, 93]. Hence, we hypothesized that privacy concerns may also be a barrier for adopting SL as well.

Our CWU study study partially supported our hypothesis. Some of our CWU participants expressed direct privacy concerns with SL. For example, P-TA-3 (female, health specialist) remarked about SL: *"This [SL] is a potential security asset because if your phone is stolen or lost, your information will not be as easily hacked as if you use the regular lock. I think that it is a toss-up between having your data for the smart lock feature recorded by your phone provider versus potential loss of information."* In general, we observed privacy concerns among our participants to be about manufacturers' unauthorized use of user's data.

Our survey study, however, did not support the role of privacy for SL adoption. Survey participants did not rank privacy as a major reason for abandoning the technology or being unwilling to adopt it (depicted in Fig 2.2).

Caution needs to be used when drawing conclusions about the role of privacy in SL adoption, however. Since we used ranking (instead of rating) in our survey to test the link between security, privacy and SL adoption, security ended up always being the most important factor, potentially masking the effect of privacy. Studies are needed to further investigate the correlation between privacy and SL use.

2.3.5 Reliability

The themes of perceived unreliability (i.e., accidental unlocking or locking the user out) of SL methods and, as a consequence, lack of trust in the technology, kept reappearing in our collected data. While interrelated with security and privacy, these themes deserve their own analysis in relation to SL and its users.

Firstly, the participants in our qualitative studies found that SL lacked reliability, precision, and accuracy. For example, several participants expressed their belief that GPS lacks the precision necessary to be trusted as an unlocking factor. One of our think-aloud participants (P-TA-16, female, unemployed) expressed such concern by stating: *"I think, for example, if I add my home as a trusted place, then if I go to a coffee shop downstairs beside my place, it might still think I'm at* *home [and unlock the phone].”*

The survey data confirmed unreliability’s role as a major adoption barrier as well. Among our SL-capable participants, those who indicated an unwillingness to adopt *BODY* ranked accidental unlocking as one of the top two (alongside security) most important reasons for their unwillingness to adopt SL (depicted in Fig 2.2a), ranking it significantly higher than other reasons ($p < 0.05$, Durbin-Conover test). This seems to be a valid concern, as observed unexpected behavior by *BODY* and accidental unlocking were the top two highest-ranking reasons that our *BODY* abandoners stopped using it too, ranking them higher than any other reason ($p < 0.05$, Durbin-Conover test). The same trend was observed with *PLACE* and *DEVICE* as well (as depicted in Fig 2.2b and c).

Overall, therefore, our results clearly suggest that a perceived lack of reliability is a deterrent of SL. This, in turn, reaffirms the importance of reliability in technology adoption, as shown by Butler and Sellbom [19].

2.3.6 Utility

It is known that perceived utility is an important factor when it comes to technology adoption. Previous studies, such as those by Zhang and Xu [126] and by Huh et al. [57], have shown that users’ perceptions of the usefulness of a new technology correlates positively with their intention to adopt it.

In the case of SL, the results of the CWU study suggest that the main utilities of SL (as perceived by our participants) are:

- **Convenience**: SL can make it easier to unlock smartphones or keep them unlocked, at least under certain circumstances, when compared to conventional explicit smartphone unlocking methods (e.g., PIN or fingerprint).

- **Speed**: SL can make it faster to unlock smartphones, at least under certain circumstances, as compared to conventional unlocking methods.

- **Redundancy (backup)**: SL can be used as a backup method to unlock phones when users cannot unlock their phones (e.g., because they forgot their PINs or patterns).

This was corroborated by the survey study, where SL adopters cited convenience and speed as the top reasons for their decision to adopt SL. These two reasons were ranked significantly higher than any of the other reasons ($p < 0.05$, Durbin-Conover test).

However, the results also showed that the majority of the participants were not convinced of the added benefits of SL. The first evidence of this was observed in the CWU study, where some participants cited a perceived lack of *added* convenience as the reason for their unwillingness to adopt. For example, one of the think-aloud participants (P-TA-11, male, language teacher) remarked about *DEVICE*: *"I probably wouldn't use it, because I can't think of a use for it. I can play music from my phone even though it is locked. So I don't need the trusted device feature."*

The survey data corroborated this finding, where those participants who either had rejected SL, or were unwilling to use it, ranked a perceived lack of utility as the most important reason (depicted in Figure 2.2) for their choice ($p < 0.05$, Durbin-Conover test).[9] In addition, when asked to rank SL against the PIN and fingerprint methods, in terms of the speed and the convenience of unlocking, these participants ranked the fingerprint method to be significantly faster and more convenient than any of the SL methods ($p < 0.05$, Durbin-Conover test). They also ranked the PIN method as significantly more convenient, but slower than SL ($p < 0.05$, Durbin-Conover test).

In summary, the results suggest that it is difficult for Android users to understand or perceive the utility of SL, especially its convenience or speed advantages over the fingerprint method. This contributed significantly to the participants' unwillingness to adopt SL. We believe that there is a potential usability misconception, in that using SL to keep phones unlocked (in trusted places or near trusted devices) will be faster than having to explicitly touch fingerprint scanners and wait a moment for phones to become unlocked. Hence, further research is necessary to determine more effective ways to communicate the usability benefits of SL (and IA in general) to smartphone users.

[9] It should be noted that it is probable that the participants put a perceived lack of utility as the most significant factor for them because they assumed SL was already security, or that they had not considered security yet. Therefore, our results should not be interpreted as meaning that if the users found a utility in SL, they would be immediately willing to adopt it.

2.3.7 Other Adoption Barriers

In addition to the discussed adoption barriers, our participants occasionally expressed other interesting reasons for not using SL. While these barriers were not perceived as important by the participants (as evidenced by how they ranked these reasons for their decision to adopt/reject SL), we believe that some of them are worth discussing, as they might have higher indirect correlations with the users' attitudes. These barriers include the following:

I) Semantics: Some of the CWU participants expressed difficulty with understanding the semantics of SL (i.e., how each SL method locked or unlocked the phone). This made them unable to correctly judge when their phone would be locked. As it has been shown that a compatibility of mental models between old and new technologies correlates positively with users' intention to switch to the latter [126], we theorize that difficulty with understanding SL semantics might have contributed to the low adoption rates we observed. However, our survey participants rarely cited difficult semantics as an important reason for being unwilling to adopt or deciding to abandon it. This is perhaps because they are unaware of the fact that they have misunderstood how SL works. Despite this, we still believe it is important that further studies be conducted to gain insight into how smartphone users understand the IA semantics. During our CWU study, we observed that a misunderstanding of semantics can lead to a misjudgment of when the phone would be locked. This can have dire consequences, as it can lead to unauthorized access to the phone. Yet, to the best of our knowledge, no study on how smartphone users understand IA semantics has been reported so far.

II) Use of Unlocking: Some of the CWU participants stated that they were unwilling to adopt SL because they were satisfied with their current unlocking method. While most of such participants were using fingerprint unlocking, we occasionally observed PIN users to express the same sentiment. As such, we hypothesized that SL might be more appealing to those who are not locking their phones. Our survey data, however, did not support this hypothesis: 17.5% (N = 55) of our survey participants reported not using any locking on their phones. A chi-squared test of association showed that such participants are not significantly more likely to be willing to adopt SL ($p > 0.05$, chi-squared test). More studies, however, are needed to fully investigate and compare the usability of SL with existing unlocking methods, such as fingerprint.

III) Usability of the SL UI: Through the CWU study, we found various usability issues with the SL UI (e.g., inconsistencies and ambiguities) that caused confusion for the participants. For example, the UI for *PLACE* lacked a tutorial screen to explain to the user how the SL method worked, whereas *BODY* and *DEVICE* had such screens. We found that such usability issues affected participants' trust in the technology, which can be a major barrier to its adoption. For example, after experiencing the SL UI, one of our CW participants (P-CW-8, male, student) mentioned: *"All these inconsistencies make you wonder if the Android team at Google cares at all ... I don't trust this [SL] at all."* While our survey participants rarely cited the usability of the SL UI as a reason for their decision to reject SL, we observed that misunderstanding the semantics can lead to users forming inadequate mental models. However, to the best of our knowledge, there are no guidelines or heuristics reported for designing or evaluating the UI for IA on smartphones. Providing such guidelines or heuristics can help system designers communicate the semantics of IA and its capabilities more efficiently, to avoid user frustration and dangerous errors. We believe this to be an important knowledge gap and a good avenue for future work.

2.4 Smart Lock Technology Acceptance Model (SL-TAM)

In the previous sections, we investigated individually factors that correlate with users' intention to adopt SL. However, our findings did not lead us to clear picture of the users' overall attitude towards adopting SL. To provide such a picture, we investigated whether there were theoretical frameworks that could explain our findings in a more structured way. We found that the technology adoption model (TAM) [28] is often used for such purposes.

Hence, in this section, we rely on TAM to structure all of our findings into a framework for reasoning about smartphone users' SL adoption decision. Providing

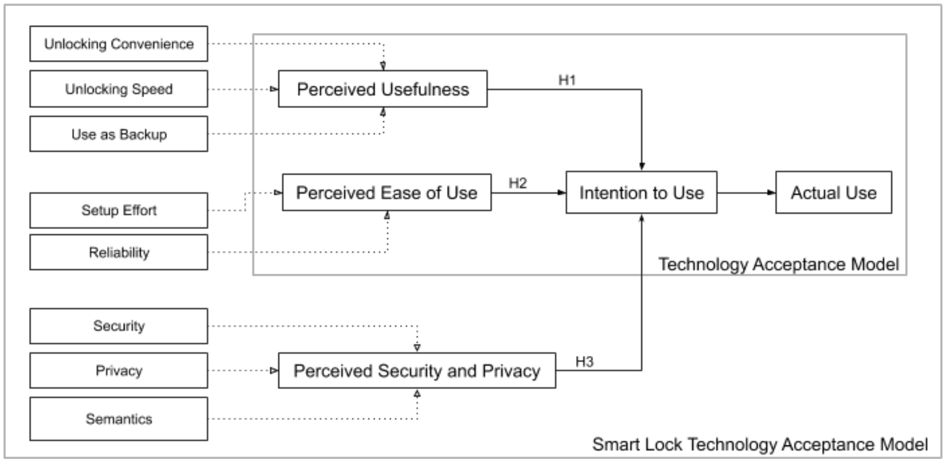


Figure 2.3: Smart Lock technology acceptance model.

such a framework is valuable because while the existing literature, e.g., TAM [28], could be used for predicting SL adoption, resulting conclusions would not be supported by existing empirical data. Also, due to the abstract definition of predictors in TAM, it would be difficult to interpret what they entail in the context of IA. As explained below, TAM would also miss the link between security and SL adoption.

To devise the framework, we first investigated how our CWU findings conformed with TAM [22, 28] and how TAM needed to be extended for the case of SL. This is presented in Section 2.4.1. Then, in Section 2.4.2, we use our survey data to test the extended TAM model.

We should note, however, that we did not use TAM in advance to inform the design of our survey study, and instead relied on the results of our qualitative study to do so. This was to prevent introducing confirmation bias into our results (i.e., trying to confirm what TAM conjectures about SL adoption, rather than confirming our own observations in the qualitative CWU study).

2.4.1 Devising SL-TAM with CWU Findings

As discussed in Section 2.1.3, TAM introduces two factors that influence adoption attitudes toward a new technology: *perceived usefulness* and *perceived ease of use*. In the following, we examined how these factors manifested themselves in our

CWU study and how they affected participants' adoption decision:

1. **Perceived usefulness**: TAM theorizes that to adopt a new technology, potential users must find it useful. Our CWU results (discussed in Section 2.3.6) showed us that the usefulness of SL translates to whether users think it can make unlocking easier or faster, or can be used as a backup unlocking method. Subsequently, as TAM would predict, we found that users' perception of these SL benefits directly correlates with their willingness to adopt it. Based on this observation, we hypothesize that *perceived usefulness* indeed is linked to the intention to use SL, and it is determined by the unlocking convenience, speed, and backup use of SL (depicted as *H1* in Figure 2.3).

As discussed before, we found another deterrent keeping our CWU participants from adopting SL to be security and sometimes privacy concerns. Such concerns, however, did not seem to be strictly related to either *usefulness* or *ease of use*. We discovered that a similar observation has been made by other studies that applied TAM to specific domains, such as Sun et al. [111] who applied it to OpenID, or Vijayasarathy [117], who applied it to online shopping. In case of SL, it could be argued that security and privacy concerns are related to the *usefulness* of the technology (after all, SL is designed to improve security). However, the significant prominence of security and privacy concerns in our findings prompted us to consider them as a separate factor, to fully capture their importance. As it is evident in the results presented so far, the *usefulness* factors we discussed before mostly deal with pragmatical hurdles of setting-up SL and/or dealing with its errors, rather that whether it is achieving its core purpose (which is what the *security and privacy* factor is aiming to capture).

To address this problem, we expanded our SL-TAM to include a new predictor called "perceived security and privacy" (depicted as *H3* in Figure 2.3), which is determined by the following factors:

1. **Security**: As we discussed in Section 2.3.3, some of our CWU participants explicitly cited insecurity as the reason they were unwilling to adopt SL.

2. **Privacy**: As discussed in Section 2.3.4, some (but not all) CWU participants cited lack of privacy as a potential drawback of SL.

Table 2.3: Results of the binomial logistic regression (BLR) tests to validate the correlation between SL-TAM predictors and survey participants' intention to adopt SL.

SL-TAM Predictor	**Dependent Variable**		
(Independent Variable)	BODY	PLACE	DEVICE

Perceived Usefulness	p-value = 0.001
	OR = 7.16
		p-value = 0.001
		OR = 3.27
			p-value = 0.001
			OR = 9.40

Perceived Ease of Use	p-value = 0.016
	OR = 9.82
		p-value = 0.275
		OR = 3.27
			p-value = 0.090
			OR = 3.29

Perceived Security and Privacy	p-value = 0.744
	OR = 1.09
		p-value = 0.001
		OR = 5.22
			p-value = 0.019
			OR = 2.34

3. **Semantics**: As discussed in Section 2.3.7, we observed that (in-)correct understanding of SL semantics can lead to misconceptions about SL security.

The final SL-TAM is depicted in Figure 2.3.

2.4.2 Testing SL-TAM with Survey Data

We tested SL-TAM by evaluating how our survey data conforms with it, for each SL methods separately. Specifically, we evaluated how our hypothesized factors ("perceived usefulness," "perceived ease of use," and "perceived security and privacy") correlate with SL-experienced participants' decision to either use SL or abandon it (to preserve ecological validity, we focus on this group of participants alone).

First, Figure 2.2 depicts the average importance score that the participants assigned to each group of reasons for their decision to adopt or reject SL. We used these scores in the following way to test SL-TAM: the utility score was used to represent "perceived usefulness," (as it is an aggregated score of perceived convenience, speed, and backup use of SL); the reliability score was used to represent "ease of use," (our survey is not capable of evaluating "set-up effort"); and the sum of the security and privacy and semantics scores was used to represent "perceived security and privacy."

To evaluate how these predictors correlated with SL adoption decisions, we report the results of three binomial logistic regression (BLR) tests (one for each method). In each test, the decision to adopt the SL method is the dichotomous dependent variable, and the hypothesized factors are the predictors. To ensure the validity of the tests, we checked for multicollinearity between variables but found that this was not an issue ($1.0 < VIF < 1.2$). Also, the model fit measures were satisfactory for all three BLR tests ($R^2 = 0.505$ for BODY, $R^2 = 0.561$ for PLACE, and $R^2 = 0.571$ for DEVICE).

Table 2.3 presents the results of the BLR tests. They show the following:

- **Testing H1**: Perceived usefulness is a statistically significant predictor for BODY, PLACE and DEVICE adoption ($p < 0.01$ in all three tests). The high odds ratios (OR) further demonstrates the high strength of this correlation. With BODY, for example, one unit increase in the "usefulness" ranking (e.g., by improving the unlocking speed of SL methods) increased the chances of adoption nearly sevenfold.

- **Testing H2**: Perceived ease of use is a significant predictor of BODY and DEVICE adoption ($p < 0.01$ in the corresponding BLR tests). The ORs further affirm that the observed correlations are strong. In case of DEVICE, for example, a one-unit increase in the "ease of use" ranking resulted in a 3.29 times higher chance of adoption. In the case of PLACE, while the correlation is not statistically significant, the OR still shows a strong association.

- **Testing H3**: Perceived security and privacy is a significant predictor of PLACE and DEVICE adoption ($p < 0.05$ in the corresponding BLR tests). Furthermore, ORs show the correlations to be strong. For example, in the case of PLACE, a one-unit increase in "security and privacy" ranking increased the chances of adoption by 5.22 times. In case of BODY, "perceived security and privacy" was not a statistically significant predictor of adoption ($p > 0.05$ in BD BLR test). As discussed in Section 2.3, the main deterrent for BODY is reliability (accidental unlocks), which is reflected in the "perceived ease of use" predictor.

In conclusion, BLR testing shows that our survey data conforms highly with the

SL-TAM model. We hope that this model can inform the design of future SL-like authentication schemes by shedding light on the important factors that can attract or deter smartphone users from IA-based unlocking.

2.5 Limitations

Any generalization of our findings needs to be made carefully, due to the following study limitations:

1. We mentioned “Smart Lock” in the study advertisement, and therefore our sample might have been skewed toward participants who are using or interested in SL. This bias could have potentially caused an overestimation of SL adoption and awareness rate. However, even with the adoption rate being overestimated, the results (14%) still suggest that SL is far from being widely adopted.

2. The cross-sectional design of our qualitative study might have prevented us from investigating the effects of prolonged use of SL on participants’ perceptions of it. We believe, however, that directly surveying SL adopters addressed this possible weakness to some extent.

3. The limited size and diversity of our sample might have prevented us from uncovering all potential factors that correlate with smartphone users’ perceptions of SL. We aimed for diversity in certain demographics that were known to be correlated with smartphone authentication perception. However, there is no guarantee that our sample has covered all possible combinations of the purpose of use of phones. This limitation could have caused us to over-estimate the adoption rate of SL (as some missed contexts of use that we missed might be more privacy-sensitive, leading to lower SL adoption rate). But an over-estimation would not affect our conclusions, as our estimated rate is already low (14%). More importantly, however, this limitation could have caused us to under-estimate the adoption rate of SL, as certain contexts of use might not require strict authentication. Such under-estimation could entail the existence of SL adoption intention factors that are missing from SL-TAM. For example, we did not aim for studying the use of SL in phone

sharing settings because, as we discussed in Chapter 1, current physical security solutions are not designed for sharing. At the same time, however, phone sharing might actually benefit from SL (e.g., when sharing happens at home, PLACE could be an apt solution to balance security and convenience of authentication). Therefore, further research is required to investigate the use of SL in phone sharing settings, and investigate whether that would lead to a further extension of SL-TAM (most likely with regards to the antecedents of the "usefulness" factor).

4. Even though SL appeared the logical choice for an IA case study at the time of writing (as it was the first and only widely deployed IA scheme), it is difficult to determine the extent to which users' perceptions of SL can be extrapolated to their perceptions of IA in general. Some SL concerns could be logically applicable to any other IA scheme. For example, security and privacy concerns could be one such case as they most likely correlate with the sensitivity of the data stored on the phone (see Chapter 1 for a detailed discussion), rather than the specific SL implementation. However, some others might be specific to SL (e.g., the difficult semantics as SL is a special case that combines IA and EA). With regards to semantics, a special case, for example, can be a hypothetical "perfect" IA solution that does not suffer from any authentication false positive or negatives. Would correct understanding of semantics be as important in the context of that system? Or can the users simply be oblivious to the semantics and always assume that the system will "do the right thing?" We can only speculate based on our results that semantics might not be as important then, as confusions in our study often happened when the system behaved differently than how it was described in the user interface (e.g., the discrepancy between the word "Smart Lock" and its ability to lock the phone). However, since SL is currently the only widely-deployed IA solution (and, hence, no comparison between different solutions is possible), it is hard to gauge what factors of adoption intention might be specific to SL (rather than IA in general). Perhaps once more IA schemes become widely-available, comparative studies can be conducted to uncover factors that are specific to this implementation (that is SL). As of

now, however, we cannot claim that our results are generalizable beyond SL.

5. Our participants self-reported their prior awareness of, and experiences with, SL. As is common with self-reported data, it is possible that the participants' answers might not be completely reflective of their real-world behavior. While we eliminated those responses that showed clear inconsistencies in the data (See section 2.2.2), there is still the chance that some participants might not have answered truthfully. This might limit the external and ecological validity of our results.

6. There are known limitations with MTurk, in regards to the diversity of the samples it can provide (e.g., in terms of tech-savviness of the participants). Such limitations limit the genralizabilty of our results beyond the the original sample. We will discuss these limitations further in Chapter 4, as we use MTurk for all of our studies.

2.6 Conclusion

Smart Lock is the first massively deployed and commercialized IA solution that allows smartphones to be automatically unlocked using a combination of contextual (e.g., location) and behavioral (e.g., body movement) authentication factors. To understand how this first widely deployed IA method is perceived by Android users, we conducted a multi-method qualitative study with 27 participants. It composed of cognitive walkthroughs, think-aloud sessions, and interviews, followed by an online survey on Amazon Mechanical Turk involving 343 Android-using participants.

The results suggest that perceived lack of reliability, utility, and security negatively correlate with Android users' intention to adopt SL, leading to a low (14%) adoption rate. Reliability-wise, participants were concerned that SL could lead to frequent accidental unlocks and pocket dialing. Utility-wise, SL was perceived as not being of enough value as it could not increase the unlocking convenience (the required physical and cognitive effort) or speed, or be used as a backup unlocking method. Finally, as far as security was concerned, participants were worried that adopting SL could lead to unauthorized access to their phone, by their family members or co-workers, for example.

To provide a framework for reasoning about SL adoption intentions, we structured our findings into an SL-specific extension of technology acceptable model (TAM). Our SL-TAM theorizes that there are three main factors affecting users' intention to adopt SL: “perceived usefulness,” ”perceived ease of use,” and ”perceived security and privacy.” “Perceived usefulness” is determined by the convenience and speed of unlocking with SL and whether it is possible to using it as a backup unlocking method. “Perceived ease of use” is determined by the amount of effort it takes to set up SL, and its reliability, and “Perceived security and privacy” is determined by the actual security and privacy of SL and how difficult its semantics are for users to grasp. We tested SL-TAM using our survey data, which showed high predictive power.

Based on the findings, we recommend that, to improve how smartphone users perceive an IA scheme like SL, its added value (in terms of speed or convenience of unlocking) needs to be communicated to users in a clear and accessible way. The scheme also needs to be reliable and trusted by the users, and the chances of malfunction (e.g., failure to lock the phone automatically) should be minimized and disclosed. In addition, to help users develop and maintain adequate mental models of the technology, the semantics of any IA scheme should be clearly communicated to users, so that they can become comfortable with it, learn how to use it effectively, and avoid dangerous errors.

Chapter 3

Investigating Smartphone Users' Difficulty with Understanding Implicit Authentication Semantics

This chapter presents the results of a mix-method study to investigate Android users' understanding of Smart Lock (SL) semantics. It is a first step towards gaining insight into smartphone users' perception of IA semantics in general. First, in section 3.1 we provide an overview of the existing literature on users' understanding of computer security semantics. Then, in section 3.2, we describe the design of our research and the studies that it comprises. Subsequently, we discuss the methodology and results of our qualitative study, in section 3.3, and those of of our quantitative study, in Section 3.4. In Section 3.5, we discuss the implications of our findings and provide design recommendations for future IA UX on smartphones. Section 3.6 discusses the limitations of our study, and, finally, Section 3.7 concludes the chapter.

3.1 Related Work

It has been well established that correct understanding of computer security semantics is of utmost importance. Firstly, a number of studies have shown how misunderstandings can lead to dangerous security errors. For example, Raja et al. [96] demonstrated that users' incorrect understanding of how the Windows firewall operates can lead to dangerous misconfigurations, potentially exposing users' PCs to remote attacks. Similarly, Chiasson et al. [20] investigated the importance of semantics comprehension when it comes to using password managers, observing that incorrect understandings lead to misconceptions about password security. Secondly, a separate line of work has demonstrated that incorrect understanding of semantics can hinder effective risk communication. It was shown that security experts and non-experts have different understandings of common security risks [8], and risk communication based on experts' understanding might be ineffective [71].

Users' understanding of smartphone security features has also been studied. For instance, Felt et al. [35] investigated smartphone users' understanding of Android permissions. They found that users have misconceptions about how the permissions work, causing them to be unable to comprehend the risks associated with installing apps. Similarly, Lin et al. [69] observed that users' decisions regarding granting permissions to Android apps depends on their understanding of mobile privacy, which is not uniform. Another example is from the secure messaging domain, where Schroder et al. [103] observed that misunderstanding encryption semantics often prevents the users of SIGNAL [106] from correctly verifying the authenticity of end-to-end encryption keys.

However, there have been no studies investigating users' comprehension of IA. To gain such insight, we use the same qualitative dataset as in the previous chapter and the same methodology for analyzing it. However, the focus of analysis is vastly different here, due to differences in research questions. Whereas in the previous chapter we investigated why users adopt or reject SL, this chapter evaluates how users understand SL semantics and what aspects of it can be confusing for them. This chapter also uses a completely different quantitative dataset (we conducted separate online surveys with different questions and participants).

3.2 Study Design

Our research study was designed to answer the following two research questions:

- **RQ1**: How well do Android users understand the semantics of SL? Particularly, what aspects of SL can cause confusion for them?
- **RQ2**: What factors (such as demographics or depth of smartphone adoption) are linked to Android users' understanding of SL?

To answer these questions, we first conducted a qualitative cognitive walkthrough with users study (also previously reported in Chapter 2). This study, which we describe in detail in Section 3.3, informed us about how well smartphone users understand SL, and what particular aspects of it might be misunderstood. We used this insight to design our second study, which was an online survey, presented in Section 3.4. The aim of the survey was to verify and expand upon the qualitative findings, leveraging a relatively representative sample of the smartphone user population. All of our data collection and analysis procedures were reviewed and approved by UBC BREB.

3.3 Qualitative Study: Expert Reviews and Think-Aloud Sessions

3.3.1 Methodology

Our first study was the cognitive walkthrough with users (CWU) [46, 70, 73] that we previously reported in Chapter 2. However, now we describe specific aspects of it that are more relevant to SL understanding, rather than SL perception. The study consisted of two separate parts:

Part I) Cognitive Walkthroughs: To evaluate the learnability of SL (and its UI), we conducted two cognitive walkthrough sessions [124] involving 10 HCI-proficient participants. To qualify as HCI proficient, participants had to have at least 4 months of formal HCI coursework.

In each session, a group of participants (6 in the first session, 4 in the second) walked through the SL UI[1] screen by screen, emulating the actions an SL user would perform to achieve a desired task (e.g., enabling BODY) and discussing any aspects of it could cause confusion.

To make sure HCI-proficient users would emulate the actions of an SL user realistically, we created a list of common tasks that an SL user might want to do with the UI, with the sequence of actions that they would perform to achieve each task. To compile a comprehensive task list, two researchers used SL on their phones for a period of two weeks. They then collaboratively composed a list of SL UI task affordances and the action sequences for each task. We compared the resulting list with official SL documentation [41], which showed conformance. For presentation to participants, we chose a wording for each task that reflected the goal of the task (e.g., "Enable On-Body Detection"). The full list of tasks is provided in Section A.1 in the Appendix.

During the sessions, we also asked the participants to give written answers to questions about how each SL method locks or unlocks the phone. These questions were added to the handout (available in Section A.1) that we provided to them as part of the cognitive walkthrough protocol.

We should note that we tested our methodology design before conducting the main 2 cognitive walkthrough sessions described above. The pilot sessions (discussed in Section 2.2) showed that our study design was effective, as it identified several confusing aspects of SL that were later confirmed by our main studies.

Part II) Think-Alouds: To strengthen our understanding of how smartphone users comprehend SL, we conducted 16 think-aloud sessions with ordinary smartphone users. We did so because, as commonly cited in the literature [73, 124], cognitive walkthroughs lack real user involvement which reduces the ecological validity. We addressed this shortcoming by combining our the cognitive walkthrough findings with the results of think-aloud sessions with ordinary (non-expert) users. We chose this method because think-aloud protocol is found to be a good tool for evaluating comprehension [118, 125].

In each session, a participant was instructed to perform tasks (the same as the ones we used for cognitive walkthroughs) while thinking out loud about their ex-

[1] Section A.4 in the Appendix provides screenshots of the main UI screens of each method

Table 3.1: Demographics of the study participants.

Parameter	Property	% (#) of participants	
		CWU ($N = 26$)	**Survey** ($N = 331$)
Gender	Female	53.8 (14)	35.3 (117)
	Male	46.2 (12)	64.0 (212)
	Other	0.0 (0)	0.7 (2)
Age	19-24	38.5 (10)	9.7 (32)
	25-34	57.7 (15)	42.6 (141)
	35-44	3.8 (1)	29.6 (98)
	45-54	0.0 (0)	10.0 (33)
	55-64	0.0 (0)	6.9 (23)
	65-74	0.0 (0)	1.2 (4)
Education	< High School	0.0 (0)	0.3 (1)
	High School	7.7 (2)	39.0 (129)
	Bachelor's	38.7 (10)	51.1 (169)
	≥ Master's	53.4 (14)	9.6 (32)
Ethnicity	White	N/A	77.3 (256)
	Black	N/A	8.0 (26)
	Hispanic	N/A	6.0 (20)
	Asian	N/A	6.0 (20)
	Other	N/A	2.7 (9)

perience and understanding of SL (we were particularly interested in whether anything confused them about SL). The participants were also given a handout and asked to submit written answers to questions about SL semantics. The sessions were audio recorded and transcribed to accurately capture participants' verbally expressed thoughts. Additionally, while the participant was thinking out loud, two researchers were taking notes about whatever caused confusion.

To test our think-aloud methodology, we conducted 2 pilot think-aloud sessions (previously described in Section 2.2). While the results showed our approach to be effective in evaluating SL understanding, we occasionally observed "search-and-click" behavior from the participants (they blindly followed instructions without trying to explore and understand SL). To mitigate this issue, we consulted literature on think-aloud protocol [118, 125] and made the following adjustments to the study design:

1. We would inform participants at the beginning that we would ask them detailed questions about SL semantics. We believed this would motivate them to understand SL.

2. We would remove the step-by-step task guidance (action sequences) from the handouts and provide participants with the goal of each task only. Note that we retained the step-by-step guidance for cognitive walkthrough (expert) participants to facilitate discussion among experts, as recommend by literature [70, 73].

3. We would remind participants to think aloud and clarify the reason for their actions, every time they went silent during the sessions.

4. We would conduct exit interviews with the participants, asking them to verbally clarify their written answers to the SL semantics questions in the handouts, to further gauge their level of SL understanding.

We then conducted two additional pilot sessions. These showed our modifications are effective, as we observed very few instances of "search-and-click."

Data Analysis: The data collected from our CWU study included the transcribed audio recordings of all sessions, participants' written answers to the questions in the handouts, and researchers' field notes. To analyze this data, we used a Thematic Analysis (TA) approach. We chose TA because it is shown to help examine emerging themes from textual data in a transparent and credible way [15, 47, 48]. We followed the steps described in Braun et al. [15]:

First, two researchers coded all the data. To do so, they examined whether a participant's understanding of SL (expressed through either written answers in handouts or verbally expressed thoughts) deviated in some aspect from the ground truth of SL semantics (which we obtained from official SL documentation and verified by our own internal testing)[2]. If it did, the researchers coded the text with a label that reflected the aspect of SL that was misunderstood.

[2] To simplify the study design, we count any deviation from the official SL documentation to be a misunderstanding. However, we acknowledge that this might not always be the case. As we will discuss later, reliability issues might interfere with a user's correct understanding of SL (i.e., the user knows the "correct" semantics of SL given in its documentation. However, their answer semantics questions based on their own experience with SL, which could contradict the documentations)

Table 3.2: Number of CWU participants who correctly understood the un(locking) semantics of each SL method.

Semantics	% (#) participants				
	BODY	**PLACE**	**DEVICE**	**FACE**	**VOICE**
Locking	76.9 (20)	73.1 (19)	57.7 (15)	34.6 (9)	30.8 (8)
Unlocking	30.8 (8)	76.9 (20)	73.1 (19)	88.5 (23)	80.8 (21)

Second, the researchers resolved differences in coding through in-person meetings. Differences happened mostly when there were discrepancies between our sources (e.g., between a participant's written and oral answers). In such cases, the researchers present at the CWU session decided which source reflected participants' understanding of SL the best and retained only the codes associated with that source. This process was done iteratively and continued until inter-rater reliability reached a satisfactory 85% (i.e., the coders agreed on which code to use for 85% of the words in the dataset).

Third, each researcher studied all the codes, merging them to draft themes of SL confusion and its antecedents. They then discussed and agreed upon the themes and drafted the results.

3.3.2 Results

We found SL misunderstandings to be prevalent and specific. Most of the CWU participants had difficulty understanding the semantics of at least some SL methods, as Table 3.2 shows.

Misunderstandings, however, were not uniformly distributed across the SL methods. As evidenced by Table 3.2, we found that PLACE and DEVICE were generally easier to understand, due to most participants having prior experience with Bluetooth and GPS. On the contrary, the unlocking semantics of BODY and the locking semantics of FACE and VOICE were the most confusing. Overall, by thematically analyzing all instances of participants being confused, we identified 4 different categories of SL misunderstanding:

Capabilities of SL

The participants were unsure what SL and its methods were capable of, in terms of locking or unlocking the phone.

Firstly, just the name "Smart Lock" was already confusing. Some participants interpreted it as the ability of the phone to lock itself when the surrounding environment is deemed insecure. However, after interacting with the UI, most of them concluded that SL was rather mostly about unlocking.

The potential for this misinterpretation was first brought up by one of the HCI-proficient participants (code-named P-CW-3) who stated:

> *"... it's called Smart Lock. But it's really more like Smart Unlock because it's not really locking your phone ... it's not as clear about when it actually locks things ... it's more clear about when it unlocks things."*

A think-aloud participant, P-TA-2 (female, 36, software engineer), voiced the same concern by explaining:

> *"... the description [of the SL UI] is really confusing to me ... because it says it keeps your device unlocked when it's safe with you ... [but] I feel like if it's keeping it unlocked, it should be called Smart Unlock, as opposed to Smart Lock. Because when I think Smart Lock, I think it knows when to lock itself. But, the first thing it [SL UI] is talking about is it knows when to keep itself unlocked."*

Interestingly, we observed this misinterpretation to be actually made by some think-aloud participants. P-TA-4 (female, 40, personal trainer) who was non-tech-savvy, believed none of the SL methods could automatically unlock the phone (even after performing all the tasks), citing this naming convention as the reason.

While this was rather a rare example in our data, our findings suggest that overall naming can have implications on users' understanding of IA. The examples above shows that the term "Lock" in "Smart Lock" might give some users the impression that the feature cannot automatically unlock the phone (because they believe it is engineered to lock the phone, not unlock it), which could lead to dangerous errors by users.

Apart from the SL naming, however, we found the discrepancy between the capabilities of SL methods to be even more confusing. The fact that BODY cannot unlock the phone automatically while PLACE and DEVICE can was startling, even to the HCI-proficient participants. Some of them incorrectly believed BODY could both lock and unlock the phone, such as P-CW-2, who stated:

> *"when there is someone carrying it, it [BODY] will unlock [the phone] automatically."*

or P-CW-4 who believed:

> *"When I am walking, moving or the phone is in motion [, the phone will unlock]."*

Unsurprisingly, we saw the think-aloud participants make the mistake as well. For example, both P-TA-12 (male, 28, flight attendant) and P-TA-3 (female, 27, unemployed) incorrectly believed that BODY could automatically unlock:

- P-TA-12: *"[BODY] unlocks the phone when you are in movement, like holding the phone."*
- P-TA-15: *"[BODY] unlocks the phone while the sensor is on your body and detects motion."*

Therefore, it seems that inconsistency in capabilities is a clear detriment to users' understanding of SL. There might be technological reasons for this inconsistency (e.g., limited accuracy of movement detection). However, as far as the users are concerned, any discrepancy between capabilities causes confusion that could lead to dangerous security errors (e.g., misjudging whether BODY would unlock the phone).

Finally, we found another capabilities-related confusion to be the mixture of IA and EA. Among our participants, this mixture often created unmet expectations about the capabilities of EA methods (FACE and VOICE), which are not capable of automatic (un)locking. Evidently, nearly 70% of our think-aloud participants mistakenly believed FACE and VOICE could also automatically lock the phone. For example, when interviewed about FACE semantics, P-TA-1 (male, 31, unemployed) remarked:

> *“[It locks the phone] when someone that’s not me looks in the camera.”*

Similarly, when asked how VOICE locks the phone, P-TA-9 (female, 22, research assistant) stated:

> *“[It locks the phone] When it doesn’t recognize my voice.”*

Our further probing with these participants showed that such understandings are caused by VOICE and FACE (which are EA methods) being packaged together with other IA methods in SL. This leads the participants to assume that since DEVICE and PLACE can automatically lock the phone, FACE and VOICE should be capable of it too. Subsequently, participants like P-TA-1 and P-TA-9 try to justify their understanding by coming up with incorrect explanations like the ones above.

The Modalities (Authentication Factors) of SL

Most participants did not fully understand what kind of data SL used to identify them. As such, they were unsure what they had to do to make SL (un)lock the phone.

In case of BODY, the culprit was the ambiguity of the term “motion.” The text presented in the BODY UI describes the feature as being able to keep the phone unlocked for as long as it is in “motion.” However, what exactly constitutes “motion” is not clearly communicated. An HCI-proficient participant (P-CW-7) voiced this concern by stating:

> *“... but also ‘motion’ seems to be the key and I don’t understand what kind of motion? ... like when I’m running?”*

Another cognitive walkthrough participant (P-CW-6) similarly stated:

> *“The text [description of ‘motion’ on [the BODY UI] is not fully clear to me what it means. It says it will be unlocked by the user holding or carrying the device. So, does it mean if I put it in my pocket and [am] moving ... it’ll be unlocked?”*

P-CW-6 was specially concerned with this scenario because she believed it could potentially lead to pocket dialing and the related privacy issues.

Another concern regarding "motion" was brought up by P-CW-4. He was wondering whether the "motion" required by BODY needed to be body specific (as the method is named "on-body detection"). He explained:

> *"I'm not sure if [BODY] will unlock with motion without being on body. It's vague as to whether the phone needs to be on the body or just in motion."*

Overall, we found the concerns with the definition of "motion" valid, as we observed think-aloud participants to exhibit such confusions. As an example, when P-TA-14 (male, 19, student) tried to make the phone lock after setting up BODY, he put it on a nearby desk, but the phone did not lock. He then voiced his frustration by stating:

> *"I expected [the phone] to [lock] if it's not in my hand ... how would it not lock if it's far away?"*

By probing further, we found that what was unclear to the participant was how BODY detects "motion," what the intensity of the motion should be, and how long the phone takes to detect it. Having this knowledge is important for the participant to make a correct judgement about the state of the phone's security.

Ultimately, as explained by several participants, why "motion" is confusing becomes clearer when we compare BODY to a conventional unlocking method, such as fingerprint. Using fingerprint is fairly straightforward—you put your finger on the sensor and the phone unlocks (and there is no automatic locking). In comparison, it is not very clear how the phone needs to be moved, with what intensity, and for how long, and whether the movement needs to resemble body movement to make BODY function. It's too nuanced. As we saw in the example with P-TA-14, this leads to confusion and frustration.

Lastly, we should note that "motion" was not the only confusing SL modality. DEVICE semantics was also found difficult to fathom. Some non-tech-savvy think-aloud participants lacked the knowledge to understand how Bluetooth devices are authenticated to each other, concluding that if there is an untrusted device around, the phone will lock. As an example, when we interviewed P-TA-3 (female, 27, unemployed) about when DEVICE would lock the phone, she answered:

> *"... it will lock the phone when you are near a device you have not added as a trusted device."*

When we asked why she thought so, the participant explained that this was simply what she expected from DEVICE, based on what she saw on the UI. She clarified that she had no prior understanding of Bluetooth, and this is what the SL UI led her to believe about DEVICE capabilities. This case showed us that correct understanding of IA methods may sometimes require deeper technical knowledge (e.g., how Bluetooth authentication works) than most ordinary users have.

The Interoperation of SL Methods

How the SL methods interoperate with each other was confusing for the participants. They did not know when the phone would be (un)locked if more than one method was enabled at the same time.

This problem was brought up by several cognitive walkthrough participants, such as P-CW-6 who stated:

> *"Will there be setting conflicts [with SL]? For example, no trusted device [is connected] to my phone, but I'm still at a trusted place; will the phone get locked?"*

Similarly, P-CW-2 remarked:

> *"I [as a user] am kind of confused about how these function[s] work together? Are they independent or overlapped in some ways?"*

This was shown to be a valid concern, as we found SL interoperation to be unclear to most think-aloud participants. P-TA-5 (32, female, health instructor), for example, said the following about PLACE capabilities and its interaction with BODY:

> *"I would guess that when I leave the trusted place, it [the phone] locks. But, I'm not sure how this [PLACE] interacts with on-body detection."*

Interoperation was specially confusing when it was about the interactions between IA and EA methods. As mentioned before, EA methods are not capable of automatically locking the phone. However, when asked when FACE and VOICE lock the phone, some participants thought of SL as a coherent entity and tied the locking capabilities of these EA methods to the IA ones. For example, when asked about how FACE would lock the phone, P-TA-3 stated:

> *"When I am not in motion, the phone would ask me to lock the phone by taking an image of my face."*

Similarly, when we asked P-TA-10 (male, 33, financial consultant) about how VOICE would lock the phone, he answered:

> *"When you set it [the phone] down, as in there is no motion."*

Overall, such observations seem to suggest that interoperation of SL methods is not a matter that is easily understandable. Misunderstandings about this can lead to dangerous security errors, such as leaving the house assuming the phone will be locked, whereas it may not be because of another method like DEVICE. Surprisingly, the SL UI does not specifically address how SL methods interoperate at all, leaving it as a guessing game for its users.

The Range Parameters of SL

Understanding when the phone would be (un)locked by PLACE and DEVICE required knowledge of their range parameters, which most of the participants lacked. Our interviews with the think-aloud participants showed that most did not recall the 80-meter operational range of PLACE or the 100-meter one of DEVICE (shown on the setup screens for each method). For example, when we asked P-TA-10 (male, 33, financial consultant) about when DEVICE locks the phone, he stated:

> *"When you take it away from the added trusted device. But how far? Nobody knows!!"*

His exclamation specifically mentioned his lack of knowledge of the range parameter. This knowledge gap may seem insignificant at first. However, it can be essential for correct IA comprehension. This is because these parameters specify the boundaries of security for users. For example, if one user does not know the 80-meter range of PLACE, they might assume that their phone would be locked when not inside their house, where, in actuality, it may not be, when left in their car parked out front.

Interestingly, even if participants knew about the range, PLACE reliability issues sometimes interfered with their correct understanding. It sometimes happened during our study sessions that PLACE failed to function as expected.[3] Such issues caused some participants to doubt their correct understanding of PLACE semantics. For example, when asked how PLACE unlocks the phone, P-TA-14 (male, 19, student) responded:

> *"I don't know. It didn't work and didn't unlock the phone at current location. I expect it to unlock in the room [where study was conducted]."*

Therefore, the data suggests that unreliability is a potential source of confusion. Evidently, even if the semantics of an IA method are clearly conveyed to the user, intermittent operational failures can cause users to doubt their correct understanding, leading to possible dangerous errors (e.g., in case of P-TA-14, the assumption that PLACE cannot automatically unlock the phone).

Regarding range parameters of SL, several of our cognitive walkthrough participants argued that the issue might be the way range parameters are communicated to the user by the SL UI. We discuss this matter further in Section 3.5.

In the end, to summarize our qualitative findings, our study suggests that SL misunderstandings are common. We found that for users to understand SL correctly, they need to know what each SL method is capable of, what data it uses for authentication, and what its operational parameters are. They also need to know how SL methods interoperate in case more than one is enabled at the same time.

3.4 Quantitative Study: Online Survey

3.4.1 Methodology

To verify and expand our qualitative findings, we conducted a survey on Amazon Mechanical Turk (MTurk) between August and September 2019. The aim was to

[3] We kept the conditions of the study as similar as possible for all participants. We believe PLACE malfunctions were mainly due to poor GPS signal.

leverage a relatively representative sample of the smartphone user population to evaluate the overall prevalence of SL misunderstandings among smartphone users (especially among SL users, RQ1) and the potential antecedents of these misunderstandings (RQ2). We chose MTurk because it is known to provide quality data for research in usable privacy and security [53, 90, 97].

To recruit participants, we advertised on MTurk, inviting participants with an Android phone to partake in a study about "Smart Lock for Android." We mentioned that participants did not need experience with SL to be eligible for the study (this was because recruiting "SL-novice" participants was necessary for verifying our hypotheses about antecedents of SL understanding, which we discuss later in this section). The survey was only visible to MTurk users who lived in North America[4] and had an approval rating higher than 90%. All those who took the survey were compensated with USD 4 (this amount was decided by assuming a minimum wage of $12 and the survey taking 15 minutes to complete, on average).

In the survey, we first asked participants about the following:

1. **Demographics**: Including age, ethnicity, level of education, and computer background.

2. **Phone Usage**: How much time they spent using their phones each day, how frequently they unlocked their phones, and the unlocking methods that they had enabled on their devices.

3. **Privacy-Sensitive App Adoption**: Consisting of 10 Likert-scale questions aimed at evaluating how often participants used privacy-sensitive apps (e.g., social networking) on their phones. The total sum of the scores was used to measure the participants' depth of smartphone adoption[5] Literature, however, suggest such cases should be rate [75]. This approach was inspired by Marques et al. [75] (our scale, which is provided in Section A.3, was a slightly simplified version of theirs).

[4] This is a limitation of our study which we will discuss in Section 3.6.

[5] It should be noted that, in this study, we are assuming that privacy-app adoption is correlated with app adoption in general, meaning a higher degree of privacy-app adoption could entail a higher degree of phone usage in general. However, it is theoretically plausible for a user to only use their smartphone for sensitive apps and nothing else. In such cases, the correlation between privacy-app adoption might not be as strong.

Next, we provided participants with a quick video introduction to SL. This was done for two reasons: (1) SL is named differently by different phone manufacturers (e.g., it is called Smart Unlock on Huawei phones). The video made it clear to participants what we refer to as SL. (2) To investigate how prior experience with SL correlates with SL understanding (i.e., whether using SL for a period of time helps users understand it better), we intended to contrast SL-experienced participants with SL novices. To this end, we used the video to introduce SL to novices[6]. To avoid inadvertently priming them, however, we carefully crafted the video to limit the amount of information it communicates and make it align with the SL UI.[7]

Afterward, we asked participants whether they knew about SL before participating in our study and, if so, if they were using any of the SL methods.

Finally, we gauged participants' understanding of SL semantics. We asked them what they thought each SL method was capable of, and how they thought SL methods interoperate (i.e., whether there is an "AND" or "OR" condition for locking or unlocking phones). We did not ask them about range parameters or modalities. This was because our CWU study, as well as our pilot surveys (explained below), showed that this quantitative cross-sectional survey could not sufficiently capture participants' understandings of these aspects, and would not lead to concrete results. The list of all survey questions is provided in Section A.3.

To increase the quality of our data, we introduced attention and consistency checks to our survey. These included putting cues in the SL introduction video and asking about them later in the survey (e.g., the video showed participants a word that they needed to input into a text box afterward), checking for inconsistencies in the phone usage answers (e.g., someone claiming to use pattern lock on an iPhone), and checking the response times to see if they deviated significantly from our pilot-based estimate of 15 minutes.

To assess the quality of our design, we consulted 7 HCI experts from our university's HCI research cluster. While most aspects of the design were well received, some experts were concerned that the video might introduce bias to the results, as

[6] The video was shown to all participants regardless, but the intent was for SL-novices to get familiarized with SL.

[7] The video is publicly available on YouTube through this link: https://www.youtube.com/watch?v=N-pC6-kWW0c

Table 3.3: Results of chi-squared tests of association between having experience with an SL method and answering the corresponding capabilities-related semantics question correctly. V refers to Cramer's V, reflecting the strength of the association.

	BODY	**PLACE**	**DEVICE**	**FACE**	**VOICE**
Experience	p = 0.227 V = 0.106	p = 0.465 V = 0.063	p = 0.067 V = 0.160	p = 0.626 V = 0.043	p = 0.228 V = 0.105

the quality of the video content might influence how SL novices understood SL semantics. To address this internal validity risk, we did the following:

Firstly, as mentioned before, we crafted the video solely based on SL UI, so it would not communicate any extra semantic information.

Secondly, we developed an alternative SL introduction medium—a text document[8] augmented with screenshots of the SL UI.

Thirdly, we conducted a pilot study with 10 participants on MTurk using the two introduction mediums (the video and the text document). Results showed that the participants' comprehension of SL semantics was broadly similar to that of CWU participants who experienced the SL UI directly.

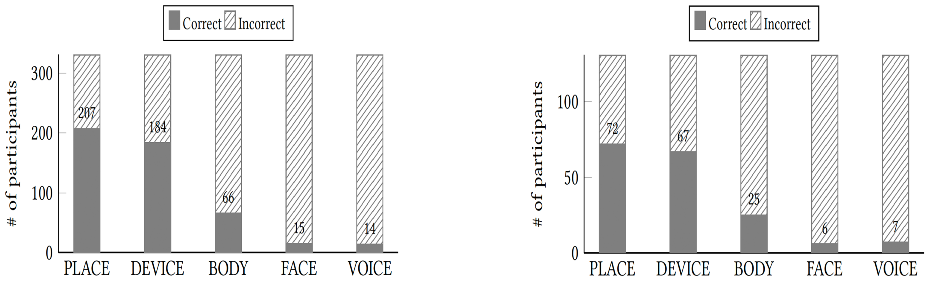
Finally, for the main survey, we randomly assigned survey takers to one of the two mediums and compared SL comprehension among the two groups. A chi-squared test of association revealed no statistically significant differences between the groups ($p > 0.05$).

Overall, the steps we took showed that our SL introduction mediums are fairly representative of the SL UI, in terms of the SL semantics information they communicate. As such, we believe we sufficiently mitigated the risk of the quality and content of the video influencing the outcome of the survey (our consultants agreed).

To test the final design, we asked the 10 pilot participants mentioned above to answer all the survey questions and provide written feedback to us. Based on their responses, we made only minor modifications to the wording of some questions.

Finally, we published the final survey on MTurk and received 382 responses. We eliminated 51 answers due to failing the data quality checks mentioned before

[8] The documentation was drafted based on Google's help page for SL. Small modifications were made to make the presentation coherent.



(a) all participants ($N = 331$)

(b) SL-aware participants ($N = 131$)

Figure 3.1: Distribution of survey participants' answers to SL capabilities questions.

(10 responses), using IP addresses outside of north America (3), using a VPN or VPS (10), or being flagged as a duplicate or bot by our survey platform Qualtrics (28). The average completion time was 12 minutes. As presented in Table 3.1, our sample was fairly diverse in age, gender, and education. We also observed diversity in occupation (not included in Table 3.1 due to great variation). Furthermore, chi-squared tests showed that the distribution of demographics (age, gender, and education) among our participants was not significantly different than that of the US smartphone user population [91]. Hence, we believe our sample to be fairly representative of that population. However, as Table 3.1 shows, the ethnicity distribution in our sample was significantly skewed toward white individuals. This is a limitation of our study that we discuss in Section 3.6.

Data Analysis: To evaluate the prevalence of SL misunderstandings (and answer RQ1), we used mostly descriptive statistics, which we present in Section 3.4.2. To identify antecedents of SL understanding (RQ2), we used chi-squared tests of association, as we did not assume any particular distribution for the data. The tests were performed for each SL method separately. For each test, correct understanding of the semantics of the SL method (i.e., whether the participant's answer conformed with our established ground truth explained before) was the dichotomous column variable, while each hypothesized antecedent (all of which were categorical) was the row variable. We used Bonferroni correction to mitigate the p-value inflation caused by multiple comparisons. The results are presented in Section 3.4.2.

Table 3.4: Results of chi-squared tests of association between SL understanding and our anticipated antecedents of it. Significant p-values are underscored (assuming $\alpha = 0.05$). V refers to Cramer's V, reflecting the strength of the association.

	BODY	**PLACE**	**DEVICE**	**FACE**	**VOICE**
Age	p = 0.756 V = 0.076	p = 0.103 V = 0.153	p = 0.705 V = 0.081	p = 0.485 V = 0.102	p = 0.527 V = 0.099
Computer Literacy	p = 0.245 V = 0.064	p = 0.390 V = 0.047	p = 0.133 V = 0.082	p = 0.642 V = 0.026	p = 0.724 V = 0.019
Privacy App Adoption	p = 0.380 V = 0.076	p = 0.017 V = 0.156	p = 0.035 V = 0.142	p = 0.356 V = 0.079	p = 0.066 V = 0.128
Security Proficiency	p = 0.122 V = 0.113	p = 0.250 V = 0.091	p = 0.379 V = 0.076	p = 0.494 V = 0.065	p = 0.171 V = 0.103

3.4.2 Results

Prevalence of SL Misunderstandings

We found SL misunderstandings prevalent, as predicted by our qualitative study. Figure 3.1a presents how many of the survey participants incorrectly answered the questions about the (un)locking capabilities of each SL method. About 80% of the participants incorrectly answered the questions about the capabilities of FACE, VOICE, and BODY. Based on our CWU findings, we predicted that the majority would be confused about these methods, due to the IA-EA mixture. The survey clearly confirmed this prediction.

Interestingly, our data also suggests that SL-aware participants (those who reported having known about SL before participating in our study) were not more likely to answer the questions correctly. As Figure 3.1 shows, the distribution of correct vs incorrect answers were nearly identical between the SL-aware subgroup and the overall sample. Furthermore, as presented in Table 3.3, chi-squared tests showed no statistically significant ($p - values > 0.05$) correlation between a participant reporting experience with an SL method and answering the corresponding semantics question correctly. This was an extension to our CWU results, as we did not have enough SL-experienced participants in our qualitative study to make such observations. This new finding suggests that SL misunderstandings are prevalent even among experienced users.

For BODY, FACE, and VOICE (where most participants answered incorrectly), the common mistake among participants was believing that these methods could unlock the phone automatically (meaning it does not prompt for explicit input for authentication).[9] This is in line with our CWU findings. As we discussed in Section 3.3.2, the confusion seems to be the mixture of IA and EA creating unmet expectations about FACE and VOICE capabilities, and the discrepancy between the capabilities of BODY and other IA methods.

As for PLACE and DEVICE, we see in Figure 3.1 that the majority (nearly 60%) answered the semantics questions correctly. This was again in line with the CWU findings (see Table 3.2), as those methods were found easier to understand. However, a new observation was that the prominent mistake among the minority was believing PLACE and DEVICE could not lock the phone automatically. Based on our CWU results, we conjecture this to be due to the participants extending their understanding of traditional unlocking methods (e.g., PIN or fingerprint) to IA, believing SL could not lock the phone. However, further studies are needed to firmly confirm this conjecture.

Finally, for the interoperation of the SL methods, we found that only 9.7% (N=32) of participants answered the corresponding semantics question incorrectly. The other 74% (N=245) correctly assumed an OR logical relationship, even though it was never communicated to them. This observation seems to suggest that OR is what the participants expect by default, which is, evidently, what SL provides. However, even though the survey data shows that interoperation misunderstandings are not highly prevalent, we still believe that the semantics should be clearly communicated by the UI. As our CWU findings show, misunderstandings can lead to various dangerous errors by the users.

To answer **RQ1** then, our survey results suggest that misunderstandings about SL are prevalent. They are mostly regarding the capabilities of SL methods, with occasional confusions about SL interoperation as well.

[9] What “automatically (un)lock” means was explained in the introductory video, which was based on how it is explained on SL UI.

Antecedents of SL Understanding

Per **RQ2**, we were interested to see if pre-existing factors (e.g., demographics) were linked to correct SL understanding. We collected a multitude of such data in our survey. However, to avoid a fishing expedition, we only examined the effect of factors that were referred to by the related work.

We should note that our list of antecedents is not comprehensive, not that we intended for it to be. We only studied how our research aligns with some notable related work from authors who investigated the antecedents of the understanding of conventional unlocking methods. While our study provides a first insight into this topic, future studies are needed to study the matter further.

Furthermore, we should also note that since we did not observe any significant differences in SL comprehension between novice and experience users, we included all participants in the following analysis. Had we found any differences, we would have included SL-aware participants only.

To start with, we examined whether age was a factor that was linked to SL understanding. Age is shown to be associated with smartphone users' perception of conventional phone unlocking methods [95]. As such, we were interested to see if it had a similar correlation with SL understanding, as well. On the contrary, our data analysis revealed no significant association between age and correct understanding of SL capabilities (See Table 3.4) or SL Interoperation (p = 0.595, Cramer's V = 0.092). This observation suggests that correct SL comprehension is not dependent solely on cognitive capability which could potentially give younger individuals an advantage.

Next, we examined the link between depth of smartphone adoption and SL comprehension. Previously, Marques et al. [75] showed that power phone users (those who use their phones for a more diverse range of applications) usually have better understandings of phone security features. We were interested to see if this was the case with SL, as well. Our analysis showed that it is, in fact, so. We observed statistically significant associations between privacy app adoption score and correct understanding of PLACE and DEVICE capabilities (see Table 3.4). However, no similar association was detected for BODY, FACE, or VOICE capabilities (Table 3.4) or SL interoperation (p = 0.339, V = 0.081). This was in line with our previous findings. Evidently, correct comprehension of PLACE and DEVICE requires deeper Bluetooth and GPS knowledge, which power phone users tend to possess.

Thirdly, we investigated the link between computer literacy and SL understanding. In particular, we examined whether those who had a computer-related occupation were less likely to exhibit the confusions we discussed in the previous section. We did this examination because such correlations have been observed by other studies [8, 69]. Our results, however, did not suggest any such association (see Table 3.4). As such, in confirmation of our CWU findings, it seems that SL semantics require specific types of knowledge that goes beyond typical computer literacy.

Lastly, we should note that we did not observe any association between medium of introduction to SL, and SL comprehension. As discussed before, we added a second SL introduction medium (a textual presentation) to our study, to make sure the video does not bias the results. Chi-squared test revealed no association between the medium of introduction and correct answers to any of the SL semantics questions (p = 0.735, V = 0.018). As such, the video seemed to had not caused any significant bias in the data.

In summary then, to answer **RQ2**, our survey data suggests that, in agreement with the related work, depth of smartphone adoption is a significant antecedent of SL understanding. However, contrary to the related work, we did not find any association between age or computer literacy with SL comprehension.

3.5 Discussion

Research suggests that IA is a promising technology for providing better physical security protection on smartphones: Not only can it make unlocking more convenient [27, 64], but it does so with only minor sacrifices to security [63]. Even more, users are actually willing to adopt this new technology, provided that it is implemented well [26, 64, 80].

However, it is becoming clear that hitting the sweet spot could be very tricky. Not only must IA schemes address numerous security challenges (e.g., resistance to mimicry attacks and minimal authentication delay) [63], but they also need to resolve the issue of intermittently available data (i.e., when to switch to explicit authentication) [64], which can adversely affect the user experience. To add to these challenges, this research discovers a new challenge for IA to overcome, which is to efficiently communicate its semantics to users. Our SL case study vividly demonstrates how important this issue is for a successful IA deployment, yet how difficult it is to address.

We found that complex SL semantics confuse users. There are a multitude of aspects to SL that can cause misunderstandings and dangerous security errors. Firstly, the users might misunderstand what type of data SL uses to authenticate them (e.g., how is motion detected by BODY?, see Section 3.3.2). Secondly, they might not know what each SL method is capable of (e.g., can BODY automatically unlock the phone?, see Section 3.3.2). Thirdly, they might not be aware of how SL methods work together (e.g., what happens if both PLACE and BODY are enabled?, see Section 3.3.2). Lastly, users probably won't have knowledge of the parameters of the SL methods (e.g., how far should the user be from their home, for PLACE to lock their phone?, see Section 3.3.2). Any of these misunderstandings can lead users to misjudge when their phone would be (un)locked, potentially leading to unauthorized access to their sensitive data.

These misunderstandings are more than hypothetical. They are prevalent. Nearly 80% of our 331 survey participants overestimated the capabilities of some SL methods, namely BODY, FACE, and VOICE (see Section 3.4.2). They incorrectly believed that these methods could automatically unlock the phone. Similarly, as also discussed in Section 3.4.2, a considerable number of the participants (40%) incorrectly believed that PLACE and DEVICE could not lock the phone automatically, expecting these unlocking methods to behave like traditional ones (e.g., PIN or fingerprint). Lastly, several participants (10%) misunderstood how SL methods would interoperate, as there was no explicit indication in SL UI.

SL misunderstandings are also universal. We could not identify any demographic group of users who would understood SL better (see Section 3.4.2). We found that younger and older adults misunderstood SL alike. We also found that computer literacy did not necessarily translate to better SL comprehension. The only factor that we found correlating with SL understanding was depth of smartphone adoption, which suggests that only those who have specific knowledge are more likely to understand SL correctly.

Our findings suggest that the SL UI is partially responsible for these confusions. It is vague about the capabilities of each method (see Section 3.3.2), it does not clearly define what “motion” entails in the context of BODY (see Section 3.3.2), and it is not clear about why it is communicating the range parameters of DEVICE and PLACE (see Section 3.3.2). Worse, it does not communicate at all how SL methods interoperate (see Section 3.3.2). These deficiencies require participants to manually explore and experiment with SL, in order to gauge the validity of their initial impressions, which our survey showed to be error-prone (as discussed in Section 3.4.2, having experience with an SL method made no difference in understanding its semantics correctly). To mitigate these misunderstandings, we rely on our findings to offer the following recommendations for IA UX design on smartphones:

1) COMMUNICATE CLEARLY WHAT DATA EACH IA METHOD USES AND HOW: We recommend that the type of data each IA method uses for authentication purposes should be clearly communicated to users. In this regard, no deep technical knowledge should be assumed on the users’ part.

Taking BODY as an example, we believe how it detects motion, how big the movement should be, and how long should this movement last are essential to convey in some way, even though these specificities might not seem important at first. For example, our results (Section 3.3.2) suggest that the lack of this understanding can result in users leaving their phone unprotected because they might not know whether being in a moving car counts as motion or not.

Another example is DEVICE, where we believe that the UI should specifically address how it would identify the trusted device (e.g., based on some hardware serial number). One might think that this information is common knowledge and need no explicit explanation. However, we found (in Section 3.3.2) this to be not the case, as some participants lacked this knowledge.

While it is out of scope of this dissertation to explore how to effectively communicate this information, the way the SL UI does it seems to be suboptimal, to say the least. In the case of BODY, for example, the main UI lacks any of the information that we suggest to communicate to the user. A help page accessible through an obscure button (that only 2 of our CWU participants clicked on) is the only place that explains how BODY works.

One efficient way of conveying the semantics could be animations, as they have been shown to be effective in communicating meaning [12], specifically in information security [9]. Animations were also suggested by some of our CWU participants (both, HCI-proficient and ordinary smartphone users), e.g.,

> *"I think for me, it would be better to have animations to show how it [BODY] works ... this graphic [still image on BD UI] doesn't tell me much as to how Smart Lock on-body detection works."* [P-TA-9]

2) CLEARLY STATE THE CAPABILITIES OF IA: We recommend that the unlocking capabilities of each IA method (i.e., whether it can automatically lock or unlock the phone) should be clearly communicated to the user by the UI.

As we saw with the case of SL, this is particularly important when IA and EA methods are mixed together (see Section 3.3.2 and 3.4.2). Evidently, most participants incorrectly believed that FACE or VOICE are capable of automatically locking the phone. As discussed in Section 3.3.2, we believe that the interplay of FACE and VOICE (EA methods) with BODY, PLACE, and DEVICE (IA methods) resulted in this incorrect understanding of EA capabilities.

The current SL UI does not communicate the capabilities of each method explicitly. The user has to deduce it from the descriptions in the UI (see Screenshots in the Appendix). Fortunately, it is rather easy to communicate whether an IA method is capable of locking or unlocking or both. Exactly how this information can be communicated effectively is of course out of scope of this research. However, a suggestion by some of our HCI-proficient participants was to name the IA methods in a manner that clearly distinguishes them from the EA ones. P-CW-2, for example, stated:

> *"I think that the names like Trusted Face and Trusted devices are similar, [even though] they function differently. So maybe you can change the name of one?"*

3) MAKE INTEROPERATION SEMANTICS CLEAR: We recommend that the UI communicate clearly how each IA scheme interoperates with other IA or EA schemes on the phone (i.e., whether this is an AND or OR logical relationship between them). Otherwise, users might mistakenly assume that one authentication method takes precedence over the other, resulting in dangerous errors.

The results of our survey study (see Section 3.4.2) showed that most users expect an OR condition by default, which is evidently what SL provides. However, the SL UI never communicates this matter explicitly, leading some users (at least 10% according to our findings in Section 3.4.2) to incorrectly assume an AND relationship, which can lead to security errors.

Additionally, the fact that 10% of survey participants expected an AND relationship suggests that the matter of how SL methods should interoperate may be subjective, and therefore should be left to the user as a configuration option. This was also suggested by some of our HCI-proficient participants. In either case, our data clearly shows that the UI cannot remain silent on this matter and should address it explicitly.

4) PROVIDE ACCURATE AND RELIABLE VALUES FOR PARAMETERS: The operational parameters of IA methods (e.g., the range of Bluetooth or GPS connections) should be clearly communicated by the UI. As discussed in Section 3.3.2, knowledge of these parameters is important for users to make sound security judgements.

However, our CWU study suggests that mere presence of these parameters in the UI is not an optimal way of communicating them. In the SL UI, this information is presented through several screens and warning messages (see screenshots in Sections A.4 and A.5), which most of our participants were either confused by or dismissed as unimportant.

As a case in point, the range parameter in DEVICE UI is communicated through a note (see Sub-figure d in Figure A.1 in the Appendix) without any context as to why this information is presented. We believe that this is why our participants often dismissed it as unimportant (Section 3.3.2).

Our findings in this regard seem to align somewhat with the case of privacy policy communications, which have been studied extensively. It has been shown that most users ignore text-based policy statements because they find them irrelevant or difficult to understand [25, 31]. Subsequently, several approaches have been proposed to improve privacy policy statements. The most notable [49, 66] provides context (e.g., how the data will be used) and appears to be somewhat successful [31]. Interestingly, we saw in our CWU study that the example use cases provided by DEVICE UI (e.g., that you can use it with your car's Bluetooth system) were well received by our participants. Therefore, providing example use cases might be a promising way to communicate range parameters.

And, lastly, we should note that, as discussed in Section 3.3.2, unreliable operation (e.g., inconsistent range of GPS) can cause users to doubt their understanding of the semantics, even when it is correct. Therefore, the UI needs to be wary of this fact as well and provide the necessary warnings to the user.

3.6 Limitations

Any generalizations of our findings need to be performed carefully due to the following study limitations:

- The cross-sectional design of our CWU study prevented us from fully investigating the effect of prolonged SL usage on participants' understating of its semantics. It is possible that continued exposure to SL can cause certain misconceptions that our CWU study was unable to capture. Our survey data bridged this gap somewhat (we surveyed long-term SL adopters). However, more longitudinal studies (e.g., diary studies) are needed to further investigate this matter.

- similar to other studies on smartphone usage [21, 29, 51, 57], our survey sample was not fully representative of the smartphone user population. It is shown that cultural factors affect smartphone users' unlocking behavior and attitude [53]. However, due to limited resources, we only included North American (NA) participants in our studies. Therefore, our results are only generalizable to NA smartphone users. Similarly, the ethnicity distribution among our participants was heavily skewed toward whites, which is a common limitation of MTurk studies [21, 29, 51, 57, 90, 97]. Finally, the limited size of our sample as well as the skewed age distribution might have prevented us from finding statistically significant correlations between age and SL understanding. Such limitations threaten the generalizability of our results. However, we still believe that our findings are valuable, as a first insight into smartphone users' understanding of IA.

- Our participants self-reported their prior experiences with SL. Thus, as is common with self-reported data, it is possible that the participants' answers were not completely reflective of their reality [39]. While we eliminated those responses that showed clear inconsistencies, it is still possible that self-reporting has affected the quality or our data, and hence the results.

- As of Android 10 (released after data collection for this study took place), Smart Lock can no longer automatically unlock the phone (any of its methods), due to changes in underlying Android APIs. This could make the capabilities of SL rather easier for users to understand, as they all have the same set of capabilities now. However, follow up studies are required to verify this conjecture.

- Similar to how it was discussed in Section 2.5, one implicit assumption of our recommendations for IA UX design is that future IA systems will not be perfect (i.e., they will incur some failures in misidentifying a user as someone else (a false positive), or in not identifying the user at all (a false negative)). If an scheme is perfect, then the user could trust that it would always "work". Thus, they might need to know about the capabilities or modalities of such IA scheme (because they would not need to make sense or be wary of of potential false positive or negatives). However, the feasibility of such perfect solution seems remote, based on what we discuss later in Chapter 5 about the availability of authentication data.

3.7 Conclusion

While implicit authentication (IA) is becoming a widely-researched approach for protecting smartphone physical security, no studies has been done in investigating users' understanding of this technology. To address this knowledge gap, we used Smart Lock as a case study and evaluated how its semantics are understood by Android users. We found SL misunderstandings to be prevalent, universal, and mostly due to insufficient communications by the SL UI. We identified four aspects of SL that have the most potential for being confusing, namely the authentication modalities, the capabilities of IA, inter-operation of IA and EA, and the operating parameters of IA. Based on the findings, we provided a set of recommendations on how to improve users' understanding of these aspects.

Chapter 4

Investigating the Efficacy of Access-Control Solutions

This chapter presents the results of a longitudinal diary study to investigate smartphone users' access control needs, and the efficacy of the existing solutions. First, in section 4.1 we provide an overview of the existing literature and a summary of the solutions. Then, in section 4.2, we describe the methodology of our study. We present our findings in Section 4.3, and, in Section 4.4, we discuss the implications of them. Lastly, in Section 4.5, we discuss the limitations of our study, and, in Section 4.6, we conclude the chapter.

4.1 Related Work

4.1.1 Smartphone Users' Access-Control Needs

Several studies have qualitatively investigated users' needs. For example, Mazurek et al. [78] interviewed 33 smartphone users. They found that users' ideal access-control policies could not be easily defined in standard role-based terms and depended on factors such as location and presence of certain individuals. They also found that, as incumbent solutions cannot uphold such complex policies, users resort to constructing ad hoc solutions, such as hiding files. Based on such observations, the authors argued that users require reactive policy creation and finer-

Table 4.1: Summary of proposed/deployed user access control solutions on smartphones.

Category	Approach	Proposed/Deployed Solutions
	Task-sensitivity	
All-or-nothing	(Un)authorized user can access none/all functionalities of the phone	Incumbent solution on current phones
Lock-screen-access	Identical to all-or-nothing, except allowing access to few apps without unlocking	Deployed on most phones [44]
App-level-access	Grant/deny access to each app individually	Progressive Authentication [98], TreasurePhone [104], ConXsense [83], CRePE [23], Secure Folder [101]
	Phone-sharing	
All-or-nothing	Allow secondary users full access to the phone	State of practice for most users [77]
Profile-switching	Each secondary user has a separate profile with isolated apps and data	Deployed on Android phones [42]
Resource-based	Deny secondary users access to “sensitive” phone resources (WiFi, Bluetooth)	DiffUser [86]
Session-based	Specify individually what apps the secondary user can access in each shared session	xShare [72], App-pinning [7]

grain access control than an all-or-nothing model. Similarly, Hayashi et al. [54] interviewed 20 users and found that their needs go beyond all-or-nothing. Their participants wanted at least one of their apps to be protected by a lock, half to be without protection, and 20% to be split (only parts of them locked). Hence, the authors found that users' authorization needs are even finer-grain than app-level.

To understand users' authorization needs in phone-sharing settings, Karlson et al. [60] interviewed 12 users to explore why they shared phones, with whom, and the concerns they had when sharing. They found that participants expressed strong preferences about which data and functionality should be available to each guest user. The authors also found that sharing preferences might be location dependent. Later, Hang et al. [50] conducted focus groups with 25 participants and found sharing to be often impromptu. They also found sharing preferences to be strongly app and data dependent.

Jacobs et al. [58] reported a similar study on couples' practices with single-user device access. They conducted 20 interviews and an 8-day diary study. They found that sharing preferences often depended on content type, making the all-or-nothing model impractical.

Lastly, Matthews et al. [77] investigated sharing across multiple devices. They conducted a survey with 99 households and a 21-day diary study. They found device and account sharing to be common. They also identified six different types of sharing, ranging from borrowing devices to getting technical help. They suggested that access control could be designed differently for each sharing type. They also found that sharing often happened in the sharer's presence.

4.1.2 Alternatives to All-or-Nothing

Several solutions have been proposed to alleviate the deficiencies of the all-or-nothing model.

The first set of solutions provides task sensitivity (requiring authentication only when the app/task being used requires it). Commercially, both Android and iOS allow *lock screen access* to some apps [44]. This solution keeps the all-or-nothing model mostly intact but allows the user to launch certain nonsensitive apps (e.g., camera, calculator) without unlocking the phone. Alternatively, Riva et al. propose a more elaborate solution called *progressive authentication* [98]. It determines a level of confidence in the user's authenticity and only prompts for authentication when a launched app requires high confidence. There have also been several context-based solutions that allow access to certain apps in specific contexts (e.g., location) without authentication [23, 83, 104].

Another set of solutions aims to provide better phone-sharing support. Commercially, *profile switching* [42] is the incumbent solution on Android. It allows the device owner to create separate profiles for secondary users. However, research has shown that users often do not utilize profiles even when set up, due to high physical and cognitive overhead [17, 32]. Android and iOS also now support *app pinning* [7, 43]. It allows the owner to limit sharing to a single app by fixing the app on the screen and preventing the switching of apps.

Researchers have proposed more elaborate solutions. Ni et al. [86] propose *DiffUser*, which implements a role-based model of access control. Secondary users are assigned to a role: administrator, normal user, or guest. Each role has restricted access to certain resources (e.g., normal users can access SMS and contacts but cannot install or uninstall apps). Liu et al. [72] propose *xShare*, which allows the sharer to quickly put the phone in a restricted mode before handing it to a sharee. The sharer must respecify what apps are available in this mode every time they enable it.

To summarize, Table 4.1 provides an overview of the existing solutions, categorizing them based on their approach to access control. We should note that in this dissertation we were focused on solutions that control a human operator's access to a phone. Therefore, OS-level solutions that control apps' or processes' access to system resources, such as FlaskDroid [18] and others [10, 11, 112, 120], were out of our scope.

4.1.3 Gaps in the Literature

Overall, we identified several gaps in the literature, which this dissertation aims to address. Firstly, while prior studies demonstrated the existence of issues with all-or-nothing, they did not offer insights into the prevalence of the issues. For example, while all-or-nothing's obliviousness to task sensitivity is suboptimal [50, 54], it remained to be investigated what proportion of users' tasks are actually sensitive or how sensitivity varies by functionality and across users. Such investigations would help researchers and developers to understand how dire these issues really are and whether there is an actual need for the other solutions. Only when the variance of task sensitivity among users is high can one argue for solutions that increase authorization granularity. Otherwise, the high configuration effort required for such solutions makes them unattractive for most users.

Furthermore, even if a need for new solutions is determined, the literature [50, 58, 60, 77] did not offer quantitative measurements of users' needs. As such, there could not have been (and was not) any attempt to evaluate and compare the efficacy of all-or-nothing and its alternatives. This left it unclear whether the proposed alternatives would succeed in addressing the issues with all-or-nothing and if/where there would be a need for further research.

Last but not least, while the existing literature showed that users' access-control preferences for phone sharing indeed depended on contextual factors (e.g., location [78] and content [54, 58, 77]), the consistency of such factors had not been demonstrated. To determine the prospects of improving access-control decisions by incorporating these factors, it is important to understand how consistent they are.

4.2 Methodology

Our study was designed to answer the following research questions:

- **RQ1**: What tasks do smartphone users perform on their phones? What are their sharing preferences for the tasks?

- **RQ2**: To what extent, in terms of false positive rate and false negative rate, do all-or-nothing and the alternative solutions meet the users' needs? How do they compare in configuration size?

- **RQ3**: How consistent are contextual factors across phone-sharing events?

We conducted a diary study with 55 participants to answer these questions. The study involved participants installing our custom Android app on their phones and using it to report the following data every day: (1) the tasks they performed with each of their apps, and whom (else) they would allow to perform each task, as well as (2) the details (e.g., time, location) of any instances of sharing their phone with others.

Inspired by the method used by Hayashi et al. [54], we used tasks as the means to collect users' authorization needs. Conceptually, we define a task as a distinct series of actions that could be performed with a mobile app, and are distinguishable in terms of purpose and sharing preferences. Technically, they comprise the functionality affordances of a mobile app, similar to affordances of real world objects or User Interfaces (UIs) [87]. For example, the app GMail, affords two separate tasks: "Sending email" and "Receiving/Reading email," according to our study participants.

To solicit tasks, we asked our participants to declare what they had used the app for (e.g., sending email) while separating actions that have different purposes. We also asked them to further break down a task if there were parts of it that they were not willing to share with others. For example, if they used the SMS app to both send and receive texts, they were asked to declare sending and receiving as separate tasks, if they were willing to let others do one (e.g., the sending) but not the other (the reading). Otherwise, they could just declare "sending and receiving sms" as the task.

We used a longitudinal design mostly in order to accommodate known limitations of human memory [16, 92]. Firstly, a cross-sectional (non-longitudinal) design [119] would have required participants to remember at once all the tasks they perform. Secondly, the participants would also have needed to remember all the instances of sharing their phone with others over a long period of time. These unrealistic expectations would have limited the internal and external validity of our results.

In the following, we will discuss in more detail our data collection tool, data analysis methodology, and a breakdown of the demographic composition of our participant sample.

4.2.1 Data Collection

Our app kept track of the participants' daily app usage and invited them every night to fill out a diary. Using an Android app, instead of paper- or web-based diaries, allowed us to avoid asking participants repetitive questions (e.g., asking them to list all the apps they had used) and to also ask detailed questions about app usage context, as explained below.

The diary questions were fully structured. First, the participant was asked if they had shared their phone during the day. If so, they were shown a list of apps launched on the phone that day and were asked to indicate which apps were shared.

Next, our app would select five apps (either shared or used by the participant themselves) to probe further about. We chose five because pilot studies (described later in this section) showed that it took ten minutes on average to complete a diary with this number of apps; taking any more time daily would result in lower data quality and higher chances of skipping diary questions or even dropping out of the study.

The five apps were selected using the following protocol (a flowchart is provided in Figure A.3):

- **Priority 1: Newly-shared apps**: We first selected apps that the participant reported sharing on that day for the first time in the study, as we anticipated them to be rarer. If there were more than five such apps, we prioritized selecting those with higher usage time (i.e., those that had more time in the foreground, based on Android's "Usage Stats" service [5]).

- **Priority 2: Newly-used apps**: If there were fewer than five newly-shared apps, we selected apps that the participant used for the first time on that day. We also prioritized selecting those with higher usage time.

- **Priority 3: Random selection**: If the two priorities above did not fill the quota, we would fill it by randomly select apps that the participant used.

Once the apps were selected, the participant was asked the following questions about each app:

- ***For both shared and non-shared apps***: They were asked to declare a list of tasks that were performed with it (either they performed themselves or they believed a sharee did while in possession of the phone). They could create new tasks or select from the ones declared previously (see Figure A.4e). Next, they were asked to specify their sharing preferences for each task, by selecting whom (else) they would allow to access it. The options included: (1) *No-one*, (2) *Any-one* or (3) *Specific People*. If they chose *specific People*, they were asked to provide a list of individuals (see Figure A.4f).

- ***Specific to shared apps***: Participant was shown a list of times the app was launched on the phone, and were asked to specify whom (either themselves or a sharee) used the app at each time (see Figure A.4b). To specify a user/sharee, the participants needed to give them a nickname (to use for any subsequent reporting of sharing with the same person) and to declare their relationship with sharee (e.g., spouse). Afterwards, for each time the app was reported to have been used by a sharee, the participant was asked to specify the location where the sharing happened, and whether they were present there (see Figure A.4d).

4.2.2 Pilot Studies

We performed three pilot studies to test our methodology and app design. The first study involved lab testing sessions and qualitative interviews to evaluate the usability of our app. We used UBC paid participant mailing list to recruit participants. They performed a series of tasks with our app (e.g., filling a daily diary), as we observed if they encountered any difficulties. We also interviewed them about their impressions of the app. Overall, 6 individuals (3 of them male) participated in this pilot, with their ages ranging from 26 to 65 and with education levels from high school to master's degree. We recruited participants one at a time and implement changes in the app to improve usability in-between the sessions. Based on the results, we improved the app set-up process, improved the clarity of our consent form and the wordings of some diary questions, and fixed a few bugs.

The second pilot study was a test run of the main study. We used the same mailing list again to recruit 7 participants (5 of them females, age from 27 to 37, education from high school to master's) who installed the app on their personal phones and used it to fill diaries for two weeks. Then, in the end, we interviewed them about their experience. Based on the results, we found that it would be optimal if the diaries would not take more than 10 minutes per day. We also improved the wordings of some questions and fixed some bugs. More importantly, we found a lot of duplication of task declarations which took unnecessary time from participants. To reduce fatigue, therefore, we decided to implement a system where each participant could see, anonymously, what tasks others had declared for each app. This way, they could either select one of those tasks or, if need be, declare a new one.

Lastly, interviews showed that some participants might have privacy concerns with installing our app on their devices. They suggested that we could provide more clarity on what exact type of data our app collects and for what purpose. We heeded this recommendation, as privacy concerns could skew any future sample towards less security-conscious smartphone users. For our main study, the study advertisement and consent form assured participants that the app would not collect any personal data, such as precise GPS location, files, and login credentials. We also invited them to ask any questions they might have about the app or the purpose of the study, and let them know that they can withdraw from the study at any time, without repercussions, if they felt uncomfortable. As a result of these changes, we observed, in our screening surveys, that very few potential participants of our main study to declare privacy concerns as a reason to be not willing to join the study.

Our third pilot study was aimed at evaluating the efficacy of our intended recruitment channel for the main study, that is Amazon Mechanical Turk (MTurk). It was also to evaluate the effects of changes we made amid the second pilot results. Specifically, we were interested to see if showing tasks defined by one participant to others would result in hive minding, reduced diversity of declared tasks, and eventually bias in the results. We recruited 8 participants from MTurk (4 of them female, age ranged from 21 to 44) and asked them to install and use the app for two to three weeks. We found MTurk to be able to provide us with sufficiently diverse samples. We also observed that we were already near theoretical saturation for task declarations after three weeks (i.e., very few new tasks were being declared towards the end), hence we decided on the length of the main study at 1 month. Lastly, results showed no signs of hive minding as we observed an even more diverse set of tasks being defined, when compared to the second pilot.

Table 4.2: Demographics of the study participants.

Variable	Category	% (#) of participants
Gender	Female	47.3 (26)
	Male	52.7 (29)
Age	18-29	9.1 (5)
	30-49	67.3 (37)
	50-64	20 (11)
	65 and higher	3.6 (2)
Education	High School	27.3 (15)
	Associate	18.2 (10)
	Bachelor	45.5 (25)
	Graduate degree	9.1 (5)
Annual income (in $1,000s)	Less than 10	5.5 (3)
	10-29	10.9 (6)
	30-60	30.9 (17)
	60-100	40.0 (22)
	100-150	5.5 (3)
	More than 150	7.3 (4)

4.2.3 Sampling

We recruited 55 participants through MTurk. We published a screening survey on the platform, inviting interested participants to complete it for a $1 compensation. The ad was only visible to Turkers who were located in North America (according to MTurk) and had an approval rate of at least 85%. We received 408 responses to the survey, and invited 226 of the respondents to install our app (The number was determined by power analysis, assuming 85% confidence level, 5% margin of error, and a target population size of 330 million for the US). We excluded those who did not have an Android phone (67 entries) or showed clear signs of duplicate or low-quality (e.g., claiming to use FaceID on an Android phone) data (20 entries). We also selected randomly from those who had a surplus of demographic quota in our sample (e.g., younger age ranges). Out of those invited, 65 installed the app, and 55 eventually completed the study. Data collection took place between May and August 2021. In the end, participants were compensated with $20 USD plus $2 for every diary they completed, paid as a gift card. This compensation model was inspired by similar longitudinal studies [24, 61, 108], and was designed to reduce the probability of early dropout.

Our final sample was fairly representative of the US smartphone-user population, which is where all of the participants were located. As Table 4.2 shows, the sample was diverse in terms of age, gender, education, and income groups. We also performed chi-square tests. They showed no statistically significant differences ($p - values > 0.05$), in terms of the distribution of the above demographics, between our sample and the general US smartphone users as of 2019, as reported by the Pew Research Center [91]. However, our sample was not very diverse in terms of ethnicity. Age distribution was also skewed, even if the difference was not statistically significant. We discuss sample limitations further in Section 4.5.

4.2.4 Ethics

Data collection was conducted according to the policies and regulations of the University of British Columbia (UBC) and Canada. All study procedures were reviewed and approved by UBC's Behavioural Research Ethics Board (Certificate ID H20-03155). All types of data that the study app would collect and the purpose for that collection were disclosed, and participants consented to its collection. To preserve participants' privacy, four measures were employed: (1) the app was not programmed to collect any data outside the scope of the study; (2) the collected data was uploaded on a daily basis through an SSL-encrypted connection to a server hosted in Canada, deleted from the phone after upload, and stored on an encrypted disk; (3) personally identifiable information (e.g., contact email) was collected as part of the web-based screening survey but not through the study app; and (4) once the study ended, the app automatically disabled itself and prompted the users to uninstall it.

4.2.5 Data Analysis

RQ1 [Tasks]: To create a cohesive list of tasks, two researchers aggregated the tasks reported by the participants to merge functionally duplicate ones. They did so by performing *qualitative inductive coding* [13]. Every day, each researcher

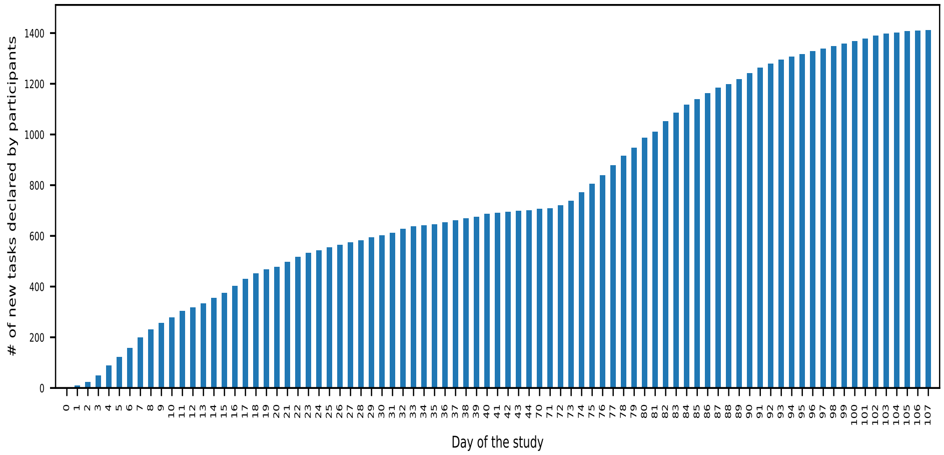


Figure 4.1: Theoretical saturation in task declaration by study participants.

mapped each newly declared task to one of the existing ones in our codebook. A mapping meant that the researcher believed the new task was the same as the previously declared one. The criteria for mapping tasks was that (1) they would either be phrased identically or were similar in functionality, and (2) the researcher could not envision a scenario in which they would require different security, based on the participant's prior data. If a new task could not be mapped to an existing one, it was added to the codebook as is. The researchers met once a week to resolve any differences (disagreement rate was 4.5%). All disagreements were resolved; if the researchers could not agree on an aggregation, both of the declared tasks would enter the codebook.

We should note that, as Figure 4.1 shows, we reached near saturation by the end of the study, as very few new tasks were being declared (most participants reported performing tasks they had already declared before). Hence, we believe our task set to be fairly representative of the participants' needs.

Once the task codebook was finalized, the two researchers performed categorization of the tasks based on functionality. The aim was to use this categorization to investigate the correlation between functionality and task sensitivity.

To perform the categorization, each researcher labelled each task with one of the 32 functionality categories the Google Play store uses for apps [45]. The researcher used Google Play's guidelines [45] to decide which category would be used for a hypothetical app only affording that task. The guidelines provide broad examples of what apps should fit in each category (e.g., for the category "Entertainment," the examples read "Streaming video," "Movies," "TV," and "Interactive entertainment.") Next, the researchers met and resolved all differences in labeling, of which there were 69 instances. Resolutions were achieved through either agreeing on one category or merging categories that had all overlapping apps.

RQ2 [Comparing Access Control Solutions]: We compared three metrics for each evaluated solution:

- ***False Positive Rate (FPR)***, also known Fraud Rate, was calculated as the ratio of the number of tasks a solution would mistakenly make available to unauthorized users (i.e., would not perform authentication before granting access), by the total number of tasks. It would directly impact the security of a solution, as higher FPR would indicate higher chances of unauthorized access to tasks.[1]

- ***False Negative Rate (FNR)***, also known as Insult Rate, was calculated as the ratio of the number of tasks that a solution would mistakenly deny access to authorized users (i.e., would ask for unnecessary authentication before it) by the total number of tasks. This measure would directly impact the usability of a solution, as it would signify how frequently the solution imposes unnecessary unlocking overhead on users.

- ***Configuration Size Rate (CSR)*** was calculated as the ratio of the number of access control preferences (e.g., indicating that an app should be available before unlocking) each user would have to explicitly specify for a solution, by the total number of the user's preferences (in our study, we assume this to be their total number of tasks). We used CSR as a basic estimate of the amount of effort it would take a user to configure a solution, without assuming any particular design for the User Experience (UX), which cloud vary

[1] To accurately take into account inconsistencies in the participants' labeling of tasks (e.g., a participant labeling a task once as being shareable with "No-One" and another time as being with "Anyone"), we treat different labels of a task as different tasks.

by phone manufacturer or academic proposal. For example, a $CSR = 100\%$ would mean that the solution expects the user to explicitly specify their sharing preference for all their tasks, where as $CSR = 1\%$ would require only 1/100 of the effort.

The metrics were calculated separately for each participant and then averaged over all participants. It should be noted that we calculated both the FPR and FNR by counting the number of tasks, not the number of times a task was performed by a participant. In the scenarios (described below), we also did not differentiate between when an unlocked phone is shared and when the passcode is shared. While these choices could make the calculated FNR less reflective of the actual user experience (e.g., the user might face more interruptions from frequently done tasks), we made them to limit the impact on the FPR by frequently done nonsensitive tasks, which could make a solution look unrealistically secure. For example, an unauthorized execution of a highly sensitive task, such as banking, could be masked by the many authorized executions of a trivial task, such as checking the weather.

Additionally, due to our methodology design, we assume that all false positives and negatives have the same weight, when computing FPR and FNR. This causes our results to be a "best case" scenario of the performance of each solution.[2]. We discuss this limitation further in Section 4.5. Also, aligned with our research objectives, we only consider physical (un)authorized access to each *Task*. This means that we assume that when a user performs a *Task*, no unauthorized operation is done by the phone, other than what necessary for accomplishing the task (for example, such operations might be injected by a malware into the app's code). As such, remote attacks (such as malware) is not considered in our evaluations.

For task sensitivity, we calculated the metrics in seven scenarios, each corresponding to one type of solution:

- ***ALL***: This scenario corresponds to one possible case of the all-or-nothing model. It represented a case where the user would not enable locking. Hence, anyone could perform any task with the phone. In such a case then, any task

[2] This is because even though some solutions might be more prone to granting unauthorized access to sensitive tasks than non-sensitive ones, we only consider the number of tasks that each solution grant such access to, rather than considering the sensitivity of each individual *Task*

that the user would label as either shareable with “No-one” or with “Specific People” would count as a false positive (i.e., fraud)[3], increasing the FPR. FNR would be zero, as there would be no unlocking. CSR would also be zero, as there would be no need for the user to either set up authentication, or specify any access control preferences. This scenario was important to consider, as it was reported in 2020 to represent the state of practice for 10% of users [80].

- ***NOTHING***: This was another possible case for the all-or-nothing model. In this case, it was assumed that the user would enable authentication. Therefore, the phone would always be locked, and no tasks could be performed before unlocking it. In such a case, any task labelled as either shareable with “Anyone” or with “Specific People” would count as a false negative, increasing FNR[4]. FPR would be zero because authentication would always be performed. The CSR would be zero as well, as the user would only need to set up authentication once, and would not specify any authorization preferences. This scenario was also important to consider, as it represented the state of practice for nearly 90% of phone users [80].

- ***LOCK_SCREEN_ACCESS***: This scenario corresponds to the lock screen app access available on most phones. Here, we assumed that all tasks by Camera, Calculator and Flashlight apps were available before unlocking. Everything else was similar to *NOTHING*.

- **APP_CONSERVATIVE**: This represented a case where users were able to choose to lock apps individually, but not specific tasks within them. It was representative of how Progressive Authentication [98] or other app-level task-sensitivity solutions (see Table 4.1) perform. Calculating FPR and FNR for it, however, required resolving conflicts in the participant’s labelling of app’s tasks. In this “conservative” scenario, we assumed that the user would want to lock an app, if they labelled any of its tasks as “No-one”.

[3] This is because the tasks would be exposed to unauthorized users, even if not necessarily performed by them.

[4] We are assuming there is no PIN sharing with the sharee.

- **APP_MAXIMAL**: Similar to the *APP_CONSERVATIVE* scenario, but used a different strategy for conflict resolution. In this case, the model would assign access to an app based on the majority vote of the user's labelling of its tasks.

- ***TASK_CONSERVATIVE***: In this conceptual scenario, we assumed that the system has perfect knowledge of the user's sharing preferences. We devised this scenario to gauge an upper bound for the efficacy of task-based access control. However, to measure FPR and FNR, similar to conflict resolution for app-level locking, in this "conservative" task-level case, we assumed that the user would prefer the most restrictive of their provided labels for each task. For example, if they would label a task once as being shareable with "No-one" and another time as being shareable with "Specific people", this model would select "No-one" as the final label. By doing so, any labels that are lower than the most conservative one would count as a false negative, due to unnecessary authentication.

- ***TASK_MAXIMAL***: This was similar to *TASK_CONSERVATIVE*, but we assumed that the user would prefer to have tasks protected according to the majority vote of their labels. For example, if the user had labelled a task twice as being shareable with "No-one" and once as shareable with "Specific People", the system would select "No-one" as the final label. As such, any instance of a task being labelled with something other than the decided final label would count as either a false positive (in case the final label was less restrictive than the current labeling), or false negative (in case of the opposite).

For phone-sharing solutions, we considered four scenarios:

- ***ALL_OR_NOTHING***: We assume that the participant either has not enabled unlocking, or has enabled it but has shared their passcode with the sharee. Hence, we calculate FPR and FNR by assuming that the sharee has access to all tasks on the phone. This scenario represents the state of practice amongst most users [77].

- ***PROFILE_SWITCHING***: We assumed a case where the participant had created a separate profile for each sharee, and has allowed them to install their own version of each app. Therefore, FPR and FNR were calculated assuming that the sharee cannot access any of the tasks not shared with them.
- ***DIFF_USER***: We assumed that the participants assigns all sharees to the "guest" group. Then, we calculate FPR and FNR by determining whether the sharee can perform a task, based on what resources (e.g., WiFi) they have access to, according to DiffUser's rules [86].
- ***X_SHARE***: We assumed that the participant would put the phone in restricted mode, before sharing it. We also assumed that they would only allow access to the apps they had reported sharing in that session. Therefore, FPR and FNR was computed holding that sharee can perform any tasks with shared apps, but none with others. In addition to xShare [72], this scenario also represents how session-based solution would perform (see Table 4.1), as (un)pinning apps mid-session can achieve the same effect.

RQ3 [Phone-Sharing Contextuality]: For each participant, we examined the variations of the following factors in all their reported cases of phone sharing: location, participant's presence, their relationship with the sharee, and functionality category of the shared tasks. We focus on these factors only, as they were reported by prior qualitative work to be correlated with sharing needs of users (see Section 4.1). Our goal was to corroborate such correlations quantitatively.

4.3 Results

4.3.1 RQ1: Performed Tasks

Overall, the participants reported performing a large and functionally-diverse set of tasks. Collectively, they declared 1,149 distinct tasks in total (after aggregation, as described in Section 4.2), for 571 distinct apps. On average, each participant reported performing 74 tasks (min = 24, max = 142), using 48 apps (min = 12, max = 103). In terms of functionality, we identified 19 separate categories, and observed that the tasks are distributed relatively evenly amongst them, as demonstrated in Figure 4.2.

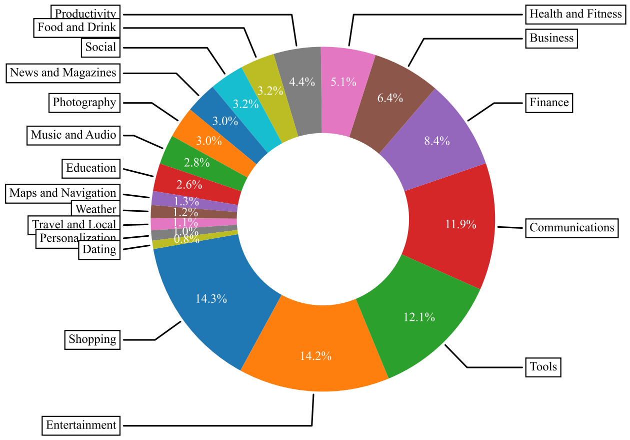


Figure 4.2: Distribution of participants' declared tasks across functionality categories.

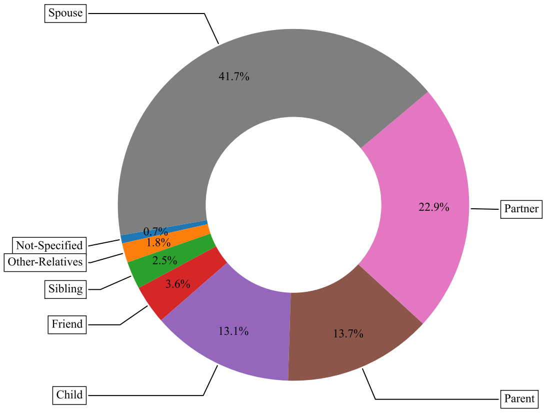


Figure 4.3: Distribution of participants' "Specific people" tasks across relationship groups.

Examining the participants' access control preferences for the tasks, we found them to be highly complex. The participants indicated to be willing to share a large portion (nearly 43%) of their tasks with others. Specifically, 23.7% of the

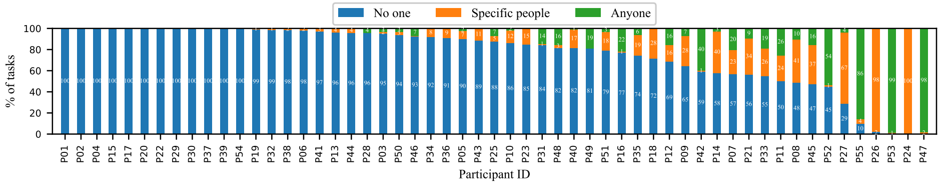


Figure 4.4: Distribution of the sharing preferences of performed tasks for each study participant.

tasks were labelled as being shareable with “Specific People,” whereas 19.4% were shareable with “Anyone”. [5] These “Specific People”, however, was not limited to a narrow group of people, confirming the conjecture of Karlson et al. [60] and others [77, 78]. Spouses and boy/girl-friends comprised the largest groups of preferred sharees (64.6% of the tasks shareable with “Specific People”). There were a myriad of other sharee groups as well, such as parents (13.7% of tasks), children (13.1%), friends (3.6%), and others (see Figure 4.3 for a more detailed breakdown).

Adding more complexity to the matter, we found the access control preferences to vary significantly in-between participants, as Figure 4.4 illustrates. About 22% of the participants were unwilling to share any tasks with anyone (e.g., participants with IDs P1, P2, and P4). We refer to them in this dissertation as *“Private” users*. A small fraction, 5% , (*“Public” users*) wanted to share everything with certain people (e.g., P24, P26). Unsurprisingly, the majority (73%) (e.g., P40 and P7) was in between these extremes. We refer to them as *“Semi-Private” users*. Obviously, for “Public“ or ”Private” users, an all-or-nothing model of access control would suffice. However, as our data shows, a more complex system is needed for *Semi-Private Users*. We will discuss this matter further in Section 4.3.2.

Moreover, we found the above-mentioned groups to be not easily distinguishable using demographic or phone usage factors. We tested the association between access control categorization (i.e., whether a participant is a “Private”, ”Public” or “Semi-Private” user) and several demographic and phone usage factors. These included age (which was suggested by Qiu et al. [95]), education level, hours of phone usage per day, depth of smartphone adoption (as measured by the privacy

[5] Note that labeling a task as either “Specific People” or “Anyone” does not indicate that the participant actually shared that task. We discuss actual sharing in Section 4.3.3

Table 4.3: Results of chi-squared tests of association between participants' access control categorization and whether the reported sharing their phones, with our anticipated antecedents of them. Significant p-values are underscored (assuming $\alpha = 0.05$).

	Access Control Categorization (Public, Private, Semi-Private)	**Whether They Share Phone**
Age	p = 0.002, V = 0.433	p = 0.048, V = 0.379
Education	p = 0.293, V = 0.258	p = 0.367, V = 0.240
Hours of Phone Usage Per Day	p = 0.350, V = 0.201	p = 0.967, V = 0.035
Privacy App Adoption	p = 0.654, V = 0.149	p = 0.279, V = 0.215
Living with Others	p = 0.577, V = 0.141	p = 0.158, V = 0.190

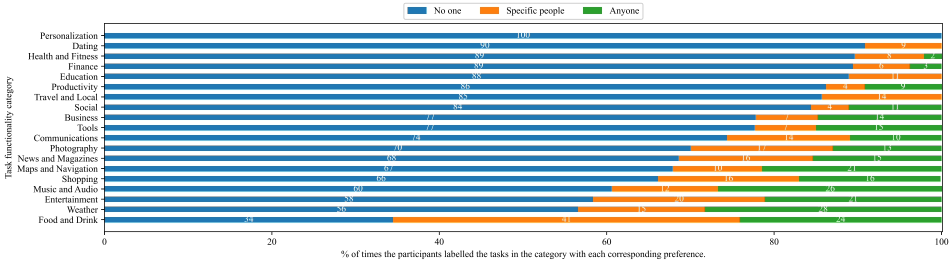


Figure 4.5: Distribution of sharing preferences across task functionality groups.

app adoption questionnaire proposed by Marques et al. [75] and modified by us as described in Chapter 2), and whether the participant lived with someone else or shared their phones with them (as we anticipated it would lead to different sharing habits). However, apart from age, none of the test results (presented in Table 4.3) were statistically significant $(p - values > 0.05)$ or strong $(Cramer'sV < 0.5)$. For age, while the link was statistically significant, it was not strong.

Finally, the non-uniformity of access control preferences was also observed in-between the categories of tasks. Specifically, we found that different functionalities have different access control needs. As Figure 4.5 illustrates, while some categories of tasks were off-limits to others (e.g., *Personalization* and *Dating*), other categories (e.g., *Food and Drink*, and *Weather*) were more often labelled as shareable, either with "Specific People" or with "Any-one."

Next, we examined the data with higher granularity, to compare participants' task and app-level preferences (the latter was previously examined by Hayashi et al. [54]). Firstly, we found that apps often afforded tasks with conflicting functionality. Our participants declared 2 different tasks per app on average, which were often from different categories. For example, "Business" apps sometimes afforded tasks from Communication, Productivity, Education, or Personalization categories (see Figure 4.5).

Secondly, in terms of sharability, we also found apps to be in-cohesive. We observed that, for each participant, apps generally fell into one of the following categories when it came to the shareability of the tasks they afford:

- **Public apps**: These were apps that afford exclusively tasks shareable with "Any-one". Hence, conceptually, these apps would require no form of authentication or authorization performed before access.

- **Private apps**: In direct contrast to public apps, private apps afford exclusively "No-one" tasks. Therefore, they would practically always require authorization before access.

- **Shareable apps**: These apps afford exclusively tasks that were shareable with specific people. Hence, they would require authorization support for different users.

- **Split apps**: These apps include tasks with different levels of shareability and, as such, do not fit in the above categories.

Obviously, for "Public" and "Private" users, all apps fell either in the Public or the Private category. However, for the majority of the participants (73%), the distribution of apps was not just between these two categories, as Figure 4.6 illustrates. The majority of the participants used Split apps, which requires a fine-grained model of access control. This result also confirms the conjectures of Hayashi et

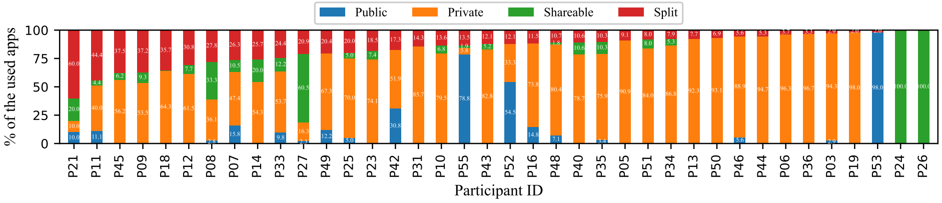


Figure 4.6: Distribution of shareability categories amongst "Semi-private" study participants.

al. [54] that smartphone users consider a significant number of apps they use as "Split", demonstrating that the all-or-nothing model lacks important sensitivity to the sharebility of individual tasks in the apps.

Lastly, we found that the perception of the shareability of an app is not consistent across study participants. Amongst the 206 apps that were used by more than 1 participant, 137 (66.5%) were categorized differently by their users. For example, for the "Amazon" shopping app, 65% of its users categorized it as "Private", 29% as "Split", 3% as "Shareable", and the rest as "Public".

In summary, and to answer RQ1, we found users to perform a large (1,149) and functionally-diverse set of tasks (19 different categories) with their phones. We also found their access control preferences for the tasks to be highly-complex and varied by sharee (see Figure 4.4 and functionality (see Figure 4.5). Lastly, in agreement with previous work, we found apps to be incoherent units in terms of access control needs, with significant variance in functionality and shareability of tasks they afford (see Figure 4.6).

4.3.2 RQ2: Comparison of Solutions

We unsurprisingly found the granularity of the incumbent all-or-nothing solution to be inadequate. As Figure 4.7 clearly illustrates, in the ALL scenario, the FPR was estimated at 90.3%, meaning that nearly 90% of the users' tasks would be exposed to unauthorized users. This result clearly demonstrates the high risk associated with this scenario, even though 10% of users choose it anyways [80]. This acceptance of risk, however, becomes somewhat justifiable when we consider the NOTHING scenario. As the figure shows, while NOTHING exposes no tasks to

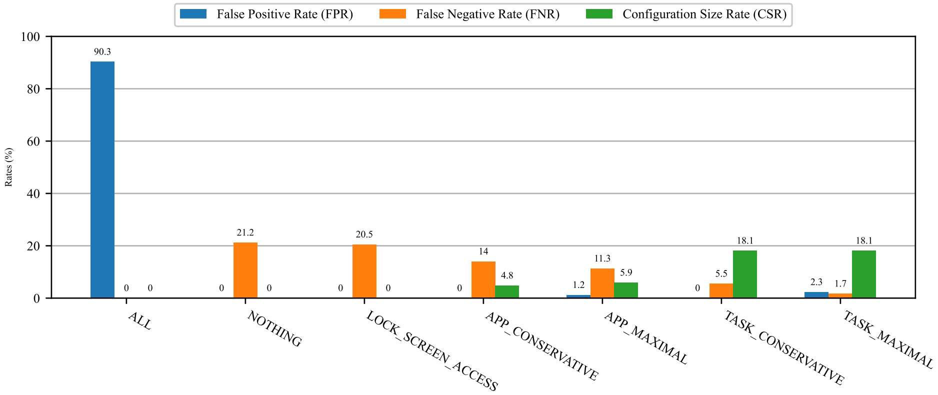


Figure 4.7: Comparison of FPR, FNR and CSR amongst solutions that aim to provide task-sensitive access control on smartphones.

unauthorized users (FPR = 0%), a 21.2% FNR would be incurred, meaning that up to 20% of the users' unlockings could be unnecessary.[6]

We found the current commercial solution (lock screen access) to not offer much improvement, either. Obviously, the ultimate goal of a task-sensitive solution would be to reduce the FNR compared to the incumbent NOTHING model, without increasing CSR or FPR substantially. Yet, as Figure 4.7 shows, the LOCK SCREEN_ACCESS solution fails to do so. It reduces FPR only by 1%, but with no change to the other metrics. This is unsurprising given the limited number of tasks afforded by the common lock screen apps (Camera, calculator, etc).

What seems to be effective, however, is increasing the granularity of access control. As the figure shows, app-level models of access control (represented by APP_CONSERVATIVE and APP_MAXIMAL) reduce FPR by 7% with only slight increases in CSR and FNR (0.5% and 5.5%, respectively). This was the case for both the CONSERVATIVE and MAXIMAL scenarios, which suggests that the increased granularity is beneficial, even app-level decisions are made based on most-restrictive tasks (which is the CONSERVATIVE case). It is out of the scope of this dissertation to gauge which scenario would be preferable to the end users (as it would require us to assume a particular UX). However, on a conceptual level,

[6] FNR of ALL and FPR of NOTHING do not necessarily add up to 100%, as "Specific People" tasks would incur False Negatives but not Positives, as we assume users do not share passcodes.

it is clear from our data that app-level control can achieve the ultimate goal stated above.

Lastly, increasing the access control granularity to task-level was found to be even more beneficial, albeit with a more noticeable trade-off. As Figure 4.7 shows, TASK_CONSERVATIVE and TASK_MAXIMAL could further reduce the FPR to as low as 1.7%, but with an 18.1% increase in CSR. This increase could be manageable, however. Considering our earlier finding that users perform on average 74 tasks on their phones, this CSR value would mean that users would need to specify as few as 13 sharing preferences, to configure such a system. However, in contrast, configuring the NOTHING model would only require a one-time set-up of unlocking, which could still be a noticeable trade-off based on user preference.

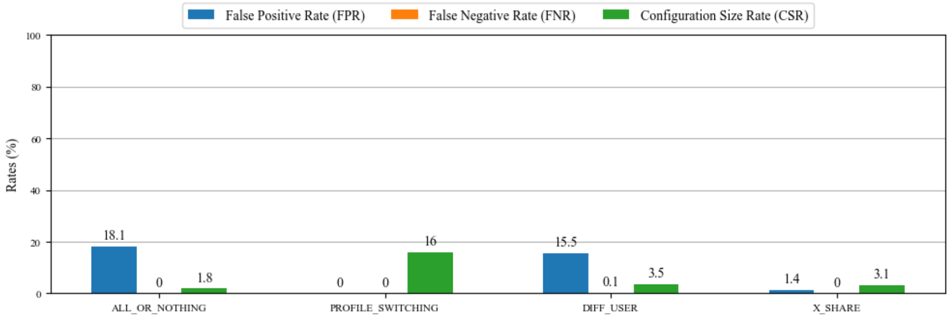


Figure 4.8: Comparison of FPR, FNR and CSR among solutions that aim to provide phone sharing support on smartphones.

As for support for phone sharing, we again found the all-or-nothing model to be inefficient. As Figure 4.8 depicts, the ALL_OR_NOTHING scenario would lead to sharees not facing any unnecessary restrictions (FNR = 0%, because we assume the PIN is shared with the sharee). However, more than 20% of the sharer's task would be exposed to unauthorized users as well, increasing the potential for frequent security/privacy violations that other studies have reported [75, 76].

In comparison, profile switching might appear to be the ideal solution. Naturally (and in direct opposition to task-sensitivity), the ultimate goal for any phone sharing solution would be to reduce the FPR compared to the ALL_OR_NOTHING model, without a substantial increase in FNR or CSR. Profile switching, as shown in Figure 4.8 seems to be achieving this goal, as it would zero FPR and FNR. However, the scheme comes with a substantial increase in CSR as well, which seems to explain why users are reluctant to use it [17, 32].

Resource-based solution, DIFF_USER, does not seem to offer much improvement either. While it can reduce FPR by 5%, it also increases FNR and CSR by 0.1% and 1.7%, respectively, which diminishes its FPR gains. This is unsurprising, given our previous finding regarding the multi-faceted nature of current apps, in terms of functionality and task shareability (see Section 4.3.1).

Finally, X_SHARE appears to provide the best balance between usability and security. As seen in Figure 4.8, it eliminates most of unauthorized accesses (FPR = 1.4%), while not increasing the FNR at all, and CSR by only 1.3%. Thus, amongst all the models, it seems that the session-based models could perform the best.

It should be noted that CSR calculations does include the effort needed to put the phone into sharing mode, each time the phone is shared. Still, the CSR for X_SHARE is substantially lower than PROFILE_SWITCHING because while X_SHARE only allows for one type of restricted sharing session (hence the user is only required to specify the apps that should be available in such sessions once), PROFILE_SWITCHING requires the primary user to not only create a separate profile for each secondary user (at least those that have different Tasks shareable with them) but also specify which apps should be available for each user. This create a much higher initial configuration burden (as seen in Figure 4.8), while providing not as substantial improvement in FPR and FNR. However, we should still note that the increase in FPR and FNR by using X_SHARE might not be acceptable to all users, and, ultimately, which solution is the best depends on the individual.

In summary, and to answer RQ2, our results demonstrate the inefficacy of the incumbent all-or-nothing model. They also offer evidence that the current commercial solutions (i.e., LOCK_SCREEN_ACCESS) do not offer much improvement. Furthermore, the findings show that more granular access control models (app- or task-level) might be the best trade-off in order to support task sensitivity. Session-based control appears to offer a better balance between security and usability for phone sharing.

4.3.3 RQ3: Contextuality of Phone Sharing

Firstly, we found actual sharing of phones (as opposed to willingness to share tasks, which was the basis of the results presented in Sections 4.3.1 and 4.3.2) to be not as prevalent as reported by previous studies. Among our participants, only 16 (29%) reported sharing their phones with others. In total, 58 sharing events were reported over the course of the study. In contrast, Jacobs et al. [58] estimated that nearly 50% of users would share phones. Mathews et al. [77] reported 14 sharing instances per participant (N=25) over 21 days. Although it is out of our scope to investigate, we anticipate the decline in sharing to be due to the increase in smartphone penetration of the consumer markets [110] since the time of those studies (2016), and the effects of the COVID-19 pandemic [88] (which we will discuss in Section 4.5).

However, similar to willingness to share, we also found the tendency to actually share phones to not to be easily distinguishable using demographic factors. As Table 4.3 shows, nearly none of our anticipated antecedents were correlated with someone reporting at least one instance of sharing ($p > 0.05$, Chi-square test).

Interestingly, the practice of sharing does not seem to be strictly aligned with access control preferences either. Evidently, amongst the participants who reported sharing, one was a *Private* user, one was *Public* and the rest (14) were *Semi-Private*. This means that, even though at least one participant ("Private" user) reported not be willing to share any of their tasks with anyone, they reported sharing their phones on one occasion anyways. Hence, it seems that sharing can indeed be impromptu sometimes, as suggested by Matthews et al. [77].

As for the conjectured contextual factors of phone sharing (the main focus of RQ3), we found them to be fairly consistent. We first examined the effect of content on sharing preferences. Previous qualitative studies of non-representative samples [54, 58, 78] conjectured that sharing preferences depend on the content being shared. Our results confirm this quantitatively. We observed that only specific functionality categories of tasks were reported to be shared by our participants. As Figure 4.9 shows, most of such tasks were from a few specific categories (e.g., Food and Drink, Entertainment), while tasks in some other categories (e.g., Personalization) were never not shared. Overall, 11 (58%) of all categories had at

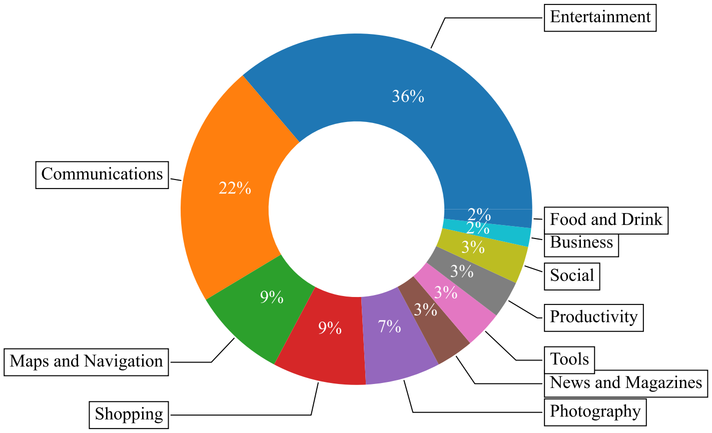


Figure 4.9: Distribution of categories of tasks that the participants reported sharing.

least one reported case of sharing, with nearly 70% all shared tasks being from three specific categories: Entertainment, Communication, and Maps and Navigation. Whereas tasks in 8 (42%) of categories were not shared at all. This aligns well with the participants' reported willingness to share tasks (see Figure 4.4).

Next, we found sharer's presence to be also a determinant factor of sharing. This was previously suggested by Mathews et al. [77]. Our data confirmed that, as the majority (96.6%, 56/58 instances) of sharing events (at least those known to the sharer, as our our study relies on self-reported data) were reported to have happened in the presence of the sharer.

We also found the relationship between the phone owner and the sharee to be a significant factor. It was suggested previously [17, 60, 78] that users' sharing preferences depend significantly on who the phone is being shared with. We found that to be the case, too, as a majority of sharing events were with people that the owner was familiar with. As Figure 4.10 depicts, more than half (52%) of all sharing happened with romantic partners, while sharing with children and parents were the next most frequent, at 21% and 16% respectively. Overall, we estimate that 89% of sharing instances happened with the immediate family of the phone owner. This finding also aligns well with the preferences reported by the participants, discussed in Section 4.3.1.

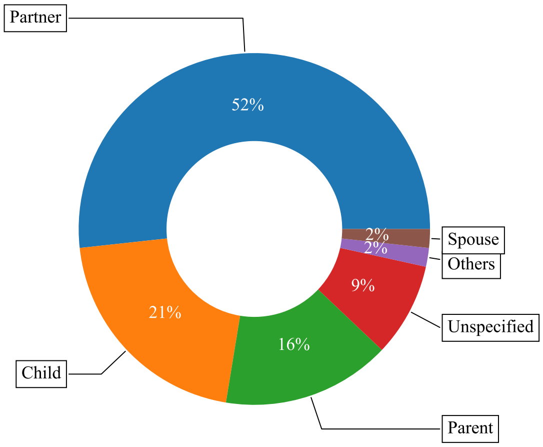


Figure 4.10: Distribution of relationships between participants and people With whom they reported sharing phones with.

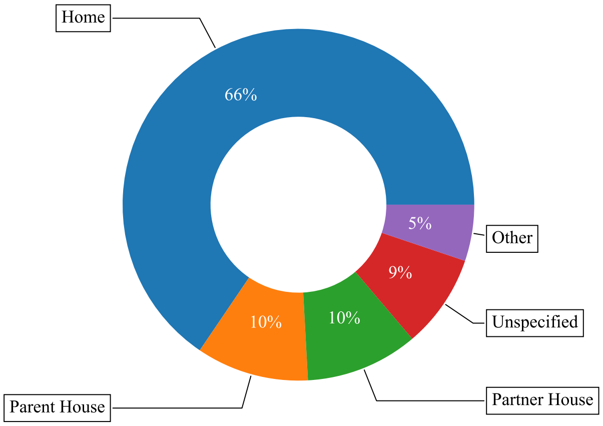


Figure 4.11: Distribution of locations in which the participants reported sharing with others.

Lastly, we found the location to also be fairly consistent across sharing events. As Figure 4.11 shows, 66% of all sharing events happened at the sharer's home, with the homes of parents and romantic partners being the next most frequent, at 10% each. Only 5% of the events happened at other locations, such as schools or public places.

In summary, and to answer RQ3, sharing is rather universal, and the contextual factors of phone sharing are highly consistent. Even though we found sharing to be not as prevalent as previously reported, we found phones to be shared often with family members and partners, at familiar locations, in the presence of the sharer, and to perform a few certain categories of tasks.

4.4 Discussion

Putting all of our results together, it is clear that modern smartphone users have diverse and complex access-control needs. They perform a functionally diverse set of tasks with a large number of apps, and they prefer to share several (sometimes even partially overlapping) subsets of tasks with different individuals (see Section 4.3.1) and in varying contexts (discussed in Section 4.3.3).

Making matters more complex, authorization needs vary significantly by task functionality (see Figure 4.4) and per user (see Figure 4.6), and it is difficult to predict users' preferences based on their demographic or phone-usage factors (see Section 4.3.1). Also, no functionality categories dominate the users' tasks to therefore limit the scope of access control to certain activities on the phone (see Section 4.3.1).

All these circumstances reduce the chance that any ad hoc solution (e.g., lock screen access) catering only to specific tasks or groups of users could achieve adequate efficacy. Hence, the need for robust and general-purpose access control on smartphones is now clearer than ever.

Yet the incumbent all-or-nothing solution falls short of this ideal. Firstly, it fails to provide timely and secure access to tasks, as its assumption that all tasks require the same level of protection is invalid. Our findings (Section 4.3.2) suggest that it could mistakenly expose more than 90% of the users' private tasks to unauthorized individuals or, conversely, unnecessarily hinder the users' access to 21% of their tasks.

Secondly, the solution fails to assert adequate control when users share their phones. We found that it could expose up to 20% of the users' private tasks to unauthorized sharees, as it forces the users to share passcodes (Section 4.3.2). This is especially important considering the clear distinction the users make between "Anyone" and "Specific people" tasks (Section 4.3.1). To them, a shared task is not necessarily a public one. So, even if they are willing to share a task, it is important to control with whom that task is shared.

As such, the dilemma users face when using the incumbent system is clearer now. They have to either (1) enable authentication and face the overhead of 20% unnecessary authentications or (2) disable unlocking altogether for more convenience and risk 90% of their tasks being available to unauthorized users (as 10% of users actually do [80]).

Making the situation direr, most of the proposed/implemented alternative solutions do not seem to provide much improvement. The lock screen access solution only provides an insignificant improvement to the FNR (0.7%), and phone-sharing solutions, such as DiffUser, rely on restricting access to system resources to assert control, which nearly doubles the CSR in exchange for reducing the FPR marginally (see Section 4.3.2).

Profile switching presents more of a conflicted situation. Even though our results show it could theoretically eliminate FPs altogether, its high CSR explains why most users decide not to adopt it [17, 32]. Moreover, we found evidence for the need to support impromptu phone sharing (see Section 4.3.3). However, profile switching does not easily support impromptu sharing, since it would require creating profiles on the spot.[7]

It seems, therefore, that users are left with neither timely access to their tasks nor proper control over whom performs them. But hope exists.

We found designs that could lead to better access control. To support task sensitivity, we found increased granularity a good starting point. The app-level models we examined were found to have nearly half the FNR of the widely used all-or-nothing system (11% vs. 21%) but with only modest increases in the CSR and FPR (see Section 4.3.2). However, the app-level model seems to be the sweet spot of granularity; creating even finer-grain task-level solutions would halve the FNR once more, while imposing a substantial increase in the CSR (18%). Thus, overall, our data supports wider adoption of app-locking solutions, which are currently only deployed on some Android phones [101].

For deliberate phone sharing, our results endorse session-based solutions. Ap-

[7] This creation would only be needed if the shared tasks were for “Specific people.” Otherwise a “Guest” profile would suffice.

proaches such as xShare [72] and app pinning, which allow users to quickly select which apps to share in each session, showed substantial reduction in the FNR (20%), only a modest increase in the CSR (1%), and no increase at all in the FPR (Section 4.3.3).

Finally, our data also shows promise for context-based phone-sharing solutions. We found several contextual factors to be highly consistent (see Section 4.3.3), which could be incorporated into future solutions:

- Content. Only certain categories of tasks were shared by our participants. Thus, future access-control schemes could further reduce the FPR with a default denial of access to task categories universally perceived as private.
- Presence of the sharer. Deliberate sharing often happens in this situation (see Section 4.3.3). Thus, detecting the owner's presence (e.g., through smartwatches or other wearables) seems to be a promising approach to detect unauthorized access.
- The sharee. Phone sharing often happened with the same group of people (see Figure 4.10). Thus, an a priori policy definition and user isolation solutions (e.g., profile switching) could indeed work but, as discussed before, need to be made more convenient. The high proportion of close family members in the sharee distributions suggests automatic user identification (e.g., through behavioral biometrics [84]) could be one way of achieving this goal and poses a promising avenue of future research.
- Location. Most of the reported sharing events happened at familiar places, such as at home (see Figure 4.11). As such, location detection (e.g., using Wi-Fi or GPS as done by Google Smart Lock for Android [41]) could provide better access control by limiting unauthorized access in unfamiliar locations.

In the end, we should note that our findings also demonstrate the need for flexibility in access control, as any one-size-fits-all solution would have unjustifiable trade-offs for some users. For example, if all users were forced to a highly granular task level, some users could benefit. But this level of access control would also impose a high CSR on semiprivate users, with no improvements in the FPR or FNR for them (e.g., P03 in Figure 4.6 who has few public apps and no split apps). For such users, an app-level system would be a better fit. Hence, it is important to consider solutions that increase granularity only when needed or only for those users who need it. This could certainly be achieved manually (e.g., by asking users directly to use a task-level system). However, as observed with profile switching (see Section 4.3.2), the increased cognitive load makes adoption an issue. An avenue for future research, therefore, can be to investigate the feasibility of automating this process (possibly like our implicit identification suggestion for profile switching).

4.5 Limitations

Firstly, similar to other studies on smartphone usage [29, 51, 57], our sample was not representative of the global smartphone user population. For example, it has been shown that cultural factors affect users' unlocking behavior [53] and privacy attitudes [100]. However, due to limited resources we only included US participants in our study; as a result, one cannot generalize the results to non-US users.

Also, our sample is not fully representative of the US smartphone user population either. While MTurk is shown to provide quality data for research in usable security [97], its known limitations (e.g., lack of diversity, tech savviness) [90] apply to our study too (for example, our participants are more than 80% white). Lack of diversity is a common limitation of smartphone studies [29, 54, 77]. However, we believe that our findings are still valuable, as they provide the very first quantitative insight into users' access-control needs.

Secondly, due to technical limitations we only included Android users in our study. Recent evidence suggests there are no significant differences in security/privacy attitudes between iOS and Android users [1]. However, cross-platform studies are required to investigate this matter further.

Thirdly, our results are based on self-reported data from the participants. Thus, as a general limitation of such studies, our data might sometimes be of lower ecological validity and not perfectly reflective of the users' true behavior [39, 80]. Also, the reported sharing events might be affected by the users' prolonged use of the all-or-nothing system. For example, the type of tasks they shared might have been influenced by what they felt comfortable sharing given the limitations of the incumbent system. Also, all-or-nothing might have caused participants to misinterpret the diary questions and believe that "Whom you'd generally allow to perform the task" meant only when they were physically presiding over the sharing. To mitigate this risk, we conducted several pilot studies (see Section 4.2) and did not find evidence of such varying misinterpretations. But they are still a possibility and could have caused us to underestimate the number of "Anyone" tasks. Having more such tasks, however, could only strengthen our argument that the incumbent system is suboptimal, as these tasks increase its FPR (already over 20%) even further.

Fourthly, in order to limit the amount of time required for the participants to fill a daily diary, we did not collect data about the perceived sensitivity of each individual *Task* (e.g., on a Likest scale). As such, one implicit assumption of our findings is that all false positive or false negatives are at the same level of importance (i.e., they degrade security/usability to the same degree). This, however, might not always be the case. For example, unauthorized access to a banking *Task* might have far worse consequences for a user, compared to such access to a gaming-related one. So, theoretically at least, the former should be assigned a higher weight in FPR or FNR calculations. However, our data collection design did not allow us to do this. As a result, while our results are beneficial in proving a first detailed look into the users' needs, they do not provide the whole picture of users' access control needs. More detailed studies are needed to fully capture the effect of individual *Task* sensitivity on the efficacy of existing systems.

Lastly, as discussed in Section 4.3.3, our results might have been affected by the COVID-19 pandemic. For example, work-from-home orders might have influenced users' security/privacy preferences for their personal devices, as they might now contain work-related information [88]. Also, the consistency of location in phone-sharing events might have been due to stay-at-home orders, and, generally, the users might have shared their phones less frequently than normal, due to not being around others.

4.6 Conclusion

Smartphones, nowadays, require strong physical security, as they host an ever-increasing array of data and services. Yet, most of the research and development so far have been towards their *"authentication"* subsystem. Their *"access control"* has remained largely unchanged, still using the all-or-nothing model. In this study, we solicited detailed task data from users, and quantified just how inefficiently this model serves the users. We found that most proposed/implemented alternatives do not provide much improvement, either. Instead, we found that increasing the granularity of access (up to a point) and session- and context-based control might be promising avenues to design future systems that better support the users' needs.

Chapter 5

Proposing Gradual Unlocking as a Solution for Task-Based Access Control on Smartphones

This chapter presents the design, implementation and evaluation of a solution for performing task-based access control on smartphones. The solution can also help with reducing unnecessary user interruptions when performing implicit authentication. First, in Section 5.1 we provide the necessary background on those parts of the Android Operating System (AOS) that are required to describe our solution. Next, in section 2.1 we provide an overview of the existing literature on how to mitigate the IA re-authentication issue. Section 5.3 describes the architecture of our solution and how we implemented a prototype of it on Android Open Source Project (AOSP) 10. In Section 5.4, we provide the results of the evaluation of our system, using data from the longitudinal diary study described in Chapter 4. Section 5.5 discusses the implications of our results. Lastly, Section 5.6 discusses the limitations of our evaluation and Section 5.7 concludes this chapter.

5.1 Background: Android Overview

5.1.1 Activities in Android Apps

Activities are one the four basic building blocks of Android applications, along with *Services*, *Content Providers*, and *Broadcast Receivers* [4]. As described by the Android Developer Guide: “An activity is the entry point for interacting with the user” [4]. Each activity roughly corresponds to a screen the user sees; it encompasses the functionality and presents the visual design of a single user interface page. Activities are independent entities, and each one often implements a different aspect of the app’s functionality [4]. As an example, a gaming app might have one Activity for representing the game-play screen and a different one for representing its settings.

5.1.2 SELinux on Android

SELinux is a an kernel extension that provides Mandatory Access Control (MAC) and Role-Based Access Control (RBAC) on Android [3]. MAC is a type of Access Control (AC) system in which both users and objects (e.g., files, documents) are assigned security levels. A user is only allowed to access an object if they have the necessary security level [102]. In RBAC, permissions are not directly assigned to users. Instead, they are associated with roles and, to acquire a certain permission, users have to be made members of the appropriate role [89].

5.2 Related Work

5.2.1 Task-Based Access Control

In the previous chapter, we used a conceptual task-level model of access control to gauge the upper limit of efficacy amongst the evaluated solutions. However, we did not discuss how such solution could be implemented or whether it was technically feasible at all. To the best of our knowledge, there exists no practical solution for performing such task-level authorization, currently. Crawford et al. [27] proposed a conceptual framework. However, as we will discuss later in this chapter, they do not provide a practical implementation.

In this chapter, we aim to propose such solution. However, we should acknowledge that, as we discussed in the previous Chapter, such solution is only necessary for certain groups of users that have highly complex access control needs. For the majority of users, an app-level model could suffice.

5.2.2 IA Re-authentication Problem

The potential efficacy of IA has been demonstrated in lab studies. For example, Crawford et al., [26] observed that their low-fidelity IA prototype resulted in their participants performing 67% fewer authentications, compared to explicit methods, such as PIN. Khan et al. [63, 64] also evaluated several proposed IA solutions for smartphones and found several to provide "acceptable" security, usability, and performance. Empirical data also suggests that there is interest in IA adoption among smartphone users, as found by our findings in Chapter 2, as well as existing literature [26, 64].

However, deploying IA in real-world has shown tricky. One of the main challenges is that IA often triggers uncalled-for re-authentication prompts when the authentication data stops being available, e.g., in case of gait-based authentication, when the user stops walking[1]. In such cases, to preserve security, the phone locks the user out mid-session and asks them to authenticate explicitly using PIN or biometrics. This results in IA behaving in an unpredictable and hectic manner, often interrupting user's tasks and causing "annoyance" [2].

Two general solutions have been proposed to solve the IA re-authentication problem. The first is to simply improve User eXperience (UX) when re-authenticate prompts are inevitability presented to the users. Agarwal et al. [2], for example, explored several strategies for how to present the prompts to the users. Each strategy involved a different visual transparency level for the prompt and a different time delay to show it. Their user study with 30 participants showed that the introduction of the time delay reduced users' "annoyance" by allowing them to finish their current task first.

Similarly, Mecke et al. [79] explored the idea of using visual cues to warn the

[1] In such case, it is reasonable to lock the phone because the situation might actually have been than an attacker has pickpocketted the phone owner as they were walking.

users of an imminent re-authentication prompt. They found that having indicators of upcoming prompts is better than no indication, as there will be less surprise for the users. However, they also discovered that, generally, the security sensitivity of the task being performed has a strong impact on user “annoyance”, and needs to be taken into account by future systems.

Overall, UX-based solutions, such as above, are only useful if the number of prompts is not too excessive. Hence, a second solution has been explored by several papers, which is to try to reduce the number of unnecessary re-authentications altogether.

Khan et al. [62], for example, proposed Itus, a framework that allows developers to designate certain Android Activities of their app as *sensitive*. An Itus-enabled phone then only prompts the user for re-authentication if a sensitive Activity is in the foreground at the time. The Itus authors showed that the framework is capable of improving usability by allowing IA to be implemented in an app-specific manner. The authors also demonstrated how this approach can be deployed with minimal performance overhead on the phone. However, deploying Itus in a large-scale requires involvement of developers to update their apps to include the framework, which is not ideal. Also, since the modifications made to the app by Itus cannot be customized per user (they are directly integrated into the code), a *task* is forced to be at the same level of sensitivity for all users, which is unrealistic, as we discussed in Chapter 4.

Earlier, Crawford et al. [27] proposed a conceptual framework for deploying IA which does not require modification to apps. This framework allows users to specify an authentication confidence threshold for each task. Re-authentication then only happens if the confidence is below the threshold, when user attempts the task. The framework, however, does not specify what exactly constitutes a “task” and how they can be distinguished amongst practically. Also, the authors do not explore how such a system can technically be implemented on the current mobile Operating Systems (OS).

Lastly, Riva et al. [98] proposed a solution called Progressive Authentication. It determines a level of confidence in the users’ authenticity, and only prompts for authentication when the app being launched requites high confidence. As opposed to Itus, however, this solution works on an app level, and does not consider tasks.

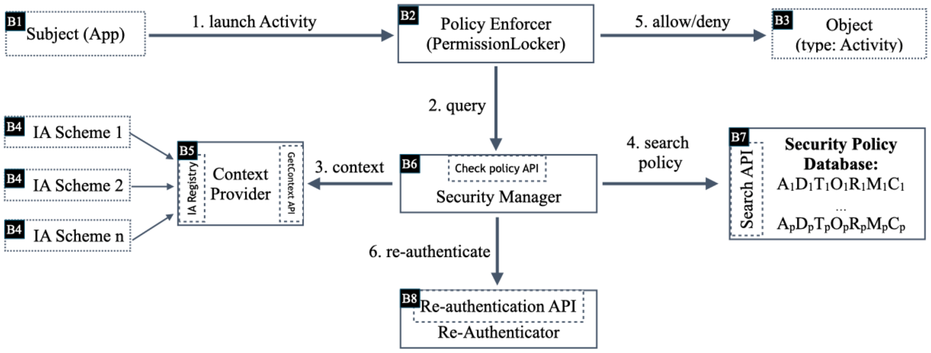


Figure 5.1: Architecture of the Gradual Unlocking System: the Components and Their Interactions.

In this chapter, we too aim at reducing the number of re-authentication prompts. However, unlike previous work, our solution does not assume a task to be at the same level of sensitivity for all users, does not require developer involvement, and provides a practical way of defining and differentiating between tasks on the phone. Like Crawford et al. [27], our solution, which we dub the Gradual Unlocking System (GUS), controls access to each task based on the confidence of authentication, However, unlike Crawford et al. [27], GUS uses the practical notion of Android Activities to represent tasks. Also, unlike Itus, the rules that determine who can access which task and under what authentication confidence are completely customizable by the user in GUS, and is not embedded in the app code. We will discuss how GUS operates in detail in Section 5.3.

5.3 System Design

In this section, we first provide an overview of how GUS operates. We then discuss in detail the architecture of GUS, including its components and how they interconnect.

5.3.1 Overview

GUS operates based on user-to-device authentication[2] and eliminates unnecessary re-authentication prompts by requiring different authentication conditions for different *Tasks* (this is as opposed to the all-or-nothing solution where there is only one authentication confidence threshold for the whole device, and the app-level systems where there is only one per app.) This makes it so unnecessary prompts are eliminated, when authentication data is lost but the task being performed is not sensitive (as designated by the user). For example, GUS allows a non-sensitive *Task*, such as checking the news, to be performed even if no authentication data is available at all. However, it denies performing sensitive *Tasks*, such as banking, unless the IA system is highly confident that the current user is authorized.[3]

To describe this approach in more detail, we formally define three concepts: *Tasks*, *Intentions*, and *Rules*.

Tasks: Conceptually, we define a *Task* as a distinct series of actions that could be performed with a mobile app, that are distinguishable in terms of purpose and access preferences. Technically, they comprise the functionality affordances of the app, similar to affordances of real world objects or User Interfaces (UIs) [87].

GUS provides a practical way of distinguishing between Tasks, which is through Android Activities. This means that GUS “sees” a Task as a series of Activities that launch on the phone. As a simple example, in case of GMail, sending email is done through the ”ConversationListActivityGmail” Activity, whereas receiving email is done by another Activity called ”ComposeActivityGmail”.

Using Activities to represent *Tasks* entails that we assume performing different *Tasks* cause the launch of distinct Activities on the phone. We discuss how the results of our user study demonstrates this conjecture to be valid, in Section 5.4.

[2] That is, the user is authenticated to the phone as a whole, not to each app or task individually. Hence, the phone decides which apps/tasks to make available to the current user, taking that burden away from apps. If any app, such a banking one, decides that the level of authentication achieved by the device is not sufficient for they purpose, they can still perform their own in-app authentication, such as asking for a password.

[3] Note that a task-level access control system still has limitations in terms of distinguishing between sensitive and non-sensitive tasks. For example, shopping can be both sensitive and non-sensitive, depending on who is present. We discuss later in this chapter how our system deals with these situations.

Intentions: They are security preferences that the users have in mind about how sensitive each *Task* is. They are a mapping between *Tasks* and three sensitivity levels defined in GUS:

- **Public**: *Tasks* that the users perceives to not be sensitive at all.

- **Semi-private**: *Tasks* that users wants only themselves or some other designated secondary users to be able to perform.

- **Private**: *Tasks* that users want only themselves (and no other secondary users) to be able to perform.

To configure GUS entails that the user makes their *Intentions* clear for each Task, by assigning a sensitivity level to it. They could do so based on the consequences of an unauthorized *Task* execution. As an example, consider a hypothetical mobile game. The owner of the phone can label the "Game Activity" as *Public*, and the "In-App Purchases Activity" as *Private*. In this case, any user (e.g, the user's children) can play the game with low authentication confidence, which provides better IA usability. However, GUS prevents financial purchases unless a high authentication confidence for the primary user can be achieved.

Rules: *Rules* are formal descriptions of the users' *Intentions*, via a machine-parseable language. They encompass conditions under which an Android Activity (which is assumed to be representative of a *Task*) can be accessed. The conditions that GUS can enforce include: confidence of authentication, the source of authentication (i.e., whether the user is authenticated using IA or other explicit authentication methods), and the identified user. An example *Rule* can be to "Deny launching any Activity of the banking app if the identified user is not the primary user or the user is identified through IA and the authentication confidence is below 99%."

Using the three concepts described above, GUS is designed to enforce *Rules* that preserve the users' *Intentions* for their *Tasks*. Specifically, once the user makes their *Intentions* clear through a UX, GUS will translate those preferences to *Rules*, which are then enforced at run-time on the phone to control access to *Tasks*. While designing such UX is out of the scope of this dissertation, we envision a simple UI that allows the user to assign sensitivity levels to Activities an app includes.

The list of Activities can be accompanied by short descriptions (provided by the developer) about what the supported functionalities of each are. Doing so seems practical as Android already allows the developer to add descriptions about how and why each requested Permission is used by the app [4].

5.3.2 Architecture

To enforce the user's *Rules*, GUS takes advantage of a compartmentalized architecture, shown in Figure 5.1. This design is inspired by SELinux [105], the FLASK architecture [107], and FLASKDroid [18]. The inspiration is in the way the *Rules* are defined, and the separation between the entities that enforce the *Rules* (we call these entities Policy Enforcers), the entity that store the *Rules* (called Security Policy Database), and the entity that authenticates the user (called the Context Provider). This design allows for (1) easy addition of extra IA schemes (e.g., in case one wants to deploy a new one on the phone); (2) increased system stability, in case of failure of one of the IA methods to authenticate the user (when there are multiple IA methods deployed); and (3) ease of defining a new way of distinguishing between *Tasks* in the future (e.g., through Android Permissions, instead of Activities).

To explain how GUS components interact, let us take an earlier example where an authorized user (e.g., a child) wants to make in-app purchase in a Gaming app. GUS access control process starts when the App (Box B1 in Figure 5.1) tries to access/launch the purchasing Activity (B3). This request is intercepted by a Policy Enforcer (B2) which consults Security Manager (B6) to see if the access is allowed. To reach a decision about whether to authorize this *Task*, Security Manager (B6) first queries the Context Provider (B5) to identify the current user and calculate the confidence of authentication. Context Provider, in turn, queries all IA schemes deployed on the phone (B4 boxes) and calculates an aggregated confidence level for each user and reports them back (let us assume the confidence is at 80% for the owner). Upon receiving this data, Security Manager (B6) contacts the Security Policy Database (B7) to see if there is a *Rule* that matches the current conditions (app, resource, user, confidence). If there is a rule to deny the access (in this example, suppose there is one that denies the access below 99% confidence), Security

Manager (B6) will respond to the Policy Enforcer (B2) to reject the access and then will contact Re-Authenticator (B8) to initiate a re-authentication prompt. If there is no rule to reject the access (or there is one, but it is masked by another rule that allows the access), Security Manager (B6) allows the access to proceed.

We now explain the details of each GUS component corresponding to boxes B2, B5-B8 in Figure 5.1:

1) Security Policy Database (SPD, B7): This component houses the *SecurityPolicy*, which consists of a set of *Rules* defined to protect the *Tasks*. To define each *Rule*, a *Policy Definition Language (PDL)* is needed. We designed a PDL (inspired by SELinux) that defines each rule as a tuple ($Action_i$, $Operation_i$, $Domain_i$, $Type_i$, $Role_i$, $Modality_i$, $Confidence_i$).

To explain what each rule component is, consider this example: the rule to deny our example gaming app access to the purchasing Activity when the phone owner is authenticated with less than 99% of confidence can be expressed as (*DENY*, *ActivityLaunch*, *GamingApp*, *com.example.gamingapp.purchasing_activity*, *PrimaryUser*, *IA*, 99%). We colloquially read this rule as ”DENY any Activity launch attempt made by the GamingApp to the Purchasing Activity if the user is not the PrimaryUser or it is but he/she is authenticated through IA with less than 99% confidence”. In general, for each rule, we have the following components:

1. $Action_i$ specifies what GUS will do (either ALLOW or DENY the *Task*) if the conditions of this security *Rule* (user, confidence) are met. In the example rule, the *Action* is *DENY*.

2. $Operation_i$ specifies the type of *Task* this *Rule* protects, Currently, GUS only recognizes Tasks as Activity launches. Therefore, the value will always be *ActivityLaunch*. However, the inclusion of this field allows the system to be expanded in the future, to explore other ways of representing *Tasks* (e.g., through Android Permissions).

3. $Domain_i$ specifies the app targeted by the *Rule*. In our example, the *Domain* is *GamingApp*. This field also allows to specify a group of apps as targets, to facilitate easier *Rule* development. We refer to a single app as *Subject*, and to a group of apps as *Domain*.

4. *Type$_i$* is the resource (the Activity that is representative of a Task) targeted by this *Rule*. In our example, the *Type* is *com.example.gamingapp.purchasing_activity*. This filed also allows for grouping of resources. We refer to a single resource as *Object*, and a group of resources as *Type*.

5. *Role$_i$* indicates the user this *Rule* applies to (e.g., either the phone owner or a secondary user). This field also allows for user grouping through implementing RBAC (explained in Section 5.1.2). In our example, the *Role* is *PrimaryUser*.

6. *Modality$_i$* indicates the scheme by which the current user was authenticated. Its value can either be IA or EA. Using this field, allows performing *Tasks* with EA (when authentication confidence does not apply).

7. *Confidence$_i$* is a threshold for authentication confidence, any value below which will result in the *Rule* being triggered (i.e., the *Action* will be performed).

2) Policy Enforcers (PE, Box B2): This component is in charge of enforcing the *Action* in each rule. Currently, GUS has only one PE–the ActivityLocker–which can prevent Activities from being launched, depending on the state of authentication.

3) Security Manager (SM, B6): This is the policy decision point of GUS, which combines the results of other components, to make a final decision as to what *Tasks* are allowed/denied in the current state of authentication.

4) Context Provider (CP, B5): This component provides a registry which allows IA schemes to provide it their confidence scores for each user. With this design, we treat IA schemes as external entities which can freely register or un-register themselves with the CP, given they are already approved to be secure by the system (e.g., their code is signed by the cryptographic public key of the phone manufacturer). This allows for easy deployment of extra schemes at any point, for research-purposes, for example.

5) Re-Authenticator (RA, B8): This component provides the ReAuth API which, when called, prompts the user for re-authentication (i.e., pops up the UX for it). The separation of RA from other components allows for easy modification of the re-authentication UX, facilitating further research in this area, by eliminating the need for user researchers to deal with low level OS APIs.

5.3.3 Prototype Implementation

We implemented a prototype of GUS on top of Android Open Source Project (AOSP) version 10. In this section, we describe the details of how we did so. Its source code can be accessed via Github[4].

1) Security Policy Database: We implemented this component by extending the Android Framework with a new Search API for our Security Policy Database. To provide the back-end implementation of the API, we developed a new system-level app, called SPD, which registers a service supporting the API with *ServiceManager*—an entity responsible for managing backend services on Android. We had to implement SDP as system-level app as it uses protected system calls which cannot be implemented in normal apps. This approach is identical to how most system services are implemented on Android, e.g., the *PackageManager* service which controls the (un)installation of apps.

We also designed an XML schema for our Policy Definition Language. The *SecurityPolicy* is stored in an XML file that uses this schema (an example file is provided in the source code). When the SPD app first starts, it parses this file and loads it into an in-memory SQLite database, used for fast lookup of *Rules*.

2) Policy Enforcers: To implement ActivityLocker, we have modified Android's *WindowManager*, which is a system service responsible for managing the visibility of Activities. Our modifications call the *Security Manager* every time an Activity is about to be made visible, to make sure showing the Activity under the current authentication state do not violate user's *SecurityPolicy*.

3) Other Components: Security Manager (SM), Context Provider (CP), and Re-Authenticator (RA) were implemented in a similar approach to SPD. In short, for

[4] https://github.com/mehrabik/GUS

each component, we added the corresponding API to Android Framework and then created a new system app that registered a back-end service for the API, with *ServiceManager*.

4) Dummy IA Scheme: To allow accurate evaluation of the usability/security of any future solutions (which might use GUS as a basis for their deployment), we have implemented and included a dummy IA scheme in our code. It allows researchers to specify and change the authentication confidence for each user manually, through a simple UI.

5.4 Evaluation

To estimate the efficacy of GUS, we sought to answer the following research questions:

- **RQ1 [To gain detailed data on users' authorization needs]**: What are the *Tasks* that smartphone users perform on their phones, and what are their *Intentions* for the tasks?

- **RQ2 [To evaluate/compare the efficacy of GUS]**: Given the users' *Tasks* and *Intentions*, how does the IA deployment solutions compare, in terms of conforming to the users' *Intentions*?

5.4.1 Methodology

To address RQ1, we use the tasks and access control needs data from the longitudinal diary study we described in Chapter 4. We treat the access control needs as *Intentions*, where we map "No-one," "Specific People," and "Any-one," labels to Public, Semi-private, and Private *Intention* designations in GUS, respectively.

To estimate the efficacy of different solutions for RQ2, we assume a scenario in which the IA data (e.g., gait data) has become unavailable mid-session. We then count for how many the user's *Tasks*, a solution would show/miss (un)necessary re-authentication prompts. We use the user's *Intentions* to decide which prompts are (un)necessary. For example, since the all-or-nothing solution always shows the re-authentication prompt (no matter what *Task* is being performed), the user would incur unnecessary prompts for any “Anyone” *Tasks*. However, they would not miss any necessary prompts, since the prompts are always shown.

Formally, we calculated two rates per solution:

- ***False Positive Rate (FPR)***, also known Fraud Rate, was calculated as the ratio of the number of *Tasks* the solution would mistakenly make available to unauthorized users (i.e., misses to show a necessary prompt for), by the total number of the participant’s *Tasks*. It would directly impact the security of a solution, as higher FPR would indicate higher chances of unauthorized access.

- ***False Negative Rate (FNR)***, also known as Insult Rate, was calculated as the ratio of the number of *Tasks* that the solution would mistakenly deny access to to authorized users (i.e., would show an unnecessary prompt for) by the total number of the participant’s *Tasks*.

The rates are calculated per participant and then averaged over all participants. Also, we also calculated the rates in two different subscenarios: *CONSERVATIVE* and *MAXIMAL*. Each scenario constituted a different way of aggregating the participants’ *Intentions* for solutions that are less granular than *Task*. We will discuss in more detail what this entails, later in this section.

We should note that there is a major difference between how the rates are calculated in this chapter, compared to how they were in Chapter 4. There, we calculated separate rates for the task-sensitivity and phone sharing scenarios. Here, we consider both scenarios together. When applicable, we assume that the primary user has shared their PIN with the secondary users, so they have access to all tasks on the phone.

To calculate the rates for the solutions that operate based on Android Activities (i.e., Itus and GUS), we needed to map *Tasks* to Activities (that is, to know what Activities are launched when each *Task* is performed). Specifically, we needed to know what *Tasks* cannot be distinguished using Activities. To obtain this, for each instance of a participant p reporting to perform a particular Task t on d-th day of the study, we created a set (denoted $A_p^d(t)$) that included all Activities launched in the app that offers the Task t, on that day. For example, if the i-th participant reported to ”Send email” using the GMail app on the 12th day of the study, we would create a set ($A_1^{12}(sendmail)$) which includes all Activities launched with GMail on the 12th day, on the i-th participant’s phone.

Once such sets were obtained for all tasks, days and participants, we calculated the intersect of all the sets corresponding to each Task t, which we denote as $A(t)$:

$$A(t) = \bigcap_{pinP}^{dinD} A_p^d(t) \tag{5.1}$$

Here, P represents the set of all participants and D represents the set of all study days. The set $A(t)$ represents the set of all Activities that would always be launched if the Task t were performed. Hence, we use $A(t)$ as a representative of the *Task* t.

We took note of any *Tasks* with $A(t) = \emptyset$, as they could not be distinguished using Activities. Additionally, depending on how the app is programmed, there might be cases that for two different *Tasks* $t1$ and $t2$, the Activity sets are the same (i.e., $A(t1) = A(t2)$). In such cases, the *Tasks* could not be distinguished using Activities either. So, we recorded all instances of such *Tasks*, as well.

Armed with this data, we calculated FPR and FNR of each solution in the following way:

- **Itus** treats each *Task* to be at the same sensitivity for all participants. Hence, we needed to aggregate all participants’ *Intentions* to have one *Intention* per *Task*. We also needed to aggregate *Intentions* in-between *Tasks* that are indistinguishable using Activities. In the CONSERVATIVE scenario, we assigned to each *Task* the most conservative of the participants’ *Intentions* for it and all other *Tasks* that have the same $A(t)$. For example, if at least once participant labelled the *Task* ”Sending email” as Private, it would be treated as Private for all participants. In the MAXIMAL scenario, majority voting was used to assign *Intentions* to *Tasks*.

  To compute the rates, any discrepancy between a participant’s individual *Intentions* for a *Task*, and the consensus *Intention* for it would count towards FPR or FNR, depending on whether the consensus was more restrictive than the individual *Intention* or vice versa.

- **ProgAuth** represents the Progressive Authentication solution proposed by Riva et al. [98]. Since it works on an app-level, we needed to aggregate each participants' *Intentions*, to have one *Intention* per app per participant. In the CONSERVATIVE scenario, we took the most conservative of the participant's *Intentions* for all *Tasks* in an app. In the MAXIMAL scenario, we took majority of the *Intentions* for the app's *Tasks*. Any discrepancy between a *Task*'s *Intentions* and that of the app that offers then would count towards FPR and FNR.

- **GUS** allows separate *Intentions* for each *Task*. Therefore, there would be no need for *Intention* aggregation, in general. However, there might be cases where two *Tasks* cannot be distinguished using Activities. In such cases, an aggregation will be performed amongst the *Intentions* of the *Tasks* that have the same $A(t)$, either in a conservative or majority voting manner. Discrepancies between individual *Intentions* for a *Task* and the aggregated one, would then count towards FPR and FNR.

- **PerfectTask** represents a hypothetical system that could perfectly distinguish between the users' *Tasks*. It is included as a lower bound of FP and FN (i.e., a hypothetical "perfect solution"), against which all other solutions will be compared. Note that the FPR and FNR of this solution are not zero because the participants sometimes provide inconsistent *Intentions* for the same *Task* (i.e., when asked multiple times, they label it once as Private and another time as Semi-private). In such cases, an aggregation would be performed, according to the scenario (CONSERVATIVE or MAXIMAL), and discrepancies between the aggregate and individual *Intentions* count towards FPR and FNR. Note that such discrepancies affect all other solutions as well.

- **All-or-nothing** always shows the prompt, and does not take *Tasks* into account. Hence, there would be no missed necessary prompts (i.e., $FPR = 0$). However, any Public or Semi-private *Task* would count towards the FNR.

In summary, our evaluation estimates how often the solutions fail to meet the

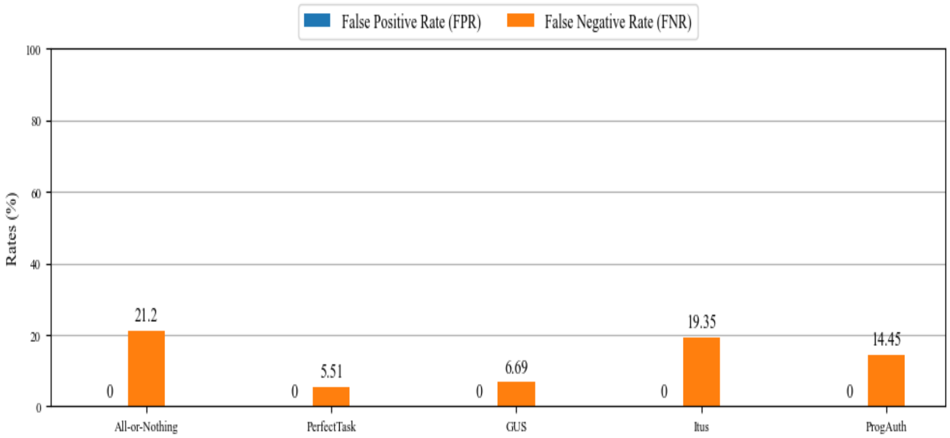


Figure 5.2: Comparison of FPR and FNR of the solutions in the CONSERVATIVE scenario.

users' *Intentions*, by either showing a re-authentication prompt when it is unnecessary (i.e., a False Negative), or not showing one when it should (a False Positive). For Activity-based solutions, failure happen when the solutions cannot distinguish between *Tasks* using Activities. For ProgAuth, failures happen when apps offer *Tasks* with different sensitivity levels. For all-or-nothing, any *Task* non-Private *Task* would be a failure.

5.4.2 Results

RQ1: As discussed in detail in Chapter 4, in total, the participants reported performing 1,149 distinct *Tasks* on their phones, using 571 distinct apps. On average, each participant reported performing 74 distinct *Tasks* (min = 24, max = 142), using 48 apps (min = 12, max = 103). The average number of *Tasks* performed by each app was 2. As for the *Intentions*, the participants indicated to be willing to share a large portion (43%) of their tasks with others. Specifically, 23.7% of the *Tasks* were labelled as Semi-private, whereas 19.3% were labelled as Public. The rest were labelled as Private. Also, each participant had conflicting *Intentions* for 6 of their apps on average, meaning the *Tasks* in these apps had different *Intentions* (e.g., some were Private, while others were Public). Moreover, 4 of the each

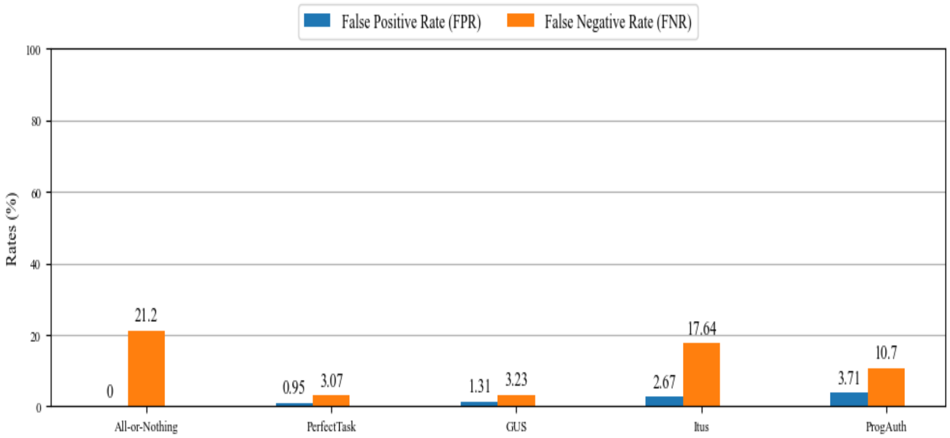


Figure 5.3: Comparison of FPR and FNR of the solutions in the MAXIMAL scenario.

participants' app had all Public *Tasks*, on average.

Additionally, our data showed that Itus's underlying design assumption (that *Tasks* are the same level of sensitivity for all users) does not hold[5]. Amongst the 1,149 *Tasks* that the participants reported performing, 618 (53.8%) were performed by more than one participant. Across these tasks, we observed that in 21% of the times (meaning 21% of the reported instances of the participants performing one of these *Tasks*), there were disagreements on the sensitivity of the *Task*. For example, for the *Task* ”Browsing/reading received emails,” with the GMail app, 76% of the participants who performed it labelled it as Private, whereas 10% labelled it as Semi-private, and the rest labelled it as Public. Hence, our data clearly demonstrated that not all users are willing to share the same *Task* with the same set of people. We will demonstrate in the next section how this negatively affects the efficacy of Itus.

RQ2: Firstly, regarding mapping *Tasks* to Activities, we found Activities to be able to successfully represent the majority of the participants' *Tasks*. In total, 655 (57%) of the reported *Tasks* caused the launch of a unique set of Activities (no overlapping $A(t)$s), allowing Activity-based solutions to differentiate between them unequivo-

[5] Although not explicitly acknowledged by the authors of Itus, this assumption is implicit in the design of the system, due to the hard-coded nature of the modifications made to the apps.

cally. Additionally, 207 (18%) of the *Tasks* caused the launch of a some Activities, but their Activity set ($A(t)$) was shared with another task. Finally, 287 (25%) of the *Tasks* did not cause the launch of any specific Activities and are indistinguishable to GUS or Itus.

As for how the solutions compare, we first discuss the CONSERVATIVE scenario. The FPR and FNR of each solution in this scenario is depicted in Figure 5.2. Naturally, the FPR of all solutions is zero, as the most restrictive *Intentions* are always applied. However, this elimination of unauthorized accesses, comes with a noticeable trade-off–any discrepancy in *Intentions* manifests itself as a rise in FNR, instead.

Unsurprisingly, all-or-nothing provides the worst, as it is the least granular. It has an FNR of 21.2%, which entails an estimated 21% of the re-authentication prompts it invokes are unnecessary (in case the users want to have the lowest chance of unauthorized access, as represented by the CONSERVATIVE scenario). This clearly justifies why Agarwal et al. [2] found their participants to perceive the all-or-nothing implementation of IA as "annoying." It disrupts them frequently.

Moving to its alternatives, Itus seems to not provide any significant improvement, due to its disregard for individual sensitivity preferences. Before, we discussed how there is often disagreement between participants as to how sensitive a *Task* is. Here, we can clearly see how extensively these disagreements affect its efficacy. Evidently, 19% of the prompts it invokes are unnecessary, because even if one single person perceives a *Task* as Private, the app developer would need to lock it behind re-authentication for all users. Our results clearly show how this can cause frequent unnecessary interruptions.

ProgAuth, conversely, seems to actually provide a noticeable improvement. Even though it is less granular than Itus, its user customizability allows it to achieve a nearly 5% lower FNR than Itus, as we see in Figure 5.2. This observation clearly suggests that increased granularity needs to be accompanied with a respect to individual preference, as well. Otherwise, the benefits of more granular control could quickly diminish. This is exactly why GUS furthers the improvements of ProgAuth. It combines the high granularity level of Itus with the user customizability of ProgAuth. As a result, it is able to reduce FNR to 6.7% (less than half of the other two), as the figure shows. Comparing GUS to PerfectTask, we can see that

GUS gets very close to the absolute best case scenario as well, with only a 1% higher FNR.

Moving to the MAXIMAL scenario, which is depicted in Figure 5.3, we can see different trade-offs between FNR and FPR. While the all-or-nothing rates does not change (as there is no *Intention* aggregation), majority voting unanimously leads to lower FNR for other solutions, albeit with small increases in FPR. In case of GUS, for example, FNR is approximately halved (From 6.69% to 3.23%), with only a 1.31% FPR trade-off. This entails that GUS is able to eliminate 97% of the unnecessary re-authentication prompts, while only eliminating 1% of the necessary ones, as well. It is out of scope of this research to evaluate which aggregation model (CONSERVATIVE or MAXIMAL) the end users actually prefer. However, our data clear shows that, if a user accepts a slightly higher risk of unauthorized access (specially for Semi-private tasks, which are not as sensitive as Private ones) MAXIMAL can lead to noticeably fewer re-authentication interruptions.

In spite of the FPR-FNR trade-off changes, however, the order of efficacy between solutions is the same as in the MAXIMAL scenario. As Figure 5.3 shows, All-or-nothing still provides the worse FNR, with Itus, ProgAuth, GUS and PerfectTask each providing improvement over the other. Interestingly, it can be seen that the difference between GUS and PerfectTask is even smaller in the MAXIMAL scenario. There is only a 0.3% difference in FPR and a 0.2% difference in FNR. So, it seems that Activity-locking can indeed provide an apt implementation of the *Task*-based unlocking model.

In summary, and to answer RQ2, our results show that, while all-or-nothing provides the worst FPR in both scenarios, Itus does not provide much improvement either. ProgAuth can provide substantial improvement. However, GUS outperforms it noticeably. Lastly, we found that GUS gets very close to the performance of a hypothetical “perfect” Task-locking system.

5.5 Discussion and Future Work

Nowadays, smartphone users perform many different tasks on their phones, with varying privacy-sensitivity levels. Based on the results of our study, each user performs 74 distinct tasks using 48 different apps, on average (see Section 5.4.2).

While 50% of these tasks are considered private (meaning the users do not want anyone else to be able to preform these tasks on their phones), 20% of the tasks are considered public (meaning the users are willing to have them available even if no authentication is performed). Additionally, 30% of the tasks are considered shareable with some specific people, such as spouses, but not the public.

This complex mixture of sensitivity levels makes providing apt physical security on smartphones challenging. On the one hand, the high proportion of private tasks (e.g., banking) makes user authentication and access control an absolute necessity. On the other hand, the high number of public tasks makes manual unlocking an unnecessary burden for performing the many insensitive tasks.

Implicit authentication (IA) solutions can be used to address this dichotomy. Instead of forcing the users to provide explicit inputs every time they want to perform any task, these solutions use behavioural or contextual data (e.g., biomedical signals [84]) to authenticate the users unobtrusively. This allows them to alleviate the cognitive and physical overhead of explicit authentication. We found that, even in the worst case scenario, any IA solution can eliminate 80% of the unnecessary unlocking attempts (see Figure 5.2).

In reality, however, these unnecessary attempts can be "annoying" [2]. So, several solutions have been proposed to eliminate them. Itus [62] is one of the most prominent such solutions, which operates based on Android Activities. It assumes that tasks can be represented by Activities, which our study showed to be valid. Evidently, we found that 57% of the users' tasks can unequivocally be distinguished using Activities, while a further 25% of them could potentially be distinguished, as well (see RQ2 in Section 5.4.2).

However, Itus also unrealistically assumes that each task is at the same sensitivity level for all users, which our study found to not be the case. We found that in at least 25% of cases, our participants disagreed on whether a task was Public, Private or Semi-private (see RQ1 in Section 5.4.2). Because of this, we found that Itus does not provide much improvement over the all-or-nothing model, and between 17% to 19% of its prompts is still unnecessary (see Figures 5.2 and 5.3).

Progressive Authentication [98] is another proposed solution which operates on an app level. Even though this solution has not been directly proposed for the IA re-authentication solution, it can be easily adopted for this purpose. Our estimation showed that even though it is less granular than Itus, it can eliminate between 5% to 7% more of the unnecessary re-authentication prompts, because it heeds individual users' sensitivity preferences.

Due to the limitations of the existing solutions, we proposed our own approach as well, which we dubbed the Gradual Unlocking System (GUS). GUS gradually lets the user to perform more tasks, as the confidence of authentication increases. Like Itus, we use Android Activities as the mean of distinguishing tasks. Unlike it, however, GUS allows the user to customize the sensitivity levels of tasks. Our study estimated that GUS can improve IA usability significantly, when compared to the other solutions. We found that it can eliminate between 94% to 97% of the unnecessary user interruptions. We also found GUS to get very close to the efficacy of a hypothetical "perfect" task-locking solution (see Section 5.4.2).

One drawback of GUS, however, is in the amount of effort the user needs to put into configuring it. While the all-or-nothing model requires no knowledge of the users' sensitivity preferences for their tasks, GUS requires perfect knowledge of them. We estimated that each user performs 74 tasks on their phones, with nearly 50% of them being Private. Assuming a default Private policy for the tasks, this means that a user need to make 37 configuration adjustments to make their sensitivity preferences clear to GUS, which might be prohibitive. In comparison, the users would need to make their intentions clear for 10 of their apps on average, as these apps have either conflicting or all Public tasks (See RQ1 in Section 5.4.2).

Additionally, the user might also want to specify their preferred authentication confidence threshold for each category of Tasks (e.g., No-one or specific people ones). Designing the UX for such case could also require careful consideration. Simply allowing the users to specify thresholds in terms of percentages might not be very comprehensible to them, which could lead to dangerous security errors (e.g., too low values being selected). Instead, one possible solution could be to use some safe defaults for the thresholds and asks users occasionally whether they are facing too many unnecessary prompts or missing too many necessary ones. Based on the received feedback then, the system can adjust the threshold values accordingly. Further studies, however, are required to evaluate the usability of this approach empirically or explore other alternative UX designs.

Based on the above revelation, it seems that the best approach for deploying

IA might be a a hybrid deployment between GUS and ProgAuth. Users can start with *ProgAuth*, as it requires the least effort. Afterwards, only if they find that an app is suffering from too many unnecessary re-authentications (as it has tasks with conflicting preferences), they can enable GUS for that particular app and assign sensitivity labels to each task in it, individually. Fortunately, GUS is designed in a way that allows for app-level control of access, as well (see Security Policy Database in Section 5.3).

Exploring ways of easing it for users to provide their *Intentions* for GUS can also be an avenue for future research. One possible solution is to include with GUS a default set of *Rules* with default sensitivity labels for common apps. The user can then customize the defaults based on their own needs. The *Rules* provided with our prototype implementation (see Section 5.3.3) can serve as examples of such defaults. They include default *Intentions* for popular apps on Google Play Store. An additional possible solution is to allow developers to include app-specific *Rule* files with their apps. These will provide default *Intentions* for tasks the apps offers.

Lastly, since studies have demonstrated how some of smartphone users' privacy and security decisions can be predictable [121–123], we believe it might be feasible for a trained machine learning classifier to learn and generate user-specific *Rules* automatically. This might be possible to do by only asking only a few questions from the user [122].

In the end, we should also note that another drawback of the more granular access control solutions, such as GUS, is the added complexity of conveying their semantics to users effectively through apt UX design (as we saw the case with IA in Chapter 3). While all-or-nothing has straight-forward operating logic (i.e., the phone is either fully locked or fully unlocked), GUS and other granular solutions have nuances about what apps are accessible to the user at any point in time. It is out of the scope of this work to investigate how such UX can be designed. Further research is needed to address this gap.

5.6 Limitations

In addition to the limitations discussed in Section 5.4, any limitations of the longitudinal user study discussed in Section 4.5 is applicable to our findings in this chapter as well. Also, since our solution requires modifying Android OS, its widespread adoption could be hindered by requiring manufacturer involvement.

5.7 Conclusion

To solve the IA re-authentication problem, and also to propose a practical solution for enabling task-based access control on smartphones, we explored the idea of gradual unlocking. The intuition was that tasks performed on a phone have different levels of sensitivity, and it might not be necessary to prompt the user for re-authentication if they are performing a non-sensitive task. We designed a system that enables such granular re-authentication prompting, by locking Android Activities to authentication confidences. We implemented a prototype of it on top of AOSP 10. We used data from our longitudinal user study to estimate the efficacy of GUS (i.e., the number of (un)necessary re-authentication prompts it shows/misses). Results showed it could eliminate up to 97% of the unnecessary interruptions, while only missing 1% of the necessary ones. We found our solution to out-perform other state-of-the-art solutions, albeit with a trade-off in the required configuration effort. We proposed several ideas for future work to reduce the configuration effort of the system. In the end, we hope our framework can pave the way for more usable IA deployment on smartphones, and improve its chances of adoption by users.

Chapter 6

Discussion and Conclusion

In this research, we aimed to better understand the user experience with smartphone physical security. The goal was to put forward a clearer picture of the state of research and practice, and provide suggestions for improving existing physical security solutions. We discussed the implications of the results of each of our studies in their respective chapters. Here, we focus on the overall takeaway messages.

Takeaway 1) We argue that, even though our results agree with the literature in that the current physical security solutions are sub-optimal, they also demonstrate the inefficacy of the existing solutions to be more nuanced than what the literature depicts. To explain why, let us discuss what the literature posits about the state of practice:

1. For authentication, the literature demonstrates qualitatively that the current knowledge- or biometric-based authentication systems are cumbersome and awkward. For example, they found that current solutions might force users to spend more time unlocking the phone, than actually using it, in short phone sessions [74]. They also found that biometric-based solutions (e.g., Face Unlock) impose negative externalities, such as the awkwardness of holding the phone in front of one's face [29]. As such, there were estimates that nearly 35% of users might completely forgo authentication[33, 51, 52, 82].

2. For access control, the literature shows that the current all-or-nothing access-control model is too simplistic to aptly meet users' needs. It lacks sensitivity

to task security requirements [33, 54], forcing users to unlock phones for even mundane activities, such as checking the weather. The system also does not enable users to assert apt control when sharing phones, despite the high prevalence of the practice [50, 54, 60, 77], which could lead to unauthorized access [75, 76].

As we can see, existing studies are mostly qualitative in both domains (authentication and access control). This prevents one from drawing any conclusion about the nuances that pertain to the prevalence or the severity of the issues. For example, while the literature provides evidence that all-or-nothing lacks the authorization sensitivity that "some" users require [54], it provides no statistics about what percentage of users' tasks might actually be sensitive and require authentication. This makes it difficult to judge whether this lack of sensitivity is truly a wide-spread problem, or rather it is relegated to corner-cases faced by only the most peculiar of users.

We, on the other hand, provide such statistics. Put simply, we re-evaluate the literature's findings in more detail:

1. For authentication, we found that, the incumbent solutions indeed fail users. However, in contrast to the literature, we found that the percentage of users adopting authentication is actually higher than previous estimates: We found nearly 90% of our participants to have adopted authentication, compared to much lower estimates of 65% before [33, 51, 52, 82]. Comparing our results to the literature (discussed in Chapter 1), there seems to be an overall trend that with the advent of better biometric unlocking solutions, such as FaceID and TouchID, more and more users are willing to adopt authentication. Hence, the adoption issue might not be as dire as it once seemed.

   Also, as discussed in Chapter 3, we identified, for the first time, specific aspects of IA that could cause users to misunderstand its semantics, leading to incorrect security perceptions, dangerous security errors, and lack of adoption. Such aspects ranged from simple information, such as range parameters (e.g., how long the range of a Bluetooth connection is), to more complex concepts, such as the logical relationship between the outcome of IA and explicit authentication systems (i.e., if there is a discrepancy between the authentication systems' decisions about the authenticity of the user, how would the phone behave?). We also found that the struggle with comprehending IA semantics is rather universal, as no group of users seems to have a significantly more accurate understanding.

2. For access control, our results agreed with the literature that all-or-nothing lacks sufficient task sensitivity and phone sharing. However, we measured the extend of this issue quantitatively and found some nuance. For example, we found that while 57% of the users tasks do indeed require authentication all the time, 24% are more nuanced and should be open to sharing, and 19% do not require any authentication. Overall, we found that for a considerable number of users, all-or-nothing might work just fine, and even those with more complex needs might be able to find a nice balance between usability and security by switching to an app-level model, relegating task-level access control to be more appropriate for only a minority of users.

Takeaway 2) We argue that the proposed alternative solutions for physical security are not as effective as the literature paints them to be. The literature advocates for implicit authentication (IA) [63] and more granular access control [72, 98]. However, to the best of our knowledge, there has never been an independent evaluation of these alternatives, to objectively investigate their merits. We set out to perform such evaluation, and we found our results to sometimes agree with and sometimes contradict the literature:

1. Regarding authentication, when we studied how users perceive and understand IA, we found many adoption hindrances that prior work failed to notice. Actually, contrary to the literature [26, 62] (as discussed in detail in Chapter 2), we observed a very low adoption rate for IA. That is, we observed how a widely deployed commercial solution (Smart Lock for Android) has failed to conjure any meaningful following (less than 14% adoption rate), after being available on more than 500 million phones for over five years [41]. We further explain this low adoption rate by proposing a new extension (SL-TAM) to the technology acceptance model. The model demonstrated strong correlations between the users' perception of usefulness, ease of use, and

security and privacy of SL, and their intention to adopt it.

We, therefore, argue that the the adoption of IA might not be for everybody, and that IA should not be viewed as a one-size-fits-all solution. This is contrary to two general beliefs in the literature [27, 62]: (1) the added value of IA will be immediately obvious to users, and (2) the reliability of current solutions are good enough for wide deployment (as done by Smart Lock). In place of these beliefs, we contribute to the community SL-TAM, a model that allows them to gauge what could potentially make an IA scheme appealing to a certain group of users, in a certain circumstance.

2. Regarding access control, we found that going more and more granular with access control might not be the best of ideas. We found that the app-level access-control systems (that some researchers have proposed before) could actually be a sweet-spot of sorts, and be the best way forward (at least conceptually). These app-level solutions could potentially reduce the percentage of the users' tasks that are exposed to unauthorized access to 3.5%, while also reducing the unnecessary explicit re-authentications by 11.3%. In contrast, w found that the task-level models could further reduce unnecessary authentications to 1.7%, but with a noticeably higher 15% increase in the required amount of configuration (which the users need to handle). This makes us advocate, overall, for wider deployment of app-level access control on smartphones, rather than task-level models or all-or-nothing ones.

   This increase in configuration effort also provided evidence why users might be unwilling to switch to such granular systems. Whereas the all-or-nothing system requires a one-time unlocking set-up, the app- and task-level solutions require detailed and lengthy configurations.

Takeaway 3) We argue that one cannot try to "fix" the authentication and access-control systems individually. They must be considered together. As discussed in Chapter 5, if one tries to switch the authentication system to IA while still preserving the all-or-nothing access-control model, it might lead to excessive user interruptions due to intermittent availability of data. In fact, it was this revelation, in addition to a lack of practical solutions for performing task-based access control (which was discussed in Chapter 4), that made us decide to propose a new access control solution in this dissertation. Dubbed the "Gradual Unlocking System" (GUS), the solution tries to facilitate IA deployment by tying the user's ability to perform tasks to the confidence of authentication. As discussed in Chapter 5, our evaluation results showed that GUS can reduce unnecessary re-authentications significantly (by 96%) with a modest security trade-off (a 1% increase in missed necessary re-authentications). However, we should note that our solution is not designed (or positioned to be) the be-all and end-all solution. It only focuses on solving the granularity issue in access control. There are many other issues with the incumbent all-or-nothing system (e.g., lack of phone sharing support) that require new ideas and further research.

Takeaway 4) We argue that access control on smartphones is quite different than on other systems, thus, it merits its own research. For example, one could have argued that instead of proposing a new access control solution, why not adopt a solution designed for other domains (e.g., personal computers), which might have years of research spent on them. We argue that it is not so straight-forward to do so. Our studies (discussed in Chapters 4 and 5), as well as existing literature [77] show that sharing habits are different per device type (e.g., even amongst tables and phones). This makes access control design requirements different per different device type. For example, the username and password based system of authentication and access control on computers has already been shown to be too cumbersome to use on smartphones, as sharing on smartphones does not happen for the same purposes as the computers (we discussed this in Chapter 4. As such, a system designed for a stationary computer might not be suitable for phones.

Takeaway 5) Based on our results, we propose the following avenues for future research or deployment in physical security on smartphones:

1. We recommend IA solution designers to pay attention to our recommendations in Chapter 3, and improve their UX to better communicate the utility and semantics of their solution to users. Otherwise they might be doing users a disservice by making their phones less secure (as discussed in Chapter 2).

2. For researchers, we provide, in Chapters 2 and 3, several ideas on how to

improve IA in both technical and non-technical ways. On the technical side, future researchers can focus on proposing solutions that would improve the speed of authentication in IA solutions, which would improve the *perceived usefulness* of the technology for users, as argued by SL-TAM. They can also focus on evaluating alternative designs for IA UX, to better convey IA utility and semantics to users. Based on SL-TAM, this could lead to improvements in perceived *security and privacy*.

3. For access control, we provide ideas, in Chapters 4 and 5, that future research can explore to automate the configuration process (e.g., through machine-learning-based approaches). Overall, we recommend that future researchers focus on improving the UX for access control, as well as facilitate better developer integration. For example, future researchers can focus on proposing and evaluating alternative designs for the UX of app- and task-based access control, in order to maximize user satisfaction. They can also work on integrating task-definition APIs[1] as well as default sensitivity levels for each Task. Doing so could reduce the configuration burden on users, by providing them with safe defaults (which they could customize if needed).

To summarize, we believe our studies clearly show the broken state of smartphone physical security. They show that the incumbent systems are insufficient and the alternatives are not yet mature enough to take over (We should acknowledge, however, that there could be a myriad of other factors—social, economic, or otherwise—that have lead to the current stagnancy of the incumbent system, which are beyond the scope of this work.) To improve this state, based on our findings we advocate for a future when authentication is done implicitly for those who need it, while its utility and semantics are clearly communicated to the user through apt UX design. This authentication would also be done in conjunction with task-level access control and an app-level default policy. That is, the phone would be able to lock tasks individually, satisfying even the most complex access-control needs; however, to reduce the required configuration effort it would ask for app-level access-control preferences from the user by default.

[1] That is, an API that allows app developers to specify what tasks their app affords.

In the end, we should also call attention to the many limitations of our studies and ask for carefulness when generalizing the findings beyond our original studies. Our participant samples are not fully representative of the global smartphone user population, especially regarding culture and ethnicity. Hence one should not assume the same trend of behaviour we observed with regards to IA and granular access control to apply anywhere in the world. Our data are also mostly self-reported, which make them lack 100% reliability all the time. As such, the actual need for IA and access control might not be fully reflected in our results. We should also note that we only focused on one implementation of IA, which is SL. We did so because it was the only widely deployed IA solution on smartphones at the time of the study. In the future, comparative studies should be conducted to investigate users' perceptions of other possible implementations and provide a more clear picture of their needs and preferences.

Bibliography

[1] D. Abrokwa, S. Das, O. Akgul, and M. L. Mazurek. Comparing security and privacy attitudes among US users of different smartphone and smart-speaker platforms. In *the 2021 Symposium on Usable Privacy and Security (SOUPS)*, 2021. → page 105

[2] L. Agarwal, H. Khan, and U. Hengartner. Ask me again but don't annoy me: Evaluating re-authentication strategies for smartphones. In *the 2016 Symposium on Usable Privacy and Security (SOUPS)*, 2016. → pages 8, 110, 125, 127

[3] Android API. SELinux on Android. https://source.android.com/security/selinux, 2020. Accessed: 2020-06-11. → page 109

[4] Android Developers Portal. Developer guides. https://developer.android.com/guide/, 2020. Accessed: 2020-04-17. → pages 109, 115

[5] Android Developers Portal. UsageStats manager. https://developer.android.com/reference/android/app/usage/UsageStatsManager, 2021. Accessed: 2021-08-06. → page 80

[6] Apple. iOS security guide, 2018. → page 2

[7] Apple. User guided access with iPhone. https://support.apple.com/en-us/HT202612, 2021. Accessed: 2021-08-13. → pages 3, 75, 77

[8] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of computer security risks. In *the 2007 Workshop on the Economics of Information Security*, 2007. → pages 46, 66

[9] C. H. Au, K. C. Lam, W. S. Fung, and X. Xu. Using animation to develop a MOOC on information security. In *the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2016. → page 69

[10] E. Bacis, S. Mutti, and S. Paraboschi. AppPolicyModules: Mandatory access control for third-party apps. In *the 10th ACM Symposium on Information, Computer and Communications Security*, 2015. → page 77

[11] M. Backes, S. Bugiel, S. Gerling, and P. von Styp-Rekowsky. Android security framework: Extensible multi-layered access control on Android. In *the 2014 Annual Computer Security Applications Conference (ACSAC)*, 2014. → page 77

[12] R. Baecker, I. Small, and R. Mander. Bringing icons to life. In *Readings in Human–Computer Interaction*. Elsevier, 1995. → page 69

[13] B. L. Berg, H. Lune, and H. Lune. *Qualitative research methods for the social sciences*. Pearson Boston, MA, 2012. → page 84

[14] R. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides. Biometric authentication on iPhone and Android: Usability, perceptions, and influences on adoption. In *the 2015 Workshop on Usable Security (USEC)*, 2015. → page 1

[15] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 2006. → pages 23, 50

[16] S. C. Brown and F. I. Craik. Encoding and retrieval of information. *Oxford handbook of memory*, 2000. → page 79

[17] A. B. Brush and K. M. Inkpen. Yours, mine and ours? sharing and use of technology in domestic environments. In *the 2007 International Conference on Ubiquitous Computing*, 2007. → pages 77, 98, 100, 103

[18] S. Bugiel, S. Heuser, and A.-R. Sadeghi. Flexible and fine-grained mandatory access control on Android for diverse security and privacy policies. In *the 2013 USENIX Security Symposium*, 2013. → pages 77, 115

[19] D. L. Butler and M. Sellbom. Barriers to adopting technology. *Educause Quarterly*, 2002. → page 33

[20] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *the 2006 USENIX Security Symposium*, 2006. → page 46

[21] G. Cho, J. H. Huh, S. Kim, J. Cho, H. Park, Y. Lee, K. Beznosov, and H. Kim. On the security and usability implications of providing multiple authentication choices on smartphones: The more, the better? *Transactions on Privacy and Security (TOPS)*, 2020. → page 71

[22] M. Y. Chuttur. Overview of the technology acceptance model: Origins, developments and future directions. *Working Papers on Information Systems*, 2009. → pages 20, 37

[23] M. Conti, V. T. N. Nguyen, and B. Crispo. CRePE: Context-related policy enforcement for Android. In *the 2010 International Conference on Information Security*, 2010. → pages 75, 77

[24] D. D. Coppersmith, E. M. Kleiman, C. R. Glenn, A. J. Millner, and M. K. Nock. The dynamics of social support among suicide attempters: A smartphone-based daily diary study. *Behaviour research and therapy*, 2019. → page 84

[25] L. F. Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal on Telecommunication & High Technology Learning*, 2012. → page 70

[26] H. Crawford and K. Renaud. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, 2014. → pages 2, 5, 13, 17, 29, 66, 110, 133

[27] H. Crawford, K. Renaud, and T. Storer. A framework for continuous, transparent mobile device authentication. *Computers & Security*, 2013. → pages 3, 17, 66, 109, 111, 112, 134

[28] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw. User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 1989. → pages 9, 20, 36, 37

[29] A. De Luca, A. Hang, E. Von Zezschwitz, and H. Hussmann. “I feel like I’m taking selfies all day!”: Towards understanding biometric authentication on smartphones. In *the 2015 CHI Conference on Human Factors in Computing Systems*, 2015. → pages 2, 30, 71, 105, 131

[30] M. O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *the 2010 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2010. → page 17

[31] N. Ebert, K. A. Ackermann, and P. Heinrich. Does context in privacy communication really matter? a survey on consumer concerns and preferences. In *the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. → pages 70, 71

[32] S. Egelman, A. B. Brush, and K. M. Inkpen. Family accounts: A new paradigm for user accounts within the home environment. In *the 2008 ACM conference on Computer Supported Cooperative Work (CSCW)*, 2008. → pages 77, 98, 103

[33] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? understanding user motivations for smartphone locking behaviors. In *the 2014 ACM Conference on Computer and Communications Security (CCS)*, 2014. → pages 1, 2, 27, 28, 131, 132

[34] N. Elenkov. *Android security internals: an in-depth guide to Android's security architecture*. No Starch Press, 2015. → page 2

[35] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner. Android permissions: User attention, comprehension, and behavior. In *the 2012 Symposium on Usable Privacy and Security (SOUPS)*, 2012. → page 46

[36] N. Firth. Touchscreen phones know it's you from taps and swipes. http://www.newscientist.com/article/dn24193-touchscreen-phones-know-its-you-from-taps-and-swipes.html, 2019. Accessed: 2019-01-26. → page 3

[37] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *Transactions on Information Forensics and Security*, 2013. → pages 3, 5, 17

[38] L. Fridman, S. Weber, R. Greenstadt, and M. Kam. Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location. *Systems Journal*, 2016. → page 17

[39] R. M. Gonyea. Self-reported data in institutional research: Review and recommendations. *New directions for institutional research*, 2005. → pages 72, 105

[40] Google. Google I/O 2014 keynote. https://www.youtube.com/watch?time_continue=1659&v=biSpvXBGpE0, 2019. Accessed: 2019-02-14. → pages 4, 5, 18

[41] Google. Use Google Smart Lock. https://support.google.com/accounts/answer/6160273?hl=en, 2019. Accessed: 2019-01-26. → pages 3, 48, 104, 133

[42] Google. Delete, switch or add users on nexus phones. https://support.google.com/nexus/answer/2865483?hl=en, 2021. Accessed: 2021-08-30. → pages 75, 77

[43] Google. Pin & unpin screens on Android. https://support.google.com/android/answer/9455138?hl=en, 2021. Accessed: 2021-08-13. → pages 3, 77

[44] Google. Google camera help. https://support.google.com/googlecamera/answer/6164997?hl=en, 2021. Accessed: 2021-08-30. → pages 75, 76

[45] Google Play Console Help. Choose a category and tags for your app or game. https://support.google.com/googleplay/android-developer/answer/9859673?hl=en, 2021. Accessed: 2021-08-24. → page 86

[46] T. Granollers and J. Lorés. Cognitive walkthrough with users: an alternative dimension for usability methods. In *the 2005 HCI International Conference*, 2005. → pages 9, 21, 47

[47] T. R. Green, M. M. Burnett, A. J. Ko, K. J. Rothermel, C. R. Cook, and J. Schonfeld. Using the cognitive walkthrough to improve the design of a visual programming experiment. In *the 2000 IEEE International Symposium on Visual Languages*, 2000. → page 50

[48] G. Guest, K. M. MacQueen, and E. E. Namey. *Applied thematic analysis*. sage, 2011. → pages 23, 50

[49] M. Hagen. User-centered privacy communication design. In *the 2016 Symposium on Usable Privacy and Security (SOUPS)*, 2016. → page 70

[50] A. Hang, E. Von Zezschwitz, A. De Luca, and H. Hussmann. Too much information! user attitudes towards smartphone sharing. In *the 2012 Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, 2012. → pages 2, 3, 76, 77, 78, 132

[51] M. Harbach, E. Von Zezschwitz, A. Fichtner, A. De Luca, and M. Smith. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *the 2014 Symposium on Usable Privacy and Security (SOUPS)*, 2014. → pages 2, 24, 71, 105, 131, 132

[52] M. Harbach, A. De Luca, and S. Egelman. The anatomy of smartphone unlocking: A field study of Android lock screens. In *the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. → pages 2, 131, 132

[53] M. Harbach, A. De Luca, N. Malkin, and S. Egelman. Keep on lockin' in the free world: a multi-national comparison of smartphone locking. In *the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. → pages 59, 71, 105

[54] E. Hayashi, O. Riva, K. Strauss, A. Brush, and S. Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *the 2012 Symposium on Usable Privacy and Security (SOUPS)*, 2012. → pages 1, 2, 76, 77, 78, 79, 94, 95, 99, 105, 132

[55] V. C. Hu, D. F. Ferraiolo, and D. R. Kuhn. *Assessment of access control systems*. National Institute of Standards and Technology, 2006. → page 1

[56] J. H. Huh, H. Kim, R. B. Bobba, M. N. Bashir, and K. Beznosov. On the memorability of system-generated pins: Can chunking help? In *the 2015 Symposium on Usable Privacy and Security (SOUPS)*, 2015. → page 2

[57] J. H. Huh, S. Verma, S. S. V. Rayala, R. B. Bobba, K. Beznosov, and H. Kim. "i don't use apple pay because it's less secure": perception of security and usability in mobile tap-and-pay. In *the 2017 Workshop on Usable Security (USEC)*, 2017. → pages 30, 33, 71, 105

[58] M. Jacobs, H. Cramer, and L. Barkhuus. Caring about sharing: Couples' practices in single user device access. In *the 2016 International Conference on Supporting Group Work*, 2016. → pages 2, 76, 78, 99

[59] S. Karatzouni, S. M. Furnell, N. L. Clarke, and R. A. Botha. Perceptions of user authentication on mobile devices. In *the 2007 The ISOneWorld Conference*, 2007. → page 32

[60] A. K. Karlson, A. B. Brush, and S. Schechter. Can I borrow your phone? understanding concerns when sharing mobile phones. In *the 2009 CHI Conference on Human Factors in Computing Systems*, 2009. → pages 2, 3, 76, 78, 92, 100, 132

[61] A. K. Karlson, S. T. Iqbal, B. Meyers, G. Ramos, K. Lee, and J. C. Tang. Mobile taskflow in context: a screenshot study of smartphone usage. In *the 2010 CHI Conference on Human Factors in Computing Systems*, 2010. → page 84

[62] H. Khan, A. Atwater, and U. Hengartner. Itus: an implicit authentication framework for Android. In *the 2014 Annual International Conference on Mobile Computing and Networking*, 2014. → pages 2, 3, 111, 127, 133, 134

[63] H. Khan, A. Atwater, and U. Hengartner. A comparative evaluation of implicit authentication schemes. In *the 2014 International Workshop on Recent Advances in Intrusion Detection*, 2014. → pages 3, 5, 13, 17, 66, 110, 133

[64] H. Khan, U. Hengartner, and D. Vogel. Usability and security perceptions of implicit authentication: Convenient, secure, sometimes annoying. In *the 2015 Symposium on Usable Privacy and Security (SOUPS)*, 2015. → pages 5, 17, 29, 31, 66, 67, 110

[65] S. Kieffer, U. B. Sangiorgi, and J. Vanderdonckt. Ecoval: A framework for increasing the ecological validity in usability testing. In *the 2015 Hawaii International Conference on System Sciences*, 2015. → page 18

[66] A. Kobsa and M. Teltzrow. Contextualized communication of privacy practices and personalization benefits: Impacts on users' data sharing and purchase behavior. In *the 2004 International Workshop on Privacy Enhancing Technologies*, 2004. → page 70

[67] Kryptowire. Kryptowire continuous authentication. https://www.kryptowire.com/continuous-authentication/, 2019. Accessed: 2019-02-20. → page 18

[68] L. Li, X. Zhao, and G. Xue. Unobservable re-authentication for smartphones. In *the 2013 Network and Distributed System Security (NDSS) Symposium*, 2013. → page 5

[69] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *the 2012 ACM Conference on Ubiquitous Computing*, 2012. → pages 46, 66

[70] W. Lira, R. Ferreira, C. de Souza, and S. Carvalho. Experimenting on the cognitive walkthrough with users. In *the 2014 International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI)*, 2014. → pages 9, 21, 47, 50

[71] D. Liu, F. Asgharpour, and L. J. Camp. Risk communication in security using mental models. *Usable Security*, 2008. → page 46

[72] Y. Liu, A. Rahmati, Y. Huang, H. Jang, L. Zhong, Y. Zhang, and S. Zhang. xShare: Supporting impromptu sharing of mobile phones. In *the 2009 International Conference on Mobile Systems, Applications, and Services*, 2009. → pages 3, 75, 77, 90, 104, 133

[73] T. Mahatody, M. Sagar, and C. Kolski. State of the art on the cognitive walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction*, 2010. → pages 9, 21, 47, 48, 50

[74] A. Mahfouz, I. Muslukhov, and K. Beznosov. Android users in the wild: Their authentication and usage behavior. *Pervasive and Mobile Computing*, 2016. → pages 1, 2, 17, 28, 131

[75] D. Marques, I. Muslukhov, T. Guerreiro, L. Carriço, and K. Beznosov. Snooping on mobile phones: Prevalence and trends. In *the 2016 Symposium on Usable Privacy and Security (SOUPS)*, 2016. → pages 1, 3, 31, 59, 65, 93, 97, 132

[76] D. Marques, T. Guerreiro, L. Carriço, I. Beschastnikh, and K. Beznosov. Vulnerability & blame: Making sense of unauthorized access to smartphones. In *the 2019 CHI Conference on Human Factors in Computing Systems*, 2019. → pages 1, 31, 97, 132

[77] T. Matthews, K. Liao, A. Turner, M. Berkovich, R. Reeder, and S. Consolvo. ”She'll just grab any device that's closer” a study of everyday device & account sharing in households. In *the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. → pages 2, 3, 75, 76, 78, 89, 92, 99, 100, 105, 132, 135

[78] M. L. Mazurek, J. Arsenault, J. Bresee, N. Gupta, I. Ion, C. Johns, D. Lee, Y. Liang, J. Olsen, B. Salmon, et al. Access control for home data sharing: Attitudes, needs and practices. In *the 2010 CHI Conference on Human Factors in Computing Systems*, 2010. → pages 1, 74, 78, 92, 99, 100

[79] L. Mecke, S. D. Rodriguez, D. Buschek, S. Prange, and F. Alt. Communicating device confidence level and upcoming re-authentications in continuous authentication systems on mobile devices. In *the 2019 Symposium on Usable Privacy and Security (SOUPS)*, 2019. → page 110

[80] M. Mehrabi Koushki, B. Obada-Obieh, J. H. Huh, and K. Beznosov. Is implicit authentication on smartphones really popular? on Android users' perception of “smart lock for android”. In *the 2020 International*

Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI), 2020. → pages 66, 88, 95, 103, 105

[81] W. Melicher, D. Kurilova, S. M. Segreti, P. Kalvani, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek. Usability and security of text passwords on mobile devices. In *the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. → page 2

[82] N. Micallef, M. Just, L. Baillie, M. Halvey, and H. G. Kayacik. Why aren't users using protection? investigating the usability of smartphone locking. In *the 2015 International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*, 2015. → pages 2, 131, 132

[83] M. Miettinen, S. Heuser, W. Kronz, A.-R. Sadeghi, and N. Asokan. ConXSense: automated context classification for context-aware access control. In *the 2014 ACM Symposium on Information, Computer and Communications Security*, 2014. → pages 75, 77

[84] A. Mosenia, S. Sur-Kolay, A. Raghunathan, and N. K. Jha. CABA: Continuous authentication based on bioaura. *Transactions on Computers*, 2017. → pages 17, 104, 127

[85] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Know your enemy: the risk of unauthorized access in smartphones by insiders. In *the 2013 International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI)*, 2013. → page 3

[86] X. Ni, Z. Yang, X. Bai, A. C. Champion, and D. Xuan. DiffUser: Differentiated user access control on smartphones. In *the 2009 IEEE International Conference on Mobile Adhoc and Sensor Systems*, 2009. → pages 75, 77, 90

[87] D. Norman. *The design of everyday things: Revised and expanded edition*. Constellation, 2013. → pages 79, 113

[88] B. Obada-Obieh, Y. Huang, and K. Beznosov. Challenges and threats of mass telecommuting: A qualitative study of workers. In *the 2021 Symposium on Usable Privacy and Security (SOUPS)*, 2021. → pages 99, 106

[89] S. Osborn, R. Sandhu, and Q. Munawer. Configuring role-based access control to enforce mandatory and discretionary access control policies. *Transactions on Information and System Security (TISSEC)*, 2000. → page 109

[90] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 2017. → pages 59, 71, 105

[91] Pew Research Center. Mobile technology and home broadband 2019. https://www.pewinternet.org/2019/06/13/mobile-technology-and-home-broadband-2019/, 2019. Accessed: 2019-07-26. → pages 27, 62, 84

[92] D. R. Pilar, A. Jaeger, C. F. Gomes, and L. M. Stein. Passwords usage and human memory limitations: A survey across age and educational background. *PloS one*, 2012. → page 79

[93] S. Prabhakar, S. Pankanti, and A. K. Jain. Biometric recognition: Security and privacy concerns. *Security & Privacy*, 2003. → page 32

[94] A. Primo, V. V. Phoha, R. Kumar, and A. Serwadda. Context-aware active authentication using smartphone accelerometer measurements. In *the 2004 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014. → page 17

[95] L. Qiu, A. De Luca, I. Muslukhov, and K. Beznosov. Towards understanding the link between age and smartphone authentication. In *the 2019 CHI Conference on Human Factors in Computer Systems*, 2019. → pages 1, 2, 27, 28, 65, 92

[96] F. Raja, K. Hawkey, and K. Beznosov. Revealing hidden context: improving mental models of personal firewall users. In *the 2009 Symposium on Usable Privacy and Security (SOUPS)*, 2009. → page 46

[97] E. M. Redmiles, S. Kross, and M. L. Mazurek. How well do my results generalize? comparing security and privacy survey results from MTurk, web, and telephone samples. In *the 2019 IEEE Symposium on Security and Privacy (SP)*, 2019. → pages 59, 71, 105

[98] O. Riva, C. Qin, K. Strauss, and D. Lymberopoulos. Progressive Authentication: Deciding when to authenticate on mobile phones. In *the 2012 USENIX Security Symposium*, 2012. → pages 2, 75, 77, 88, 111, 122, 127, 133

[99] E. M. Rogers. *Diffusion of innovations*. Simon and Schuster, 2010. → page 20

[100] N. Sambasivan, G. Checkley, A. Batool, N. Ahmed, D. Nemer, L. S. Gaytán-Lugo, T. Matthews, S. Consolvo, and E. Churchill. "Privacy is not for me, it's for those rich women": Performative privacy practices on mobile phones by women in south asia. In *the 2018 Symposium on Usable Privacy and Security (SOUPS)*, 2018. → page 105

[101] Samsung. What is the secure folder and how do i use it? https://www.samsung.com/uk/support/mobile-devices/what-is-the-secure-folder-and-how-do-i-use-it/, 2021. Accessed: 2021-09-14. → pages 75, 103

[102] R. S. Sandhu and P. Samarati. Access control: principle and practice. *Communications magazine*, 1994. → page 109

[103] S. Schröder, M. Huber, D. Wind, and C. Rottermanner. When SIGNAL hits the fan: On the usability and security of state-of-the-art secure mobile messaging. In *the 2016 European Workshop on Usable Security*, 2016. → page 46

[104] J. Seifert, A. De Luca, B. Conradi, and H. Hussmann. TreasurePhone: Context-sensitive user data protection on mobile phones. In *the 2010 International Conference on Pervasive Computing*, 2010. → pages 75, 77

[105] SELinux. SELinux Project. https://github.com/SELinuxProject, 2020. Accessed: 2020-04-20. → page 115

[106] Signal. Signal app. https://signal.org/en/, 2020. Accessed: 2020-06-12. → page 46

[107] S. Smalley, P. Loscocco, M. Hibler, D. Andersen, J. Lepreau, R. S. Stephen, R. Spencer, et al. The Flask security architecture: System support for diverse security policies. In *the 1998 USENIX Security Symposium*, 1998. → page 115

[108] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. A diary study of mobile information needs. In *the 2008 CHI Conference on Human Factors in Computing Systems*, 2008. → page 84

[109] A. Sotirakopoulos, K. Hawkey, and K. Beznosov. On the challenges in usable security lab studies: Lessons learned from replicating a study on SSL warnings. In *the 2011 Symposium on Usable Privacy and Security (SOUPS)*, 2011. → page 18

[110] Statista. Global smartphone penetration rate as share of population from 2016 to 2020. https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/, 2021. Accessed: 2021-09-01. → page 99

[111] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov. What makes users refuse web single sign-on?: an empirical investigation of OpenID. In *the 2011 Symposium on Usable Privacy and Security (SOUPS)*, 2011. → pages 20, 38

[112] N. K. Thanigaivelan, E. Nigussie, A. Hakkala, S. Virtanen, and J. Isoaho. CoDRA: Context-based dynamically reconfigurable access control system for Android. *Journal of Network and Computer Applications*, 2018. → page 77

[113] J. Tiongson. Mobile app marketing insights: How consumers really find and use your apps, 2015. → page 21

[114] UnifyID. Unifyid implicit authentication platform. https://unify.id/product/, 2019. Accessed: 2019-02-20. → page 18

[115] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 2000. → page 20

[116] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, 2003. → page 20

[117] L. R. Vijayasarathy. Predicting consumer intentions to use on-line shopping: the case for an augmented technology acceptance model. *Information & management*, 2004. → pages 20, 38

[118] S. E. Wade. Using think alouds to assess comprehension. *Reading Teacher*, 1990. → pages 48, 49

[119] X. Wang and Z. Cheng. Cross-sectional studies: strengths, weaknesses, and recommendations. *Chest*, 2020. → page 79

[120] X. Wang, K. Sun, Y. Wang, and J. Jing. Deepdroid: Dynamically enforcing enterprise policy on Android devices. In *the 2015 Network and Distributed System Security (NDSS) Symposium*, 2015. → page 77

[121] P. Wijesekera, A. Baokar, A. Hosseini, S. Egelman, D. Wagner, and K. Beznosov. Android permissions remystified: A field study on contextual integrity. In *the 2015 USENIX Security Symposium*, 2015. → page 129

[122] P. Wijesekera, A. Baokar, L. Tsai, J. Reardon, S. Egelman, D. Wagner, and K. Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *the 2017 IEEE Symposium on Security and Privacy (SP)*, 2017. → page 129

[123] P. Wijesekera, J. Reardon, I. Reyes, L. Tsai, J.-W. Chen, N. Good, D. Wagner, K. Beznosov, and S. Egelman. Contextualizing privacy decisions for better prediction (and protection). In *the 2018 CHI Conference on Human Factors in Computing Systems*, 2018. → page 129

[124] C. Wilson. *User interface inspection methods: a user-centered design method*. Newnes, 2013. → pages 21, 47, 48

[125] K. A. Young. Direct from the source: the value of think-aloud data in understanding learning. *Journal of Educational Enquiry*, 2005. → pages 48, 49

[126] W. Zhang and P. Xu. Do I have to learn something new? mental models and the acceptance of replacement technologies. *Behaviour & Information Technology*, 2011. → pages 33, 35

Appendix A

Supporting Materials

A.1 Cognitive Walkthrough With Users Material

A.1.1 Persona, Task and Action Sequence Definitions

- **User Profile**: It is important to define the profile correctly because any unreasonable assumption about the abilities of the user can negatively influence the CW findings to be unrealistic, and not representative of the interaction between a user and the UI in reality. In case of SL, based on Google's description of the technology, we assumed that the UI caters for users that have some experience with a smartphone but no particular knowledge or training in SL authentication, or any other aspect of computer security or privacy. Therefore, throughout the course of the CW sessions, we asked our HCI-proficient participants to put themselves in the shoes of a computer-science-illiterate regular smartphone user.

- **Task List**: We chose those tasks that are more likely to be performed by a first-time SL user. The tasks included enabling/disabling on-body detection, adding/removing a new trusted place, adding/removing a trusted device, adding/removing a new trusted face, setting up/removing a voice model for Voice Match.

- **Action Sequences**: To the best of our ability, we designed the action sequences to be as simple as possible and to resemble real usage. When possible, we also used steps outlined in the SL help page to design the action sequences.

- **CW Problem Reporting Form**: We designed a form that allowed each participant to answer typical CW Yes/No questions, in addition to providing space for them to write down their comments about each step in the action sequence and each task in general. The handouts also include all the tasks we defined and (in case of CW) the action sequence for each task.

- **UI Representation**: For participants in the group sessions, we projected in real-time the UI of an Android phone (Google Nexus 6P) on a large TV screen using a Google Chromecast. This was done so that all the participants can see on the screen the task being carried out and the corresponding UI. For think-aloud sessions, we handed the same Android smartphone to participants.

A.1.2 Cognitive Walkthrough Handout

Consent: Please read the consent form carefully and sign it before starting with the study. Feel free to ask any questions you might have.

Task scenario: You just heard about Android's Smart lock feature, you want to explore it and set it up for use on your mobile device.

Instructions: For each Action Sequence below:

- Look at the UI on the TV and pretend to do the action and ask yourselves Q1-Q4; write down Yes/No

- If answer is No for any question:
  - Write down the problem (Possible solutions if you have ideas)
  - Then assume it's fixed; go on to next step

Answer these question after you've gone through all the action sequences:

- What do you think Smart Lock is supposed to do? What is it good for?

- How do you think each of the smart lock methods (On-body detection, Trusted devices, Trusted places, Trusted face and Voice match) function?

In the end, please write any comments or suggestions you have regarding this study, the Smart Lock functionality or the UI.

Questions:

- **Q1**: Will the user try to achieve the right action? (Does the user know what to do?)
- **Q2**: Will the user notice that the correct action is available? (Is the action e.g menu/button/... visible to the user?)
- **Q3**: Will the user associate the correct action with the effect that the user is trying to achieve? (Does the action have good labeling and suitable signifiers?)
- **Q4**: If the correct action is performed, will the user see that progress is being made toward solution of the task? (Will the user understand the system's response? Is the feedback understandable? And will the interpretation be correct?)

Action sequence for opening Smart Lock settings:

1. Click on "Settings" on your smart phone
2. Click on "Lock screen and security"
3. Click on "Smart Lock"
4. Draw the current security pattern

Action sequence for enabling On-Body Detection:

1. Open Smart Lock settings
2. Click on "On-body detection"
3. Slide the slider to "On" for Task 1 or to "Off" for Task 2

4. Click on “Continue” (Only for Task 1)

5. Click on the “Back arrow”

6. Click the “Home” button

Action sequence for adding a new trusted place:

1. Open Smart Lock settings

2. Click on “Trusted Places”

3. Click on “Add Trusted Place”

4. Select a location

5. Click on the “Back arrow”

6. Click the “Home” button

Action sequence for deleting a previously added trusted place:

1. Open Smart Lock settings

2. Click on “Trusted Places”

3. From the list of locations, click on the one you want to delete

4. Click on “Delete”

5. Click on the “Back arrow”

6. Click the “Home” button

Action sequence for adding a new trusted device:

1. Open Smart Lock settings

2. Click on “Trusted Devices”

3. Click on “Add Trusted Device”

4. Choose a device

5. Click on the “Back arrow”
6. Click the “Home” button

Action sequence for deleting a previously added trusted device:

1. Open Smart Lock settings
2. Click on “Trusted Devices”
3. From the list of devices, click on the one you want to delete
4. Click on “Remove Trusted Device”
5. Click on the “Back arrow”
6. Click the “Home” button

Action sequence for adding trusted face:

1. Open Smart Lock settings
2. Click on “Trusted Face”
3. Click on “Setup”
4. Click on “Next”
5. Hold your face inside the circle drawn on screen
6. Click on “Done”
7. Click on the “Back arrow”
8. Click the “Home” button

Action sequence for improving trusted face detection:

1. Open Smart Lock settings
2. Click on “Trusted Face”
3. Click on “Improve face-matching”

4. Click on “Next”

5. Hold your face inside the circle drawn on screen

6. Click on the “Back arrow”

7. Click the “Home” button

Action sequence for deleting trusted face:

1. Open Smart Lock settings

2. Click on “Trusted Face”

3. Click on “Remove trusted face”

4. Click on “Remove”

5. Click the “Home” button

Action sequence for enabling Voice Match:

1. Open Smart Lock settings

2. Click on “Voice Match”

3. Slide the slider for “Say ‘Ok Google’ any time”

4. Click on “Next”

5. Say “Ok Google” three times

6. Click on “Yes, I’m in”

7. Draw the security pattern

8. Click on the “Back arrow”

9. Click the “Home” button

Action sequence for deleting voice match:

1. Open Smart Lock settings

2. Click on “Voice Match”

3. Click on “Delete voice model”

4. Click on “Ok”

5. Click on the “Back arrow”

6. Click the “Home” button

Followup Questions:

1. What do you think Smart Lock is supposed to do? What is it good for?

2. How do you think each of the smart lock methods (On-body detection, Trusted devices, Trusted places, Trusted face and Voice match) authenticate you?

3. Please write any comments or suggestions you have regarding this study, the Smart Lock functionality or the UI.

A.1.3 Think-aloud Handout

Consent: Please read the consent form carefully and sign it before starting with the study. Feel free to ask any questions you might have.

Task scenario: Imagine that you just heard about Android’s Smart lock feature, you want to explore it and set it up for use on your mobile device. Note that the we are evaluating the system, not you. As such, there is no right or wrong answer for any of the questions asked. Follow your intuition whenever you are in doubt.

Instructions:

- For each task:

    1. Perform each task using the phone that is temporarily handed out to you.

    2. Speak your thoughts about the functionality, look and feel, difficulties, possible changes to improve the user interface, or any other aspect of the experience out loud as you are interacting with the phone.

3. Remember to return to phone's home screen after you finish each task.

- Answer additional questions presented after each set of tasks.
- After going through all the tasks, answer the follow-up questions on the last page of the handout.

On-body detection tasks:

1. Open "Smart Lock" settings.
2. Enable "On-Body detection".

On-body detection questions:

1. In general, when does "On-Body detection" unlock your phone?
2. In general, when does "On-Body detection" lock your phone?

Trusted places tasks:

1. Add current location as a "Trusted place".
2. Remove current location as a "Trusted place".

Trusted places questions:

1. In general, when does "Trusted places" unlock your phone?
2. In general, when does "Trusted places" lock your phone?

Trusted devices tasks:

1. Add "Mi Band 2" as a trusted device to unlock the phone.
2. Remove "Mi Band 2" as a "Trusted device" to unlock the phone.

Trusted devices questions:

1. In general, when does "Trusted devices" unlock your phone?
2. In general, when does "Trusted devices" lock your phone?

Trusted face tasks:

1. Add your face as the “Trusted face” to unlock the phone.
2. Improve the accuracy of face detection.
3. Remove your face as the “Trusted face” to unlock the phone.

Trusted face questions:

1. In general, when does “Trusted face” unlock your phone?
2. In general, when does “Trusted face” lock your phone?

Voice Match tasks:

1. Add your voice as the trusted voice to unlock the phone.
2. Remove your voice as the trusted voice.

Voice Match questions:

1. In general, when does “Voice Match” unlock your phone?
2. In general, when does “Voice Match” lock your phone?

Followup Questions:

1. In 2-3 sentences, tell us what you think Smart Lock feature it good for.
2. Would you consider using any of the following Smart Lock methods on a regular basis? Please explain in 1-2 sentences.
3. Please write down any comments or suggestions you have regarding this study, the Smart Lock functionality or the Smart Lock UI.

A.2 SL Perception Survey Questions

Demographics

1. What is your age?

2. What is your gender?
3. What is the highest level of education you completed?
4. What is your current occupation?
5. What is your ethnicity?

Smartphone Usage:

1. What is the model number of your phone?
2. On average, how many hours do you spend on your phone each day?
3. Which of the following unlocking methods do you use on your phone?
    - PIN / Password
    - Fingerprint
    - Face detection
    - Iris scanning
    - None
4. On average, how frequently do you unlock your phone?

Smart Lock Intro:

1. Please watch the video below about“Smart Lock for Android.”

Smart Lock Familiarity

1. Prior to taking part in this study, how familiar were you with Smart Lock for Android?
    - I had no idea what Smart Lock was and how it worked.
    - I knew what Smart Lock was, but didn't know how it worked.
    - I knew what Smart Lock was, and had some idea as to how it worked.

- I knew what Smart Lock was, and had a good understanding of how it worked.

2. If you knew about Smart Lock prior to this study, how did you learn about the existence of Smart Lock?

    - I didn't know about Smart Lock before this study.
    - Through notification about Smart Lock on my phone.
    - Through browsing in the settings menu on my phone.
    - On the internet.
    - Through friends and family.
    - Other (please specify).

3. For each Smart Lock method (On-body Detection, Trusted Places, Trusted Devices, Trusted Face, and Voice Match), please select an option which best describes your past or potential experience with Smart Lock:

    - Never used it, and would not use even if it were available on my phone.
    - Never used it, but would use if it were available on my phone.
    - Experimented with it, but never fully set it up and used it.
    - Used it for a while, but stopped using it.
    - Am currently using it.
    - I don't know what this means.

On-Body Detection:

1. (Only if willing to use or are using) Previously, you answered that either you are using or would use On-Body Detection. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

    - It is secure.
    - It makes unlocking the phone easier for me.

- It provides an additional way of unlocking my phone.
- It makes unlocking the phone faster.
- Other [can add reason].

2. (Only if not willing to use or experimented but did not set up) Based on your previous answers, you decided not to use On-Body Detection. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

   - It is not secure (a.k.a., can be tricked into unlocking the phone).
   - It might cause accidental unlocks and pocket dialling.
   - I don't think it makes unlocking easier for me.
   - I don't understand how it works.
   - Other [can add reason].

3. (Only if were using before but stopped) Based on your previous answers, you were using On-Body Detection before, but have stopped. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

   - It was not secure (a.k.a., could be tricked into unlocking the phone).
   - It caused accidental unlocks and pocket dialling.
   - I didn't make unlocking easier for me (a.k.a., wasn't useful).
   - I didn't understand how it worked.
   - I found my phone to be locked when I expected it to be unlocked or vice versa.
   - I changed phones.
   - Other [can add reason].

4. (Only if were using before but stopped) For how long were you using On-Body Detection before stopping?

5. (Only if were using before but stopped) When did you stop using On-Body Detection?

6. (Only if using) For how long have you been using On-Body Detection?

Trusted Places:

1. (Only if willing to use or are using) Previously, you answered that either you are using or would use Trusted Places. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

   - It is secure.
   - It makes unlocking the phone easier for me.
   - It provides an additional way of unlocking my phone.
   - It makes it easier for me to share my phone with others.
   - It makes unlocking the phone faster.
   - Other [can add reason].

2. (Only if not willing to use or experimented but did not set up) Based on your previous answers, you decided not to use Trusted Places. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

   - It is not secure (a.k.a., can be tricked into unlocking the phone).
   - It might allow my family members or co-workers to access my private information.
   - It might cause accidental unlocks and pocket dialling.
   - I don't understand how it works.
   - I don't think it can make unlocking the phone easier for me.
   - I can't think of a place to add as a trusted place.
   - Other [can add reason].

3. (Only if were using before but stopped) Based on your previous answers, you were using Trusted Places before, but have stopped. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

- It was not secure (a.k.a., could be tricked into unlocking the phone).
- It caused accidental unlocks and pocket dialling.
- I didn't make unlocking easier for me (a.k.a., wasn't useful).
- I didn't understand how it worked.
- I found my phone to be locked when I expected it to be unlocked or vice versa.
- I changed phones.
- Other [can add reason].

4. (Only if were using before but stopped) For how long were you using Trusted Places before stopping?

5. (Only if were using before but stopped) When did you stop using Trusted Places?

6. (Only if using) For how long have you been using Trusted Places?

Trusted Devices:

1. (Only if willing to use or are using) Previously, you answered that either you are using or would use Trusted Devices. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

    - It is secure.
    - It makes unlocking the phone easier for me.
    - I provides an additional way of unlocking my phone.
    - It makes it easier for me to share my phone with others.
    - Other [can add reason].

2. (Only if not willing to use or experimented but did not set up) Based on your previous answers, you decided not to use Trusted Devices. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

   - It is not secure (a.k.a., can be tricked into unlocking the phone).
   - It might allow my family members or co-workers to access my private information.
   - It might cause accidental unlocks and pocket dialling.
   - I don't understand how it works.
   - I don't think it can make unlocking the phone easier for me.
   - I can't think of a Bluetooth device to add as a trusted device.
   - Other [can add reason].

3. (Only if were using before but stopped) Based on your previous answers, you were using Trusted Devices before, but have stopped. Why did you choose so? Rank the options below in order of importance to you (with 1 being the most important):

   - It was not secure (a.k.a., could be tricked into unlocking the phone).
   - It caused accidental unlocks and pocket dialling.
   - I didn't make unlocking easier for me (a.k.a., wasn't useful).
   - I didn't understand how it worked.
   - I found my phone to be locked when I expected it to be unlocked or vice versa.
   - I changed phones.
   - Other [can add reason].

4. (Only if were using before but stopped) For how long were you using Trusted Devices before stopping?

5. (Only if were using before but stopped) When did you stop using Trusted Devices?

6. (Only if using) For how long have you been using Trusted Devices?

Smartphone Unlocking Convenience, Speed and Security:

1. Please rank the following screen unlocking methods in order of how convenient you think they make smartphone unlocking: (With 1 being the most convenient)
    - PIN / Password / Pattern Unlock
    - Fingerprint
    - On-Body Detection
    - Trusted Places
    - Trusted Devices
    - Trusted Face
    - Voice Match
2. Please rank the following screen unlocking methods in order of how fast you think they make smartphone unlocking: (With 1 being the fastest)
    - PIN / Password / Pattern Unlock
    - Fingerprint
    - On-Body Detection
    - Trusted Places
    - Trusted Devices
    - Trusted Face
    - Voice Match
3. Please rank the following screen unlocking methods in order of how secure you think they make smartphone unlocking: (With 1 being the most secure)
    - PIN / Password / Pattern Unlock
    - Fingerprint
    - On-Body Detection
    - Trusted Places

- Trusted Devices
- Trusted Face
- Voice Match

A.3 SL Understanding Survey Questions

Demographics

1. What is your age?
2. What is your gender?
3. What is the highest level of education you completed?
4. What is your current occupation?
5. What is your ethnicity?

Smartphone Usage:

1. What is the model number of your phone?
2. On average, how many hours do you spend on your phone each day?
3. Which of the following unlocking methods do you use on your phone?
    - PIN/Password
    - Pattern unlock
    - Fingerprint
    - Face detection
    - Iris scanning
    - Other
    - None
4. On average, how frequently do you unlock your phone?

Smartphone Security and Privacy:

1. For each statement, please select one of the options that best describes your smartphone usage habit (from never to daily):
   - I use my smartphone to check my personal email account.
   - I use my smartphone to take pictures of myself or people close to me.
   - I use my smartphone to go on social networks.
   - I use my smartphone to exchange instant messages.
   - I use my smartphone to look up information about health conditions.
   - I use my smartphone to do online banking.
   - I use my smartphone to look up jobs or submit job applications.
   - I use my smartphone to look up government services and information.
   - I use my smartphone to look up direction to places.
   - I use my smartphone to organize personal affairs.
2. On smartphones, a simple passcode is the default 4- or 6-digit number that is used to lock the screen of the phone. Do you know how to enable longer or alphanumeric passcodes on your phone?
3. On your phone, do you know how to turn off an app's access to the camera in the settings?

Smart Lock Intro:

1. Please watch the video/read the text below about "Smart Lock for Android."

Smart Lock Familiarity

1. Prior to taking part in this study, how familiar were you with Smart Lock for Android?
   - I had no idea what Smart Lock was and how it worked.
   - I knew what Smart Lock was, but didn't know how it worked.

- I knew what Smart Lock was, and had some idea as to how it worked.
- I knew what Smart Lock was, and had a good understanding of how it worked.

2. If you knew about Smart Lock prior to this study, how did you learn about the existence of Smart Lock?

    - I didn't know about Smart Lock before this study.
    - Through notification about Smart Lock on my phone.
    - Through browsing in the settings menu on my phone.
    - On the internet.
    - Through friends and family.
    - Other (please specify).

3. For each Smart Lock method (On-body Detection, Trusted Places, Trusted Devices, Trusted Face, and Voice Match), please select an option which best describes your past or potential experience with Smart Lock:

    - Never used it, and would not use even if it were available on my phone.
    - Never used it, but would use if it were available on my phone.
    - Experimented with it, but never fully set it up and used it.
    - Used it for a while, but stopped using it.
    - Am currently using it.
    - I don't know what this means.

Smart Lock Semantics

1. For each Smart Lock method, select one option that best describes what you think that method is capable of:

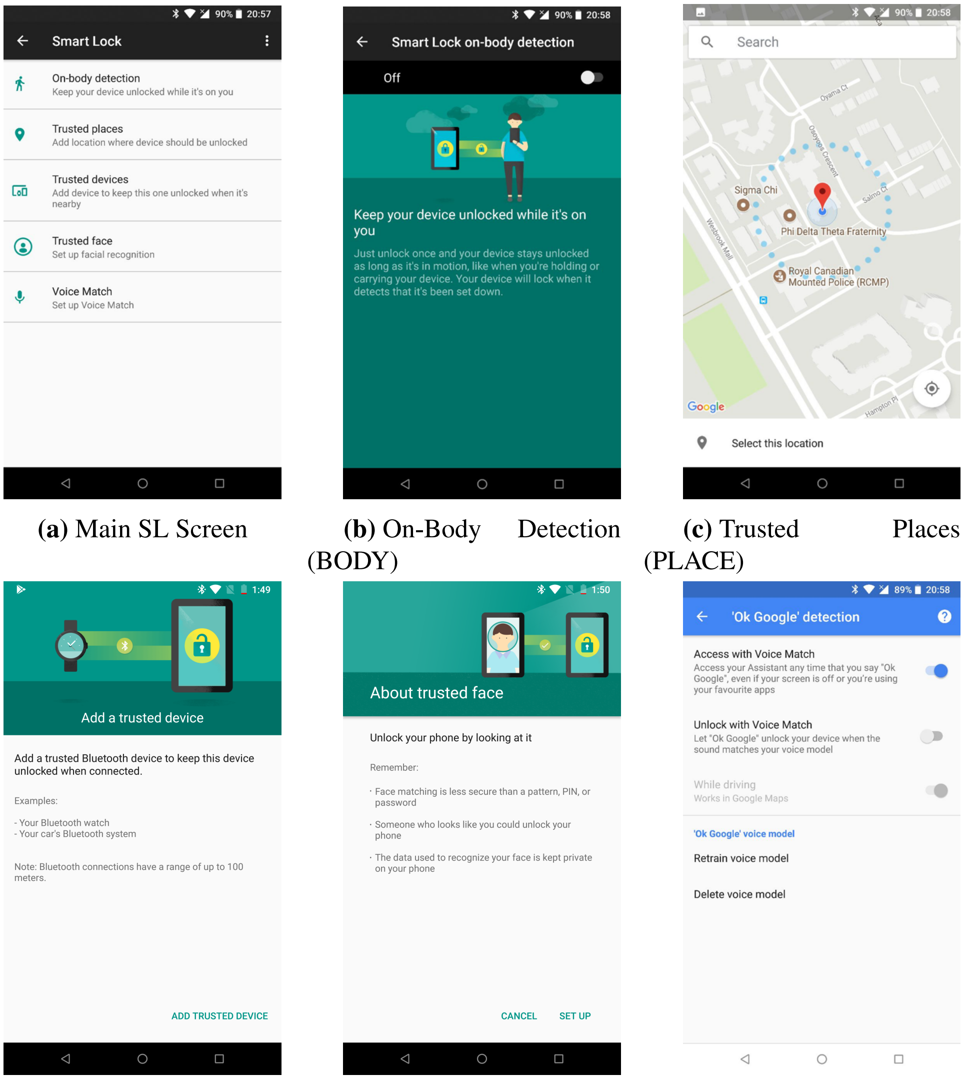
    - Can automatically lock the phone but not unlock it.
    - Can automatically unlock the phone but not lock it.
    - Can automatically lock and unlock the phone.

   - Cannot lock or unlock the phone automatically.
   - I have no idea.
   - I didn’t understand the question.

2. How do you think Smart Lock behaves if multiple methods (e.g., Trusted Places and Trusted Devices) are enabled at the same time?

   - Smart Lock locks/unlocks the phone when any one of the enabled methods (e.g., Trusted Devices, On-Body Detection) tell it to.
   - Smart Lock locks/unlocks the phone only when all of the enabled methods tell it to.
   - I have no idea.
   - I didn’t understand the question.

A.4 Screenshots of SL UI



20:57

Smart Lock

On-body detection
Keep your device unlocked while it's on you

Trusted places
Add location where device should be unlocked

Trusted devices
Add device to keep this one unlocked when it's nearby

Trusted face
Set up facial recognition

Voice Match
Set up Voice Match

(a) Main SL Screen

20:58

Smart Lock on-body detection

Off

Keep your device unlocked while it's on you

Just unlock once and your device stays unlocked as long as it's in motion, like when you're holding or carrying your device. Your device will lock when it detects that it's been set down.

(b) On-Body Detection (BODY)

20:58

Search

Sigma Chi
Phi Delta Theta Fraternity
Royal Canadian Mounted Police (RCMP)
Google

Select this location

(c) Trusted Places (PLACE)

1:49

Add a trusted device

Add a trusted Bluetooth device to keep this device unlocked when connected.

Examples:

- Your Bluetooth watch
- Your car's Bluetooth system

Note: Bluetooth connections have a range of up to 100 meters.

ADD TRUSTED DEVICE

(d) Trusted Devices (DEVICE)

1:50

About trusted face

Unlock your phone by looking at it

Remember:

· Face matching is less secure than a pattern, PIN, or password

· Someone who looks like you could unlock your phone

· The data used to recognize your face is kept private on your phone

CANCEL SET UP

(e) Trusted Face (FACE)

20:58

'Ok Google' detection

Access with Voice Match
Access your Assistant any time that you say "Ok Google", even if your screen is off or you're using your favourite apps

Unlock with Voice Match
Let "Ok Google" unlock your device when the sound matches your voice model

While driving
Works in Google Maps

'Ok Google' voice model

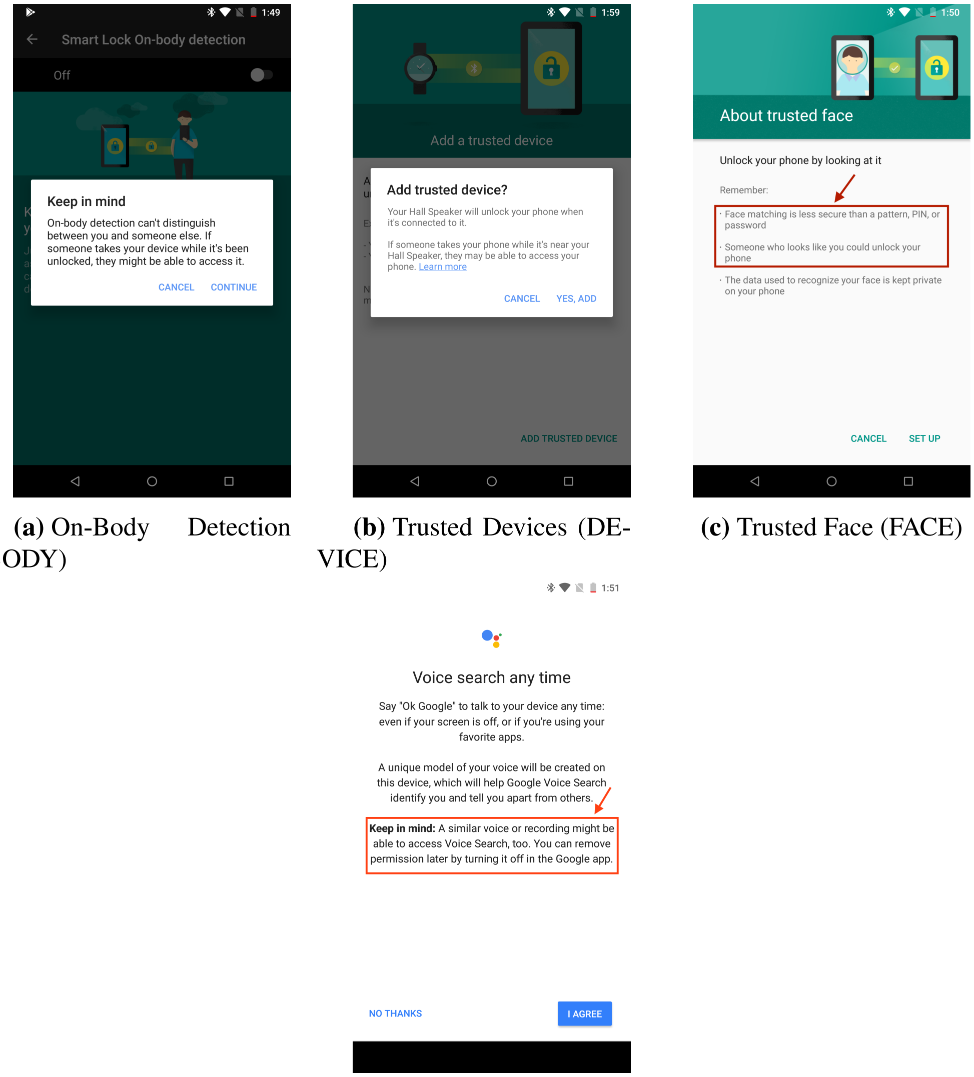
Retrain voice model

Delete voice model

(f) Voice Match (VOICE)

Figure A.1: UI of SL methods as presented to the user on a Google Nexus 6P Smartphone.

A.5 Screenshots of the Warning Messages by SL UI



Smart Lock On-body detection

Off

Keep in mind

On-body detection can't distinguish between you and someone else. If someone takes your device while it's been unlocked, they might be able to access it.

(a) On-Body Detection (BODY)

Add a trusted device

Add trusted device?

Your Hall Speaker will unlock your phone when it's connected to it.

If someone takes your phone while it's near your Hall Speaker, they may be able to access your phone. Learn more

(b) Trusted Devices (DEVICE)

About trusted face

Unlock your phone by looking at it

Remember:

- Face matching is less secure than a pattern, PIN, or password
- Someone who looks like you could unlock your phone
- The data used to recognize your face is kept private on your phone

(c) Trusted Face (FACE)

Voice search any time

Say "Ok Google" to talk to your device any time: even if your screen is off, or if you're using your favorite apps.

A unique model of your voice will be created on this device, which will help Google Voice Search identify you and tell you apart from others.

Keep in mind: A similar voice or recording might be able to access Voice Search, too. You can remove permission later by turning it off in the Google app.

(d) Voice Match (VOICE)

Figure A.2: SL warning screens as presented to the user on a Google Nexus 6P.

A.6 Materials for the Longitudinal Diary Study

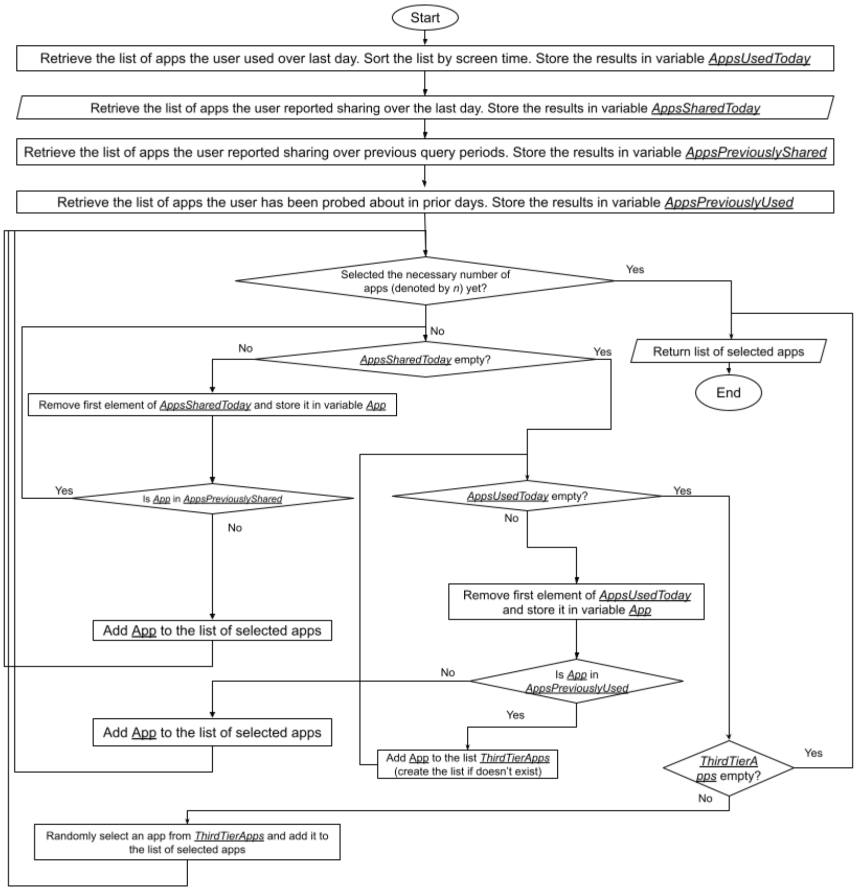
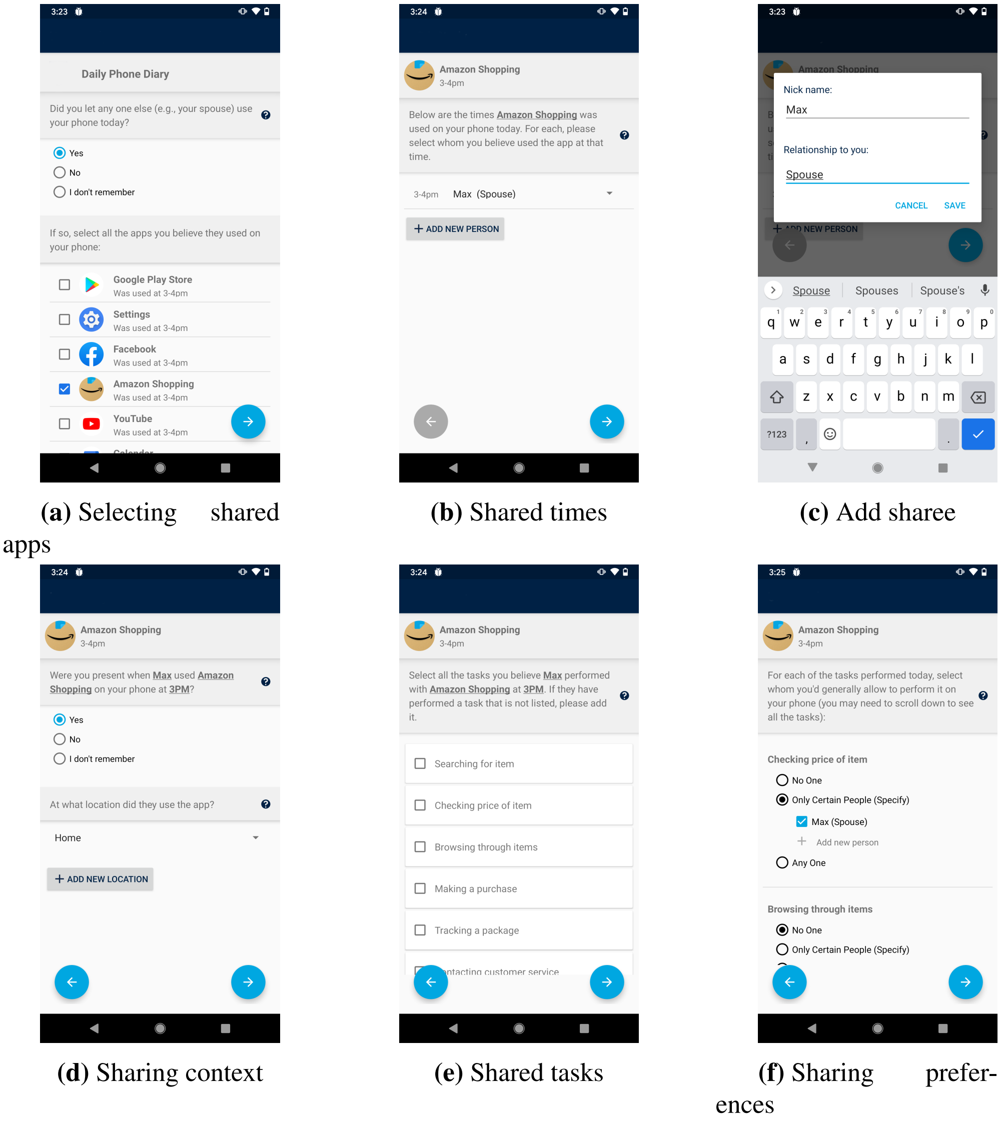

Start
Retrieve the list of apps the user used over last day. Sort the list by screen time. Store the results in variable AppsUsedToday
Retrieve the list of apps the user reported sharing over the last day. Store the results in variable AppsSharedToday
Retrieve the list of apps the user reported sharing over previous query periods. Store the results in variable AppsPreviouslyShared
Retrieve the list of apps the user has been probed about in prior days. Store the results in variable AppsPreviouslyUsed
Selected the necessary number of apps (denoted by n) yet?
Yes
No
Return list of selected apps
End
AppsSharedToday empty?
No
Yes
Remove first element of AppsSharedToday and store it in variable App
Is App in AppsPreviouslyShared
Yes
No
Add App to the list of selected apps
AppsUsedToday empty?
No
Yes
Remove first element of AppsUsedToday and store it in variable App
Is App in AppsPreviouslyUsed
No
Yes
Add App to the list of selected apps
Add App to the list ThirdTierApps (create the list if doesn't exist)
ThirdTierApps empty?
Yes
No
Randomly select an app from ThirdTierApps and add it to the list of selected apps

Figure A.3: Flowchart of the algorithm used to select apps for inclusion in the diaries.



Daily Phone Diary

Did you let any one else (e.g., your spouse) use your phone today?

● Yes
○ No
○ I don't remember

If so, select all the apps you believe they used on your phone:

□ Google Play Store
Was used at 3-4pm

□ Settings
Was used at 3-4pm

□ Facebook
Was used at 3-4pm

☑ Amazon Shopping
Was used at 3-4pm

□ YouTube
Was used at 3-4pm

(a) Selecting shared apps

Amazon Shopping
3-4pm

Below are the times Amazon Shopping was used on your phone today. For each, please select whom you believe used the app at that time.

3-4pm Max (Spouse)

(b) Shared times

Nick name: Max

Relationship to you: Spouse

(c) Add sharee

Amazon Shopping
3-4pm

Were you present when Max used Amazon Shopping on your phone at 3PM?

● Yes
○ No
○ I don't remember

At what location did they use the app?

Home

(d) Sharing context

Amazon Shopping
3-4pm

Select all the tasks you believe Max performed with Amazon Shopping at 3PM. If they have performed a task that is not listed, please add it.

□ Searching for item
□ Checking price of item
□ Browsing through items
□ Making a purchase
□ Tracking a package
□ Contacting customer service

(e) Shared tasks

Amazon Shopping
3-4pm

For each of the tasks performed today, select whom you'd generally allow to perform it on your phone (you may need to scroll down to see all the tasks):

Checking price of item
○ No One
● Only Certain People (Specify)
☑ Max (Spouse)
+ Add new person
○ Any One

Browsing through items
● No One
○ Only Certain People (Specify)

(f) Sharing preferences

Figure A.4: Screenshots of the study app.