

RUHR
UNIVERSITÄT
BOCHUM

RUB

Securing
Knowledge-Based Authentication
Against Online Attackers

DISSERTATION

zur Erlangung des Grades eines Philosophiae Doktor (Ph.D.)
der Fakultät für Computer Science
an der Ruhr-Universität Bochum

vorgelegt von

Daniel Vernon Bailey

geboren in Wilmington, Delaware, USA

Bochum, 16. Juli 2023

Tag der mündlichen Prüfung: 6. Juli 2023

Gutachter:

Prof. Dr.-Ing. Markus Dürmuth, Leibniz Universität Hannover

Zweitgutachter:

Prof. Dr.-Ing. Christof Paar, Ruhr-Universität Bochum

Abstract

User authentication remains a critical building block for trust online. Despite their widely-known shortcomings, knowledge-based authenticators (KBAs) including passwords and Personal Identification Numbers (PINs) are in daily use for access control, and they appear to still be relevant for years to come. This dissertation examines several aspects of their shortcomings, along with concrete suggestions on how to improve them. Most of these shortcomings stem from a simple fact: users are commonly relied upon to select, recall, and enter their own KBAs, resulting in systems with poor usability that also poorly protect their users. Such KBAs are often drawn from a narrow probability distribution making them easy for an attacker to guess. Moreover, this distribution is affected not only by users, but also a number of factors under the control of service providers and system designers.

We will return to these themes throughout this dissertation: we will characterize the probability distributions for several classes of KBAs (PINs/passwords) in various settings and we will explore system and service design factors that broaden (harder to guess) or narrow (easier to guess) the distributions. Put simply, in this dissertation we show how systems may enable the selection of KBAs that better protect the user. For example, we show the influence of account type on this distribution and we characterize the PINs used in situations of mobile device and app unlock.

Our attacker model crucially focuses on *online* guessers that are limited in the number of allowed attempts. After all, given unlimited time and guesses, an attacker would easily guess practically all human-chosen and human-memorable PINs and passwords. As the purpose of a KBA in the end is to provide security against an adversary, we will necessarily characterize adversary capabilities in detail. Results in this dissertation optimize guessing attacks for various situations, but especially in the context of mobile device unlocking.

Given the sheer number of KBAs managed by a user and their difficulties in selecting and remembering them, it is interesting for system designers to see how KBAs are reused across accounts and services. Where the previous work would rely on self-reports to estimate the prevalence of password reuse, we use a novel methodology to *directly* measure both resistance to password guessing and the incidence of reuse, using the first dataset of its kind in the literature. Our results show that users select more-secure passwords for their financial accounts, but critically, are more likely to reuse these passwords, perhaps as a means of coping with the added complexity.

Continuing with this theme, we study 4- and 6-digit PINs in detail. We gather a collection of new datasets in user studies: first from participants primed for the scenario of smartphone unlocking, then investigating how user understanding of the PIN feature in the Signal encrypted messaging app correlates with PIN complexity, then finally by priming participants to guess the PINs of others. Through these new datasets, we show that 6-digit PINs are in many cases not more secure than 4-digit, that participants can guess the 4-digit PINs of roughly 1 in 8 smartphones, and that those who understand PINs are more likely to select more-complex PINs. Taken together, these PIN datasets are the largest ever assembled in formal user studies.

Through this observation that user comprehension can lead to increased PIN complexity, we suggest user education as a means to increase PIN security, among other potential interventions by system designers.

Kurzfassung

Die Authentifizierung von Personen ist ein entscheidender Baustein für Vertrauen im Internet. Trotz ihrer allgemein bekannten Schwächen, werden wissensbasierte Authentifizierungsmethoden (Englisch: knowledge-based authenticators, KBAs), wie Passwörter und Persönliche Identifikationsnummern (PINs), täglich zur Zugangskontrolle eingesetzt und scheinen auch in den kommenden Jahren relevant zu bleiben. Diese Dissertation untersucht verschiedene Aspekte ihrer Schwächen und erarbeitet konkrete Verbesserungsvorschläge. Ein Großteil der Schwächen ergibt sich aus einer einfachen Tatsache: Zumeist müssen die Nutzenden von Systemen ihre KBAs selbst auswählen, erinnern und erneut eingeben. Dies führt nicht nur zu einem Mangel an Benutzerfreundlichkeit, sondern auch dazu, dass die Nutzenden nur unzureichend geschützt sind. Das grundlegende Problem solcher selbstgewählten KBAs liegt darin, dass diese oft aus einer eingeschränkten Wahrscheinlichkeitsverteilung stammen, welche sie für Angreifende leicht erratbar machen. Darüber hinaus wird diese Verteilung nicht nur von Benutzenden, sondern auch von verschiedenen anderen Faktoren beeinflusst, die von Diensteanbietern und Systemdesignern festgelegt werden.

Im Verlauf dieser Dissertation werden die folgenden Themen im Detail behandelt: Neben den Wahrscheinlichkeitsverteilungen verschiedener Klassen von KBAs (PINs/Passwörtern) in unterschiedlichen Situationen charakterisieren wir auch die Faktoren im System- und Dienst-Design die die Verteilungen vergrößern (schwerer zu erraten) oder verkleinern (leichter zu erraten). Einfach ausgedrückt zeigen wir in dieser Arbeit, wie Systeme die Auswahl von KBAs beeinflussen können, um Nutzende besser zu schützen. Zum Beispiel zeigen wir, wie die Art des Kontos die Verteilung beeinflusst und charakterisieren die PINs, die typischerweise beim Entsperren von Mobilgeräten und Apps verwendet werden.

Unser Angriffsmodell konzentriert sich auf *online Rateangriffe*, die in der Anzahl der erlaubten Versuche eingeschränkt sind. Schließlich würden Angreifende bei unbegrenzter Zeit und unbegrenzten Versuchen praktisch alle von Menschen wähl- und merkbaren PINs und Passwörter erraten. Da das Ziel einer KBA letztendlich darin besteht, Sicherheit gegenüber bestimmten Angreifenden zu bieten, werden wir zwangsläufig die Fähigkeiten von Angreifenden detailliert charakterisieren.

Angesichts der schieren Anzahl von KBAs, die von Benutzenden verwaltet werden und der Schwierigkeiten, sie auszuwählen und sich an sie zu erinnern, ist es für Systemdesignende interessant zu verstehen, wie KBAs über Konten und Dienste hinweg wiederverwendet werden. Während frühere Arbeiten sich hauptsächlich auf Eigenangaben aus Umfragen stützen, um die Häufigkeit der Wiederverwendung von Passwörtern abzuschätzen, verwenden wir eine neuartige Methodik, um sowohl die Erratbarkeit von Passwörtern als auch die Häufigkeit der Wiederverwendung *direkt* zu messen, und zwar anhand des ersten Datensatzes dieser Art in der Literatur. Unsere Ergebnisse zeigen, dass Benutzende sicherere Passwörter für ihre Finanzkonten auswählen, aber leider auch dazu tendieren diese Passwörter eher wiederzuverwenden, vielleicht um mit der zusätzlichen Komplexität besser zurechtzukommen.

In diesem Zusammenhang untersuchen wir auch 4- und 6-stellige PINs im Detail. Hierfür sammeln wir eine Reihe neuer Datensätze in Benutzerstudien: zunächst von Teilnehmenden, die auf das Szenario des Entsperrens von Smartphones vorbereitet sind, dann untersuchen wir, wie das Verständnis der Benutzenden für die PIN-Funktion in der verschlüsselten Messaging-App „Signal“ mit der PIN-Komplexität korreliert, und schließlich indem wir Teilnehmende darauf vorbereiten, die PINs anderer zu erraten. Durch diese neuen Datensätze zeigen wir, dass 6-stellige PINs in vielen Fällen nicht sicherer sind als 4-stellige, dass Teilnehmende die 4-stelligen PINs von etwa 1 aus 8 Smartphones erraten können und dass diejenigen, die PINs verstehen, eher komplexere PINs auswählen. Insgesamt sind diese PIN-Datensätze die größten, die jemals in formalen Benutzerstudien gesammelt wurden.

Aufgrund der Beobachtung, dass ein besseres Verständnis der Benutzenden zu stärkeren PINs führen können, schlagen wir Schulungen für Benutzende, neben anderen möglichen Interventionen durch Systemdesignende, als Mittel zur Erhöhung der PIN-Sicherheit vor.

Acknowledgements

I find it improbable and thrilling beyond compare that this lifelong dream of mine has come true, and I owe it all to the support and encouragement of so many people. First and foremost, I dedicate this work to my wife Melissa, for believing in me. Her boundless joy, care, patience, compassion, and love have been the bedrock of my journey. Especially when, in April 2022, I wanted to quit my corporate job to focus on finishing my degree, she agreed — knowing she might have to support us financially if there wasn't enough consulting work.

My deep gratitude goes to my advisor Markus Dürmuth, who graciously took me on as a non-traditional student and allowed me to pursue this path. His guidance, both in minor and major matters, played a pivotal role in shaping the research contained in this thesis. Chapter 6 stands as a testament to how his prior work directly inspired my own.

Over the years, Christof Paar has been a pillar of support, from our days at Worcester Polytechnic Institute to the present at the Ruhr University Bochum. His leadership qualities continue to inspire me, and I hope to emulate them in the future.

Adam J. Aviv taught me a great deal about how to structure, perform, and ultimately write about user-centered research. Working with him on three of the papers that make up this thesis ended up being something of an apprenticeship: I now have the tools and techniques to conduct independent research in this area. His mentorship has been transformative, and I am thankful for the knowledge he imparted.

Maximilian Golla is an extraordinary friend, colleague, and collaborator. In countless ways, he was instrumental in bringing this thesis to a conclusion. He deeply reviewed and commented on earlier drafts of the dissertation, leading to this final document you hold in your hands. Moreover, he handled logistics “on the ground” in Bochum, even printing and submitting the review copies to the department. Simply put, I cannot imagine this journey without his presence.

For all of my other co-authors over the years, Philipp and Collins especially, I am extremely grateful for their collaboration and the collective effort that went into producing this work.

Certainly not least, I want to express my heartfelt thanks to my family, especially my Mom, for all the love, inspiration, and hard work they invested in my education and dreams. Like Minnie Pearl said, “I’m just so proud to be here.”

With deepest gratitude and excitement for the next adventure!!

Dan

July, 2023

Contents

1. Introduction	1
1.1. Challenges in Knowledge-Based Authentication	2
1.2. Human Factors and Conflicting Advice	3
1.3. About this Dissertation	5
1.4. Summary of Contributions	7
1.5. List of Publications	9
2. Preliminaries	11
2.1. Knowledge-Based User Authentication	12
2.2. Attacker Models	19
2.3. Re-Using Authentication Secrets	23
2.4. Ethics in Knowledge-Based Authenticator Research	26
3. Password Re-use	31
3.1. Introduction	32
3.2. Malware-derived Dataset	33
3.3. Correlation of Password Strength and Account Value	37
3.4. Password Re-use	44
3.5. Conclusion	48
3.6. Author Contribution	49
4. Security of Smartphone Unlock PINs with Blocklists	51
4.1. Introduction	52
4.2. Background	54
4.3. User Study	57
4.4. PIN Selection on Smartphones	66
4.5. Blocklists and PIN Selection	72
4.6. Conclusion and Recommendations	84
4.7. Author Contribution	85
5. Users' Understanding of Signal PINs	87
5.1. Introduction	88
5.2. Enhancing Signal User Authentication	90
5.3. Related Work on Secure Usable Messaging	93
5.4. User Study Design and Methodology	95
5.5. Results	99
5.6. Discussion of Recommended Enhancements	110
5.7. Conclusion	112
5.8. Author Contribution	113

6. Analyzing How Untrained Attackers Guess PINs	115
6.1. Introduction	116
6.2. Related Work on Smartphone Threat Models	117
6.3. Methodology	118
6.4. PIN Characteristics	124
6.5. Novice Attackers' Performance	126
6.6. Context for Smartphone Access	130
6.7. Discussion and Conclusion	133
6.8. Author Contribution	135
7. Summary and Future Work	143
7.1. Summary and Key Results	144
7.2. Thesis Statement Evaluation	145
7.3. Outlook and Future Work	146
List of Figures	149
List of Tables	151
A. PIN Blocklists	153
A.1. Survey Instrument	154
A.2. Demographics	158
A.3. Device Usage	159
A.4. PIN Selection and Changing Strategies	160
A.5. Feelings and Sentiments	161
B. Signal PINs	163
B.1. Additional Pre-Screening Study	164
B.2. Survey Instrument of the Main Study	164
B.3. Additional Figures	167
B.4. Codebooks	169
C. Novice Attackers	179
C.1. Survey Instrument	180
C.2. Qualitative Codes	183
Bibliography	187

1

Introduction

Contents

1.1. Challenges in Knowledge-Based Authentication	2
1.2. Human Factors and Conflicting Advice	3
1.3. About this Dissertation	5
1.3.1. Thesis Statement	5
1.3.2. Password Re-Use for Financial Accounts on the Web	6
1.3.3. Analyzing the Security of Smartphone Unlock PINs	6
1.3.4. User Understanding of Signal PINs on Mobile Devices . . .	7
1.3.5. How Novice Attackers Guess Smartphone Unlock PINs . . .	7
1.4. Summary of Contributions	7
1.5. List of Publications	9

1.1. Challenges in Knowledge-Based Authentication

Decades of research in computer security have delivered ever-greater cryptographic algorithms, protocols, and deployments. At the same time, the increase in processing and communication performance has enabled new and more personal devices that gather more sensitive personal data. As of 2022, Google reports that about 95% of the traffic across its products and services is now encrypted [1]. Despite this massive technological advancement, in which extraordinary amounts of data are communicated and stored each day, encrypted and decrypted on demand, controlling access remains a challenge. To be useful, the data must be available to legitimate parties for legitimate purposes while protected from attackers. Unfortunately, despite all our technological prowess, the steady stream of data breaches continues, in which some form of access control has failed.

User authentication is a necessary building block to limit access to legitimate parties. For human users, authentication often relies on Knowledge-Based Authentication (KBA), colloquially referred to as “something you know.” Although other factors like tokens (“something you have”) and biometrics (“something you are”) are also important, our focus here is on KBA due to its extremely widespread use, as it is often combined with biometrics. We further focus on situations of mobile device unlocking again due to their widespread nature. Observe that even if a user adopts a biometric to unlock their iOS or Android phone, a PIN is still required especially when the device restarts — and obviously can be attempted by any adversary with access to the device.

In this familiar arrangement, to prove their identity to a system (website, PC, or smartphone), the human user provides some piece of presumably secret information like a password or Personal Identification Number (PIN). It is of course possible that a human user creates and flawlessly remembers a long, random password or uses a password manager. But overall, in this simple interaction, we must contend with many very human limitations and we aim to study and improve upon real-world security. In particular, this research is concerned with understanding — and potentially improving — the PINs and passwords *chosen by users* and used to access websites and devices. Moreover, we make concrete suggestions for service providers to enhance the security of their use of KBA. In so doing, we contribute to the state of knowledge in this area by studying user and attacker behaviors. Our results can lead to improved system designs, user education, and user communication.

1.2. Human Factors and Conflicting Advice

Here the challenge is not entirely technological: human factors impose very real limits on what can be achieved in KBA. Specifically, a large body of literature has shown that humans have a hard time choosing and remembering secrets.

However secure it might be, if a solution is too hard to use, it will remain mostly unused, as previously demonstrated in a classic usability study by Whitten et al. [2] on the difficulties non-expert users faced trying to secure their email with PGP 5.0.

It is not only users who are frustrated by this state of affairs. Service providers and system/design engineers also face challenges since they must deal with customers who have difficulties accessing systems, such as forgotten passwords and passwords guessed or stolen by attackers. Just like a user, a service provider is usually not primarily concerned with authentication, but rather providing some useful service.

By way of example, consider some difficulties faced by a *service provider* in the adoption of a KBA regime. Some services require users to change their PIN/password periodically, say every 90 days. We observe here that government and industry regulations often provide incomplete and contradictory advice. For example, the U.S. National Institute for Standards and Technology (NIST) Special Publication 800-63B [3] provides this guidance to service providers it frames as “verifiers:”

Verifiers SHOULD NOT impose other composition rules (e.g., requiring mixtures of different character types or prohibiting consecutively repeated characters) for memorized secrets. Verifiers SHOULD NOT require memorized secrets to be changed arbitrarily (e.g., periodically).

Although this guidance was issued more than five years ago, based on research published in 2010 and even earlier from Inglesant et al., Komanduri et al., Chiasson et al., [4–6] and others, many service providers still require these “password complexity rules” and “password rotation.”

On the other hand, it’s not difficult to see why this is the case. For one thing, older editions of the NIST 800-63B standard from 2004 required composition rules and password rotation [7]. The changes between password composition rules in 2010 and 2016 are studied by Mayer et al. [8]. For another, the Payment Card Industry Data Security Standard (PCI-DSS) sets the rules for service providers that process credit, debit, and other payment card transactions worldwide. Service providers must be audited against these rules annually. PCI-DSS version 3.2.1, section 8.2.3-8.2.4, published September 2022 [9] contains the following excerpt:

Customer passwords/passphrases are required to meet at least the following strength/complexity: require a minimum length of at least seven characters, contain both numeric and alphabetic characters ... [and] require users to change passwords at least once every 90 days.

Though PCI 4.0 was published in March 2022 [10], service providers may still choose to be validated against PCI 3.2.1. For its part, PCI 4.0 requires passwords to increase in length to 12 characters, but also says:

Passwords/passphrases are changed at least once every 90 days, or the security posture of accounts is dynamically analyzed, and real-time access to resources is automatically determined accordingly.... [Passwords must] contain both numeric and alphabetic characters.

We focus on the issue of conflicting regulation only to draw attention to our broader point: if even regulators provide conflicting advice to service providers, it is unsurprising that users continue to face challenges. Users, after all, are fundamentally trying to complete some task: sending email, conducting business, or recreation. Their fundamental goal is not necessarily security. So users have a tendency to skip extra security-oriented steps if possible. Moreover, if it is easier for a user to misuse a mechanism, it will be misused — not by every user, but by a large population of users. This dissertation, therefore, focuses on ways in which commonly-used KBA mechanisms are selected in various scenarios, and proposes concrete, tangible steps to increase the protections afforded to large populations of users in practice.

Our work lies at this intersection of security and usability, gathering direct evidence that currently-deployed systems fail to adequately address the human factors involved. We present new datasets comprising passwords and PINs. Taken together, to our knowledge this is the largest collection of PINs gathered in formal user studies. By their nature, passwords and PINs are difficult to study due to a general lack of data. System designers strongly discourage users from revealing their passwords or PINs. Over time the research community has accumulated more passwords than PINs: while the study of the probability distribution of alphanumeric passwords is aided by periodic large-scale password leaks from websites [11], such leaks do not exist for PINs as they are locally stored and validated on user devices. Observe also that these website leaks expose user credentials from a website, but do not expose all of a given user’s credentials used elsewhere. We provide one of the first datasets with this focus, gathered from password-stealing malware.

Studies suggest more than 60% of users secure their smartphone with a PIN [12]. Since there are billions of smartphone subscribers worldwide, weaknesses in this scheme would affect many millions of users. Understanding and improving upon this situation is vital to understand and improve upon the security of KBA, which is the focus of the present work.

As in many areas of security, we face a careful balance between security and usability on top of an endless progression of attacks, countermeasures, and attackers with their counter-countermeasures. It is well-known that users commonly practice poor PIN and password hygiene.

As noted above, passwords created for financial-services accounts commonly require a mixture of numbers, mixed-case letters, and symbols. Unfortunately, as we show below with a unique dataset, these higher-complexity passwords tend to be reused more often. Users appear to cope with higher-complexity passwords by remembering fewer of them and reusing them. Though the resulting passwords might be harder for an attacker to outright guess, attackers have also taken advantage of this situation by replaying lists of stolen credentials from one site across other websites. These lists are traded among groups of attackers or sold through websites focusing on cybercrime.

One suggested approach for service providers to increase the cost of attack is to specifically ban the use of PINs/passwords that are frequently used. Indeed, advice from the U.S. National Institute for Standards and Technology has recommended that users should be prevented from choosing one of these [3]. As before, there is much more data available about these overused passwords than there are for PINs. Previous researchers along with our own work, due to this lack of data, often used data of uncertain provenance. One of our contributions is the collection and analysis of new PIN and password datasets, with an emphasis on mobile unlock for PINs and an emphasis on desktop website login for passwords.

Afterward, we focus on the specific case of PINs on mobile devices. We explore a variety of blocklisting approaches and make specific recommendations to improve blocklists. Following this, we study a particular case of user comprehension of PIN usage in the Signal mobile app. Finally, we conclude by evaluating the ability of untrained users to guess the mobile-unlock PINs of others.

1.3. About this Dissertation

In this thesis, we make several contributions to the state of the art in knowledge-based user authentication. This dissertation compiles research previously published and to appear. These collaborative works appear in summarized and annotated form below and would not have been possible without the contributions of the co-authors.

1.3.1. Thesis Statement

This dissertation examines pressing issues faced by users, designers, and administrators of secure websites and mobile devices. The central thesis of this dissertation is: The widespread use of knowledge-based authentication means a shared responsibility for security in the face of usability challenges. Adapting authentication to the way users comprehend, select, use, and re-use their passwords and PINs along with studying the way attackers guess online can lead to improved outcomes for all stakeholders. From our work, future knowledge-based authentication systems can provide a better

user experience and improved attack resistance. Our data show that increased user comprehension of the purpose of the PIN and enhanced user understanding of the value of their accounts can lead to more diverse PIN composition.

1.3.2. Password Re-Use for Financial Accounts on the Web

We compiled and analyzed a unique corpus of data: actual user passwords intercepted by the Zeus banking malware. Although researchers commonly ask study participants about their password re-use habits and draw conclusions, we were able to directly measure this behavior. This analysis is made possible by the fact that the gathered dataset includes multiple leaked passwords per user. We find that in our sample, financial-account passwords are stronger than others, such as social-media passwords. Nevertheless, these stronger passwords are *more prone to re-use*: valuable passwords are identical to 21% of the remaining passwords of a user. This fact suggests a need for user education around password re-use. Before the publication of our study, little was known about password re-use for different account values.

1.3.3. Analyzing the Security of Smartphone Unlock PINs

In this paper, we provide the first comprehensive study of user-chosen 4- and 6-digit PINs ($n = 1705$) collected on smartphones with participants being explicitly primed for device unlocking. We find that against a throttled attacker (with 10, 30, or 100 guesses, matching the smartphone unlock setting), using 6-digit PINs instead of 4-digit PINs provides little to no increase in security, and surprisingly may even decrease security. We also study the effects of blocklists, where a set of “easy to guess” PINs is disallowed during selection. Two such blocklists are in use today by iOS, for 4-digits (274 PINs) as well as 6-digits (2910 PINs). We extracted both blocklists and compared them with six other blocklists, three for each PIN length. In each case we had a small (4-digit: 27 PINs, 6-digit: 29 PINs), a large (4-digit: 2740 PINs, 6-digit: 291 000 PINs), and a placebo blocklist that always excluded the first-choice PIN. For 4-digit PINs, we find that the relatively small blocklist in use today by iOS offers little to no benefit against a throttled guessing attack. Security gains are only observed when the blocklist is much larger. In the 6-digit case, we were able to reach a similar security level with a smaller blocklist. As the user frustration increases with the blocklists size, developers should employ a blocklist which is as small as possible while ensuring the desired security. Based on our analysis, we recommend that for 4-digit PINs a blocklist should contain the 1000 most popular PINs to provide the best balance between usability and security, for 6-digit PINs the 2000 most popular PINs should be blocked.

1.3.4. User Understanding of Signal PINs on Mobile Devices

We conducted an online study with ($n = 235$) Signal users on their understanding and usage of PINs in Signal. In our study, we observe a split in PIN management and composition strategies between users who can explain the purpose of the Signal PINs (56%; enthusiasts) and users who cannot (44%; casual users). Encouraging adoption of PINs by Signal appears quite successful: only 14% opted-out of setting a PIN entirely. Among those who did set a PIN, most enthusiasts had long, complex alphanumeric PINs generated by and saved in a password manager. Meanwhile more casual Signal users mostly relied on short numeric-only PINs. Our results suggest that better communication about the purpose of the Signal PIN could help more casual users understand the features PINs enable (such as that it is not simply a personal identification number). This communication could encourage a stronger security posture.

1.3.5. How Novice Attackers Guess Smartphone Unlock PINs

This chapter provides experimental justification for the attacker models studied above. Previous studies on smartphone unlock PINs including earlier chapters of this thesis primarily consider very well-informed attackers that guess PINs in frequency order using PINs collected through experiments or extracted from password leaks. In this study, we consider a more commonplace, untrained attacker who simply tries to guess a PIN on another person's smartphone. To simulate such a scenario, we adapt a methodology used by Uellenbeck et al. [13] that directs participants to select a 4- or 6-digit (secret) PIN, and then enter 5 PIN guesses that they believe others in the study selected as their secret PIN. In an online survey ($n = 210$), we find that 10% of participants' secret PINs are guessed, with 85% of participants successfully guessing at least one other participant's PIN.

1.4. Summary of Contributions

In short, this dissertation carefully gathers and analyzes new data on KBAs: passwords and especially 4- and 6-digit PINs. The analysis shows precise circumstances when particular measures like blocklisting are of benefit when applied to smartphone unlocking. In addition, we study the success rate of novice PIN guessers and show that improved user comprehension of PIN usage leads to more-diverse PIN composition, suggesting a means to improve PIN security. In summary:

- We gather, analyze, and report on several new datasets: one focused on passwords that directly shows higher strength, but also a higher incidence of re-use

among online financial passwords. Each of our datasets is the largest of its kind presented and analyzed in the literature.

- Our data suggest users choose financial-account passwords that are stronger than others, such as social-media passwords. Nevertheless, these stronger passwords are *more prone to re-use*: valuable passwords are identical to 21% of the remaining passwords of a user.
- We report on the security and composition of user-chosen four- and six-digit PINs applied to smartphone unlocking. In this throttled setting, the benefit of six-digit PINs is marginal at best.
- We show how different blocklisting approaches influence the PIN selection process for both security and usability, finding that blocklists in use today offer little to no added security.
- We explore users' perception of security, memorability, and ease-of-use of PIN-based authentication, finding that participants perceive that blocklisting will improve their PINs without impacting usability, except for very large blocklists.
- We provide guidance for developers on choosing an appropriately-sized PIN blocklist that can influence users to select better PINs. An effective blocklist for 4-digit PINs should consist of about one thousand PINs, while an effective blocklist for 6-digit PINs should contain about two thousand PINs.
- Our data show that 31% of participants in our study have attempted to access someone else's smartphone in the previous year. Through an online simulated-attack study ($n = 210$) where participants try to guess the PINs set by other participants, we find that 85% of participants successfully guess the PIN of a stranger, suggesting the need for more design interventions and user education to encourage users to select more secure PINs.
- We provide experimental justification for the data-driven attackers studied above. To our knowledge, our study is the first to gather and analyze PINs chosen by participants primed to guess smartphone PINs. Our earlier studies speculated that their guessing proxies were representative of real guessers. We can now answer that question in the affirmative. At least in the aggregate, we show that novice attackers perform similarly to our earlier data-driven attackers. This fact suggests that in our setting focused on the first few attempted guesses, data-driven attackers model a real-world threat.
- We explore novice attackers for PINs, as this category represents attackers with the most opportunities to attempt PIN guessing. In our experiment, 85% of

participants successfully guess the PIN of another participant. In addition, 37% of participants admit trying to access the smartphone of another person in the previous year. Since 14% of 4-digit PINs were guessed, our data suggests a novice can unlock 1 in 8 smartphones locked with a 4-digit PIN. Our results indicate the need for more design interventions and user education to nudge users towards more secure PINs, and emphasize risks from “insider” novices with temporary access to unattended devices.

- We highlight the threat models that users are most concerned about with access to their smartphones, showing that most participants feel their close social connections are the ones most likely to attempt to access their smartphone. Further, users indicate a need to delegate access to their devices, suggesting the need for system designers to deploy and further educate users about device sharing options.
- We further surveyed Signal users ($n = 235$), asking about their understanding, usage of the Signal PIN feature, and response to Signal PIN verification. For example, we asked participants to explain the purpose of Signal PINs, in their own words. We additionally asked participants about the composition of their PIN (e.g., length, character set), if they reuse the PIN in other contexts (e.g., phone lock, in another messenger app), if they have opted out of selecting a PIN, and their response to periodic PIN verification.
- We find that only 14% ($n = 33$) of respondents opted out of setting a Signal PIN, and also we find a large disparity between the practices of participants who can explain the purpose of the in-app PIN authentication (who we term *Signal enthusiasts*; $n = 132$; 56%) and those who cannot (dubbed *casual* Signal users; $n = 103$; 44%).

1.5. List of Publications

The content of this thesis is drawn in part from peer-reviewed publications. The research described in those publications was carried out in collaboration with students, colleagues, and other members of the respective research projects. Except as described below, I was lead author on these publications. The author’s contribution to each publication is described in detail in the respective chapters. As an editorial note, this dissertation makes use of the academic “we” throughout.

The interested reader may find these publications as shown:

- **Chapter 3: Password Reuse in Financial Accounts**

D. V. Bailey, M. Dürmuth, and C. Paar, “Statistics on Password Re-use and Adaptive Strength for Financial Accounts,” in *Security and Cryptography for Networks (SCN '14)*. Amalfi, Italy, September 2014.

- **Chapter 4: Analyzing the Security of Smartphone Unlock PINs**

P. Markert, **D. V. Bailey**, M. Golla, M. Dürmuth, A. J. Aviv, “On the Security of Smartphone Unlock PINs,” in *ACM Transactions on Privacy and Security*, Volume 24, Issue 4, November 2021.

- **Chapter 5: Users’ Understanding of Signal PINs**

D. V. Bailey, P. Markert, A. J. Aviv, “‘I have no idea what they’re trying to accomplish’: Enthusiastic and Casual Signal Users’ Understanding of Signal PINs,” in *Symposium on Usable Privacy and Security (SOUPS '21)*. Virtual, August 2021.

- **Chapter 6: Analyzing How Untrained Attackers Guess PINs**

D. V. Bailey, C. W. Munyendo, H. Dyer, P. Markert, M. Grant, A. J. Aviv, “‘Someone Definitely Used 0000’: Analyzing How Novice Attackers Guess Unlock PINs,” to appear at European Symposium on Usable Security, Oct. 2023.

*The men of Gilead said vnto him,
Art thou an Ephraimite?
If he said, Nay; then said they vnto him,
“Say now Shibboleth:” and he said “Sibboleth:”
For hee could not frame to pronounce it right.*

— The Book of Judges (KJV)

2

Preliminaries

Contents

2.1. Knowledge-Based User Authentication	12
2.1.1. Physical Setting: Historical Perspective	12
2.1.2. Online Setting	14
2.1.3. The Mobile Device Setting	15
2.1.4. Measuring Guessability	17
2.2. Attacker Models	19
2.2.1. Online Trawling Guessing Attackers	19
2.2.2. Attacks Beyond Our Scope	21
2.3. Re-Using Authentication Secrets	23
2.3.1. Users’ Tendency to Re-Use	23
2.3.2. Responses to Re-Use	25
2.4. Ethics in Knowledge-Based Authenticator Research	26
2.4.1. Taxonomy of Password Lists	27
2.4.2. Discussion	29

2.1. Knowledge-Based User Authentication

This section reviews the most basic concepts of user authentication. It can come as something of a surprise that such a seemingly familiar topic can be the subject of so much current research. After all, if humans could readily generate and enter long cryptographic keys, this problem would be seemingly solved.

It is well-known that users generally pick passwords and PINs poorly. Exactly how poorly — and the advantage conferred to the attacker — depends on the particular situation, or the human factors considered in the system design. It depends also on the objective of the attacker. This section reviews some perspectives on knowledge-based user authentication, especially arising from the literature. Throughout this section, we will note that a wide variety of attacker models has been considered in the literature. The chapters of this dissertation examine particular settings for this problem. While this section gives a general overview necessarily including a number of different attacker models, the chapters of this dissertation will explain the particular attacker model under study.

2.1.1. Physical Setting: Historical Perspective

Problems of human authentication have been known for recorded human history. Here we focus on their mechanical realization: a physical lock that latches to keep some people out of important boxes, doors, or structures. Combination locks that open after the entry of a secret have appeared from antiquity. Recovered artifacts like the Antikythera Mechanism [14] show the ability to design and manufacture complex geared machines by the second century BCE. Hoepfner described an example surviving from ancient Kerameikos [15], which opens when two knobs are rotated to the correct position. It contained a provision for the owner who may have forgotten the combination: a hidden additional means of pressing the latch, besides entering the unlock code.

Considerably more sophisticated combination locks with longer combinations were available from at least the 12th century. A notable example is dated 1197 and attributed to Muhammad al-Asturlabi. This device is held today in the Museum of Fine Arts in Boston [16] and notably requires the entry of a secret code of length 8 drawn from an alphabet of 16. The Arabic letters are laid out in their Abjadi numerically ascending order, so these locks easily accommodate PINs or passwords. Two later examples of al-Asturlabi's locks keeping the same basic layout survive today in museums. In a widely-distributed manuscript attributed to Ismail al-Jazari published in 1206 [17], a diagram is reproduced that depicts the same internal lock mechanism used by al-Asturlabi's locks along with the alphabet of 16, but the dia-

gram shows a secret code of length 12 instead of 8 in the artifacts. Neither of these designs accommodate an owner who has forgotten the PIN/password.

Although the manuscript's author is silent on this point, one can only imagine that no 12-character locks survive today from ca. 1200 since they were destroyed by owners who could not remember the secret codes. In truth, we cannot know today if the artifacts influenced the design in the manuscript or perhaps the manuscript describes a type of lock that was widely used, but we do know that al-Jazari's diagram influenced generations of system designers including in Europe [17]. Moreover, we can see that while lock designs have long since improved, discussions on PIN and password lengths, including the challenges around memorability, continue to the present day, some 800 years later.

Military use of PINs and passwords for authentication has also been handed down from antiquity [18]. In the twentieth century, U.S. atomic weapons used literal mechanical combination locks to prevent explosive or fissile materials from coming together. Bellovin reports that this practice continued into the 1980s and that in 1981, more than half of the so-called "Permissive Action Links" (PALs) in use were still mechanical combination locks with various configurations and combination lengths [19].

It has been reported that the code to enable at least part of a U.S. Minuteman atomic missile launch sequence was set to 00000000 [19] to aid usability by Air Force crews facing an emergency. This characterization has been disputed by U.S. Air Force officials and we cannot be sure of the veracity or practical impact of this claim as of this writing [20]. What is clear from an examination of the available sources is that there is at least some evidence for this claim in field service manuals, and military commanders had concerns about on the security-usability tradeoff.

KBAs necessarily represent a usability tradeoff, and these difficulties have long been observed in Information Technology. In a classic study from 1999, Adams and Sasse identified issues including a basic lack of awareness on the part of users as to what constitutes a "good" password along with the sheer number of passwords users must remember [21]. Although many systems and proposals exist to incorporate software and hardware into authentication protocols, system designers are still left with a need to verify the "right" person and not a thief intends to use the software/hardware. Even though a modern protocol like WebAuthN aims to improve upon the problems of passwords on the scale of the Web, it still relies on a "User Verification" [22] step involving a PIN, password, or biometric.

Proposals to eliminate KBAs have come and gone, and yet PINs and passwords are still very much in use. The difficulty of migrating away from KBAs has been studied in 2012 by Bonneau et al. [23], who presented a framework of requirements that a Web authentication scheme must satisfy. After evaluating twenty years' worth

of alternative schemes from the literature and industry, passwords still provided an unmatched set of benefits.

2.1.2. Online Setting

Since at least the 1970s, computer systems have required a KBA before use. That is, a piece of software is responsible for controlling access with the (presumed) user tasked with creating and recalling the KBA. This piece of software is obviously responsible for taking KBAs as input and producing an output of True or False. To do this, it must also be able to check against a known-good version of the user's password.

In this thesis, we are focused on how hard it is for attackers to guess passwords and PINs: our attackers are *online* or *UI-bound* and have an objective to correctly guess a PIN or password to log into a device or service provider. The attacker makes a *guess*: a tuple (u, p) sent to a verifier as part of a protocol. Unless otherwise noted, our attackers are also *trawling* in the sense of being satisfied to gain access to any account(s) on the system. Contrast this situation with a *targeted* attacker who wants access to one particular user's account.

In this situation, our attacker has an obvious winning strategy: simply try all possible passwords. To offer some protection, the system must also implement a limitation on guessing. Online, this is often done by either a throttling mechanism to slow an attacker trying all possible KBAs, or simply locking an account after a certain number of bad guesses.

If users chose truly-random long unique passwords, this arrangement would make UI-bound attacks infeasible. Instead, as early as 1979 it has been observed [24] that users tend to choose weak passwords that are susceptible to so-called *dictionary attacks* where the attacker knows that certain passwords are more likely to be chosen by users. By making use of such a dictionary, the attacker increases their chance of success before the system slows or stops the acceptance of new guesses.

Especially with PINs, this lockout or throttling behavior is directly responsible for any security provided. Without the throttling safeguards, an attacker can very easily run through all possible 4- or 6-digit PINs. These basic facts about the need to limit KBA guessing have long been known from the realm of payment cards. An excellent survey and history of PINs used in payment cards appears in the work of Bonneau et al [25]. Systems obviously vary, but generally less than 10-100 guesses are accepted before the online service/mobile device either slows or locks out further guesses. As memorably observed by Florencio et al. [26] beyond this point lies a large "online-offline chasm" for guess counts out of reach for an online guesser, but still less than the 10^{15} guesses available to a moderately capable offline attacker not bound by the UI, who can independently make a guess and check its validity.

2.1.3. The Mobile Device Setting

Mobile devices bear some similarities to our situation in the online setting. Here, the attacker makes a *guess*: a value p sent to a verifier as part of a protocol. In the mobile device setting, the username is generally omitted. But otherwise we again have a verifier that can accept guesses, and will respond by either slowing the acceptance of new guesses or locking the device completely.

Given their ubiquity, we will also treat the special case of 4- and 6-digit PINs on mobile devices. Obviously, research on PIN authentication for mobile devices is related to the larger area of mobile authentication. User preferences for different unlock methods for Android devices were studied by Harbach et al. [27] in 2014. It would be possible for users to select and enter alphanumeric passphrases as their KBA, but very few do so as reported by Harbach et al. and our own work. It is not difficult to see why: given the number of situations in which a mobile KBA is entered, usability suffers with longer or more-diverse PINs.

Biometrics are quite popular for mobile-device unlock, but their use still requires a KBA on the device. For one thing, the biometric might not be read properly, so devices allow a user or attacker the option of using the biometric or presenting the KBA. Moreover, KBAs have found new uses in encrypting data held on mobile devices [28–30]. This usage means devices require a PIN as part of the keying material used when a device reboots. Even beyond these uses, as of this writing Apple for example just implemented *Advanced Data Protection* (iOS 16.3, 13 Dec 2022) worldwide which relies on the device passcode (PIN or password) to encrypt data backed up to iCloud [31].

The fact that users are aware of their insecure behavior when asked about it was shown in a study by Harbach et al. [27] where most participants reported that unwanted access to their smartphone would have been possible. Similarly, Mahfouz et al. [32] observed an average auto-lock timeout of 65 s in their study, enabling an attacker to access a smartphone if the owner leaves it unattended and does not lock it manually. To mitigate such risks, Kraus et al. [33] recommended offering simple mechanisms which are secure by default. This could also prevent potential social pressure toward bad security behaviors.

Matthews et al. [34] presented a taxonomy of device-sharing scenarios while Alabayram et al. [35] highlighted risk awareness as a driving aspect for insecure behavior rather than inconvenience, suggesting the need to effectively communicate risks. A follow-up study with users in Saudi Arabia by Al Qahtani et al. [36] confirmed these results. We will continue this theme by showing that users of the Signal app with increased comprehension of the PINs' purpose selected more diverse PINs.

Shi et al. [37] presented a two-fold threat model for mobile authentication based on an informed and uninformed stranger. Muslukhov et al. [38] extended Shi et

al.’s model with capabilities such as physical access and shoulder surfing. They conducted a survey of threats perceived by smartphone users and find that participants were equally concerned about strangers and “insider” threat actors such as friends. Moreover, 12% of the participants in their study had experienced someone accessing sensitive data without their permission; 9% had “sneaked into the smartphone” of someone else in the previous year.

Levy et al. [39] further explored privacy threats in intimate relationships like families or partnerships. They categorized attackers and victims based on their relationship and identify features of threats across these relations. At this point, Levy et al. noted that the attacker’s motivation may be based on multiple factors, and they are not inevitably intended to cause harm.

Marques et al. [40] quantified the prevalence of snooping on mobile devices in a survey designed to minimize social-desirability bias. The survey defined snooping as “looking through someone’s phone without their permission,” finding 31% of participants had done so in the previous 12 months.

In subsequent work, Marques et al. [41] explicitly recruited participants who have past experience with unauthorized mobile device access. They found that distilling stories of unauthorized access into identifying the familiar who, what, and why categories led to interesting insights. Participants felt that making themselves vulnerable to unauthorized access was necessary to sustain relationships with friends, partners, co-workers, and others. In explaining their past experience with unauthorized access, participants rarely blamed themselves, instead blaming circumstances or the other person’s shortcomings.

In this thesis, when we consider mobile attackers, we are primarily interested in attackers who:

- Gain brief physical access to an unattended smartphone locked with a 4- or 6-digit PIN.
- Have no hints to rely upon, such as smudges on the screen, shoulder surfing or targeted personal knowledge about the victim.
- Have no malware or vulnerability that would allow them to bypass the lock screen without the PIN.
- Have a limited number of guesses before the device slows or stops its acceptance of PIN guesses.
- Know the PIN length, because devices like iOS provide a visual hint on the lock screen.

2.1.4. Measuring Guessability

Because this thesis focuses on guessability, it is helpful to review some basic definitions [42]. Likely the most influential notion of information entropy is *Shannon entropy*, introduced in seminal work by Claude Shannon in 1948 [43]. For a discrete random variable X with finite domain $D = \{d_1, \dots, d_n\}$ and $p_i := \Pr(X = d_i)$, *Shannon entropy* is defined as

$$H_1(X) := - \sum_{i=1}^n p_i \log(p_i).$$

In general, Shannon entropy is most relevant in the context of compressibility of data, as it can be misleading when it comes to the specific case of PIN/password guessability. Massey [44] showed that there are (artificial) distributions with low Shannon entropy and high guessing entropy. In addition, he also proved that high Shannon entropy implies high guessing entropy, see [45] for more bounds.

Another useful notion of entropy is *min-entropy*, which is a measure for security under a specific subcategory of online guessing attacks called “one-guess attacks” [46]. Applied to our application, min-entropy identifies the most-common PIN/password. For a random variable X and monotonically decreasing probabilities $p_1 \geq p_2 \geq \dots \geq p_N$, min-entropy is defined as

$$H_\infty(X) = -\log_2(p_1).$$

Min-entropy only takes into account the most frequent event, so it is not a good estimation for most password distributions. For an idealized version of an *online website attack*, min-entropy turns out to be the right notion: In an online website attack, the adversary has a virtually unlimited number of accounts at hand, assuming that they can easily guess new usernames. Then, the most efficient strategy to maximize their success is to guess the most likely password for each account, and this is exactly what is addressed by min-entropy. Observe that in the case of mobile devices, which we will study in some detail, the attacker generally can only guess against one account. Additionally, in reality the attacker gets a small number of guesses before the device throttles (delays) additional guesses or locks entirely. Therefore, we stress again that these notions of entropy can be highly misleading: despite their theoretical foundations, they only apply to an unthrottled, perfect knowledge attacker that will exhaustively guess the PIN/password space.

Both Shannon entropy and min-entropy are special cases of R enyi entropy [47], which is defined as

$$H_n(X) = \frac{1}{1-n} \log \sum_{i=1}^n p_i^n$$

for $n \geq 0, n \neq 1$. For $n \rightarrow 1$ this converges to Shannon entropy, for $n \rightarrow \infty$ this converges to min-entropy.

Guessing entropy [44,47] measures the expected number of guesses that the optimal attack needs in order to find the correct password. For a random variable X with countable domain D and $P(X = d_i) = p_i$, ordered with decreasing probabilities $p_i < p_j$ for $i < j$, *guessing entropy* $G(X)$ is defined as

$$G(X) = \sum_{i=1}^{|D|} i \cdot p_i \quad (2.1)$$

which is the expected number of guesses to guess a password. However, a practical attacker is generally satisfied with breaking into a certain fraction of accounts, which guessing entropy does not take into account. For this usage we turn to *partial guessing entropy* [48] (or α -*guesswork*) takes this into account.

For $0 \leq \alpha \leq 1$ let

$$\mu_\alpha = \min\{i_0 \mid \sum_{i=1}^{i_0} p_i \geq \alpha\}, \quad (2.2)$$

be the minimal number so that the guesses cover at least a fraction α of the passwords, and let

$$\lambda_\alpha = \lambda_{\mu_\alpha} = \sum_{i=1}^{\mu_\alpha} p_i \quad (2.3)$$

be the actual sum (which is greater or equal to α). With these, partial guessing entropy is defined as

$$G_\alpha(X) = (1 - \lambda_\alpha) \cdot \mu_\alpha + \sum_{i=1}^{\mu_\alpha} i \cdot p_i \quad (2.4)$$

Intuitively, the first term is contributed by those passwords that weren't guessed in the allotted number of guesses, and the second term is contributed by those passwords that were. We want to express this in “bits of information” to be able to compare it with other measures more easily. This is done as follows:

$$\tilde{G}_\alpha(X) = \log \left(\frac{2 \cdot G_\alpha(X)}{\lambda_\alpha} - 1 \right) + \log \frac{1}{2 - \lambda_\alpha} \quad (2.5)$$

where the “correction term” $\log \frac{1}{2 - \lambda_\alpha}$ is used to make the metric constant for the uniform distribution (see [48] for a more detailed explanation).

2.2. Attacker Models

No practical system resists all attacks. Instead, we focus on attackers with specific capabilities. Doing so means adopting a specific set of assumptions about the system and the attacker. Fundamentally, in our work we explore the ability for attackers who try to guess KBAs. Moreover, this guessing is limited by the system under examination to prevent trivial exhaustive search. Of course, system designers may need to consider other threat models to adequately protect their real-world systems. In later sections, we will make this general theme much more concrete; we found for example, that an attacker who is limited to only 10 guesses is more effective at guessing 6-digit PINs than 4-digit PINs, while a guesser with 100 guesses is more effective at 4-digit PIN-guessing. See Section 4.4.1 for more details; this finding underscores the importance of careful threat modeling.

2.2.1. Online Trawling Guessing Attackers

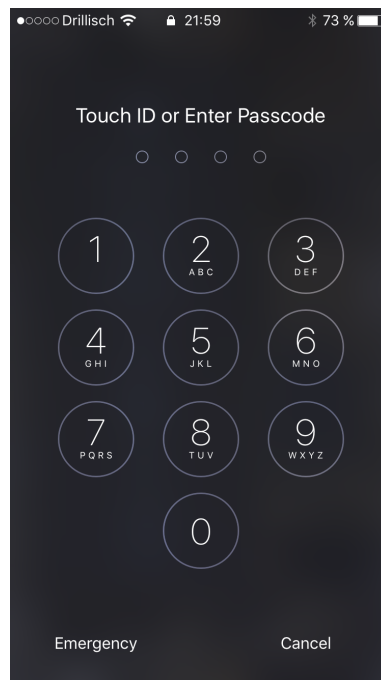


Figure 2.1.: iOS lock screen shows the number of digits in a user’s PIN

Our attack model is *trawling*. This model means the attacker has no personal knowledge, “hints” about the account owner, or physical side-channel measurements that might make guessing easier. One advantage of this model is simplicity, allowing it to act as something of a lower bound on guessability. Another advantage is its conceptual accessibility: that an attacker could try to guess someone’s PIN or password is understood by practically all users. That a phone could be snatched

out of someone’s hands in public, or left behind in a taxicab is also well-known. Moreover, systems have long employed measures like blocklists in an attempt to get users to pick better PINs; one of our main contributions is to analyze how effective these countermeasures are. Obviously, any personal hints can only serve to help the attacker. Our advice to system designers can therefore represent a minimum viable security objective — a real-world attacker will do no worse at the task of guessing.

Another setting optimized to resist online guessing attackers can be found in the context of *Chip-and-PIN* systems. Bonneau et al. [25] considered 4-digit PIN creation strategies for banking customers for use with ATMs/credit cards. Bonneau et al. identified techniques adopted by users for selecting PINs, where choosing (birth) dates/years was the most popular strategy—also true in our setting. An attacker can leverage the skewed distribution of PIN choices to improve their guessing strategy, and precisely quantifying this advantage in certain circumstances is one of our main contributions.

The distribution of PINs has received previous study. As noted, PINs are effective only in the case of an online guessing attacker, since an offline attacker can easily attempt all possible PINs. This means the distribution of PINs is of vital importance. Wang et al. [49] have also analyzed the distribution of PINs – in this case without any specific usage context. They report on comparing 4- and 6-digit PINs created by English and Chinese users. One counter-intuitive finding is that 6-digit PINs are less resistant to online attacks, despite the key space expansion from 4- to 6-digit PINs. Our results support the observation that in a rate-limited guessing scenario there may actually be no benefit of using 6-digit PINs at all and in certain cases security even decreases. Wang et al. used PINs extracted from leaked, text-based password datasets whereas we tend to increase the ecological validity of our results by collecting new PINs specifically primed for mobile authentication and the smartphone form-factor with its standard PIN layout.

Blocklists have been considered in the context of PINs by Kim et al [50]. They tested blocklists for both 4-digit as well as 6-digit PINs, and concluded that a reasonably-sized blocklist could indeed increase the security. Kim et al. used Shannon entropy and guessing entropy as the strength metric and thus only consider an unthrottled, perfect knowledge attacker that will exhaustively guess the PIN space [45]. This is a questionable attacker model especially given the sparsity of their dataset. Kim et al. compared blocklists representing 2% and 32% of the possible PIN space and found the large blocklist led to lower Shannon-entropy and lower offline guessing-entropy PINs, perhaps due to the composition of Kim et al.’s large blocklist.

In contrast, we show in our analysis of 4-digit PINs that with a more realistic rate-limited, online attacker, a larger blocklist containing 27.4% of all possible PINs provides a benefit over a smaller one that blocks only 2.7%, differing from the sug-

gestion of Kim et al. regarding the effect of the size of the blocklist. We also make similar observations in our analysis of 6-digit PINs.

Bonneau et al. [25] also proposed the use of a blocklist containing the 100 most popular PINs. From our analysis, it seems that their suggestion may have formed the basis for Apple iOS’s 4-digit blocklist.

Our work differs from Bonneau et al. in two significant ways. Foremost, Bonneau et al. were primarily concerned with payment cards, not smartphone unlock authentication. Second, Bonneau et al. did not collect new PINs but instead relied on digit sequences found in leaked passwords along with PINs collected without the benefit of a controlled experiment [51]. Our work aims for greater ecological validity by specifically priming users for this task. Our data suggests that using password leaks may be an imperfect approximation for how users choose PINs for unlock authentication.

2.2.2. Attacks Beyond Our Scope

We emphasize that many other attack types besides guessing are possible that are beyond the scope of the present work. We will summarize some of these briefly in order to draw out the contrast. An attacker could simply intercept a tuple (u, p) over the network (if unencrypted) and steal it that way, or convince the user to hand it over as in a *phishing* attack. Of course, an attacker might conduct an attack using a software vulnerability in an online service and therefore not need to submit any guesses. Naturally, the compromise of a user’s account with an online service like iCloud that stores unencrypted data, would have many of the same deleterious effects as guessing a smartphone’s PIN.

Closer to our area of inquiry is the fact that smartphone lockscreens may have software vulnerabilities allowing them to be bypassed. One of these appeared in June, 2022 as Schültz discovered a peculiar set of steps involving swapping SIM cards on an Android phone that exposed a flaw in the state machine controlling the lockscreen [52]. Using this flaw and any SIM card owned by the attacker, the attacker could unlock the smartphone — no guessing needed.

Obviously the flaw found by Schültz requires physical access to the smartphone. An attacker who has physical access has other options as well. Complex systems have physical characteristics that can be measured, as in a side-channel attack, and we stress that any additional information the attacker has is virtually certain to help their guessing performance.

For instance, Abdelrahman et al. [53] employ thermal imaging attacks on PINs and Android patterns. After entering a PIN or pattern, the portions of the screen touched by the user are slightly warmer. Abdelrahman et al. found that thermal attacks are indeed viable on mobile devices, with PINs being significantly more susceptible than patterns that contain overlaps. For another example of a side-

channel attack Zarandy et al. [54] showed that voice assistants can be used to extract a PIN by collecting acoustic signals of the user typing. A hardware power-analysis side channel was discovered in iPhone 4 (and possibly up to iPhone X) by Lisovets et al. [55] that allows an attacker to conduct an offline search for the PIN that can be run in parallel. A timing side channel was discovered by Haas et al. in the Apple A10 Fusion chip first introduced in 2016 with the release of the iPhone 7 and iPhone 7 Plus [56]. There are many more side channels possible but these are out of scope for the present work. One can readily see that additional signals serve only to help the attacker. Any of these attacks could be combined with our work on guessing.

We focus on passwords for online services and PINs for mobile devices. Beyond PINs, another common knowledge-based mobile authentication mechanism is the Android unlock pattern, whereby a user selects a secret pattern that connects points on a 3x3 grid. Uellenbeck et al. [13] showed that user selection of unlock patterns is highly biased: most patterns start in the upper left corner giving a resulting construction that is weaker than a PIN. These results have been confirmed by other works [57–60]. We will compare the security of mobile unlock PINs to that of patterns and have obtained datasets from the related work [13, 57–59].

As noted above, we focus in this thesis on trawling attacks. Of course, there are also scenarios where the attacker has information about the victim, like if their birthday or anniversary is known [61, 62]. In other cases, the attacker obtains information about the victim’s PIN via shoulder surfing [63–69], or smudges on the screen [70, 71].

In contrast to our attack model, an *offline* attacker is limited only by available computation. In this attack setting, typically as the result of a SQL injection attack on an online service, an attacker has a list of hashed passwords. Assuming the cryptographic hash algorithm is one-way, it is infeasible for an attacker to directly invert the function. Since passwords are selected in a highly non-uniform manner, the attacker can prioritize guessing higher probability passwords. This problem has been studied extensively and led to development of automated tools such as *John the Ripper* [72] and *Hashcat* [73] which combine password dictionaries with “mangling rules” to guess passwords in a heuristic priority order. More advanced password guessers based on Markov models have also been presented [74, 75] to list only a few. These approaches could be used to generate guesses in an online attack, but observe that because smartphones and services can either slow or stop accepting guesses, the first few guesses are crucial. In this dissertation, we are primarily concerned with understanding the attacker’s success rate in these first few guesses, before the mangling rules are applied.

For our purposes, the crucial distinction is that in an online attack the attacker cannot independently check if a guess is correct: they need to rely on a device or online service. In an offline attack, the attacker has a leaked password file with

hashed passwords or other means to check if a guess is correct. An online device or service can decide to stop or delay accepting any new guesses. Especially with PINs, this lockout or throttling behavior is directly responsible for any security provided. Without these safeguards, an attacker can very easily run through all possible 4- or 6-digit PINs. While it is possible to optimize the guessing order for an attacker who has thousands of guesses available, we are unaware of a practical setting where this would matter.

2.3. Re-Using Authentication Secrets

We will also investigate PIN/password reuse as a recurring theme throughout this dissertation, both through direct observation and survey questions.

2.3.1. Users' Tendency to Re-Use

Several previous studies have examined password reuse. Ives et al. give an interesting high-level overview of password reuse [76], including some examples of actual damage done by password reuse. Florencio and Herley [77] present a large-scale user study on passwords including password reuse, where they collected their data from browsers running the Windows Live toolbar (from consenting participants). They find that each user has, on average, 25 accounts and 6.5 passwords. In other words, each password is used for 3.9 accounts. This study has some limitations: First, as they hash the stored passwords for security reasons, they can only detect exact password reuse, not reuse of very similar passwords. Based on the design of the study, it is difficult to detect if a password was entered at the same site under two different URLs.

In a lab study, Gaw and Felten ask participants to conclude when groups of passwords are similar [78]. This approach is adopted to preserve confidentiality of participant passwords, but the resulting similarity measure is vague. They find between 2.2 and 3.2 accounts per password.

Bonneau [79] used two password lists that both included usernames, allowing reuse measurement between these two sets. Both lists were hashed, so the hashes first needed to be cracked. From those accounts cracked in at least one list, 49% of users used the same password for accounts on both sites, however, this does not take into account those accounts that weren't cracked, and thus we cannot say what the actual reuse rate is. It may be that those passwords that weren't cracked belong to more security-savvy users and that those have a lower rate of password reuse, so 49% most likely constitutes an upper bound. Furthermore, in the same text Bonneau recognizes the need for a study on password reuse based on *account value*.

Of the 456 common users, 161 had their password cracked in both datasets, 46 only had their rootkit.com password cracked and 77 only had their Gawker password cracked, leaving 172 with neither password cracked. Of the accounts for which passwords were cracked at both sites, 76% used the exact same password. A further 6% used passwords differing by only capitalisation or a small suffix (e.g. password and password1?).

An industry advisory [80] considers password reuse by utilizing a browser plug-in intended to warn about phishing attempts against banking passwords that also detects reuse. They report that “73% of users share the online banking password with at least one nonfinancial website” [80]. However, not many details are given about the exact setup and distribution of the plug-in. In addition, to compare the results with other work we would require at least the average number of accounts per user they recorded.

Perhaps as a result of its inclusion in PCI-DSS [9], forcing users to periodically change their passwords remains a common technique to prevent attackers from using leaked passwords. Besides conflicting standards, it should be clear from this discussion that many factors go into a given website’s password policy. Sahin et al. identify others including usability, requirements from auditors, and organizational inertia [81].

Zhang et al. use a database of 7700 accounts to examine the difficulty in guessing the replacement password given the expired one [82]. They found in this attack model that 41% of replacement passwords could be guessed in a few seconds. More recently in a similar attack model applied to PINs, Munyendo et al. [83] found that mostly, if an attacker knows a user’s 4-digit PIN, a replacement 6-digit PIN is easily guessed.

Our study on password reuse first appeared in print in 2014. In the intervening years, this topic has seen a great deal of additional attention. For example, the service called haveibeenpwned.com (HIBP), started by Troy Hunt in 2013 [84] has matured into a resource on the web for users to find if their account has been part of a breach. Leaked password lists from Adobe, LinkedIn, and others contain username (typically email address) and password hash. HIBP stores username-site name pairs so that users can check if their account data was leaked. Separately from the username-site name pairs, leaked passwords are stored so that, using an online challenge-response protocol, users can check if a given password has been previously leaked [85]. HIBP stores the hashed password separately to partially insulate from being an unwitting accomplice in further attacks. Though HIBP cannot be said to be completely privacy preserving, it serves a useful purpose in notifying users of breaches. Otherwise, the only users of this data would be criminals trading or selling this data. The security community generally appears to have accepted the trade-off. To see this,

one needs only to observe the number of tools including password managers and browser extensions that make use of HIBP or similar functionality. Thomas, et al. [86] report on one such system implemented as “Google Password Checkup” that is now available to anyone with a Google Account.

Since the first publication of our work, several studies have appeared in this vein. Wash, et al. [87] in 2016 combined survey methods with a browser plug-in that 122 participants used for six weeks. Their study also found that more-complex passwords were more likely to be re-used, but were additionally able to show that the frequency of a given password’s entry was a stronger predictor of re-use likelihood.

A study from Wang et al. [88] examined the patterns of password re-use and modification among a large dataset of millions of users across 107 services. Fully 52% of the users engaged in password re-use. Interestingly, even after a data breach, users continue to use the same leaked passwords for other online services. The study further showed that more than 16 million password pairs, including 30% of the modified passwords, can be easily cracked with just 10 guesses.

On the topic of PIN re-use, recently, Khan et al. [89] and Casimiro et al. [90] studied PIN re-use across different contexts. Both find that re-use is rampant, and that users tend to have a small set of PINs they use regularly. In our work we also find that certain kinds of PIN re-use are common, such as for an ATM/Credit/Payment card.

2.3.2. Responses to Re-Use

Understanding the exact motives that lead to the observable differences both in password strength and password re-use is important. A reasonable method seems to be user interviews, which also might inform efforts to influence users towards better behavior, i.e., choosing strong passwords for those accounts that have high value, and to re-use only those passwords that have low value or are sufficiently protected on the server.

Many large services now offer some form of message to the user about re-use. In particular, it is common for the largest sites to monitor underground forums and the like for appearances of their users’ passwords. If found, the site will send a password-reuse notification. A 2018 study by Golla et al. [91] studies the approach of providers notifying users that their KBA has been compromised elsewhere. The notifications themselves are troublesome. The study found that only about 20% of participants understood what is something of a complex situation: that the breach of this password took place on another service. Only 18.8% mentioned password re-use as a contributing factor. It therefore remains an active area of research to find interventions to reduce re-use.

An important tool to prevent password re-use is the password manager. This type of software tool does not require changes to how servers prompt for and accept KBAs. In addition, they are now available in most Web browsers by default as well as a host of third-party add-ins. These use a master password to encrypt a database (called a “vault”) of actual site passwords, decrypting them and replaying them to sites as needed. Many studies have appeared on this topic. For example, Lyastani et al. [92] report on an online study with 170 participants using a browser plugin to examine usage patterns. In early 2018, the publication date of that study, Chrome did not yet create new passwords for users as needed, but merely auto-filled existing choices. That approach tended to exacerbate the problem of re-use with more than 80% of passwords auto-filled by Chrome re-used. This situation was addressed in Fall of 2018 when Chrome 69 implemented the ability to create new passwords [93].

Password managers are certainly a welcome development as they can generate and store into a “password vault” KBAs that are difficult to guess. Given that most users access the Internet on multiple devices, most password managers additionally offer a cloud service that can synchronize a user’s password vault across all their devices. The appeal for a user is clear: they need to select and recall only one “master password” to decrypt the vault, rather than a password for each site they access. On the downside, attackers now have a new target: a cloud service that holds all of a user’s passwords. A successful attack on the cloud service would have very grave consequences, as they could now start an offline guessing attack to find the master password.

Unsurprisingly, we have seen a string of exactly these attacks online targeting the LastPass service. In 2015, hackers gained access to user email addresses, password reminders, and encrypted master passwords. The breach affected all users of the LastPass service, and the company recommended that all users change their master password in response, and enable two-factor authentication. LastPass took steps to enhance its security measures, including increasing its infrastructure security and implementing scrypt to hash master passwords [94]. More recently in 2022, a similar incident transpired in which attackers once again gained access to encrypted vaults [95]. It bears repeating that the “last line of defense” in this setup remains a KBA. If a user’s site password leaks, and they re-use that password for their vault, the attacker can easily compromise all of the user’s online accounts. Research into KBAs and re-use is therefore of vital importance to the overall online economy.

2.4. Ethics in Knowledge-Based Authenticator Research

A central theme in KBA research is: “what do users pick for passwords/PINs in various situations?” Research in this area raises a few obvious ethical issues. As

researchers we aspire to discover information that is of benefit to the computing community. From previous work we know that the probability distributions of passwords and PINs are skewed. Humans are simply not well-equipped to generate, retain, and recall truly-random numbers. Understanding this problem is crucial to devising advice for system designers. Among other objectives, KBA research aims to quantify the ground truth under various conditions. This goal is in natural tension with the constant exhortations to users not to reveal their KBAs to anyone.

We stress here that this is for good reason: though we intend to not cause harm, collections of KBAs are a bit like lockpicking tools: having dual uses in understanding the limitations of locks, but also facilitation of theft. Cybercriminals already have sophisticated tools, value chains, and networks to distribute credentials all on their own, without needing the help of researchers. Still, it's possible that a researcher would violate that trust and use the KBAs themselves or resell them for profit. Similarly, it's possible that an attacker would steal credentials from where researchers have stored them. As we have seen in Section 2.3.2, even well-funded security vendors in industry have difficulty keeping hackers at bay.

Each of the main chapters of this thesis grapples with these questions in their respective Ethical Considerations section. We examine three main types of leaked credential sets used in this thesis, each with unique characteristics.

2.4.1. Taxonomy of Password Lists

We refer in this section to “password lists,” as no comparable PIN lists have surfaced as of this writing, save, of course, any numeric entries appearing in an otherwise alphanumeric list. Password leaks tend to arise from a single service with many user passwords revealed, sometimes with usernames. It is not hard to see that these pairs would be valuable to criminals, as they can test these not only on the indicated site, but also on other sites in what's known as a *credential-stuffing* attack. PINs, meanwhile, belong to a different threat model. Since they are validated locally on devices instead of by a remote service, no large-scale databases of PINs have leaked. In addition, these are less immediately valuable to criminals, as even knowing the PIN will not allow access to any data without the device.

Quasi-public sets These have been in wide circulation and have been extensively studied in the literature. In particular, the password leaks from RockYou (2009), LinkedIn (2012), and Amitay's 2011 PIN list [51] and others in no small measure were responsible for kickstarting the ability of researchers to study user password choices at all, and we have made extensive use of them. Most distributions of RockYou contain some personally-identifiable information (PII). As the file itself is of uncertain provenance with no canonical distribution, we cannot know why it contains bits of

what appear to be HTML. Worse still, we cannot be sure that these passwords actually came from RockYou given the uncertain chain of custody. Given the length of time since 2009, we can hope that any affected users would have changed their passwords by now. On top of this, RockYou lacks usernames/email addresses. For those lists that have these, it would be possible, of course to actually mount a credential stuffing attack whereby we automate the testing of username/password combinations on various sites, but this is beyond the pale from an ethical and legal standpoint. This state of affairs brings us to a crucial point: even though these sets are commonly available online, the affected users did not consent to this use of their passwords/PINs. This is true even though the resulting password research has undoubtedly made services safer in general. A major advantage of these quasi-public sets is that they allow other researchers to independently reproduce claimed results, without the need to trust the researchers involved.

Private sets Many research articles, including our own Chapter 3 were made possible by controlled access to a unique and revealing data set. This approach is not uncommon in the literature. For instance, Bonneau et al. [45] used a list of 70 million passwords from Yahoo users, carefully documenting the steps taken to prevent such a list from leaking. Contrasting this approach with for example RockYou and LinkedIn, that were made possible by a SQL injection attack on those particular sites, we can have more confidence that the Yahoo passwords are genuine than those from the RockYou and LinkedIn attacks. These private sets offer an improved *chain of custody* compared with the quasi-public sets. Most sets will not have multiple entries per username. As of its first publication, the material in Chapter 3 was to our knowledge the first report in the literature on a dataset including multiple passwords per user. That fact allowed that study to directly examine the re-use patterns of users. We were able to use this data only by special circumstances that preclude independent validation by others. This dataset was captured only incidentally by a honeypot operated by my then-employer, RSA Security, examining the behavior of the Zeus trojan in an attempt to possibly develop specific countermeasures for its two-factor authentication (2FA) tokens. Access was tightly restricted and the dataset was destroyed shortly after. RSA Security’s Legal department approved this experiment. Although efforts were made to prevent further leakage of this data by sanitizing it and accessing it only through scripts, the affected individuals did not consent to this use of their data. Due to its sensitivity, there was no possibility of sharing this data with other researchers, which harms the reproducibility of these results.

User surveys The third category of dataset we use in this dissertation comes from recruiting participants for online surveys. In this case, participants explicitly gave

consent for this use of their data. Moreover, we have made these datasets available so other researchers may examine them and draw their own conclusions. User study data, though, is not without its own drawbacks. In particular, it is always possible that participants' responses do not reflect their real-world choices suggesting possible difficulty with ecological validity. Here again we face a challenge: we explicitly do not want participants to enter their actual PINs since we do not want to serve as a secondary vector for data leakage. Despite this, substantial numbers (up to 25%) of participants indicate that they use their actual device PIN.

2.4.2. Discussion

Our aim is to discover the ground truth of KBA selection, but much of our data has an uncertain chain of custody. Of course, this is a deep topic for which we can merely scratch the surface. As we have seen, user surveys excel in their careful gathering of participant consent. But even here, participants say they reveal their actual PIN, which in turn is likely to correspond to a birthdate, anniversary, or other important milestone. We used only a random identifier to identify a participant wherever possible.

In our case, private sets can offer unique insights that other datasets do not offer. Because the sets are tightly controlled, they are unlikely to assist attackers. On the downside, because they are secret, reproducibility by others in the community is practically impossible. Others in the community have weighed in on using datasets of illicit origin. Thomas et al. [96] evaluate a number of papers from the KBA literature and find that ethical statements and practices vary widely. To define what constitutes "illicit," Thomas et al. gave the following criteria: an unintended disclosure by the data owner; an unauthorized leak by someone with access to the data or the exploitation of a vulnerability in a computer system. From these criteria we can see immediately that illicit data is not limited to only that obtained through breaking specific laws, but expands further to unintended disclosure, as when data is de-anonymized.

The Menlo Report [97] provides a basic framework balancing benefits and risks to human subjects and society as a whole. It sets forth four ethical principles: respect for persons, beneficence, justice, and respect for law and public interest.

Respect for persons certainly includes protecting the privacy and confidentiality of subjects. Beneficence maximizes the benefits of research while minimizing any harm to participants. Justice is concerned with issues of equity: ensuring that the benefits and risks of research are distributed fairly. Respect for law and public interest involves ensuring that research is conducted legally and following regulations.

Ienca and Vayena specifically treat the question of how to do research on hacked data sets [98] in the context of machine learning. Bringing this framework to bear on future KBA research yields the following guidelines to consider.

1. **Uniqueness:** We know that KBA selection depends on many design factors [99], so researchers should show that the research question under consideration could not be studied using data obtained any other way.
2. **Risk–Benefit Assessment:** Researchers should provide a candid assessment of the benefits of the research as well the risks to data subjects and steps taken to address the threats.
3. **Consent:** In the context of password leaks, informed consent cannot be obtained. This fact simply means the benefits must be compelling and the threats of additional harm minimized.
4. **Traceability:** The chain of custody for password leaks is often obscured. Researchers should provide a thorough account of how the data was obtained. Of course, if doing so would itself cause secondary harm, such as in the case of revealing someone to be part of a vulnerable population, these competing interests must be carefully addressed.
5. **Data Loss Prevention:** The specific technology and process controls that govern the use of data should be documented, with a particular focus on minimizing the data required to answer a given research question, and confidentiality for the rest of the data.
6. **IRB Approval:** An IRB can help formalize and record the results of this process. It can certainly help to overcome bias on the part of the researcher.

These come with some important caveats: IRBs may not understand the technical details of a proposal to protect data, or might not understand the current best practices in data-loss prevention.

Although we have sought and received IRB approval where it was possible, software-development companies in industry rarely have IRBs. Since RUB had no IRB at the time of their publication, our research on password re-use was reviewed by RSA Security’s Legal department instead, while our research on Signal PINs aimed to follow the procedures typical of IRB-approved research. As valuable as IRBs are, they are not necessarily available to all researchers.

In general, the research community needs to improve its ethical approaches. Many research disciplines involving human subjects have existing codes of conduct and perhaps it is time for one treating KBAs.

3

Password Re-use

Contents

3.1. Introduction	32
3.1.1. Related Work	33
3.1.2. Chapter Outline	33
3.2. Malware-derived Dataset	33
3.2.1. The Malware Dataset	33
3.2.2. More Password Sets	36
3.2.3. Ethical Considerations	37
3.3. Correlation of Password Strength and Account Value . .	37
3.3.1. Measures for Password Strength	38
3.3.2. Malware Dataset Password Strength	40
3.3.3. Comparing with Other Datasets	43
3.3.4. Comparing with MtGox	43
3.4. Password Re-use	44
3.4.1. Measuring Re-Use from Random Samples	45
3.4.2. Re-Use Rates Across Accounts	47
3.4.3. Discussion	48
3.5. Conclusion	48
3.6. Author Contribution	49

3.1. Introduction

In this chapter, we focus on the classic setting of authentication to an online service using a username/password combination, and the widespread problem of password re-use. *This study was first published in 2014. This version expands on the original by presenting additional explanatory text, tables, and results that were cut due to space considerations. We will conclude with a discussion of the subsequent work.*

Most online services rely on users to choose passwords for authentication. Conventional wisdom holds that users generally do not choose passwords that are difficult to guess. Several alternatives to passwords have been proposed, but none has found widespread use, as passwords are easy/“free” to deploy, scale to an Internet-wide user-base, and are easy to understand for the users. Alternative technologies have a number of drawbacks: hardware like *smart cards* and *security tokens* can be expensive to procure and manage for Website operators and can be perceived as an impediment to usability. *Biometric* identification systems also require extra hardware, can raise privacy issues, and many biometrics are not secret (e.g., we leave fingerprints on many surfaces we touch).

One important aspect is password re-use: As user accounts proliferate, users are forced to remember more and more passwords that must also remain confidential and hard to guess. In response, users often re-use the same password for multiple logins to keep the number of passwords they have to remember low [100]. When a re-used password leaks, then the security of all accounts using the same password is at risk. Even worse, a rogue service could collect login credentials (typically usernames and corresponding passwords) and test those at other sites, which is hard to detect for the user.

While it is known from leaked password lists that users choose weak passwords on average, there is some hope in the community that users choose stronger passwords for those accounts that are valuable. The question of which accounts have high value is another topic which is out of the scope of this text. We will use financial-related sites as high-value sites, which we believe reflects the intuition of most users. While from a security point of view, email accounts might be at least as valuable, as they are often used as fall-back security mechanism for other sites, it is unknown how many users take this into consideration. (see, e.g., [101]). However, this belief has never been justified with real-world data. Actually, there is very little data available on high-value passwords at all, which is most likely the reason why so little research has been conducted on the topic. However, this question is of importance, as a number of studies in the literature use low-value passwords as input. Arguably, research on password security is most interesting for high-value passwords, as these are most likely the target of actual attackers.

The lack of available data is one of the main problems in password research as, by their nature, passwords are meant to be confidential. For password re-use, most available studies use data collected in user surveys, where great care has to be taken to ensure ecological validity, see Section 3.1.1 for more details. Our data show the type of site influences the password strength chosen by a user – at least, for users of malware-infected PCs. As explained later, we feel our data provides insight into the behavior of average users as well. This work is the first, to our knowledge, studying real-world password data collected by malware. As our dataset consists of passwords “in the clear” and unhashed, we can make certain unique measurements, finding for example, empirical evidence that among people that re-use their password, most re-use it in exactly the same form — without site-specific modifications.

The contribution of this work is the result of a publication in *Security and Cryptography for Networks (SCN '14)* in collaboration with Markus Dürmuth and Christof Paar.

3.1.1. Related Work

A good overview on password guessing can be found in Bonneau’s thesis [48]. With very few exceptions in the older literature, relevant research was conducted on passwords for low-value sites, and it is not known if users choose stronger passwords for more valuable sites.

3.1.2. Chapter Outline

In Section 3.2 we describe our datasets and the preprocessing steps we used. Section 3.3 studies the relation between password strength and account value. In Section 3.4 we study password re-use, concluding with some final remarks in Section 3.5.

3.2. The Datasets

This section describes our dataset along with discussing some important limitations. The dataset has multiple password-account combinations per user and allows us to study password re-use and the relation between account value and password strength.

3.2.1. The Malware Dataset

A username-password combination allows a thief to log into an online-banking account and, depending on further security measures, drain it of funds. Malware such as Trojans specifically target Web browsers and aim to capture the data entered in HTML forms. In its simplest form, an unsuspecting user enters her username and password in a browser, and the malware silently relays the data. Once obtained, the

Table 3.1.: Overview of the password lists

	Abbrev.	Size	Users	PWs/User	Avg. Length
Malware-List					
– total	MW	3,531	1,721	2.05	9.01
– financial	MW-Fin	177	134	1.3	9.1
– rest	MW-Btm	3,354	1,686	2.09	9.01
Mt. Gox (Bitcoin)	BITC	61,020	61,020	1	–
RockYou	RY	32 M	32 M	1	7.89
Carders.cc	CC	5,062	5,062	1	7.59

criminal can redistribute these assets at a profit. Many organizations attempt to monitor this situation, working with law enforcement, alerting affected banks, and publishing reports on emerging threats. To do so, they obtain some of this data for forensic purposes. As the malware captures all of the HTTP POST data, IP address, operating system version and so on can prove to be valuable clues on infection rates and locations. One of these organizations (RSA Security) allowed us limited access to this data. No additional malware output was collected to enable the present work. The dataset contains thousands of passwords captured by the Zeus Trojan in late 2012. Our main result in this chapter compares data about these passwords to previously-leaked lists of passwords. While those lists contain many passwords for a particular site, the malware list contains several passwords for a given user. This aspect of the dataset allows us to additionally measure the number of times a particular password is re-used by a user. We partition the Malware list (MW) into two (disjoint) subsets according to the perceived value to a user.

- **High-value accounts: Financial passwords (MW-Fin)** The first sample includes passwords for accounts at banks, insurers, brokers, and related financial services. An attacker takeover of one of these accounts has obvious financial consequences and therefore heightened risk perception on the part of the user. We selected the accounts by searching the domain names for financial-services related keywords in a variety of languages, as well as a number of known banks. In addition, we manually inspected the domain names to ensure accuracy. This yielded a set of 177 passwords from 95 different domains, however, the number of distinct entities/sites/... is smaller as a single bank may service several domains.
- **Lower-value accounts: Remaining passwords (MW-Btm)** This group includes all other passwords. This sample includes well-known email providers and social networks. This yielded a set of 3354 passwords from 1134 different sites; Facebook is the largest subset with 1163 passwords.

Perceived value of accounts The perception of security risk is known to be subjective and based on several factors including dread of consequences [102]. The compromise of a user’s financial account obviously carries real financial consequences for a user. Malware-promulgating attackers generally aim to take over an online account and drain it of funds – or perhaps to gather enough sensitive personal information to fraudulently apply for a credit card or loan (often called identity theft). We therefore group these financial-site passwords together (similar to [103]). This classification includes sites likely to directly enable transfer of funds including banks, credit-card issuers, stock brokers, and insurers. In addition, we include those housing sensitive information that would enable identity theft such as payroll processors and tax collectors. In fact, other accounts can be quite valuable to attackers as well: email accounts can be used for password recovery, for example. However, for the overwhelming majority of users (except maybe politicians, celebrities, bloggers, and corporations) the compromise of a user email or social-networking account leads to practically no direct financial consequences. A common sentiment seems to be that “Nobody wants to read my private email.”

A potential objection to this approach is that intuitively, restricting the category of high-value passwords to only financial passwords leaves out other valuable passwords. However, we show in Section 3.3.2 that the passwords in MW-Fin are significantly stronger than those in MW-Btm. Even if some high-value passwords (not from financial sites) are still contained in MW-Btm, this means that the real difference is even stronger than we measured. So the error incorporated from this rather narrow interpretation would lead us to underestimate the disparity, reinforcing our main point.

Bias in the dataset There are two potential sources of bias in the dataset: First, we have a subset of the total set of passwords collected by the Zeus trojan, and second, this bigger set could be biased as it is collected by malware and infections are not necessarily uniform across all users. The sub-sample contains a wide variety of sites in many countries and languages, and represents a snapshot of the actual data available to criminals. The passwords in our case were captured by malware, rather than reported in response to a user survey, eliminating a common source of bias in studies. Second, only those users infected by malware are included in our dataset. We feel the results will likely hold true for many other users given the widespread nature and infection methods of Zeus. According to industry reports around 2014, Zeus variants have been observed in the wild on Windows (IE, Firefox, and Chrome browsers), and Android, including one of every 3000 computers worldwide [104]. Most Zeus infections occurred on PCs with up-to-date antivirus software. Zeus spread through email attachments as well as “drive-by infection,” where a user need

only visit a website to become infected, thanks to a malicious JavaScript redirection. These properties to a certain extent dispel the misconception that malware afflicts only unsophisticated or careless users.

Once a PC is infected and a user is browsing the web, Zeus injects HTML that causes usernames and passwords to be silently sent to command and control servers (C & C), see for example Grammatikakis et al. [105] which details exactly the format of the malware strings we analyzed. The malware dataset does not include any captures from MacOS or Linux, which induces some amount of bias. However, Windows represents more than 85% of desktops accessing the Internet in both 2014 and 2022, so the bias due to operating system choice is expected to be small [106, 107].

Furthermore, we expect the comparison of the strength of passwords in MW-Fin and MW-Btm (see Section 3.3.2) to be largely unaffected by these biases, as both lists are sampled with the same bias, and there is no indication that the bias is such that it affects both subsets in a different way.

3.2.2. More Password Sets

To relate our findings to previous work, we compare against several other sets.

- **RockYou (RY)** One of the largest lists publicly available is the RockYou list (RY), consisting of 32.6 million passwords that were obtained by an SQL injection attack in 2009. The passwords were leaked in plaintext, but all metadata like username was stripped from the list before it was leaked to the public. This list has two advantages: First, its large size yields precise information even about less-common passwords; second, it was collected via an SQL injection attack therefore affecting all the users of the compromised service, basically removing sample bias due to user selection. These advantages have made RockYou studies quite popular in the literature, so we use it to compare our findings with previous work.
- **MtGox/Bitcoin (BITC)** Bitcoin is a digital asset/currency. It was created in 2009 and has gained widespread recognition and adoption as a means of exchange and a store of value. Bitcoins can also be exchanged for other currencies. One of the biggest websites (at the time) providing this service was Mt.Gox. The password file containing over 61 000 hashed passwords leaked online in 2011 [108].
- **Carders.cc (CC)** Carders.cc is an online forum where hackers would negotiate stolen assets like passwords and credit-card account numbers. In 2010, Carders.cc was itself subject to a hacking attack that exposed its database of 5,062 passwords [109]. Most interesting about this list for our purposes is the

user population. Unlike general social-networking sites, this one catered to users who are (on average) both technology-savvy and security aware.

3.2.3. Ethical Considerations

All passwords analyzed in this paper were leaked by attacks in 2012 and collected as a side effect of industry efforts on risk-based authentication. The collected dataset included the HTTP POST data for username/password logins from malware-infected PCs. No additional data was collected specifically to enable the present work. This fact means that practical attackers already had independent access to our datasets for more than two years as of the first publication of this work in 2014. It is not expected that the present work aids actual attackers, though the data was retained about two months longer than it would otherwise have been.

Nevertheless, special care was taken to avoid our work leading to a new consolidated source of passwords for actual attackers. The malware passwords (and the rest of the HTTP POST data) themselves were stored in a private enclave segmented away from typical corporate or academic networks. They were only available to researchers through a chain of proxies (jump hosts) with a full complement of firewalls, network monitoring, and data-loss prevention tools meant to stop data exfiltration. Then, direct access was eschewed in favor of scripts that returned only statistics to the researchers. On completion of the original paper, all of the data collected by malware was deleted. Given the sensitivity of this data set, prompt deletion is the better option, but it limits the ability to reproduce the results with, say, a later version of John the Ripper, see Section 3.3.1.

There was no Institutional Review Board (IRB) at Ruhr University Bochum at that time. The author was employed by RSA Security at the time, whose Legal department reviewed and approved this additional use of the data, subject to the limitations set out in this section. Further studies I performed below were all conducted in conjunction with a university that has an IRB. Each of these has its own Ethical Considerations section to speak to its unique issues.

3.3. Correlation of Password Strength and Account Value

One unique aspect of the Malware password list is that it contains passwords for multiple accounts per user, and those are sampled in the same way and with the same bias. A closer inspection reveals that it often contains passwords for accounts that are more valuable than others, which allows us to compare the strength of those passwords. These findings are relevant for several reasons: First, it allows us to test if “users choose more secure passwords for accounts of value,” which is often expressed in the literature when weak passwords are discovered. Second, password

studies are by their nature limited to the available data: collections of passwords from social networks or portals like Yahoo! [45]. By contrast, our study includes passwords directly used to protect financial transactions.

3.3.1. Measures for Password Strength

At a high level, we can distinguish measures that evaluate *resistance against a specific password cracker* (either by directly attacking them, or by using mathematical models to estimate their effectiveness), and approaches that consider the *distribution of passwords*. While the former are motivated by practice and model common attacks pretty well, they depend on the specific software tool and do not necessarily generalize well. The latter are based on mathematical models and thus have a clearly defined meaning and are (in some sense) optimal, but not necessarily relevant for practice.

Entropy measures A number of different entropy measures have been used to measure the security of passwords. For an overview, as well as more details about the one presented here, see Bonneau et al. [45, 48] as well as Section 2.1.4.

We have two reasons to deviate from classical definitions of entropy. First, to approximate the distribution of passwords X (the probabilities p_i) requires a *large sample set size* which is much larger than the Malware dataset; second, one can be interested in getting a more *comparable metric for a specific attack*. Typically the success of actual attacks is measured in “ x passwords were guessed after y guesses.”

Intuitively, we can re-use Equation 2.1, provided we do not assume the probabilities p_i to be sorted according to their numerical value, but sorted according to the order in which a specific password guesser outputs the guess.

John the Ripper A well-known and wide-spread tool for password cracking is John the Ripper (JtR) [73]. JtR uses a number of heuristics that show good performance in practice. It can be configured in a wide range, but in the standard mode of operation it performs the following steps. (i) *Single crack mode*. In the first step, JtR tries items like username and home-directory name both as-is and after simple “mangling” modifications like appending digits or reordering letters. (ii) *Wordlist mode*. JtR comes with a dictionary of 3557 common passwords to try, along a set of mangling rules that are applied. (iii) *Incremental mode*. A mode that can try all possible combinations, sometimes called “odometer mode” to provide some intuition.

We used John the Ripper 1.7.8-jumbo-5, instrumented with an additional patch that logs the number of passwords tried. The Jumbo version supports counting of plaintext guesses as well as hashed passwords. The number of guesses by JtR is often seen as a good approximation for the practical strength of a password. As we

are only interested in comparing the strength of different password lists, the specific choice does not make a substantial difference. As of this writing, the latest version of JtR is 1.9.0. It is not possible to re-run this experiment with the latest release, since the malware list passwords were deleted in 2014. We measured the number of guesses needed for passwords in each of our lists. JtR can run for a very long time generating every possible password of a given length, so for practical considerations, we aborted JtR after a given amount of time.

For the Malware datasets, we ran JtR against every password. As the other lists contained substantially more passwords, we randomly sampled 1024 from each. The BITC list consists of salted, hashed passwords and so required substantially more computation time to check the validity of a guess. The plaintext lists required only the generation of a guess and not the hash. Experimentally, approximately the same number of hashed guesses can be checked in 10 hours of CPU time vs. one minute of CPU time for plaintext.

Experimental entropies We combine the theoretical entropy measure with real-world password-guessing tools to yield what we will call *experimental guessing entropy*. As discussed before, there are two main reasons why we do not use the above measures directly, namely that entropy measures require substantial knowledge about the distribution and thus a large number of samples to approximate it with sufficient precision, and second that the output of guessing tools is specific to that tool and hard to compare with other results.

To calculate the *experimental partial guessing entropy (EPGE)*, we use JtR to determine the proportion of passwords cracked for a given number of guesses (see, e.g., Figure 3.1). We then use these probabilities in Equations 2.4 and 2.5 instead of the optimal attack considered originally. That is, we replace the optimally ordered p_i 's with probabilities from a realistic attack with JtR. Note that the resulting entropy values depend on the guessing tool used, and are in general higher than the true partial guessing entropy, which assumes an optimal guesser. As our main objective is to compare different distributions, the EPGE suffices.

Statistical significance One potential concern about the Malware dataset is that the set is rather small (at least when compared with password lists such as the RockYou list with 32.6 million passwords), which can lead to a higher variance of the results. We used an approach similar to that by Bonneau [45] to determine bounds on these effects. We sampled more than 80 uniformly chosen subsets of the RockYou password list of the appropriate size (3354 and 177, respectively), ran password guessing and entropy estimation (for both $\alpha = 0.05$ and $\alpha = 0.2$) just as for the Malware dataset, and measured the confidence interval for the level 95%. We find that the confidence intervals for a sample size of 177 passwords (as in MW-Fin)

	$\alpha = 5\%$	20%
MW-Btm (n=3,354)	± 0.35	± 1.18
Malware-Btm-rest (n=2,186)	± 1.4	± 2.8
Malware-Fin (n=177)	± 0.7	± 4.4

Table 3.2.: Accuracy of the experimental partial guessing entropy for several success probabilities

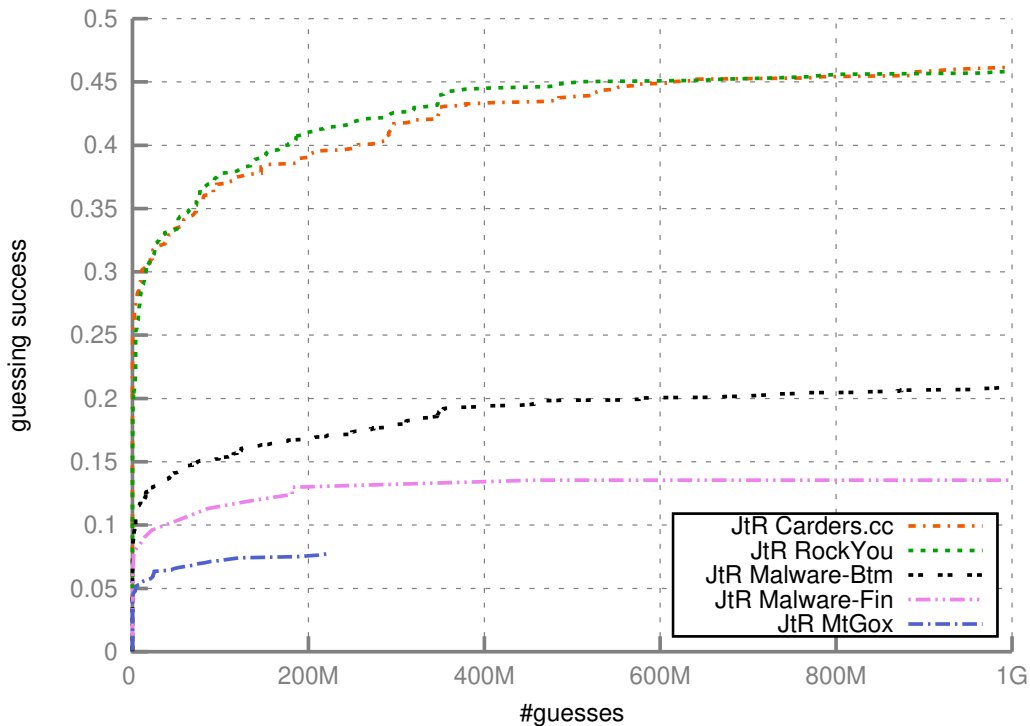


Figure 3.1.: Fraction of passwords successfully guessed when running JtR against various password lists

is ± 0.7 for $\alpha = 0.05$ and ± 4.4 for $\alpha = 0.2$, and for a sample size of 3354 samples (as in MW-Btm) is ± 0.35 for $\alpha = 0.05$ and ± 1.18 for $\alpha = 0.2$.

These (empirical) confidence intervals are determined from another list of passwords that might have different characteristics compared to the Malware list, and thus have to be considered carefully. However, as the differences in entropy that we will encounter later are substantially larger than these confidence intervals, they give us a reasonable level of trust.

3.3.2. Malware Dataset Password Strength

In the first experiment, we compare the strength of the financial passwords (MW-Fin) compared to the others (MW-Btm).

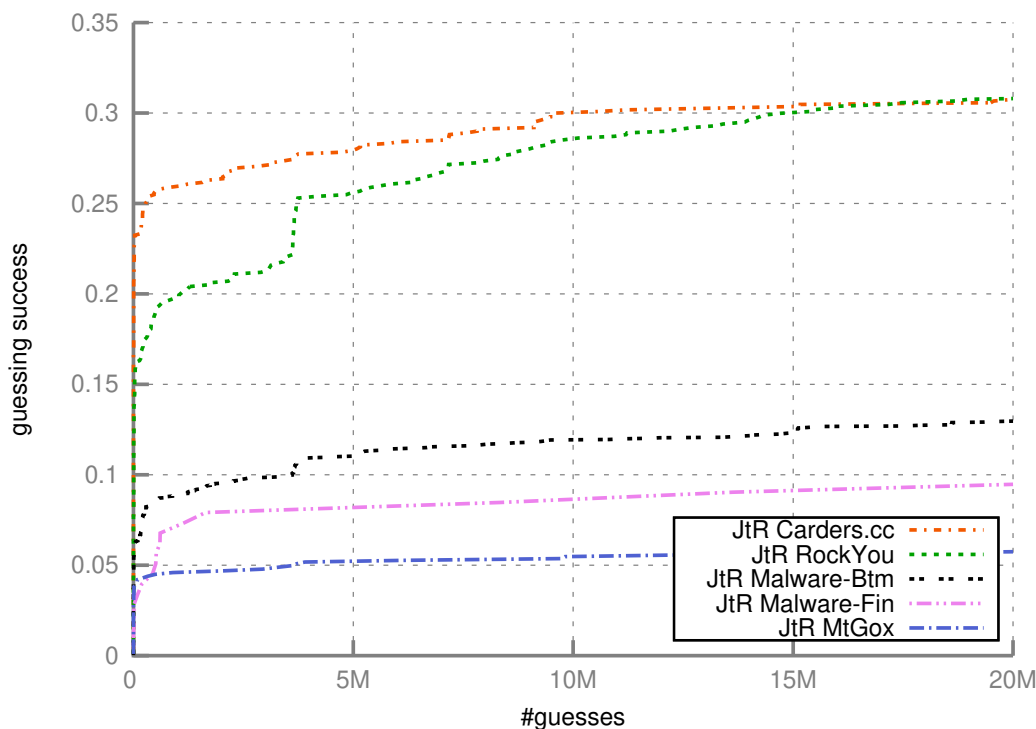


Figure 3.2.: Fraction of passwords successfully guessed when running JtR against various password lists (zoomed in)

Running the experiments We run JtR as described in Section 3.3.1 against the two Malware sub-lists (see Section 3.2), i.e., the Malware list filtered for financial passwords (MW-Fin) and the remaining (MW-Btm), the most interesting and directly comparable set of passwords. All passwords in these lists are available in plaintext, so no hash operations need to be performed and running time is no concern. Note that John the Ripper is highly customizable, with the potential for dictionaries and rules tailored for particular lists. This approach clearly gives the best performance in practice. As our purpose here is simply to *compare* guessing success among the various lists, the default settings will suffice. Our presented results do not reflect JtR’s performance potential in absolute terms, especially considering these reflect JtR’s behavior in 2014.

Figure 3.1 shows the resulting graphs, plotting the number of password guesses on the x -axis and the fraction of accounts guessed successfully on the y -axis, Figure 3.2 shows a more detailed view for fewer guesses. Table 3.3 gives the experimental guessing entropy for $\alpha \in \{5\%, 10\%, 15\%, 20\%\}$ (along with the entropy values for other password lists we will evaluate in the following). From Figure 3.1 one can already see quite clearly that the different lists have different strengths. This is substantiated by the entropy values in Table 3.3, where we see that, e.g., for $\alpha = 5\%$ we get entropies

Table 3.3.: Experimental partial guessing entropy for several success probabilities, using John the Ripper as baseline as explained in Section 3.3.1. A dash means that fewer passwords have been cracked for the respective list, so the respective value cannot be computed from the data at hand.

	$\alpha = 5\%$	10%	15%	20%
RockYou	15.1	15.0	17.4	22.2
Malware-Btm	16.2	25.0	28.8	–
Malware-Fin	23.3	28.6	–	–
MtGox	26.1	–	–	–
Carders	14.4	13.6	13.8	14.0

of 16.2 and 23.3, respectively. From the measurements in Section 3.3.1 we conclude that this difference is significant.

While this result is not surprising, prior to the present work limitations in the lists available to researchers served as a hindrance. This is because the differences may be more due to userbase, differing password policies, or other causes than a specific behavior on the part of a user population. With the Malware dataset and the two subsets MW-Fin and MW-Btm, we are finally in the situation to have several passwords sampled under comparable situations. In addition, we believe that the dataset has less bias than lists obtained by phishing. But even though the data is somewhat biased, both sublists MW-Fin and MW-Btm are biased in the same way, so the results for both are comparable.

One explanation for the difference in password strength could be that different password rules were deployed. This is hard to verify, as the passwords are from a wide variety of different accounts, and there is no efficient method to obtain the password rules that were in place at the time a password was changed. However, we are convinced that password rules do not explain the differences for a few reasons: First, in general password rules are known to be a bad indicator for password strength [110, 111], so we would not expect such a strong impact on password security. Other studies in the literature [103] find password rules are determined more by a site’s need to be usable than the extractable financial value. As we additionally note in Section 2.3, financial sites may be subject to regulation like PCI-DSS [10].

At first glance, we can see substantial differences among the guessing success for different lists: At 20 million guesses, the success rate varies between less than 6% for the BITC list and more than 30% for RY and CC, which is reflected in the experimental guessing entropies.

3.3.3. Comparing with Other Datasets

More interesting insights come from comparing the results for the Malware lists MW-Fin and MW-Btm with other lists of passwords that are publicly available; this also allows us to relate our results to previous research. With the same parameters as in the previous section, we run the experiments again for other password lists: (i) the RockYou (RY) list as examples for a list of weak passwords that is regularly used in the literature that allow us to compare our results with other work, and (ii) the carders.cc (CC) list which represents a list of low-value passwords for a technology-savvy userbase (on average). Again, these lists are available in plaintext, so no hashing is required.

We see that even the weaker passwords in MW-Btm are significantly harder to guess than those in the lists RY and CC (and MW-Fin is even more secure). For $\alpha = 0.1$ the entropy of MW-Btm is 25.0, whereas entropies for RY and CC are below 15.0, and similarly for $\alpha = 0.15$. (For $\alpha = 0.5$ the entropy values are still somewhat similar, which means that the weakest passwords are similarly weak in those lists.) This can also be seen in the graphs in Figures 3.1 and 3.2. (Our estimates from Section 3.3.1 suggest that most differences are significant.)

An additional difference between those lists of weak passwords and the MW-Btm list is that the former contain passwords from a single low-value site only, whereas MW-Btm contains a mix of low- and medium-value (and potentially even some high-value) sites. Another factor that needs to be taken into account is that the Malware list contains data that was collected in 2012, while the RockYou list leaked in 2009. The enforced password rules as well as user's perception of password security have improved over those years, which explains the difference at least in part.

The list RY is regularly used in the literature both as example for weak passwords and as benchmark for work on password security, which might not be an optimal choice in light of our results.

3.3.4. Comparing with MtGox

In a third experiment, we compare our results with the only other list of high-value passwords that was available in 2014, the MtGox list, which is a representative for a list of high-value passwords for another technology-savvy userbase (on average). This list is, however, not available in plaintext, but in hashed form, which is a likely explanation why it has only rarely been studied in the literature. As only some passwords can be guessed in a reasonable amount of time, this results in a sample bias towards weaker passwords. In fact, this is one of the reasons why we use JtR, as we can directly compare results without additional bias. Running time for these tests is substantial. As we need to compute a hash to check the validity of a guess, it takes about ten hours of CPU time to check the same number of guesses as in one

minute of CPU time when passwords are in plaintext. For this reason we only make 345,000,000 guesses per password hash, which limits the resulting graphs.

We can see that the passwords in the BITC list are substantially more secure than those from any other list we consider. For $\alpha = 0.05$, we estimate an entropy of 26.1 for BITC, which is only moderately harder than the estimate of 23.3 for MW-Fin, but substantially harder than all other estimates which fall in the range from 14.4 bits to 16.2 bits. There are two potential explanations for these differences: First, these passwords (often) protect direct monetary value, so users could be inclined to protect that money and choose strong passwords, and second, the userbase of the Bitcoin system and thus MtGox in 2014, were more technology-savvy, and were likely to choose stronger passwords. When additionally considering the CC list, which is the least secure one we tested, the following explanation seems likely: Technology-savvy users might differentiate between high-value accounts (BITC, 26.1 bits for $\alpha = 0.05$) and low-value accounts (CC, 14.4 bits for $\alpha = 0.05$), whereas the average user differentiates less between high-value (MW-Fin, 23.3 bits for $\alpha = 0.05$) and low-value accounts and low-value passwords (MW-Btm, 16.2 bits for $\alpha = 0.05$).

3.4. Password Re-use

Several studies show that users often re-use passwords for several accounts, to decrease the amount of information they need to memorize, recall, and enter. However, re-use can be problematic, because single passwords leak quite frequently, which then puts a number of accounts at risk given the prevalence of re-use attacks in the wild. Even worse, malicious website operators have direct access to a user's login credentials, and misuse will go unnoticed.

However, the studies available so far suffer from two problems: Most work uses *surveys* to answer such questions about re-use, which requires great care to avoid biased data caused by the observer-expectancy effect. Moreover, people might not recall every site where they have registered (see Section 3.1.1 and Table 3.4). As of 2014, we are aware of two studies not using surveys: one [77] uses data that was collected for another purpose and was available only hashed (so similarity or edit distance could not be measured). The other [79] used two leaked password lists that both contained usernames, however, both were hashed and only those could be compared that were broken by a brute-force attack. This fact leads to a bias towards weak passwords, which might also have higher re-use.

A crucial aspect that has not been considered prior to the present work is that the security implications of re-using a password depend on the value of an account/password. (The only exception being an industry advisory [80] with unclear methodology and little explanation.) Re-using a low-value password at another low-value site can

Table 3.4.: Comparing our results on password re-use with previous work. (A dash means that the values are not given/cannot be computed from the data.)

Source	#accts	#pwds	$\frac{\#accts}{\#pwds}$	re-use rate (RR)
Previous work				
Florencio/Herley [77]	25	6.5	3.9	12%
Gaw/Felten [78]	–	–	2.3–3.2	–
Komanduri [110]	–	–	–	27% to 52%
Dhamija/Perrig [112]	10–50	1–7	–	–
Brown et al. [113]	8.18	4.45	1.84	12%
Trusteer Inc. [80]	–	–	–	(73%) ¹
Our work				
RR_0^{all}	–	–	–	14%
$RR_{0.2}^{\text{all}}$	–	–	–	19%
RR_0^{fin}	–	–	–	21%
$RR_{0.2}^{\text{fin}}$	–	–	–	26%

often be seen as a rational choice by the user, as creating, remembering, and recalling a unique password for a large number of low-security sites is practically infeasible. What really constitutes a problem is re-using a password from a high-value site (such as a bank) on a low-value site, as the low-value site is often easier to compromise. Attackers will obtain lists of username/password pairs from low-value sites, and then indiscriminately try to use them on other sites in an attack dubbed “credential stuffing.” We will study this form of re-use in the remainder of this section.

3.4.1. Measuring Re-Use from Random Samples

Previous work on password re-use often gives results as *average number of passwords per user* and *average number of accounts per password*. This is less than ideal, as it does not differentiate between the case where each cluster has the same size, or where the size of clusters is heavily skewed, which can make a big difference in practice. In addition, to make such a statement one needs complete knowledge of the participant’s accounts and passwords. This is problematic because offhand, a participant probably won’t know the exact number of accounts they have. This fact seems to play out depending on survey design and exactly how many aids to recall are given to the participant. In the end, when working with randomly sampled data there is no way to compare the results with confidence.

¹This number is not directly comparable to the other numbers, as they only measured *any* other password matched, which yields (much) higher percentages than the re-use rate.

Therefore, we introduce a new measure for password re-use that we call *re-use rate*. The re-use rate gives the following probability: Choosing a user at random, and choosing two of their accounts at random, what is the probability that the two passwords for the two accounts are identical? As one would expect, a re-use rate of 0 means that no passwords are re-used, and a re-use rate of 1 means that for each user, all passwords are identical. Note that this measure can handle very well the situation when one has access to a subset of a user's passwords, provided that this sample is randomly chosen: Choosing a random password from *all* passwords or from a randomly sampled subset does not make a difference. Hence the re-use rate is a suitable measure for our dataset, where only a random sample of passwords for each user is available.

We are not only interested in exact re-use of passwords, but also in re-use of similar passwords. In practice, tools like JtR implement established concepts like (*normalized*) *edit distance*. The edit distance of two strings s_1 and s_2 is the minimal number of weighted edit operations required to transform s_1 to s_2 . Typical edit operations are *delete/insert/substitute character* (weight 1); we add *prepend/append character* (weight 0.5) to approximate JtR's mangling rules. We normalize the resulting value by dividing by the length of the longer string.

To compare our results with previous work, we convert the numbers from previous work to re-use rate. Here, we have to make assumptions on the sizes of the clusters, which we assume to be of the same size. Writing A for the number of accounts and B for the number of accounts per password, the probability that we get the same password is $RR = \frac{B-1}{A-1}$. The results are shown in Table 3.4.

In addition, the expected value for the number of passwords is

$$E_2 = (1 - R) + 2 \cdot R = 1 + R$$

3.4.2. Re-Use Rates Across Accounts

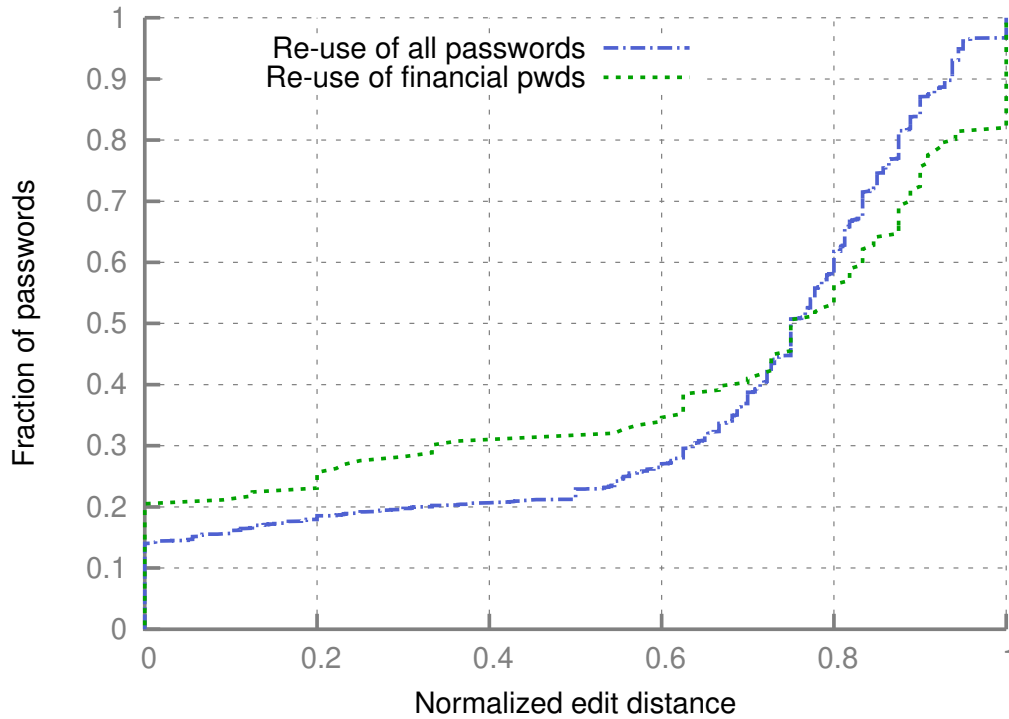


Figure 3.3.: Measuring the re-use of passwords for variable levels of similarity, given by their edit distance

- First, we measured *re-use across all passwords* of a user, regardless their assignment to MW-Fin or MW-Btm. These measurements allow us to compare the results with previous work.
- Second, we measured *re-use of financial passwords on other sites*, i.e., the re-use rate when, for a fixed user, we select one password randomly from MW-Fin and one from MW-Btm. Such results have never been obtained before and are enabled by the specifics of our dataset.

For both scenarios, we considered both exact re-use as well as approximate re-use (such as “password” and “password1” for instance). For exact re-use across all passwords we got 14%, for a (normalized) edit distance of 0.2 we have 19%, and for re-use of financial passwords we got 21% and 26% percent, respectively. The results are summarized in Table 3.4, which also gives figures from previous work for comparison. The detailed graphs are given in Figure 3.3, where we plot the normalized edit distance on the x -axis, and the fraction of password pairs with normalized edit distance up to that bound on the y -axis.

3.4.3. Discussion

We can see that the re-use rates only increase slightly between the distance 0 and 50%, which is already larger than what is usually considered “similar”. For example, the strings “*password*” and “*password-123*” have edit distance 20% while the strings “*use*” and “*re-use*” have edit distance 50%. This means that among people that re-use their password, most re-use it in the exactly same form. (Re-using with even small modifications would be a much wiser choice than exact re-use, as this would already prevent credential stuffing.)

Surprisingly, we find that re-use is more common for financial passwords than for the set of all passwords, 21% vs. 14% for exact re-use and 26% vs. 19% for approximate re-use. We speculate that financial passwords are re-used more frequently because their increased strength represents a cognitive burden on the user, and this is something of a maladaptive coping strategy.

When we compare these results with the work of Florencio and Herley [77], we see that our results are very similar; because they determined a re-use rate of 12% compared with our 14%, we feel confident that these results are correct. Comparison with the study by Trusteer Inc. [80] is not easy, as they do not describe their methodology. They state that “73% of users share the online banking password with at least one nonfinancial site.” How this relates to our results depends on the number of accounts they observed per user, and it is not clear how they handle the case where one user has multiple banking passwords.

3.5. Conclusion

In this work we studied two important aspects of password security that have received little attention previously. We used a dataset obtained by malware, which has passwords for multiple accounts for most users. This allowed us to compute meaningful statistics on two aspects of password security: first if users choose stronger passwords for accounts that are more valuable, and second on the re-use of passwords from high-value accounts on low-value accounts.

We found that password strength indeed *does* correlate with account value, a result we also were able to confirm with other lists of leaked passwords. This means that high-value real-life passwords are stronger than widely suspected, even though more work is required to see if they are actually strong enough. We were also able to show that users *do* re-use their high-value password on low-value accounts, a practice which must be considered unsafe, and we were able to confirm previous results on password re-use.

Our work also hints at further interesting research topics, which have in fact been explored in the intervening years. First, it is interesting to find other meaningful

sources for passwords that have multiple passwords for the same user, that are either larger or have a different/less bias than our present dataset. Evaluating these datasets would further increase the trust and the understanding of our results.

Our focus in this study was on alphanumeric passwords, but the re-use of PINs should be treated in future work. Innovative methodologies could lead to better understanding of re-use as it applies in new settings such as *local authentication*, [114] also called *app unlock*, or *in-app authentication*. For example, we will explore below how often do people re-use their mobile device unlock KBA as their Signal PIN? What exactly is the unaddressed risk and what could be done about it? We also explored the impact of password policies on re-use, finding that financial-site passwords were more likely to be re-used. Since that initial experiment, prevailing guidance on password composition rules in NIST SP 800-63B [3] has changed. Password rotation and complexity rules involving specific characters, such as requiring a number or symbol, are no longer recommended. As the research community delves deeper into this area, guidance is certain to change again. What effect do these new rules have on re-use for accounts of different value? Are there targeted interventions that could assist users in selecting better passwords? KBA re-use is clearly a ripe area for future work.

3.6. Author Contribution

In this paper appearing in Security and Privacy for Networks, I was the first author and I personally contributed most elements. I was first told by a colleague at RSA Security of the inadvertent collection of these usernames and passwords, and I suggested the idea of performing a study. I alone had access to the data and therefore contributed the analysis plan and implemented all of the data analysis tools we needed. That included developing a patch for the open-source John the Ripper package that would give us the guess number we needed to estimate the strength of each password, as well as producing the figures and tables in the final manuscript. In addition, I wrote the code that performed the edit-distance calculations found in the paper. I took the lead drafting the paper as well as our presentation and talk based on our work which appeared at the conference.

4

Security of Smartphone Unlock PINs with Blocklists

Contents

4.1. Introduction	52
4.2. Background	54
4.2.1. Attacker Model	54
4.2.2. Datasets	56
4.3. User Study	57
4.3.1. Study Protocol and Design	57
4.3.2. Treatments	61
4.3.3. Recruitment and Demographics	63
4.3.4. Ethical Considerations	65
4.3.5. Limitations	65
4.4. PIN Selection on Smartphones	66
4.4.1. Strength of 4- and 6-digit PINs	67
4.4.2. Selection Strategies	71
4.5. Blocklists and PIN Selection	72
4.5.1. PIN Creation and Entry Times	74
4.5.2. Attacker's Knowledge of Blocklists	74
4.5.3. Blocklisting Impact on Security	75
4.5.4. Enforcing the Blocklist	79
4.5.5. PIN Changing Strategies	80
4.5.6. User Perception	81
4.5.7. User Sentiment	83
4.6. Conclusion and Recommendations	84
4.7. Author Contribution	85

4.1. Introduction

We provide the first study focused on the selection of Personal Identification Numbers (PINs) based on data collected from users specifically primed for the smartphone setting. While authentication on mobile devices has been studied in several contexts, including patterns [13] and passwords [115], little is known about PINs used for mobile authentication. Despite the rise of biometrics, such as fingerprint or facial recognition, devices still require PINs, e.g., after a restart or when the biometric fails. That is because the biometric does not replace knowledge-based authentication; access to a device is still possible with a PIN even when using a biometric. Moreover, the presence of a biometric may actually lead to a false sense of security when selecting knowledge-based authenticators [116].

Our study focuses on the PINs users choose to unlock their mobile devices. Previous work on PINs was primarily focused on the context of banking, e.g., as part of the *Chip-and-PIN* system [25] and also mainly relied on the analysis of digit sequences found in leaked text-based password datasets since this data is more readily available [49]. Given the sparsity of information about PINs in the context of mobile authentication, we sought to fill this vital knowledge gap by conducting the first study ($n = 1705$) on the topic where participants either selected a 4- or 6-digit PIN, the two predominant PIN lengths used for device unlock. In addition to only allowing participants to complete the study on a smartphone, we also primed them specifically for the mobile unlock authentication setting, reminding participants that the selected “PIN protects [their] data and is used to unlock [their] smartphone.” While our study cannot speak to the memorability of the selected PINs due to the short time duration, our qualitative feedback suggests that participants took this prompt seriously and selected relevant PINs.

PINs of 4 and 6 digits only provide security when paired with system controls like lockouts and delays that limit offline (or *unthrottled*) guessing. An unthrottled attacker who can bypass these controls can quickly guess all PIN combinations. We instead consider a *throttled* attacker model to empirically analyze the security of PINs when the system limits the guessing rate. This is usual in the smartphone-unlocking setting where pauses are enforced after a certain number of wrong guesses in order to slow attacks down. Guessing is then limited (or throttled) to, e.g., just 10, 30, or 100 attempts in a reasonable time window, such as a few hours. In such a model, it is essential to prioritize guessing resistance in the first few guesses. Our study found little benefit to longer 6-digit PINs compared to 4-digits. In fact, our participants tend to select more-easily guessed 6-digit PINs when considering the first 40 guesses of an attacker.

As a mechanism for improving PIN selection, we also studied how PINs are affected by blocklisting. A blocklist is a set of “easy to guess” PINs, which triggers a warning

to the user. Apple iOS devices show the warning “*This PIN Can Be Easily Guessed*” with a choice to “*Use Anyway*” or “*Change PIN*.” Previous work in text-based passwords has shown that users choose stronger passwords due to a blocklist [117, 118], and recent guidance from NIST [3] concurs. To understand selection strategies in the presence of a blocklist, we conducted a between-subjects comparison of PIN selection using a number of different blocklists. This included two small (27 4-digit PINs and 29 6-digit PINs), two large (2740 4-digit PINs and 291 000 6-digit PINs), and two blocklists (274 4-digit PINs and 2910 6-digit PINs) in use today on iOS devices, which we extracted for this purpose. To determine if the experience of hitting a blocklist or the content of the blocklist itself drives the result, we included *placebo* blocklists that always excluded the participants’ first choice. Finally, we included both enforcing and non-enforcing blocklists, where participants were able to “click through” and ignore the blocklist, the approach taken by iOS. Despite the popularity of blocklists and the positive impact on textual passwords, our results show that currently employed PIN blocklists are ineffective against a throttled attacker, in both the enforcing and non-enforcing setting. This attacker performs nearly as well at guessing PINs as if there were no blocklist in use. To be effective, the blocklist would need to be much larger, leading to higher user frustration. Our results show that for 4-digit PINs a blocklist of about 10% of the PIN space may be able to balance the security and usability needs, for 6-digit PINs the same effect can be achieved by blocking 0.2% of the keyspace.

Finally, we collected both quantitative and qualitative feedback from our participants about their PIN selection strategies, perceptions of their PINs in the context of blocklists, and their thoughts about blocklisting generally. Overall, we find that despite having mostly negative sentiments about blocklist warnings, participants do perceive the PINs they select under a blocklist as more secure without impacting the memorability and convenience, except in situations of a very large blocklist.

To summarize, we make the following contributions:

1. We report on the security of 4- and 6-digit PINs as measured for smartphone unlocking, finding that in the throttled setting, the benefit of 6-digit PINs is marginal and sometimes worse than that of 4-digit PINs.
2. Considering a realistic, throttled attacker model, we show how different blocklisting approaches influence PIN selection process for both security and usability, finding that blocklists in use today offer little to no added security.
3. Through quantitative and qualitative feedback, we explore users’ perception of security, memorability, and ease-of-use of PIN-based authentication, finding that participants perceive that blocklisting will improve their PINs without impacting usability, except for very large blocklists.

4. We provide guidance for developers on choosing an appropriately-sized PIN blocklist that can influence the security in the throttled scenario, finding that a blocklist for 4-digit PINs should consist of ~ 1000 PINs to have a noticeable impact while minimizing the negative effects. To achieve the same results in the 6-digit case, ~ 2000 PINs should be blocked.

The contribution of this work is the result of a publication in *ACM Transactions on Privacy and Security*, Volume 24, Issue 4, November 2021 in collaboration with Philipp Markert, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. This version is based on that publication and additionally edited to remove some results that were not my work.

An earlier version of this work appeared as “This PIN Can Be Easily Guessed: Analyzing the Security of Smartphone Unlock PINs” published at the IEEE Symposium on Security and Privacy in May, 2020. Compared to that version, we collected and analyzed new data (6-digit PINs) to get a better understanding of the security of user-chosen smartphone unlock PINs. Combined with other changes such as additional investigation on the impact of biometrics and use of additional external PIN data sets, our analysis of 6-digit PINs is now as comprehensive as our analysis of their 4-digit counterparts. *Note: We responsibly disclosed all our findings to Apple Inc.*

4.2. Background

In this section, we define our attacker model, describe the used datasets, and outline the extraction of the two iOS PIN blocklists which we evaluate in our user study.

4.2.1. Attacker Model

When studying guessing attackers, there are two primary threat models. An *unthrottled* attacker can guess *offline*, indefinitely, until all the secrets are correctly guessed, while a *throttled* attacker is limited in the number of guesses, sometimes called an *online* attack. Google’s Android and Apple’s iOS, the two most popular mobile operating systems, implement real-world rate limiting mechanisms to throttle attackers because otherwise, it would be possible to simply guess all PIN combinations. In our attacker model, we assume the rate limiting works as designed, and as such, it is appropriate to consider a throttled attacker when evaluating security as this best matches the reality of the attacks PINs must sustain for the mobile unlock setting.

The choice of the throttled attack model is further justified when considering mobile devices’ *trusted execution environments* (TEE), where the key for device encryption is stored in “tamper resistant” hardware and is “entangled” with the user’s unlock secret [119]. This forces the attacker to perform decryption (unlock) attempts

Table 4.1.: Rate limiting on mobile operating systems

To Make n Guesses	Accumulated Waiting Time	
	Android 7, 8, 9, 10, 11	iOS 9, 10, 11, 12, 13, 14
1-5 guesses	0 s	0 s
6 guesses	30 s	1 m 0 s
7 guesses	30 s	6 m 0 s
8 guesses	30 s	21 m 0 s
9 guesses	30 s	36 m 0 s
10 guesses	30 s	1 h 36 m 0 s
30 guesses	10 m 30 s	-
100 guesses	10 h 45 m 30 s	-
200 guesses	67 d 2 h 45 m 30 s	-

on the device itself in an online way. Moreover, the TEE is used to throttle the number of decryption attempts tremendously by enforcing rate limiting delays which also survive reboots.

An overview of the currently enforced limits is given in Table 4.1. Apple’s iOS is very restrictive and only allows up to 10 guesses [119] before the iPhone disables itself and requires a reset. Google’s Android version 7 or newer are less restrictive with a first notable barrier at 30 guesses where the waiting time increases by 10 minutes. We define the upper bound for a reasonably invested throttled attacker at 100 guesses when the waiting starts to exceed a time span of 10 hours on Android [120], but we also report results for less determined attackers at 10 guesses (30 s) and 30 guesses (10.5 m) for Android. The iOS limit is 10 guesses (90 m) [119].

In our attacker model, we assume that the adversary has no background information about the owner of the device or access to other side-channels. In such a scenario, the best approach for an attacker is to guess the user’s PIN in decreasing probability order. To derive this order, we rely on the best available PIN datasets, which are the Amitay-4-digit and RockYou-6-digit datasets as defined below. Again, we only consider an *un-targeted attacker* who does not have additional information about the victim. If the attacker is targeted, and is able to use other information and context about the victim, e.g., via shoulder-surfing attack [63, 66, 121] or screen smudges [70], the attacker would have significant advantages, particularly in guessing 4- vs. 6-digit PINs [66].

In other parts of this work, we make use of blocklists. In those cases, we consider an attacker that is aware and in possession of the blocklist. This is because the attacker can crawl the system’s blocklist on a sample device, as we have done for this work. Hence, with knowledge of the blocklist, an informed attacker can improve the guessing strategy by *not* guessing known-blocked PINs and instead focusing on common PINs not on the blocklist.

Table 4.2.: Datasets for strength estimations and comparisons

Kind	Dataset	Samples
4-digit PINs	Amitay-4-digit [51]	204 432
4-digit PINs	RockYou-4-digit [49]	1 780 587
6-digit PINs	RockYou-6-digit [49]	2 758 490
3x3 Patterns	“All” unlock patterns [122]	4 637
Passwords	LinkedIn [123]	10 000
Passwords	Pwned Passwords v7 [124]	<i>Top</i> 10 000

4.2.2. Datasets

Perhaps the most realistic 4-digit PIN data is from 2011 where Daniel Amitay developed the iOS application “Big Brother Camera Security” [51]. The app mimicked a lock screen allowing users to set a 4-digit PIN. Amitay anonymously and surreptitiously collected 204 432 4-digit PINs and released them publicly [51]. While collected in an uncontrolled experiment, we apply the dataset (Amitay-4-digit) when guessing 4-digit PINs, as well as to inform the selection of our “data-driven” blocklists. As there is no similar 6-digit PIN data available to inform the attacker, we rely on 6-digit PINs extracted from password leaks, similar to Bonneau et al.’s [25] and Wang et al.’s [49] method. PINs are extracted from consecutive sequences of exactly n -digits in leaked password data. For example, if a password contains a sequence of digits of the desired length, this sequence is considered as a PIN (e.g., PW: `ab3c123456d` \rightarrow PIN: `123456`, but no 6-digit PINs would be extracted from the sequence `ab3c1234567d`). By following this method, we extracted 6-digit PINs from the *RockYou* password leak, which we refer to as RockYou-6-digit (2 758 490 PINs). We also considered 6-digit PINs extracted from other password leaks, such as the *LinkedIn* [123] dataset, but found no marked differences between the datasets.

To provide more comparison points, we consider a number of other authentication datasets listed in Table 4.2. For example, we use a 3x3 Android unlock pattern dataset described by Golla et al. [122], combining four different datasets [13, 57–59]. It consists of 4637 patterns with 1635 of those being unique. In addition, we use a text-password dataset. Melicher et al. [115] found no difference in strength between passwords created on mobile and traditional devices considering a throttled guessing attacker. Thus, we use a random sample of 10 000 passwords from the LinkedIn [123] leak and use the *Pwned Passwords v7* [124] list to simulate a throttled guessing attacker to estimate the guessing resistance for the sampled LinkedIn passwords as a proxy for mobile text passwords.

As part of our set of blocklists, we also consider a blocklist of “easily guessed” 4/6-digit PINs as used in the wild by Apple, which we obtained via brute-force extraction from an iPhone running iOS 12. We were able to verify that blocklisting of PINs

is present on iOS 9 through the latest version iOS 16, and we also discovered that Apple updated their blocklist with the deployment of iOS 10 (for example, the PIN 101471 is blocked on iOS 10.3.3, but is not on iOS 9.3.5). In theory, it is possible to extract the blocklist by reverse engineering iOS, yet, we found a more direct way to determine the blocklist via brute-force: During device setup, when a PIN is first chosen, there is no throttling. To test the membership of a PIN, one only needs to enter *all* the PINs and observe the presence of the blocklist warning, and then intentionally fail to re-enter the PIN to be able to start over.

The extraction of all 10 000 4-digit PINs took ~ 9 hours. Testing all 1 million 6-digit PINs took about 30 days using two setups. We repeated the process for 4-digit PINs multiple times, tested lists of frequent 6-digit PINs, and verified the patterns found in the PINs. Moreover, we validated all blocked PINs multiple times. We refer to these two lists as the iOS-4 and iOS-6 blocklists.² In total, the 4-digit blocklist contains 274 PINs and includes common PINs as well as years from 1956 to 2015, but its composition is mostly driven by repetitions such as `aaaa`, `abab`, or `aabb`. The 6-digit blocklist contains 2910 PINs and includes common PINs as well as ascending and descending digits (e.g., `543210`), but its composition is, again, mostly driven by repetitions such as `aaaaaa`, `abcabc`, or `abccba`. The common PINs blocked by Apple overlap with a 4-digit blocklist suggested by Bonneau et al. [25] in 2012 and the top 6-digit PINs reported by Wang et al. [49] in 2017.

4.3. User Study

In this section, we outline the treatment conditions, the user study, and the collected data. We also discuss limitations and our ethical considerations. Appendix A.1 outlines the entire questionnaire.

4.3.1. Study Protocol and Design

We conducted a user study of 4- and 6-digit PINs using Amazon Mechanical Turk (MTurk) with $n = 1705$ participants. To mimic the PIN creation process in our browser-based study, participants were restricted to mobile devices by checking the user-agent string. We applied a 12-treatment, between-subjects study protocol for the PIN selection criteria, e.g., 4- vs. 6-digit with or without blocklisting. The specifics of the treatments are discussed in detail in Section 4.3.2. At the end of the study, we collected 851 and 854 PINs, 4- and 6-digits respectively, for a total of 1705 PINs as our core dataset. These PINs were all selected, confirmed, and recalled. We additionally recorded all intermediate PIN selections, such as what would happen if a

²To foster future research on this topic, we share the described blocklists and the PIN datasets at: <https://this-pin-can-be-easily-guessed.github.io>.

selected PIN was *not* blocked and the participant did not have to select a different PIN. For more details of different kinds of PINs collected and analyzed, refer to Table 4.7. All participants were exposed to a set of questions and feedback prompts that gauged the security, memorability, and usability of their selected PINs, as well as their attitudes towards blocklisting events during PIN selection.

The survey itself consists of 10 parts. Within each part, to avoid ordering effects, we applied randomization to the order of the questions that may inform later ones; this information is also available in Appendix A.1. The parts of the survey are:

1. *Informed Consent*: All participants were informed of the procedures of the survey and had to provide consent to continue. The informed consent notified participants that they would be required to select PINs in different treatments, but did not inform them of any details about blocklisting that might be involved in that selection.
2. *Agenda*: After being informed, participants were provided additional instructions and details in the form of an *agenda*. It stated the following: “You will be asked to complete a short survey that requires you to select a numeric PIN and then answer some questions about it afterwards. You contribute to research so please answer correctly and as detailed as possible.”
3. *Practice*: Next, participants practiced with the PIN entry screen, which mimics typical PIN selection on mobile devices, including the “phoneword” alphabet on the virtual PIN pad. The purpose of the practice round was to ensure that participants were familiar with the interface prior to selecting a PIN. There was clear indication during the practice round that this was practice and that participants would begin the primary survey afterwards.
4. *Priming*: After familiarization and before selection, participants were further primed about mobile unlock authentication and PINs using language similar to what iOS and Android use during PIN selection. A visual of the priming is in Figure 4.1. A lock icon was used to prime notions of security, and users were reminded that they will need to remember their PIN for the duration of the study without writing it down. Participants must click “I understand” in order to continue. The qualitative feedback shows that the priming was understood and followed with some participants stating that they reused their actual PIN.
5. *Creation*: The participants then performed the PIN creation on the page shown in Figure 4.2. The PIN was entered by touching the digits on the virtual PIN pad. As usual, participants had to enter the PIN a second time to confirm it was entered correctly. Depending on the treatment (see Section 4.3.2), the users either selected a 4- or 6-digit PIN and did or did not experience a blocklist event. In Figure 4.3 and Figure 4.4 we depicted the two blocklist warnings which either allowed participants to “click through” the warning (or not). This

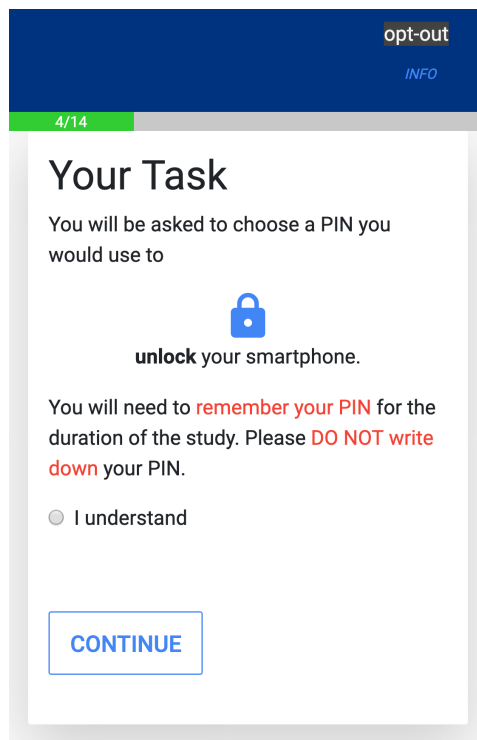


Figure 4.1.: Priming information provided before the participants were asked to create a PIN

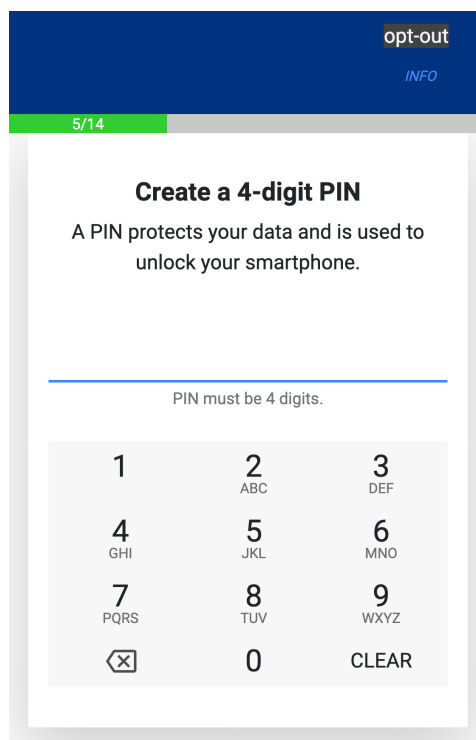


Figure 4.2.: The design of the page on which we asked the participants to create a PIN

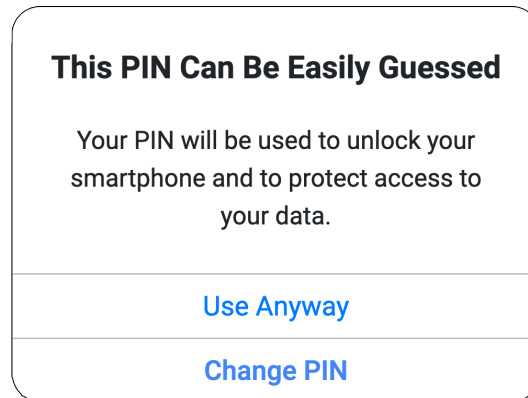


Figure 4.3.: Blocklist warning **with** the ability to “click through”

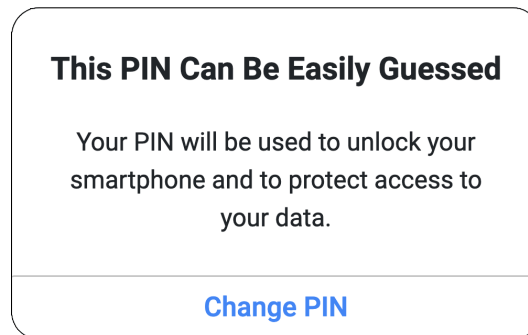


Figure 4.4.: Blocklist warning **without** the ability to “click through”

feedback was copied to directly mimic the wording and layout of a blocklist warning used by Apple since iOS 12.

6. *Blocklisting Followup*: After creation, we asked participants about their attitudes and strategies with blocklisting. If the participants experienced a blocklist event, we referred back to that event in asking followup questions. Otherwise, we asked participants to “imagine” such an experience. These questions form the heart of our qualitative analysis (see Section 4.5.7).
7. *PIN Selection Followup*: We asked a series of questions to gauge participants’ attitudes towards the PIN they selected with respect to its security and usability, where usability was appraised based on ease of entry and memorability (see Section 4.5.6). As part of this questionnaire, we also asked an attention check question. We excluded the data of 19 participants because we could not guarantee that they followed our instructions completely.
8. *Recall*: On this page, participants were asked to recall their earlier selected PIN. Although the two prior parts formed distractor tasks we do not expect that the recall rates measured here speak broadly for the memorability of these PINs. As expected, nearly all participants could recall their selected PIN.

Table 4.3.: Overview of studied treatments

	Treatment	Short Name	Blocklist	Size	Click-thr.
4 digits	Control-4-digit	Con-4	–	–	–
	Placebo-4-digit	Pla-4	First choice	1	✗
	iOS-4-digit-wCt	iOS-4-wC	iOS 4-digit	274	✓
	iOS-4-digit-nCt	iOS-4-nC	iOS 4-digit	274	✗
	DD-4-digit-27	DD-4-27	Top Amitay	27	✗
	DD-4-digit-2740	DD-4-2740	Top Amitay	2740	✗
6 digits	Control-6-digit	Con-6	–	–	–
	Placebo-6-digit	Pla-6	First choice	1	✗
	iOS-6-digit-wCt	iOS-6-wC	iOS 6-digit	2910	✓
	iOS-6-digit-nCt	iOS-6-nC	iOS 6-digit	2910	✗
	DD-6-digit-29	DD-6-29	Top RockYou	29	✗
	DD-6-digit-291000	DD-6-291000	Top RockYou	291000	✗

9. *Demographics*: In line with best practice [125], we collected the demographics at the end, including participants’ age, gender, IT background, and their current mobile unlock scheme.
10. *Honesty/Submission*: Finally, we asked if the participants provided “honest” answers to the best of their ability. We informed them that they would be paid even if they indicated dishonesty. Using this information in combination with the attention check described above, we excluded the data of 19 participants to ensure the integrity of our data. After affirming honesty (or dishonesty), the survey concluded and was submitted.

4.3.2. Treatments

We used 12 different treatments: 6 treatments for 4-digit PINs and 6 treatments for 6-digit PINs. The details for each treatment can be found in Table 4.3.

Control Treatments

For each PIN length, we had a control treatment, **Control-4-digit** and **Control-6-digit**, that primed participants for mobile unlock authentication and asked them to select a PIN without any blocklist interaction. These PINs form the basis of our 4- and 6-digit mobile-authentication primed PIN dataset. In total, we have 231 control 4-digit PINs and 236 control 6-digit PINs. We also created two additional datasets, **First-Choice-4-digit** (851 PINs) and **First-Choice-6-digit** (854 PINs), by combining the control PINs with those chosen by participants from other treatments in their “first attempt” before having been subjected to any blocklist.

Blocklist Treatments

The remaining treatments considered PIN selection in the presence of a blocklist. There are two types of blocklist implementations: *enforcing* and *non-enforcing*. An enforcing blocklist does not allow to continue as long as the selected PIN is blocked; the user *must* select an unblocked PIN. A non-enforcing blocklist warns the user that the selection is blocked, but the user can choose to ignore the feedback and proceed anyway. We describe this treatment as providing the participant an option to *click through*. Otherwise, the treatment uses an enforcing blocklist. Visuals of the non-enforcing and enforcing feedback can be found in Figure 4.3 and 4.4.

Placebo blocklist As we wanted to determine if the experience of hitting a blocklist or the content of the blocklist itself influenced the results, we included a *placebo* treatment for both 4- and 6-digit PINs (**Placebo-4-digit** and **Placebo-6-digit**). In this treatment, the user’s first choice PIN was blocked, forcing a second choice. As long as the second choice differed from the first, it was accepted.

iOS blocklist For this treatment, we included the blocklists used on Apple’s iOS 13. The 4-digit iOS blocklist contains 274 PINs (2.74 % of the available 4-digit PINs), and the 6-digit iOS blocklist contains 2910 PINs (0.291 % of the available 6-digit PINs). These blocklists provide measurements of real scenarios for users selecting PINs on iOS devices. As iOS allows users to “click through” the blocklist warning and use their blocked PIN anyway, we implemented our blocklisting for the iOS condition in the same way (i.e., conditions **iOS-4-digit-wCt** and **iOS-6-digit-wCt**). To understand the effect of non-enforcing blocklists, we also tested enforcing versions of the iOS blocklists (**iOS-4-digit-nCt** and **iOS-6-digit-nCt**).

Data-driven blocklists We considered two blocklists for each PIN length that are significantly smaller and larger than the iOS blocklist. The blocklists were constructed using the most frequently occurring PINs in the Amitay-4-digit and RockYou-6-digit dataset. We refer to the 4-digit treatments as **DD-4-digit-27** and **DD-4-digit-2740** because the blocklists contain 27 and 2740 PINs respectively. Following this, we blocked the 29 most frequent PINs in the treatment **DD-6-digit-29** while 291 000 were blocked in **DD-6-digit-291000**.

When comparing the two data-driven 4-digit blocklists and the one used in iOS, it can be seen that they are differently composed. While 22, or 82 % of the PINs contained in DD-4-digit-27 are blocked in iOS, there are also 5 PINs which are not. Surprisingly, these PINs correspond to simple patterns like 0852 which is a bottom-up pattern across the PIN pad or 1379, the four corners of the pad chosen in a left-to-right manner. Now, when extending the comparison to the DD-4-digit-2740

blocklist we see that 258 of the 274 PINs from the iOS blocklist, or 92%, are also blocked by our large data-driven blocklist. The remaining 16 PINs all follow the same repetitive *aabb* scheme, e.g., 0033, 4433, or 9955. Interestingly, only one of those PINs, 9933, was selected in our study which shows that double repetitions are presumably not as common as Apple expects.

Similar observations can be made in the 6-digit case when comparing the iOS blocklist with the two data-driven versions. There are 3 PINs (159357, 147852, 246810) in our DD-6-digit-29 blocklist with only 29 PINs which are not rejected by Apple’s blocklist with 2910 entries. Of those 3 PINs, at least 159357 and 147852 follow straightforward patterns which one may expect to be blocked. The intersection with the large data-driven blocklist covers 2314 PINs, or 80% of the iOS blocklist. The 596 PINs which are solely rejected by Apple follow three schemes: *ababac* (323 PINs), *abccba* (258 PINs), and *abcabc* (15 PINs). Again, those schemes are not very popular across our participants: only 7% of the PINs which were selected in our study follow them.

4.3.3. Recruitment and Demographics

We recruited a total of 1944 participants using Amazon’s Mechanical Turk (MTurk). After excluding a portion due to invalid responses to attention tests or survey errors, we had 1705 participants remaining. We required our participants to be 18 years or older, reside in the US (as checked by MTurk), and have at least an 85% approval rate on MTurk. The IRB approval required focusing on participants residing in the US, but there may be a secondary benefit to this: US residents often do not have *chip-and-PIN* credit cards (although, they do use 4-digit ATM PINs), in contrast to residents in Europe or Asia, and thus may associate PIN selection more strongly with mobile device locking. In any case, participants were explicitly primed for the mobile device unlock setting. Participants indicated they understood this instruction, and their qualitative responses confirm their understanding.

We also reviewed all of the participants’ responses for consistency, including answers to attention-check questions, the honesty question, and speed of entry. We removed 19 participants who provided inconsistent data but did not “reject” any participants on Amazon Mechanical Turk. Participants were compensated with \$1 (USD) for completion; the survey took on average 5 minutes for an hourly rate of \$12.

Demographics and background As typical on MTurk, our sample is relatively young and better educated than the general US population. Of the participants, 923 identified as male (54%) while 768 (45%) identified as female (1% identified as other or preferred not to say), and the plurality of our participants were between 25 and 34 years old (48%). Most participants had some college (21%) or a bachelor’s degree

Table 4.4.: Usage of mobile unlock authentication schemes

Primary Scheme	No.	%	Secondary Scheme	No.	%
Fingerprint	779	46 %	4-digit PIN	387	50 %
			6-digit PIN	215	28 %
			Pattern	109	14 %
			Other	68	8 %
Face	263	15 %	4-digit PIN	113	42 %
			6-digit PIN	104	40 %
			Pattern	23	9 %
			Other	23	9 %
Other Biometric	33	2 %	4-digit PIN	10	30 %
			6-digit PIN	3	9 %
			Pattern	16	49 %
			Other	4	12 %
4-digit PIN	218	13 %	<i>No secondary scheme used.</i>		
6-digit PIN	59	4 %			
Pattern	88	5 %			
Other	76	4 %			
None	189	11 %			

(42%), and few (11%) had a master’s or doctoral degree. While 28% described having a technical background, 69% said they did not. We have the full details of the demographics responses in Appendix A.2.

Smartphone OS We asked participants which operating system they use on their primary smartphone. Slightly more than half, 1008 (59%), of the participants were Android users, while 676 (40%) were iOS users. We collected browser user-agent strings during the survey, and confirmed similar breakdowns, suggesting most participants used their primary smartphone to take the survey. A detailed breakdown can be found in the Appendix A.3.

Unlock schemes usage As we focus on mobile authentication, we were interested in learning about the kind of mobile authentication our participants use, recalling both biometric and knowledge-based authentication may be in use on a single device. We first asked if a biometric was used and then asked what authentication scheme participants use instead of, or as a backup for the biometric, when for example it fails. While Table 4.4 shows a compact description, a detailed breakdown can be found in the Appendix A.3. Among KBAs considered here, PINs are the most common: 43% described using a 4-digit PIN, 22% using a 6-digit PIN, and 3% using a PIN of another length. The second most common form of KBA are Android unlock patterns at 14%, and 57 participants (or 3%) reported using an alphanumeric password. In our study, 189 participants (11%) reported not using any locking method.

4.3.4. Ethical Considerations

All of the survey material and protocol detailed in this chapter were approved by The George Washington University Institutional Review Board (IRB). Beyond meeting the approval of the IRB, we worked to uphold the ethical principles outlined in the Menlo Report [97].

In practicing *respect for persons* and *justice*, beyond informing and getting consent, we also sought to compensate participants fairly at least at the minimum wage of the municipality where the oversight was performed. Since some of our treatments may frustrate participants, including where the blocklist was comparatively large (DD-4-digit-2740 & DD-6-digit-291000), we also compensated those who returned the survey and notified us of their frustration.

Additionally, as we are dealing with authentication information, we evaluated the ethics of collecting PINs and distributing blocklists in terms of *beneficence*. With respect to collecting PINs, there is risk in that participants may (and likely will) expose PINs used in actual authentication. However, there is limited to no risk in that exposure due to the fact that PINs are not linked to participants and thus cannot be used in a targeted attack. A targeted attack would also need proximity and awareness of the victim, of which, neither is the case for this study. Meanwhile, the benefit of the research is high in that the goal of this research is to improve the security of mobile authentication. Similarly, distributing blocklists increases social good and scientific understanding with minimal risk as a determined attacker likely already has access to this material.

Finally, we have described our procedures transparently and make our methods available when considering *respect for law and public interest*. We also do not access any information that is not already publicly available.

4.3.5. Limitations

There are a number of limitations in this study. Foremost among them is the fact that the participant sample is skewed towards mostly younger users residing in the US. However, as we described previously, there may be some benefit to studying PINs from US residents as they are less familiar with *chip-and-PIN* systems and may be more likely to associate PINs directly with mobile unlocking. We argue that our sample provides realizable and generalizable results regarding the larger ecosystem of PIN selection for mobile authentication. Further research would be needed to understand how certain populations, for example, more age-diverse ones select PINs [126]. For populations from different locations, there is some knowledge about the differences between English-speaking and Chinese users [49], but other populations have also not been studied yet.

Another limitation of the survey is that we are asking participants to select PINs while primed for mobile authentication and there is a risk that participants do not act the same way in daily life. We note that similar priming is used in the authentication literature for both text-based passwords for desktop [127, 128] and mobile settings [115], and these results generalize when compared to passwords from leaked password datasets [129]. We have similar results here. When compared to the most realistic dataset previously available, Amitay-4-digit, the most common 4-digit PINs collected in our study are also present in similar distributions to Amitay [51]. Also, in analyzing the qualitative data, a number of participants noted that they re-used their actual unlock PIN.

While this presents strong evidence of the effectiveness of mobile unlock priming, we, unfortunately, do not have any true comparison points, like what is available for text-based passwords. There is no obvious analog to the kinds of attacks that have exposed millions of text-based passwords that would similarly leak millions of mobile unlock PINs. Given the available evidence, we argue that collecting PINs primed for mobile unlock authentication provides a reasonable approximation for how users choose PINs in daily life.

Due to the short, online nature of our study, we are limited in what we can conclude about the memorability of the PINs. The entirety of the study is only around 5 minutes, while mobile authentication PINs are used for indefinite periods, and likely carried from one device to the next. There are clear differences in these cases, and while we report on the recall rates within the context of the study, these results do not generalize.

Finally, we limited the warning messaging used when a blocklist event occurred. We made this choice based on evaluating the messaging as used by iOS, but there is a long line of research in appropriate security messaging [130–133]. We do not wish to make claims about the quality of this messaging, and a limitation of this study (and an area of future work) is to understand how messaging affects changing strategies and click-through rates.

4.4. PIN Selection on Smartphones

In the following section, we discuss the security of both 4- and 6-digit PINs. Unless otherwise stated, our analyzed dataset consists of the PINs entered before any blocklist warning in Step (5) of the study. These so-called “first choice” PINs (cf. Table 4.7) are unaffected by the blocklists.

Table 4.5.: Guessing difficulty for a perfect knowledge attacker

Dataset	Size	Online Guessing (Success %)			Offline Guessing (bits)			
		λ_3	λ_{10}	λ_{30}	H_∞	$\tilde{G}_{0.05}$	$\tilde{G}_{0.1}$	$\tilde{G}_{0.2}$
First-4	851	3.41 %	6.23 %	11.75 %	5.65	7.07	7.81	-*
Amit-4 [†]	204 432	9.28 %	15.28 %	22.91 %	4.52	4.82	5.20	6.68
Rock-4 [†]	1 780 587	8.23 %	17.63 %	30.67 %	4.73	5.00	5.42	5.94
First-6 [†]	854	5.05 %	7.99 %	13.04 %	4.73	5.88	7.43	-*
Rock-6 [†]	2 758 490	13.04 %	15.51 %	19.27 %	3.10	3.10	3.10	7.41

†: For a fair comparison we downsampled the datasets to the size of First-4 (851 PINs).

*: We omit entries which are not sufficiently supported by the underlying data.

4.4.1. Strength of 4- and 6-digit PINs

Entropy-based strength metrics We analyzed PINs in terms of their mathematical metrics for guessing resistance based on entropy estimations. For this, we consider a *perfect knowledge* attacker who always guesses correctly (in perfect order) as described by Bonneau et al [45]. The advantage of such an entropy estimation approach is that it always models a best-case attacker and does not introduce bias from a specific guessing approach. Our results are given in Table 4.5.

We report the β -success-rate, which measures the expected guessing success for a throttled adversary limited to β -guesses per account (e.g., $\lambda_3 = 3$ guesses). Moreover, we provide the Min-entropy H_∞ as a lower bound estimate that solely relies on the frequency of the most common PIN (1234, 123456). Finally, we present the partial guessing entropy (α -guesswork) G_α , which provides an estimate for an unthrottled attacker trying to guess a fraction α of all PINs. In three cases, the calculation of $\tilde{G}_{0.2}$ is based on PINs occurring only once, due to the small size of the datasets. This constraint would result in inaccurate guessing-entropy values which is why they are not reported.

For a fair comparison among the datasets which all differ in size, we downsampled all datasets to the size of the smallest dataset First-4 (851 PINs). We repeated this process 500 times, removed outliers using Tukey fences with $k = 1.5$. In Table 4.5 we report the median values. The low Min-entropy of the Rock-6 dataset is due to the fact that the PIN 123456 is over-represented. It is $21\times$ more frequent than the second-most popular PIN. In contrast, the most common 4-digit PIN occurs only $1.7\times$ more often, leading to a higher H_∞ value. Overall, the PINs we collected, specifically primed for mobile authentication, have different (and *stronger*) strength estimations than PINs derived from leaked text-based password datasets. This is true for both the 4- and 6-digit PINs, which supports our motivation for conducting studies that collect PINs directly.

Guess number-driven strength estimates Next, we estimate the security of the PINs in regard to real-world guessing attacks. Our attacker guesses PINs in decreasing probability order based on the Amit-4, Rock-4, and Rock-6 datasets. When two or more PINs share the same frequency, i.e., it is not possible to directly determine a guessing order, we order those PINs using a Markov model [99]. We trained our model on the bi-grams (4-digit PINs) or tri-grams (6-digit PINs) of the respective attacking datasets which simulates the attacker with the highest success rate for each case without overfitting the problem.

In the throttled scenario, depicted in Figure 4.5, we find that guessing 4-digit PINs with the Amitay-4-digit dataset (\triangle) is the most effective attack. In contrast to the RockYou-4-digit dataset (∇) for which we extracted PINs from a password leak, the Amitay dataset consists of actual PINs (cf. Section 4.2.2). We observe that guessing based on the RockYou-4-digit is less effective than Amitay-4-digit. This finding suggests we should use the actual PIN data contained in the Amitay set whenever possible to simulate the attacker.

When comparing 4- (\triangle) and 6-digit PINs (\times), we see that guessing performance varies. For 10 guesses (the maximum allowed under iOS), we find 4.6% of the 4-digit and 5.7% of the 6-digit PINs are guessed. For 30 guesses (a less determined attacker on Android), 7.6% of the 4-digit and 8.8% of the 6-digit PINs are guessed and for 100 guesses (a reasonable upper bound on Android), 16.2% of the 4-digit and 12.5% of the 6-digit PINs.

Somewhat counter-intuitive is the weaker security for 6-digit PINs for the first 40 guesses. Upon investigation, the most-common 6-digit PINs are more narrowly distributed than their most-common 4-digit counterparts. The most common 6-digit PINs consist of simple PINs, such as 123456 as defined in Appendix A.4, and repeating digits. In contrast, the most common 4-digit PINs consist of simple PINs, patterns, dates, and repeating digits. As a result, the most common 6-digit PINs may actually be easier to guess and less diverse than the most common 4-digit PINs.

There could be many explanations for this counter-intuitive finding. One explanation may be that users have more 4-digit PIN sequences to draw on in choosing a PIN, such as dates, but have fewer natural 6-digit analogs, and thus revert to less diverse, more easily guessed choices. We will present some evidence for this hypothesis in Section 4.4.2 where we analyze the selection strategies of 6-digit PINs. Another explanation may be that users have a false sense of security that comes with 6-digit PINs as they are “two digits more secure” than 4-digit PINs. Thus, users do not feel that they need more complexity in their 6-digit PIN choices. Either way, future research is needed to better understand this phenomenon, which has also been observed by Aviv et al. [57] in the context of increasing the size (3x3 vs. 4x4) of Android graphical unlock patterns.

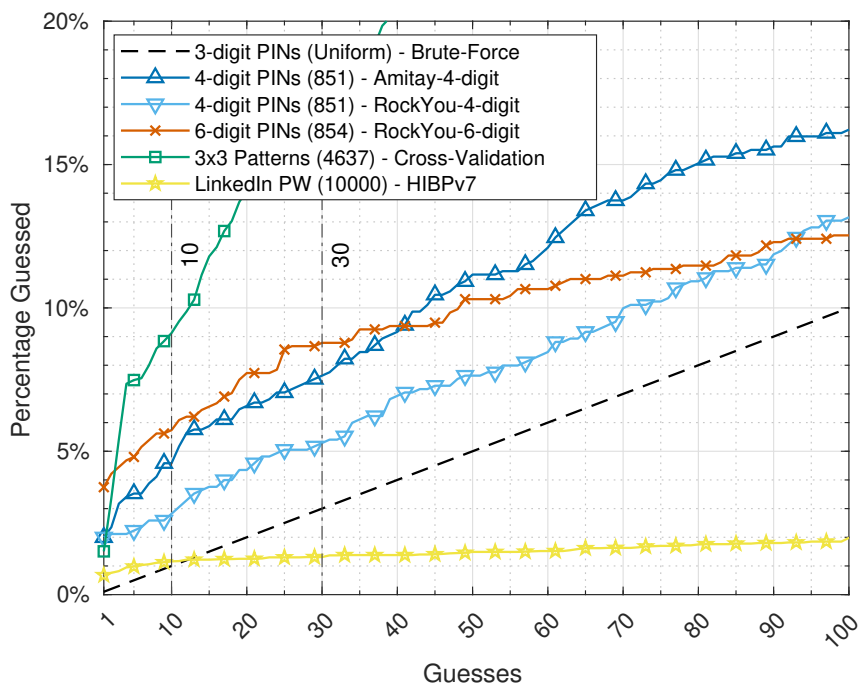


Figure 4.5.: Guessing performance against mobile authentication systems based on the number of guesses

Finally, we compare guessing resistance with other mobile authentication schemes including Android’s graphical unlock patterns drawn on a 3x3 grid (\square) and alphanumeric passwords (\star), along with a uniform distribution of 3-digit PINs ($-$). In theory, a 3x3 grid allows 389 112 unique patterns, yet, the distribution of patterns is highly skewed [13]. When considering an attack throttled to 100 guesses, 35.5% of the patterns will be guessed. Against this attack, 4- and 6-digit PINs are twice as good. Password-based authentication, on the other hand, is the most secure scheme. After 100 guesses only 1.9% of the passwords are recovered.

Figure 4.6 shows the guessing time of an attacker due to rate limiting based on Table 4.1 for iOS and Android. iOS has stricter rate limiting with a maximum of 10 guesses that can be completed in 1h 36m, at which point an attacker compromises 4.6% of the 4-digit PINs and 5.7% of the 6-digit PINs. At the same time limit of roughly 1.5h, an attacker on Android is able to compromise 13.6% of the 4-digit PINs and 11.0% of the 6-digit PINs because of less restrictive rate limiting.

Especially on iOS, rate limiting becomes more aggressive after the initial guesses. For example, the first 6 guesses on iOS can be done within a minute, while the first 8 guesses take 21 minutes. An attacker with only one minute on iOS can compromise 3.5% of the 4-digit PINs and 5.2% of the 6-digit PINs. However, for 10 guesses which take 1h 36m on iOS, there are only marginal gains with 4.6% of the 4-digit PINs and 5.7% of 6-digit PINs compromised. Hence, after the first minute with

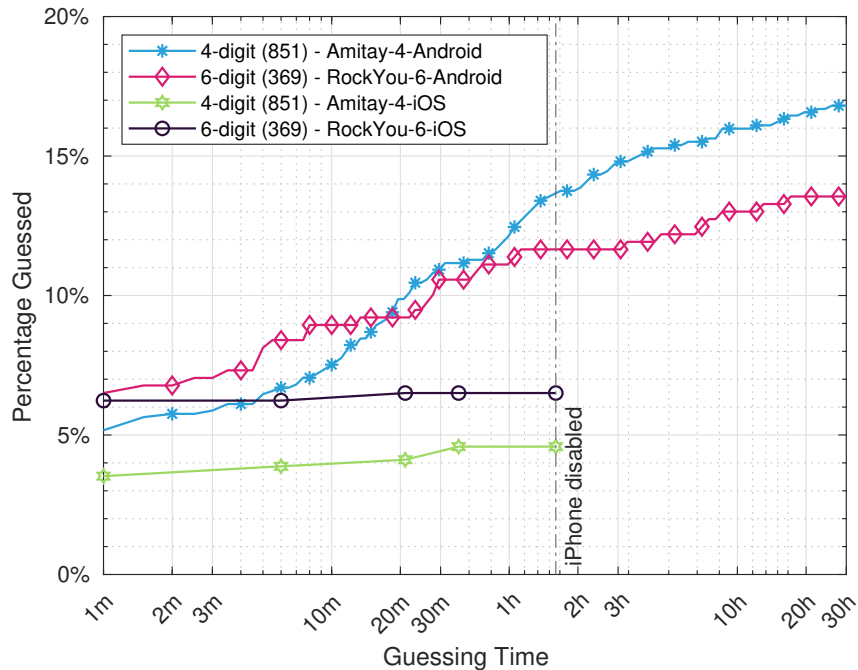


Figure 4.6.: Guessing performance against 4- and 6-digit PINs on Android and iOS based on the required time. For 4-digit PINs, we only show the success rate of an attack with Amit-4 as it outperforms Rock-4.

6 guesses on iOS, it does not greatly benefit the attacker to continue through the aggressive timeouts for 4 more guesses at 1h 36m. In contrast, an attacker on Android would benefit more from continuing to guess beyond the initial large increases in rate limiting. Note, in a targeted attack, there may be additional information or other motivations for the attacker not modeled here.

To summarize, we confirmed previous work from Wang et al. [49] that there is no evidence that 6-digit PINs offer any security advantage over 4-digit PINs considering a throttled guessing attacker with up to 40 guesses, which covers most mobile unlock authentication settings. Only when considering threat models where the attacker is allowed to make more guesses, 6-digit PINs start to exceed 4-digit PINs in terms of their guessing resistance. To support this claim, we performed χ^2 tests ($\alpha = 0.05$) for both the 4- and 6-digit PINs guessed within 10 [4.6 %, 5.7 %], 30 [7.6 %, 8.8 %], and 100 guesses [16.2 %, 12.5 %]. The test for 10 ($p = 0.28$) and 30 guesses ($p = 0.39$) did not show a significant difference in PIN strength. For 100 guesses, on the other hand, we were able to observe that the 6-digit PINs are significantly stronger than the 4-digit ones ($p = 0.03$). This again highlights the importance of clearly defining threat model in terms of how many guesses the attacker is able to make when deciding on a certain PIN length.

Effect of biometrics Users who employ a biometric, cf. Table 4.4, do not need to provide their KBA as often as users who solely rely on a PIN, pattern, or password. This choice may shift users towards more complex choices which are more cumbersome to type, but, owing to the biometric, only need to be provided on rare occasions like a device restart. Hence, the question arises: do users who authenticate with a biometric select more secure PINs?

To test this hypothesis, we split each of the First-4 and First-6 dataset into two datasets, based on if participants said they used a biometric. As we primed our participants to select a PIN they would use to unlock their smartphone (cf. Figure 4.1), we have all the information required for this type of analysis. The security metrics for the “Biometric-used” and “No-biometric-used” datasets are shown in Table 4.7. The results do not support the hypothesis, but instead, participants who do not use a biometric tend to create more secure PINs. However, while the success rates of the attacker differ by up to 3% for 30 guesses when comparing Biometric-used-4 and No-biometric-used-4, we were not able to observe any significant differences using a χ^2 test ($\alpha = 0.05$).

4.4.2. Selection Strategies

In Step 6 of our study, we asked participants about their “strategy for choosing” their PIN. We analyzed the free-text responses to this question by building a codebook from a random sample of 314 PIN selection strategies using two coders. Inter-rater reliability between the coders measured by Cohen’s kappa was $\kappa = 0.90$. Appendix A.4 shows the 10 most popular strategies.

While users have many different selection strategies, most participants chose PINs that they perceive as memorable in general or based them on personally-important dates, especially birthdays and anniversaries. Other popular strategies are PIN-pad patterns, choosing randomly, or selecting other kinds of meaningful numbers to the participants, like a ZIP Code or a favorite number. While most of those strategies are common across both PIN lengths, most participants who said they chose randomly (26 of the 33, or 79%) were asked to create a 6-digit PIN. This result supports the intuition that users have less experience with 6-digit PINs and have fewer meaningful sequences at hand.

To further understand how users create 6-digit PINs and to see if users take their 4-digit selection strategy and just extend it to create a longer version, we now look at the 4-digit substrings of the 6-digit PINs. For this, we took the 855 First-Choice-6-digit PINs and created three lists extracting the 4 leftmost, the 4 middle, and the 4 rightmost digits of each PIN. For comparison, we overlapped those lists with the First-Choice-4-digit PINs. As can be seen in Table 4.6, the greatest overlap with 23%, occurs for the leftmost substring PINs, followed by the rightmost (14%).

Table 4.6.: Overlap of the First-4 PINs with the three substring lists extracted from the First-6 PINs

Substring	Overlap		Top 5 PINs			
	No.	%	PIN	No.	%	Most Common Addition
Leftmost	196	23 %	1234	17	9 %	123456 (91 %)
			2580	7	4 %	<i>not distinct</i>
			6969	5	3 %	696969 (80 %)
			1212	4	2 %	121212 (50 %)
			1379	3	2 %	<i>not distinct</i>
Middle	74	9 %	0000	2	3 %	000000 (100 %)
			1111	2	3 %	111111 (100 %)
			2121	2	3 %	121212 (100 %)
			7777	2	3 %	777777 (100 %)
			9898	2	3 %	898989 (100 %)
Rightmost	116	14 %	6969	5	4 %	696969 (80 %)
			4321	4	4 %	654321 (100 %)
			1212	3	3 %	121212 (67 %)
			4578	2	2 %	124578 (100 %)
			7777	2	2 %	777777 (67 %)

The substring PINs consisting of the 4 digits in the middle only overlap with the First-4 PINs by 9%. Moreover, all of the PINs we extracted for this list follow simple repetitions. These strategies are not specific to a certain PIN length. A similar conclusion can be drawn from the rightmost PINs, there is no indication that participants started with a 4-digit PIN and added two digits on the left. Again, we see that the creation strategies can be used to create PINs of arbitrary length, mostly repetitions (e.g., 1212/121212), and sequences (e.g., 4321/654321). However, Table 4.6 also depicts two exceptions: 2580, and 1379. The former is a top-down walk, which allows for a simple 4-digit PIN, yet, each of the 7 participants who started a 6-digit PIN this way ended up differently. A similar observation can be made for 1379, where each of the four corners is selected without an apparent addition for a 6-digit PIN. Both cases suggest that there are participants who did not have an actual 6-digit strategy but used one they had in mind for 4-digits and added two digits. This also fits the overall impression that users are more familiar with 4-digit PINs.

4.5. Blocklists and PIN Selection

We now present results on our ten blocklist treatments: five for each PIN length as shown in Table 4.7.

Table 4.7.: Security metrics and usage times for PINs considering different datasets and treatments

	Name	Participants	Blocklist Hits	10 Guesses No. %	30 Guesses No. %	100 Guesses No. %	Guess No. Median	Creation Time	Entry Time	Number of Attempts
Datasets	First-Choice-4-digit	851	-	39 5%	65 8%	138 16%	1330	-	-	-
	Clicked-through-4	19	19	5 26%	6 32%	13 68%	50	-	-	-
	Biometric-used-4	533	-	28 5%	47 9%	91 17%	1347	-	-	-
	No-biometric-used-4	318	-	11 4%	18 6%	47 15%	1257	-	-	-
Treatments	Control-4-digit	231	-	11 5%	19 8%	39 17%	1185	7.9s	1.5s	1.01
	Placebo-4-digit	122	122	5 4%	11 9%	19 16%	2423	21.8s	1.5s	2.15
	iOS-4-digit-wCt	124	28	5 4%	8 6%	18 15%	1405	10.4s	1.4s	1.17
	iOS-4-digit-nCt	126	21	4 3%	10 8%	14 11%	1747	9.3s	1.6s	1.29
	DD-4-digit-27	121	5	4 3%	7 6%	18 15%	1928	8.8s	1.5s	1.11
	DD-4-digit-2740	127	88	0 0%	0 0%	1 1%	2871	25.4s	1.6s	2.98
Datasets	First-Choice-6-digit	854	-	49 5%	75 9%	107 13%	49021	-	-	-
	Clicked-through-6	10	10	9 90%	9 90%	9 90%	1	-	-	-
	Biometric-used-6	542	-	33 6%	51 9%	68 13%	47773	-	-	-
	No-biometric-used-6	312	-	16 5%	24 8%	39 13%	50922	-	-	-
Treatments	Control-6-digit	236	-	15 6%	26 11%	35 15%	42584	11.0s	2.5s	1.01
	Placebo-6-digit	117	117	3 3%	6 5%	10 9%	154521	28.5s	3.0s	2.17
	iOS-6-digit-wCt	125	15	9 7%	9 7%	13 10%	40972	11.9s	2.6s	1.06
	iOS-6-digit-nCt	125	16	2 2%	4 3%	6 5%	61036	12.2s	2.8s	1.22
	DD-6-digit-29	126	12	1 1%	2 2%	7 6%	82373	11.1s	2.5s	1.23
	DD-6-digit-291000	125	90	0 0%	0 0%	0 0%	324621	45.2s	3.5s	3.94

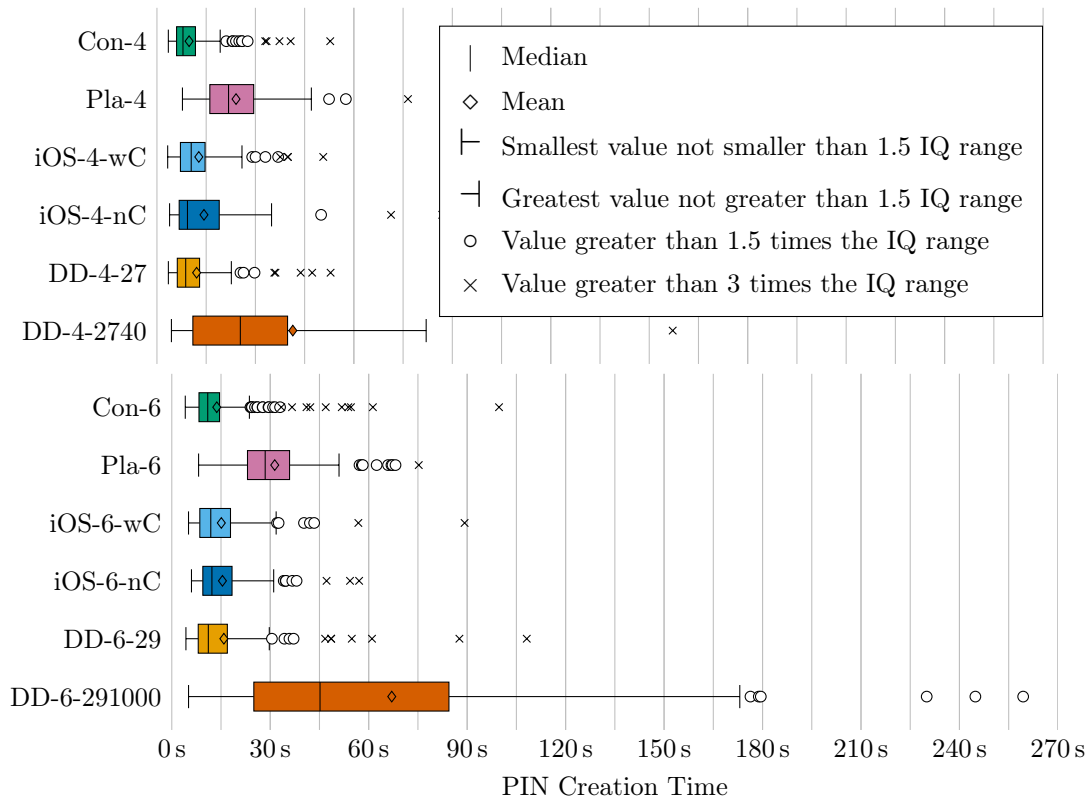


Figure 4.7.: PIN creation times for the different treatments. For the sake of clarity, we excluded two extrema from the plot: 1542.32s from the DD-4-2740 treatment and 1105.13s from DD-6-291000.

4.5.1. PIN Creation and Entry Times

The blocklist has an impact on the PIN creation time: increase in the number of blocklist messages leads obviously to increased creation time. The median creation time when receiving a blocklist message can be found in Table 4.7; a more detailed breakdown for each treatment can be seen in Figure 4.7.

In the 4-digit case, there are obvious differences between the control treatments and the placebo and the large data-driven treatment DD-4-2740. The median creation time increases from 7.9s for the Con-4 treatment to 21.8s for Pla-4 and 25.4s for DD-4-2740. Both differences are significant ($p < 0.001$) using a Kruskal-Wallis test followed by a Bonferroni-corrected pair-wise Wilcoxon test. The differences for the remaining 4-digit treatments iOS-4-wC, iOS-4-nC, and DD-4-27 are more subtle. The median creation time for the small data-driven treatment DD-4-27 only increases by 0.9s to 8.8s, followed by the iOS-4-nC treatment (9.3s), and iOS-4-wC (10.4s). Moreover, we were able to observe significant differences for the latter comparison, i.e., Con-4 vs. iOS-4-wC ($p < 0.01$), whereas we were not for the comparisons of iOS-4-nC and DD-4-27 with the control treatment. We did not observe any significant differences between the 4-digit treatments for entry time.

The situation is similar for 6-digit PINs. As can be seen in both in Table 4.7 and Figure 4.7, the creation times for the Pla-6 and DD-6-291000 treatment increase compared to the control treatment, but both iOS treatments (iOS-6-wC & iOS-6-nC) and the small data-driven treatment DD-6-29 show minimal differences compared to control. We observed significant differences for both Pla-6 and DD-6-291000 ($p < 0.001$) using a Kruskal-Wallis test followed by Bonferroni-corrected pair-wise Wilcoxon tests. We did not find significant differences among iOS-6-wC, iOS-6-nC, and DD-6-29.

The entry times are again not affected with one exception: the 6-digit case. Participants required more time to enter the PIN they created in the large data-driven treatment DD-6-291000. The median here is 3.5s compared to 2.5s in the respective control treatment and this difference is also significant ($p < 0.001$) using the same statistical tests.

This suggests that blocklists, when properly sized, can lead to significant increases in creation time which may in turn frustrate users, as we will explore in Section 4.5.6 and 4.5.7. However, the subsequent usage of the PIN, as evidenced by the entry time, is unaffected. Only in the case of a very large blocklist with 6-digit PINs do we observe any meaningful increase in entry time.

4.5.2. Attacker's Knowledge of Blocklists

As described in Section 4.2.1, we assume the attacker knows which blocklisting strategy is used by the system and can optimize their guessing strategy by *not* guessing

Table 4.8.: Attacker’s gain from blocklist knowledge

Treatment	10 Guesses		30 Guesses		100 Guesses		Guess No. Median	Knowledge Beneficial
	No.	%	No.	%	No.	%		
Pla-4	±0	±0 %	±0	±0 %	±0	±0 %	±0	–
iOS-4-wC	-3	-2 %	-4	-2 %	-9	-8 %	-303	✗
iOS-4-nC	+3	+2 %	+7	+6 %	+3	+2 %	+245	✓
DD-4-27	+4	+3 %	+7	+6 %	+5	+4 %	+27	✓
DD-4-2740	±0	±0 %	±0	±0 %	+1	+1 %	+2740	✓
Pla-6	±0	±0 %	±0	±0 %	±0	±0 %	±0	–
iOS-6-wC	-9	-7 %	-5	-4 %	-8	-6 %	-7322	✗
iOS-6-nC	+2	+2 %	+2	+2 %	+2	+2 %	+1524	✓
DD-6-29	+1	+1 %	+2	+2 %	+2	+2 %	+29	✓
DD-6-291000	±0	±0 %	±0	±0 %	±0	±0 %	+291000	✓

items on the blocklist. Here, we consider how much benefit this optimization provides. Table 4.8 shows the net gains and losses for guessing PINs when considering a blocklist-informed attacker.

Knowledge of the blocklist is unhelpful when considering the placebo (Pla-4 and Pla-6) and the click-through treatments (iOS-4-wC and iOS-6-wC). The blocklist is effectively of size one for the placebo as the first choice of a participant is dynamically blocked. Merely knowing that a PIN was blocked is of little help to the attacker. As there is no clear gain (or harm), we model a blocklist-knowledgeable attacker for the placebo treatments (see Table 4.7).

The case with a non-enforcing blocklist where users can click through the warning message is more subtle. If the attacker is explicitly choosing not to consider PINs on the blocklist, even though they may *actually* be selected due to non-enforcement, the guessing strategy is harmed (negative in Table 4.8). None of the tested modifications of this strategy, such as by incorporating the observed click-through rate, lead to an improvement. Therefore, we consider an attacker that *does not* use the blocklist to change the guessing strategy for the click-through treatments (iOS-4-wC and iOS-6-wC). In the remaining treatments (iOS-4-nC, DD-4-27, DD-4-2740, iOS-6-nC, DD-6-29, DD-6-291000), there are clear advantages when knowing the blocklist.

4.5.3. Blocklisting Impact on Security

We now consider how the different blocklists perform in terms of improving security. The primary results are in Table 4.7 where we report on the guessing performance against each treatment. As described in Section 4.2.1, there are certain rate limits implemented on Android and iOS which is why we report on throttled attacks with 10, 30, and 100 guesses in terms of the number and percentage of correctly guessed PINs (No. and % columns). Furthermore, we provide the attacker’s performance

in an unthrottled setting based on the median guess number. The 4-digit attacker is informed by the Amit-4 dataset, while the 6-digit attacker employs the Rock-6 dataset. Both attackers guess in frequency order with knowledge of the blocklist where appropriate (see Section 4.5.2).

To analyze the security, we performed a multivariate χ^2 test comparison ($\alpha = 0.05$) for the PINs guessed within 10, 30, and 100 guesses across treatments. The test for 10 suggested some significant differences in the data ($p = 0.007$), however, we did not find any actual significant differences in the post-hoc analysis (Bonferroni-corrected). For 30 guesses and 100 guesses the test also showed significant differences ($p < 0.001$); the results of the post-hoc analyses are described below.

Smaller blocklists When looking at the 4-digit treatments, there is little difference among Placebo-4-digit, iOS-4-digit-wCt, iOS-4-digit-nCt, and DD-4-digit-27, compared to Control-4-digit or First-Choice-4-digit. In our post-hoc analyses (Bonferroni-corrected), we found no significant difference.

For our 6-digit treatments, the situation is similar, yet, there is one exception: for 30 guesses we observed a significant difference between the small data-driven blocklist and the control ($p < 0.01$). While this implies that it can make sense to employ a small blocklist in certain cases, we will show in Section 4.5.7 that blocklist warnings are associated with negative sentiments. Hence, it is hard to justify the combination of throttling and blocklists in general.

In the unthrottled setting, we see differences between the smaller and placebo blocklists. Notably, the smallest (DD-4-digit-27, DD-6-digit-29) outperforms the larger iOS blocklists (iOS-4-digit-nCt, iOS-6-digit-nCt). We conjecture this may be due to iOS' inclusion of PINs based on repetitions which were chosen less often by our participants. As a result, in an unthrottled setting, blocklisting can offer real benefits. The median guess numbers for both 4- and 6-digit placebos suggest that just pushing users away from their first choice can improve security. Unfortunately, direct use of a placebo blocklist is unlikely to be effective and is problematic in practice as users will quickly figure out the deception.

Finally, these improvements to the unthrottled attack setting appear to be only of academic interest: given the small key space, any attacker that is able to bypass the enforced rate limiting is able to exhaustively test all possible combinations [134]. For example, a tool from Elcomsoft is able to bypass the rate limiting on Apple's iPhone 5 and 5c. In this case, guessing all 4-digit PINs takes about 12 minutes while enumerating all 6-digit PINs takes 20.5 hours [135].

Large blocklist We also consider very large blocklists in the DD-4-digit-2740 and DD-6-digit-291000 treatment containing 2740 PINs and 291 000 PINs respectively. These blocklists are bigger than their iOS counterparts, blocking 27.4% in the 4- and

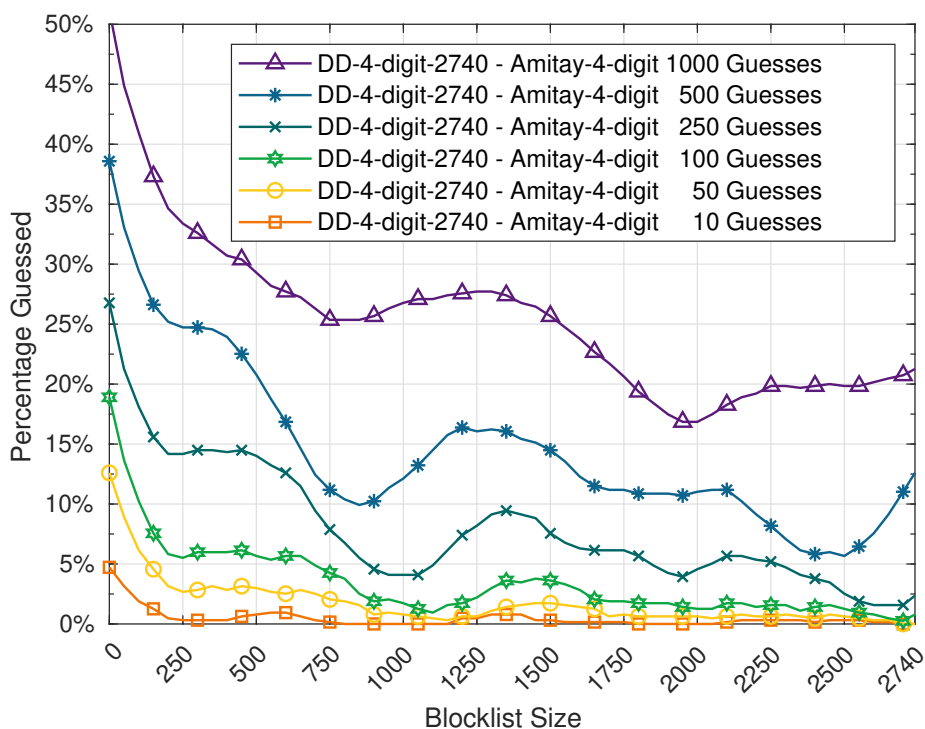


Figure 4.8.: [4-digits PINs: For throttled attackers, limited to 100 guesses, a blocklist of $\sim 10\%$ of the key space (~ 1000 PINs) is ideal

29.1% of the key space in the 6-digit case. At this scale, we do see noticeable effects on the security in the throttled setting. In the 4-digit case, the attacker finds only 1% of 4-digit PINs after 100 guesses. Our χ^2 tests support this, for 100 guesses we found a significant difference ($p < 0.001$). For post-hoc analyses (Bonferroni-corrected) we found significant differences between the large DD-4-2740 blocklist and Con-6 ($p < 0.01$) as well as the treatments: Con-4 ($p < 0.001$), Pla-4 ($p < 0.01$), iOS-4-wC ($p < 0.05$), and DD-4-27 ($p < 0.05$).

In the 6-digit case, we make similar observations for the guessing routine although we already start to see significant differences for 30 guesses when comparing the DD-6-291000 and the control treatment ($p < 0.01$). For 100 guesses the guessing success of the attacker in the DD-6-291000 treatment is significantly lower than for all 4-digit treatments: Con-4 ($p < 0.001$), Pla-4 ($p < 0.01$), iOS-4-wC ($p < 0.05$), DD-4-27 ($p < 0.01$), as well as the 6-digit control treatment ($p < 0.001$). All of this suggests that a larger blocklist can improve security in a throttled setting.

While similar positive security results are present for the unthrottled setting, we show in Section 4.5.6 that the larger blocklist also leads to perceived lower usability, and thus it is important to balance the user experience with security gains.

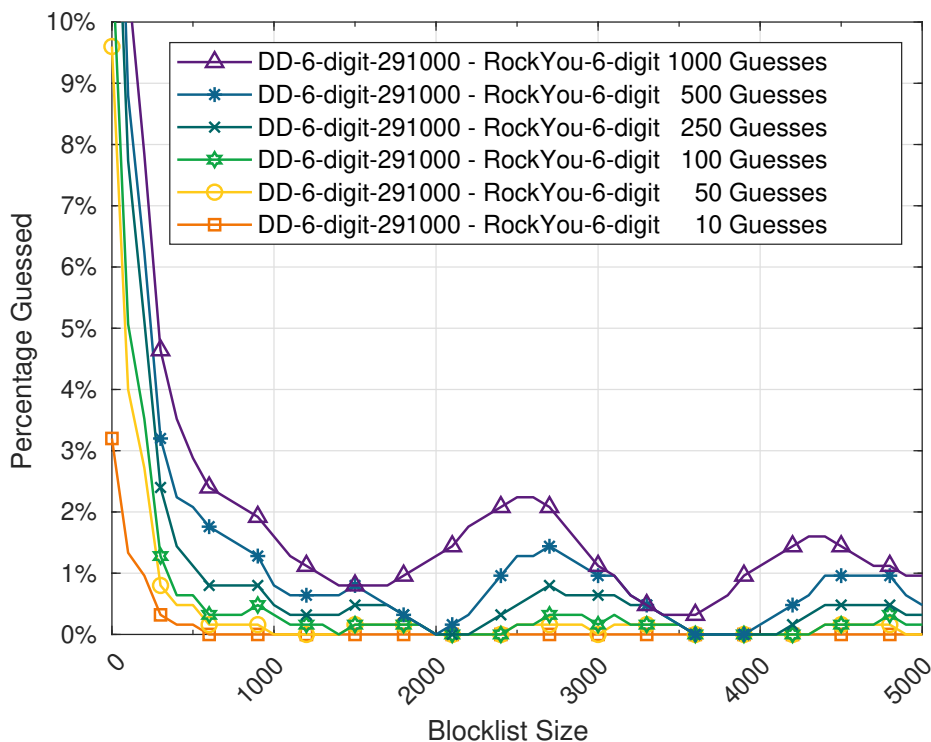


Figure 4.9.: [6-digits PINs: For throttled attackers, limited to 100 guesses, a blocklist of $\sim 0.2\%$ of the key space (~ 2000 PINs) is ideal

Correctly sizing a blocklist While there is a clear security benefit to having a large blocklist, it is important to consider the right size of a blocklist to counteract negative usability and user experience issues. This leads to the question: *Can a smaller blocklist provide similar benefits in the throttled setting and if so, what is an appropriately sized blocklist?* Data from the large data-driven treatments enable us to simulate how users would have responded to shorter blocklists. In our user study, we collected not only the final PIN accepted by the system, but also all $n-1$ intermediate (first-choice, second-choice, and so on) PINs rejected due to the blocklist. Consider a smaller blocklist that would have permitted choice $n-1$ to be the final PIN, rather than n . To simulate that smaller blocklist size, we use choice $n-1$.

The results of the simulation are shown in Figures 4.8 and 4.9. We observe that there are several troughs and peaks in the curves in both figures. We speculate that these relate to changes in user choices as they move from their first choice PIN to their second choice PIN, and so on due to the expanding blocklist restrictions. For example, entering the first trough, the attacker is most disadvantaged when it is no longer possible to rely on guessing only first choice PINs and second choice PINs need to be considered. Eventually, the blocklist has restricted all first choice PINs, whereby the attacker can now take advantage of guessing popular second choices which results in a peak. These cycles continue until the blocklist gets so large that

few acceptable PINs remain, and the attacker’s advantage grows steadily by guessing the remaining PINs not on the blocklist.

Based on these cycles, we conclude that an appropriately-sized blocklist should be based on one of the troughs where an attacker is most disadvantaged to maximize the security gained in the throttled setting. As we are also concerned about minimizing user discomfort and frustration (e.g, PIN creation time, see Section 4.5.1), the first trough appears the most ideal. As can be seen in Figure 4.8, for 4-digit PINs the first trough occurs at about 1000 PINs (10 % of the 4-digit PIN space) throttled at 100 guesses. A similar suggestion can be drawn from the simulation for 6-digit PINs in Figure 4.9, however, due to the overall larger key space, a blocklist with 2000 PINs only corresponds to 0.2 % of all possible selections. We do not observe equally aligned distribution, but the attacker’s success rate is sufficiently low when blocking only 0.2 % of the keyspace. In contrast, the ideal 4-digit blocklist rejects 10 % of all possible PINs.

4.5.4. Enforcing the Blocklist

To test the effect of a click-through option, we compared the enforcing treatment for each length (iOS-4-nC / iOS-6-nC) with its non-enforcing counterpart (iOS-4-wC / iOS-6-wC). Neither showed significant differences. This absence of evidence suggests that using a click-through option does not reduce security in the throttled attacker setting despite the fact that clicked-through PINs are extremely weak (see row Clicked-through-4 in Table 4.7). These results seem to be driven by the fact that it is uncertain whether the user clicked through (see Table 4.8). In an enforcing setting, the attacker can leverage the blocklist but is equally challenged in guessing the remaining PINs.

We also investigated why participants chose to ignore and click through the warning. From 28 participants who saw a blocklist warning in the iOS-4-wC treatment, we observed a click-through-rate of 68 % (19 participants). In the respective 6-digit treatment iOS-6-wC, 10 out of 15, i.e., 67 %, ignored the warning. This is twice the rate at which TLS warnings are ignored ($\sim 30\%$) [136]. Furthermore, we asked the 29 participants who pressed “*Use Anyway*” about their motivations. The 3 most observed answers are *Memorability Issues*: “Because this is the number I can remember,” *Incomplete Threat Models*: “Many people don’t tend to try the obvious PIN as they think it’s too obvious so people won’t use it,” and *Indifference*: “I don’t give [sic] about the warning. Security is overrated.” These findings are similar to prior work where users do not follow external guidance for a number of reasons [137, 138]. In older versions of iOS, the blocklist warning message was “*Are You Sure You Want to Use This PIN? This PIN is commonly used and can be easily guessed.*” with the safe option “*Choose New PIN*” in bold and the unsafe click-through option saying

Table 4.9.: Changes in participants’ PIN selection strategies across treatments

Treatment	Hits	Selection vs. Changing Strategy			Edit Distance		
		Sample	Same	Minor	New	Mean	SD
Pla-4	122	29	35 %	24 %	41 %	3.20	0.90
iOS-4-wC	9*	9	0 %	44 %	56 %	3.11	0.87
iOS-4-nC	21	21	19 %	29 %	52 %	3.24	0.92
DD-4-27	5	5	40 %	40 %	20 %	3.20	0.75
DD-4-2740	88	29	14 %	24 %	62 %	3.39	0.76
Pla-6	117	28	28 %	18 %	54 %	4.59	1.41
iOS-6-wC	5*	5	0 %	40 %	60 %	4.40	1.20
iOS-6-nC	16	16	6 %	50 %	44 %	4.00	1.54
DD-6-29	12	12	33 %	33 %	33 %	5.25	0.72
DD-6-291000	90	29	14 %	21 %	65 %	4.82	1.13

*: Hit blocklist, and did not click-through.

“Use PIN.” We observed that Apple changed this wording with iOS 11 to what is depicted in Figure 4.3. Considering that TLS warning design research started with similarly high click-through-rates of around 70% [131], we hope that new designs can also improve blocklist warning CTRs [136].

4.5.5. PIN Changing Strategies

In our study, we asked 485 participants who faced a blocklist how their creation strategy changed in response to the warning. We sampled 183 responses ($\sim 10\%$ of our total number of participants) and grouped them into three categories: participants who continued using the “Same” strategy, participants who made “Minor” changes to the strategy, and participants who came up with a completely “New” strategy. Two coders independently coded the data. Inter-rater reliability between the coders measured by Cohen’s kappa was $\kappa = 0.92$. The detailed results for each treatments are shown in Table 4.9.

About 50% of the participants choose a new strategy when confronted with a blocklist warning. Only participants of the DD-4-27 and DD-6-29 treatment with a very small blocklist, tended to keep their pre-warning strategy. The edit distances vary slightly across the treatments and support this self-reported behavior: participants in the 4-digit scenario changed on average 3 digits with the standard deviation showing that some participants changed their PIN completely while some participants only changed 2 digits. The same conclusion can be drawn from the edit distances in the 6-digit case with one difference: participants in the DD-6-29 treatment changed more digits on average. This is particularly interesting because the blocklist is by far the smallest which suggests that users may be more willing to change their PIN if the warning does not appear to be arbitrary.

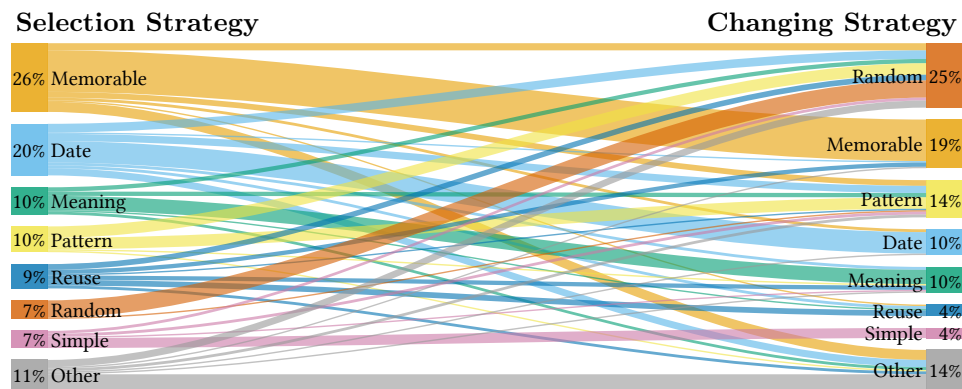


Figure 4.10.: Participants' PIN selection (first choice) and changing strategies (final choice) for $n = 183$

To analyze how participants changed their PIN selection, we mapped the initial selection strategies to the final ones. The result is shown in Figure 4.10. First of all, 25 % of the participants stated to have changed their PIN into something random (cf. Table A.3 in Appendix A.4). While there are 7 % of the participants who already had this strategy, we observe a shift from all of the other selection strategies to a random PIN which shows the effectiveness of the blocklist warnings. Moreover, we see that participants usually do not change their PIN to be “memorable,” a “date,” or “simple.” Furthermore, we also see that a certain number of participants stick to their strategy. While we already described that this decision is influenced by the treatment (cf. Table 4.9), we are now able to see that the selection strategy also influences this decision. For example, nearly all participants who initially selected a random PIN, held on to this approach. This is less distinct across other strategies, yet, participants who stuck to their selection strategy are always the largest group. The only two exceptions are participants who reused a PIN or selected it based on a pattern, they tended to change their strategy after seeing a blocklist warning.

4.5.6. User Perception

We analyzed participants' perceptions regarding PIN selections with respect to security and usability. Participants were asked to complete the phrase “*I feel the PIN I chose is*” with three different adjectives: “*secure, memorable, and convenient.*” The phrases were displayed randomly and participants responded using a Likert scale. The results are shown in Figure 4.11. To compare these results, we converted the Likert responses into weighted averages on a scale of -2 to +2. As the weighted averages are not normally distributed, tested using the Shapiro-Wilk test ($p < 0.001$), we tested for initial differences using a Kruskal-Wallis test, followed with post-hoc, pairwise tests using Dunn's-test comparisons of independent samples with a Bonferroni correction.

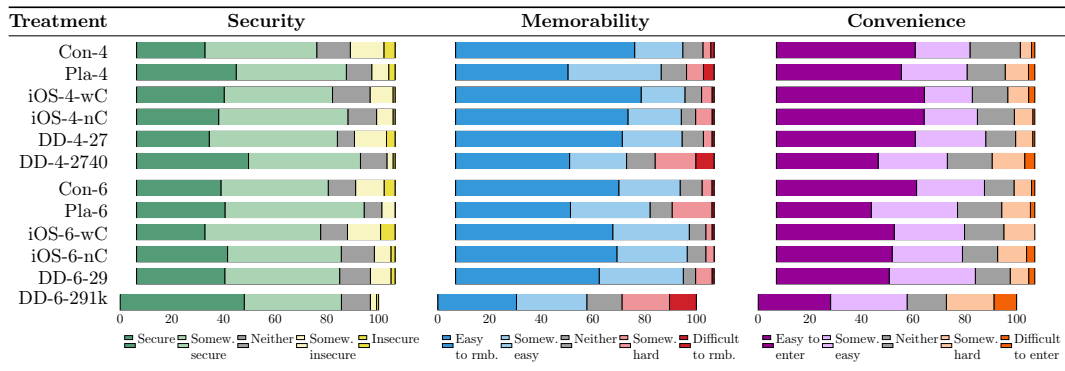


Figure 4.11.: Participants' perception of their PIN's security (*Secure – Insecure*), memorability (*Easy to remember – Difficult to remember*), and convenience (*Easy to enter – Difficult to enter*).

We found that there are significant differences across treatments when considering Likert responses for *security*. For the 4-digit PINs, post-hoc analysis did not indicate any significant differences. One explanation for this overall high confidence in the security of the PIN choice, may be the familiarity with 4-digit PINs. In contrast to this, participants in the DD-6-291000 treatment perceive their PINs as more secure compared to participants of the 6-digit control ($p < 0.05$), and iOS-6-wC treatment ($p < 0.01$). Here, the large portion (72%) of participants who encountered the blocklist may have lead to a change in the overall perception.

For *memorability* we also found significant differences among the treatments. In post-hoc analysis we found that increased interaction with the blocklist led to lower perceived memorability of PINs, as evidenced by the Pla-4 ($p < 0.001$), DD-4-2740 ($p < 0.05$), Pla-6 ($p < 0.001$), and DD-6-291000 ($p < 0.001$) treatments compared to their respective control treatments. The DD-4-2740 and DD-6-291000 showed the most significant differences with other treatments. Again, this is likely due to the fact that many participants encountered a blocklist warning sometimes even for multiple PIN choices and were thus relying on not just second-choice PINs, but also third- and fourth-choice, etc. PINs that are perceived to be less memorable.

The responses to perceived *convenience* also show significant differences, however, post-hoc analysis revealed limited effects when considering pair-wise comparisons. In general, participants perceived their 4-digit or 6-digit PINs at the same convenience level across treatments. However, there is one exception: PINs created in the DD-6-291000 treatment are perceived as significantly more difficult to enter than PINs in the 6-digit control treatment ($p < 0.01$), iOS-6-wC ($p < 0.05$), DD-6-29 ($p < 0.05$), and all 4-digit treatments ($p < 0.001$). As for the memorability, this suggests that while users may be comfortable with their first-choice 6-digit PIN, there is much higher perceived *inconvenience* when having to conform with a large blocklist.

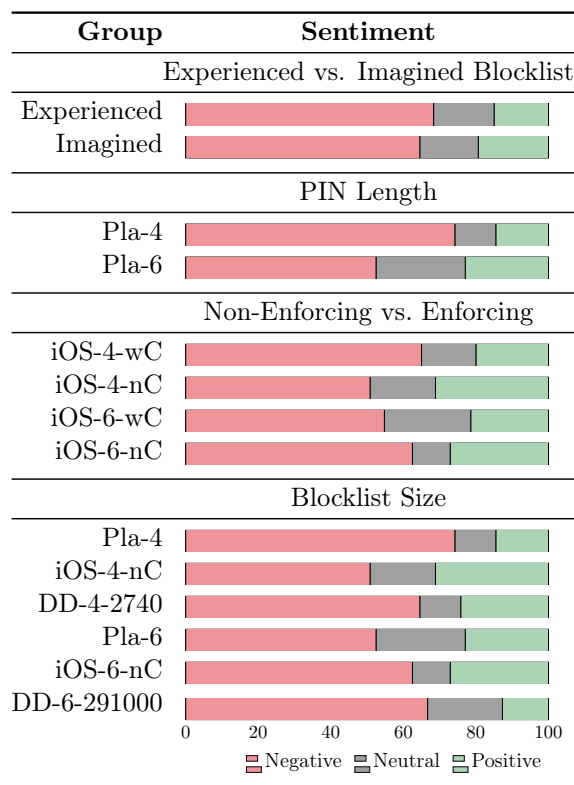


Figure 4.12.: Participants’ sentiment: We split the participants into four categories and classified their feelings in terms of sentiment using EmoLex [139]

4.5.7. User Sentiment

To gain insight into participants’ sentiments regarding blocklisting, we asked “*Please describe three general feelings or reactions that you had after you received this warning message*” or “*would have had*” if the participant did not encounter a blocklist. Accompanying the prompt are three free-form, short text fields. A codebook was constructed by two individual coders summarized in Appendix A.5 in Table A.5. For each of the four categories (blocklist hit experienced vs. imagined, 4- vs. 6-digit PINs, non-enforcing vs. enforcing, different blocklist sizes), 21 individuals’ responses were randomly selected. Again, two individual raters were tasked with coding the responses. The inter-rater reliability, computed using Cohen’s kappa, was $\kappa = 0.98$.

Using the NRC Word-Emotion Association Lexicon [139], we classified assigned codes in terms of sentiment (positive, negative, or neutral) for Figure 4.12. EmoLex maps individual English words (in this case, codes assigned by our coders) to exactly one sentiment. For example, “indifference,” is labeled with the “negative” sentiment. As expected, participants generally had a negative reaction to the blocklist warning.

While overall, participants expressed negative sentiments towards blocklist messages, which may be expected as warning messages are not often well received by users [131], we only observed significant differences in a single comparison. Using a

χ^2 test, we found that there was significant difference ($p < 0.05$) in the proportion of negative sentiment when considering PIN length for the two placebo treatments. As both groups always experienced a blocklist event, a higher negative sentiment exists for the placebo blocklist with 4-digits. This might be because users were confused by the warning as the blocklist event was arbitrary. However, in the 6-digit PIN case, less familiarity with 6-digit PINs may have led to less negative reactions.

Interestingly, participants in general consider displaying warnings about weak PIN choices to be appropriate although they cannot imagine that their own choice might be considered insecure. Moreover, sentiments are similar for those who hit the blocklist and those who imagined having done so. This suggests that future work on blocklist warning design may benefit from asking participants to imagine such events.

4.6. Conclusion and Recommendations

This paper presents the first comprehensive study of PIN security as primed for the smartphone unlock setting. In the smartphone unlock setting, developers have adopted notable countermeasures—throttling, blocklisting, PIN length—which we consider as part of our analysis. Using a throttled attacker model, we find that 6-digit PINs offer little to no advantage, and sometimes make matters worse. Also, we find that blocklists in use on today’s mobile operating systems are not designed reasonably. In some cases, they need to be larger in order to affect security at all, while they are oversized in other cases, needlessly impairing the user experience.

Given this information, we offer a number of recommendations to developers.

- In a throttled scenario, simply increasing the PIN length is of little benefit. In our results, we were only able to observe a significant difference between 4- and 6-digit PINs for an attacker that performs at least 100 guesses. As this is beyond reach for mobile attackers of interest, developers should not merely adopt longer PINs without a concomitant change in threat model. Observe that without throttling, an attacker could quickly try all 4- and 6-digit PINs.
- On iOS, with only 10 possible guesses, we could not observe any security benefits when a blocklist is deployed, either for 4- or 6-digit PINs. On Android, where 100 guesses are feasible, we find that a blocklist would be beneficial. Such a blocklist would need to contain the 1000 most popular PINs in the 4-digit case or the 2000 most popular for 6-digit PINs, in order to increase the security of the chosen PINs while minimizing user frustration.
- We observe that the increase in terms of the perceived security is only significant when users are forced to conform with a large 6-digit blocklist as compared to selecting a PIN in presence of a large 4-digit blocklist (as was the case in the data-driven treatments). This may suggest users are less familiar with

selecting 6-digit PINs, an observation our analysis of the selection strategies supports. Yet, a detailed exploration of the reasons for this are left to future investigation.

- While we observed advantages for using a placebo blocklist in the unthrottled settings, we do not recommend implementing a placebo blocklist, as users will simply game it once the deception is known.

4.7. Author Contribution

In this paper appearing in ACM Transactions on Privacy and Security, I personally contributed a number of elements. Firstly, I designed and implemented an exhaustive literature review on the topic, critically analyzing previous studies and delineating the scope of our investigation. This review identified several gaps in the literature including the need to study PINs on smartphones and the lack of public PIN datasets. In turn, I contributed the Research Questions meant to reflect this delineation, and our inquiry.

Secondly, I played a pivotal role in the development of our research methodology, I contributed the design of a rigorous user study, including the survey instrument as well as development of an analysis plan to interpret our findings.

Thirdly, I contributed an early version of the data collection website. Additionally, I contributed the analysis and coding of the qualitative data, including identification of participant perceptions of usability and security around selecting a smartphone PIN. I contributed the primary codebook, and additionally recruited, supervised, and collected data from an independent second coder to ensure the reliability of our findings. This effort led directly to my contribution of the qualitative figures and tables in the final manuscript.

Fourthly, I contributed our responsible disclosure message to Apple, drawing on my experience responding to these messages in industry.

Fifthly, I designed and presented a poster based on our work. This presentation took place during the poster session of IEEE Security and Privacy 2022. Finally, I contributed extensively to the writing and revision process of the submitted manuscript, helping to communicate our findings in a clear and concise manner. I focused particularly on the 'Introduction,' 'Methodology,' and 'Discussion' sections, with particular emphasis on communication of our qualitative results and design and review of our charts and figures.

5

Users' Understanding of Signal PINs

Contents

5.1. Introduction	88
5.2. Enhancing Signal User Authentication	90
5.3. Related Work on Secure Usable Messaging	93
5.4. User Study Design and Methodology	95
5.4.1. Study Design	95
5.4.2. Recruitment and Demographics	96
5.4.3. Limitations	97
5.4.4. Ethics	99
5.5. Results	99
5.5.1. RQ1: Comprehension	100
5.5.2. RQ2: PIN Recall and Reminders	103
5.5.3. RQ3: PIN Re-use and Composition	106
5.6. Discussion of Recommended Enhancements	110
5.7. Conclusion	112
5.8. Author Contribution	113

5.1. Introduction

Signal is an encrypted messaging application that is dedicated to preserving the privacy of its users. It implements features toward that end, such as not centrally storing users' contact lists, messages, or location histories unencrypted. Signal has historically relied only on users' telephone numbers for identification, authentication (via SMS), and contact discovery. Unfortunately, these methods are insufficient against attacks, including SIM-swapping [140–142]. In addition, these have some usability issues such as users who lose access to their telephone numbers also lose their Signal contact lists. Finally, they hamper additional features requiring additional metadata, like user profiles.

To improve the app in terms of these shortcomings, Signal released two new features: *Secure Value Recovery* (SVR) [143] and *registration lock* [144]. Both features require the user to select a PIN, which can be a sequence of numbers, like a traditional PIN, but also include letters and symbols. SVR uses the PIN to recover encrypted backups of contacts and settings stored on Signal servers. The registration lock aims to prevent anyone but the original user from creating a Signal account for a phone number without the associated PIN. The security of both features relies solely on the assumption that an attacker does not have access to the victim's phone number *and* cannot guess the PIN.

Signal's choice of naming the credential a "PIN" (as in, personal identification number) may not clearly indicate to the user the importance of the PIN in the Signal ecosystem. Unlike device or screen lock which is familiar to users, the in-app use of the Signal PIN is meant to achieve an app-specific purpose not satisfied by the device or operating system's features. Observe that the phrase "local authentication" is also used for this case of an app-specific KBA [114]. A banking app for example might mostly be using local authentication to protect access to an OAuth token, while Signal has a different goal.

As Signal represents one of the first, large-scale usages of in-app or "local" PINs, in this chapter we investigate to what extent do participants, both the security-/privacy-savvy and the general population, understand the PIN feature and what effect does this have on their choice and usage? Additionally, we also investigate how participants react to Signal's PIN verification reminders that encourage users to not only select a complex PIN but regularly remind users to reenter it for verification. This feature may have been implemented because the PIN is not meant for daily use, but instead only needed in acute moments of setting up a new device with the Signal app. Finally, we examine the way participants select and compose their Signal PINs and the effect of their general understanding of the underlying Signal features to make these decisions. To this end, we consider the following research questions:

- RQ1** Are participants aware of how and why in-app PINs are used in Signal?
- RQ2** How effective are PIN reminders assisting participants to remember PINs?
- RQ3** How do participants choose and compose a PIN for Signal, and does their understanding of how these PINs are used affect that choice?

We surveyed Signal users ($n = 235$), asking about their understanding, usage of the Signal PIN feature, and response to Signal PIN verification. For example, we asked participants to explain the purpose of Signal PINs, in their own words. We additionally asked participants about the composition of their PIN such as the length and character set, if they reuse the PIN in other contexts like phone lock or in another messenger app, if they have opted out of selecting a PIN altogether, and their response to periodic PIN verification.

We find that only 14% ($n = 33$) of respondents opted out of setting a Signal PIN, and also we find a large disparity between the practices of participants who can explain the purpose of the local PIN authentication (who we term Signal *enthusiasts*; $n = 132$; 56%) and those who cannot (dubbed *casual* Signal users; $n = 103$; 44%).

Many enthusiasts set PINs because they thought it was required — initial communication from Signal indicated that it was, although it is not in current versions of the app. Many enthusiasts also specifically mentioned registration locking and cloud backups. Interestingly, when enthusiasts did not set a PIN, 44% cited anti-cloud storage sentiments, indicating that they are aware of the features Signal PIN provides such as cloud backups of profiles but felt that this metadata storage did not sufficiently guard their privacy. Among casual users, 25% set a PIN for generalized security reasons although they are not able to clearly articulate what those might be. Moreover, 13% set a PIN simply because they were prompted by Signal or do not know why they actually set a PIN (16%). If casual users did not set a PIN, they typically indicate that it was inconvenient (18%) or they did not see the necessity (18%). Their inaccurate understanding also affects this decision: 24% state that they do not need an additional safeguard to secure access to their Signal app although the PIN is not used for this purpose.

Very few participants who set a PIN indicated that they had difficulty remembering their PIN; only 12% said they *occasionally*, *frequently* or *very frequently* have difficulty remembering. When interacting with the periodic reminders to verify their PIN, 59% confirm their PIN *frequently* or *very frequently*. Only 24% of all participants confirm their PIN *rarely*, *very rarely*, or *never* when prompted, yet, here the behavior of enthusiasts and casuals diverges: 16% of the latter tend to ignore the reminder prompt compared to 28% of the enthusiasts. In addition, 45 or 24% of the participants who currently use a PIN disabled these reminders. When asked why,

67% of the enthusiasts mention that they use a password manager while casuals are mostly annoyed (42%) or do not feel it is necessary to be reminded (33%).

We also find that enthusiasts' PINs are more password-like, often containing numbers, letters and symbols. Compared to casuals, enthusiasts on average choose PINs with an additional 1.3 digits, 3.0 letters, and 1.3 special characters. Moreover, many participants, particularly enthusiasts, use a password manager to store their Signal PIN, which additionally increased the complexity of their PIN: password manager users selected PINs with an additional 2.1 digits, 5.3 letters, and 3.1 special characters compared to non-password manager users. A number of participants, both enthusiasts and casuals, noted the reuse of their Signal PIN in other contexts, apps, and as their screen lock, yet, 76% of the participants who use a PIN within Signal said they do not reuse it.

In short, it appears Signal's core audience of privacy-conscious enthusiasts is using the PIN effectively, however, this roll-out may have been affected by inconsistent communication. Some earlier versions of the app made PIN creation a requirement. In addition, Signal PINs can contain letters and special characters. Weak Signal PIN choices can have consequences for those that choose secure PINs as secure communication requires both parties to be secure. We would recommend that Signal consider adding features to encourage better choices, like an improved blocklist, or even re-branding Signal PINs to more accurately depict their use, like "Account Recovery Passwords," which could help users apply the right context during selection and storage of this credential. Though our focus is on Signal, our results may inform communication strategies of other app developers, since account recovery and registration lock features are common in secure messaging.

All our findings were shared with the Signal developers.

This work appeared in *Symposium on Usable Privacy and Security (SOUPS '21)* in collaboration with Philipp Markert and Adam J. Aviv.

5.2. Enhancing Signal User Authentication

Signal is an open-source app and service, developed and operated by the non-profit Signal Technology Foundation. Signal implements the underlying Signal protocol which includes forward secrecy [145, 146] and is used by other secure messaging clients, like WhatsApp [147] and Facebook Messenger's secret conversation feature [148]. Signal boasts more privacy consciousness in its design and implementation, eschewing linkages to an identity or collection of metadata, as compared to its competitors, like Telegram, WhatsApp, or Threema [149, 150]. Hereinafter, when we refer to *Signal*, we mean the app/service and not the protocol unless otherwise specified.

Given its focus on privacy, Signal historically relied on a user's mobile phone number as an identifier, reasoning that this system was already in place. This approach also makes migrating to a new device easier for users when using the same phone number, as long as the user's contacts were already backed up by other means. Other app settings, e.g., groups and blocked contacts, were formerly not backed up.

Additionally, receiving a valid SMS with a security code was sufficient to re-establish an account with Signal to send/receive encrypted messages. Unfortunately, phone numbers can be subject to SIM-swapping attacks [140–142], whereby an attacker is able to register an existing phone number with a new mobile SIM card, effectively stealing a user's account on Signal.

To address both backing up device settings and preventing account hijacking, Signal introduced two new features: *Secure Value Recovery* [143] and *registration lock* [144]. Both services require an additional authentication check, namely a PIN, and in the rest of this section, we describe Secure Value Recovery, registration lock, and how Signal rolled out PINs.

Secure Value Recovery *Secure Value Recovery* (SVR) enables encrypted backup and recovery of the Signal app settings, including contacts, profile, and group memberships. The backup data is encrypted and stored on Signal's servers. When a user migrates to a new device, the goal is to restore this data into the new app installation. As the decryption needs a key, the user has to choose, recall, and enter a PIN which is input to a key-derivation function. The resulting symmetric master key is used to further derive the backup encryption key.

Registration lock The registration lock is an optional feature that binds the Signal PIN to the user's phone number. This way knowledge of the PIN is required as a second authentication factor in addition to the ability to receive an SMS with a one-time security code. This approach protects Signal from attacks like SIM swapping [140–142] where an attacker can obtain the SMS code.

To realize this functionality, the protocol uses the symmetric master key that is calculated as part of SVR, this time to derive a 32-byte registration lock hash. This value is used similarly to a password: it is sent to the server to authenticate the user. If the calculated registration-lock hash matches the one that is stored on the Signal server, the SMS code is sent. If not, the SMS code will not be sent and the registration of the phone number cannot be completed.

On the other hand, if an account needs to be migrated to a new device and the user does not know the PIN, setting up the account with the phone number is only possible after 7 days of inactivity. After this time span, the server's registration lock hash (of the PIN) expires and a new account can be created. However, the counter will be reset each time the client connects to the Signal server which happens when

receiving or sending messages. Additionally, the iOS or Android apps make requests on a regular basis to keep the PIN hash alive even if the app itself is used infrequently.

Signal PINs Unlike PINs used to authenticate to gain access, like unlocking your phone, the Signal PIN is used as a secondary authentication factor when moving an account from one device to another. A user does not need to enter the PIN to use Signal once it is installed on a particular device. However, Signal has a separate setting that locks the application from unauthorized access by forcing the user to verify their mobile phone's unlock authentication, such as the PIN used to unlock the device.

Also different than unlock authentication PINs, if a user forgets their Signal PIN while maintaining access to the Signal app, it can be reset without any repercussions as the current secure messaging keys can serve the purpose of authentication. After resetting the PIN, the SVR-encrypted backup can be re-encrypted and uploaded to Signal's servers, and the registration lock hash can be regenerated.

Communicating the purpose of the PIN to users, including all the features it does and does not support, is not a straightforward task. While Signal published an article explaining the technical details of SVR and registration lock [144], explaining it to all users remains a challenging task. Signal also originally required a user to establish a PIN, but later made that choice optional.

Finally, as the Signal PIN is only needed at acute moments, Signal employs periodic PIN reminders to help users memorize their PIN. These reminders to verify a PIN are spaced at regular intervals, starting at 12 hours, then 1 day, 3 days, 7 days, and every 14 days. Figure 5.1 shows the prompt that is shown to users for this purpose.

To encrypt the backup, the PIN is first turned into a 32-byte key k using the key derivation function Argon2. Afterwards, k and a label are used as input to HMAC-SHA256 to calculate an 32-byte identifier `auth_key`. The key k and a different label are input to HMAC-SHA256 to generate a 32-byte pseudorandom value `c1`. Observe that knowledge of this value along with a copy of an encrypted backup would enable an attacker to recover all other values.

A second 32-byte value `c2` is generated by a secure random-number generator. Subsequently, `c1` and `c2` are used as the input to calculate the 32-byte master key `master_key` using HMAC-SHA256. Finally, the master key is used as the input to an HMAC-SHA256 to calculate an application key which is used to encrypt the backup. By saving the random value `c2` along with the identifier `auth_key` in a secure enclave on the Signal server, a client is able to redo the whole computation as long as the PIN is known. In case the user forgets the PIN but still has access to the device, a new PIN can be set which will create a new backup. If the user forgets the PIN

and does not have access to the device, a new account can be created as long as the registration lock has not been set.

We observe here that most apps deploying new features collect plaintext metadata about the usage of the features, making tasks like these easier. Owing to its desire for user privacy, Signal has limited ability to gauge user behavior.

SVR was introduced by Signal to store metadata about user accounts in an encrypted way on the Signal servers. Thereby, it allows users to recover this information which includes the profile settings, contacts, and the list of blocked contacts, if their device gets stolen or lost.

Furthermore, SVR is necessary for Signal’s plan to support non-phone number based identifiers, such as a username, so that is possible to use Signal without revealing a phone number. Observe that as long as the contact list in the Signal app is based on the contact list of a user’s smartphone, Signal does not need this kind of additional backup mechanism. Now, if Signal allows users to register solely with a username, all contacts would be lost when switching to a new device or reinstalling the application.

5.3. Related Work on Secure Usable Messaging

In our study, we note a number of casual users with limited comprehension, a theme also observed in other circumstances of secure messaging. Abu-Salma et al. [151] noted that security and privacy is not always a leading driver in the adoption of a secure messenger like Signal, but rather community pressure of wanting to be able to reach specific contacts. De Luca et al. [152] and Das et al. [153, 154] come to a similar conclusion and show that the influence of social factors is not only limited to the adoption of messengers but security tools in general. Abu-Salma et al. [151] further note that many users have misconceptions about the security of messaging, such as the perception of SMS as secure for sensitive communication. Oesch et al. conduct a user study confirming user misconceptions and finding that group-chat users tend to manage security and privacy risks using non-technical means such as self-censorship and manually inspecting group membership [155].

The work of Unger et al. [156] divides secure messaging protocols into trust establishment, conversation security, and transport privacy. Observe that while Signal addresses each of these, the Signal PIN – and therefore our work, falls squarely into the problem of trust establishment. SIM-swapping attacks are one of the prime attacks against Signal’s former design. This class of attacks has been studied by Lee et al. and others [140, 142].

In general, Signal and other secure messaging services often face the problem of explaining secure protocols, however, authentication ceremonies are challenging for

users to understand [157, 158]. To address this issue, Wu et al. offered a redesign of the authentication ceremony that emphasizes comprehension [159]. Vaziripour et al. [160], on the other hand, suggested to partially automate the ceremony by using social media accounts. The Signal PIN is used for key derivation and is an example of a *usable encryption* scheme in the real world. These have been previously studied by Ruoti et al. [161] who propose a secure email system and study varying levels of user transparency and automation. While Signal aims for automatic key management and automatic encryption, Ruoti et al. find that users had more trust in an approach that emphasized manual steps and therefore comprehension. While our research aims to understand how comprehension affects users' PIN practices, similar efforts to better communicate about this feature would likely help users.

Recently, Khan et al. [89] and Casimiro et al. [90] studied PIN reuse across different contexts. Both find that re-use is rampant, and that users tend to have a small set of PINs they use regularly. In our work we also find that certain kinds of PIN re-use is common for Signal PINs, such as for an ATM/Credit/Payment card. As Signal PINs are generally chosen and entered on mobile devices, users may be less inclined to choose hard-to-guess, full-fledged, alphanumeric passwords with special symbols. (Recall that a Signal PIN can have numbers, letters, and special symbols.) Melicher et al. studied user selection of passwords on mobile devices [115], finding that the limitations of the keyboard setting may lead to more easily guessable and weaker passwords.

In our work, we find that participants using a password manager are more likely to select strong Signal PINs. Unfortunately, in the mobile setting, users remain challenged in using password managers. Seiler-Hwag et al. investigated common password managers on smartphones [162], finding that all score poorly on standard usability metrics. Even when a password manager is adopted, using the password generation feature is not a given for all users. Pearman et al. [163] studied why users do (and do not) adopt a password manager and find that even those that do use a password manager may not use the password generation feature.

To the credit of the Signal team, they understood that the Signal PIN is unlike the case of mobile unlock authentication where a typical user unlocks the device multiple times per day. Instead, they realized an infrequently-used PIN is much more subject to being forgotten by the user. So they employ the well-known technique of *graduated interval recall* (also called *spaced repetition*). While the positive effects on recall rates have been shown in multiple studies [164–167], including the memorability of passwords [168–173], the usage of it in this context is novel. The deployment of Signal's periodic reminders to verify the PIN offers a real world example of the effectiveness of this strategy.

Our work is related to the area of messaging and usability of secure systems. For example, a lot of work has been done on the design of security warnings and the decisions made by users. Considering the browser setting, Egelman et al. investigate the effectiveness of browser phishing warnings [174]. Biddle et al. report on the feature added to browsers to communicate that a site has an Extended Validation TLS certificate [175]. Sunshine et al. and Felt et al. report on the sometimes-bad decisions made by users in the face of repeated warnings about TLS certificates [130, 132, 176]. Reeder et al. and Akhawe et al. continue this work studying user reactions to browser warnings [131, 177]. Other research also covered the design of warning messages in general [178].

Bravo-Lillo et al. investigate the design of security decision user interfaces and build a mental model of how and why users respond to or ignore warnings [179–181]. Shay et al. focus specifically on guidance given to users during the password creation process [118].

5.4. User Study Design and Methodology

We conducted a user study of $n = 235$ Signal users recruited to complete a survey about their understanding and strategies for managing their Signal PINs. In this section, we provide details of the survey, recruitment, limitations, and ethics.

5.4.1. Study Design

We recruited participants in two samples. The first sample was from Reddit, the Signal Community Forum, and snowballing; the second sample via Prolific. For participants completing the study on Prolific, we first used Prolific’s built-in screening to only recruit participants who use Signal, and as this pool was still insufficient, we used a single screener question (Appendix B.1) as part of a two-part recruitment, where participants noted which messaging app they used. Those using Signal were invited to the main study. The entire survey is provided in Appendix B.2, and it took participants 7 minutes, on average, to complete.

1. *Informed Consent*: All participants were informed of the procedures of the survey and provided consent. The informed consent notified participants that they would be asked to complete a short survey that asks questions about how they select PINs and how they feel about Signal’s implementation.
2. *Signal Usage*: Participants must indicate they are a Signal user answering the question: “Do you use Signal?”(Q1) All participants who responded in the affirmative continued with the survey.
3. *PIN Comprehension and Usage*: Participants were now prompted with the text: “PINs are a new feature provided by Signal. In your own words, please

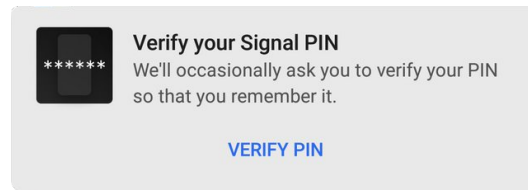


Figure 5.1.: Prompt used by Signal to occasionally ask users to verify their PIN

explain how PINs are used by Signal,” (Q4) followed by “Did you set a Signal PIN?”(Q5) and why they did (Q6a) or why they did not (Q6b). Those who did not set a PIN skipped ahead to Q25.

Those who did set a PIN were asked if the PIN was since disabled (Q7), and if so, why (Q8). We also asked participants who still had their PIN enabled if they have difficulty remembering their PIN (Q9), and what they would do if they forgot their PIN (Q10).

4. *PIN Reminders*: We then asked a series of questions (Q11-Q14) on Signal’s periodic PIN reminders (cf. Figure 5.1), including if participants currently have the reminder set; for those who do, how frequently they verify the PIN when prompted; and if they disabled it, why.
5. *PIN Reuse and Sharing*: Participants were asked to report if they reuse their Signal PIN in other contexts, such as mobile device unlock (Q15), ATM and other payment cards (Q16), and other mobile applications (Q17). In addition, participants were also asked if they have shared their PIN with friends or family (Q18). These questions were derived from related work on PIN usage [89, 90].
6. *PIN Selection and Composition*: The survey continued with a series of questions about PIN length and composition, as well as the perceived strength of the PIN (Q19-Q24).
7. *Other Messengers*: The survey continued by asking about the use of PINs in other messengers, including Facebook, Skype, Telegram, WeChat, and WhatsApp (Q29). We also asked if the Signal PIN is reused in any other messenger as well as the reasons for doing or not doing so (Q30a/Q30b).
8. *Demographics*: Finally, we asked about demographics (D1-D5), including age, gender, and IT background.

5.4.2. Recruitment and Demographics

We recruited a total of $n = 235$ participants. Of those 170 were recruited from Reddit, the Signal Community Forum, and snowballing, and 69 were recruited on Prolific. We posted to Reddit’s r/SampleSize and r/Signal forums; and the Signal Community Forum. We decided against a fixed payment for these participants in favor of not collecting any personally identifiable information, such as an email ad-

dress to offer a gift-certificate via a raffle, and thus these participants took the survey voluntarily without compensation.

We used Prolific’s built-in custom prescreening filters, which allow researchers to post a study to participants that meet specific criteria, such as residing within the US. We applied the custom prescreening for Prolific members who indicated Signal is one of the “chat apps” they use regularly. We were able to recruit 69 participants this way, each paid GBP 1.50. To expand the Prolific pool, we also employed a custom screening survey to find other Signal users, recruiting 500 responses (paying GBP 0.15). Those who indicated that they used Signal were invited to the main study (paying GBP 1.50). We were able to recruit an additional 11 participants this way.

As shown in Table 5.1, the demographics of our sample is skewed toward a younger, more male-identifying, and more IT-oriented group. On the other hand, our participants reside in many different countries increasing the generality of our results. Of the 235 participants, Germany accounted for (68; 29%), the USA for (61; 26%); the UK for (24; 10%). The rest of the world was the largest group with (82; 35%). The actual demographics of the Signal community at large are unknown, so the skew towards a certain participant pool may reflect our recruiting strategy or may be influenced by the makeup of the underlying community. We observe that at 75%, males make up the largest cohort. Similarly, at 64%, those with IT-focused education or employment make up a majority of participants. In terms of education, bachelor’s and master’s groups combined account for 55% of participants. Our group of enthusiasts is also male-dominated: self-identified males outnumber females more than 8:1. Finally, we note that among enthusiasts, the IT-focused group is substantially larger at 3:1, while the figures are more balanced for casuals: about 1.2:1. It is reasonable to surmise that an IT background makes one more likely to be an enthusiast — put another way, Signal’s existing communication strategy about the Signal PIN appears to be more effective for those with an IT background.

5.4.3. Limitations

As this study took place online, it shares the usual limitations of many online studies, such as finding a representative recruitment. On the one hand, our sample may not be a representative sample of all Signal users. Though we did not explicitly sample enthusiasts and casuals separately, we found that comparatively more enthusiasts were recruited via Reddit and Signal Community Forum, which led us to perform additional sampling from Prolific.

As an online survey, this study necessarily relies on self-reported data. With regard to security and privacy user studies, Redmiles et al. [182] show online-survey responses generalize quite readily to the broader population. Additionally, we con-

Table 5.1.: Demographics of participants divided by subgroups

	Enthusiasts		Casuals		Total	
	No.	%	No.	%	No.	%
Gender	132	56 %	103	44 %	235	100 %
Male	106	45 %	71	30 %	177	75 %
Female	13	6 %	24	10 %	37	16 %
Non-Binary	1	0 %	1	0 %	2	1 %
Other	1	0 %	0	0 %	1	0 %
Prefer not to say	11	5 %	7	3 %	18	8 %
Age	132	56 %	103	44 %	235	100 %
18–24	31	13 %	14	6 %	45	19 %
25–34	57	24 %	53	23 %	110	47 %
35–44	29	12 %	17	7 %	46	20 %
45–54	7	3 %	10	4 %	17	7 %
55–64	5	2 %	3	1 %	8	3 %
65–74	0	0 %	3	1 %	3	1 %
75 or older	1	0 %	0	0 %	1	0 %
Prefer not to say	2	1 %	3	1 %	5	2 %
Education	132	56 %	103	44 %	235	100 %
Some High Sch.	0	0 %	3	1 %	3	1 %
High School	31	13 %	12	5 %	43	18 %
Some College	0	0 %	0	0 %	0	0 %
Trade	0	0 %	4	2 %	4	2 %
Associate's	3	1 %	6	3 %	9	4 %
Bachelor's	35	15 %	32	14 %	67	29 %
Master's	38	16 %	25	11 %	63	27 %
Professional	9	4 %	3	1 %	12	5 %
Doctorate	10	4 %	12	5 %	22	9 %
Prefer not to say	6	3 %	6	3 %	12	5 %
Country	132	56 %	103	44 %	235	100 %
Germany	48	20 %	20	9 %	68	29 %
USA	25	11 %	36	15 %	61	26 %
United Kingdom	7	3 %	17	7 %	24	10 %
Other	52	22 %	30	13 %	82	35 %
Background	132	56 %	103	44 %	235	100 %
Technical	96	41 %	54	23 %	150	64 %
Non-Technical	33	14 %	44	19 %	77	33 %
Prefer not to say	3	1 %	5	2 %	8	3 %

ducted extensive pilot testing among members of our research groups and trusted colleagues to identify any ambiguities in our survey questions.

Another limitation is that participants' responses may suffer from the well-known tendency toward providing socially-desirable answers [183,184]. For example, it is possible that PIN reuse is more prevalent than our study suggests, or that people

choose PINs that are shorter and have less-diverse composition. The same holds for questions where we asked participants about their own understanding, where they might have looked up answers on Signal’s website. Despite this possibility, the answers provided appeared unique and participants provided many apt phrases to describe the situation. Additionally, we did not find responses that were directly cut and paste from Signal’s website.

5.4.4. Ethics

The study was administered at an institution that does not have an Institutional Review Board (IRB), but we still followed all appropriate study procedures similar to studies that obtained IRB approval. For example, participants were informed about the nature of the study, participated voluntarily, and could opt-out at any time. Additionally, we conformed with the ethical principles laid out in the Menlo Report [97], for example, we minimized any potential harm by not collecting any personally-identifiable information from our participants.

As described above, we completed two recruitments, one with paid and one with unpaid participants. Unpaid participants were recruited via Reddit, Signal Community Forum, and snowballing. We decided not to pay those participants as paying a comparatively small amount did not appear to withstand the harm that went along with collecting email addresses. Additionally, as this community tends to be more privacy-conscious, doing so might have depressed participation. Participants recruited via Prolific were paid GBP 1.50 for successfully completing the main survey, as this amount is in line with the recommended rewards on Prolific [185].

5.5. Results

In this section, we present the results of our study of $n = 235$ Signal users. For the structure of the section, we follow our three research questions: we start by analyzing the comprehension of the usage of PINs in Signal (**RQ1**), continue with user responses to the reminder feature (**RQ2**), and conclude with PIN selection and composition (**RQ3**).

For qualitative analysis, we had a primary coder code all the qualitative responses, producing an initial codebook. A secondary coder used that codebook to independently code the same responses, and afterward, the two coders met to resolve differences to produce a final codebook. The primary coder then used that final codebook to re-code the data. The codebook used for each qualitative question can be found in Appendix B.4.

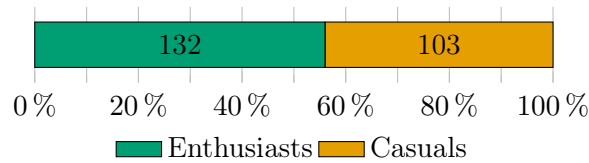


Figure 5.2.: Classification of the participants based on the participants' ability to explain the usage of PINs in Signal (Q4).

5.5.1. RQ1: Comprehension

As part of **RQ1**, we seek to understand Signal users' awareness and understanding of PINs and how they fit into the Signal ecosystem. To answer this question, we divide the participant pool by those that have or have not adopted a Signal PIN, and also by those that demonstrate understanding of how Signal uses the PIN.

Understanding Signal PINs After indicating if they are a Signal user (Q1–Q3), we first ask participants to describe how Signal PINs are used in their own words (Q4): *PINs are a new feature provided by Signal. In your own words, please explain how PINs are used by Signal.* These responses were coded by comprehension and accuracy; specifically, we seek to understand if the participants recognized that PINs are used for SVR and registration lock. Participants who accurately described the usage of Signal PINs were coded as *enthusiasts* ($n = 132$; 56%), and those who could not describe Signal PIN usage were coded as *casual* Signal users ($n = 103$; 44%).

We observed many different ways of capturing the main elements of how PINs are used by Signal. Many of the enthusiasts were even able to demonstrate a deep understanding, for example P10 said:

“It protects data like settings and group membership and signal [sic] contacts that will be stored on Signal’s servers using SVR. Previously this was only stored locally on a user’s device and was lost upon device reset or getting a new device unless a full backup was made on Android.”

Participant responses were assigned one or more codes based on the aspects correctly described. Overall among enthusiasts, the most popular codes were backup (65; 49%), encryption (45; 34%), contacts (31; 24%), and registration (23; 17%). Some also noted settings (8; 6%), profile (4; 3%) or groups (3; 2%), which are also secured via a Signal PIN during backup, and a few specified key derivation (7; 5%). Some also mentioned that PINs were part of a process for Signal to move away from using phone numbers for identity (6; 5%). A handful of enthusiasts also expressed anti-cloud sentiments when asked about Signal PINs (2; 2%), suggesting that they

understood that the PINs play a role in the encrypted cloud backup functionality of SVR, and that they are opposed to that design direction.

For the casual users, a majority (57; 55%) provided non-answers, or answers that do not indicate any understanding of the way the Signal PIN is used. The answer of P47 accurately summarized the reasoning we observed for many casual users:

“I don’t understand their purpose very well. I thought that they might be using the PIN system to verify the identity of the person using signal (if for instance someone unauthorized gained access to the phone), but the way that pin entry is optionally offered every few weeks doesn’t align with such a purpose. as such, I have no idea what they’re trying to accomplish.”

As the majority of casual users didn’t know or provided non-answers, there are many other examples to choose from, including “I initially thought it was used as a local PIN to unlock the app on my phone. It doesn’t do that so I have no idea how it works,” from P62. Additionally, many casual users falsely associated PINs with securing messages (21; 20%) although messages are not part of the backed-up data and are not protected by the PIN, as explained by P183: “Keep your messages on Signal encrypted via use of the PIN.”

An equal number felt that the PIN locks the Signal app (21; 20%), while in fact that functionality is called Signal Screen Lock and is not related to the Signal PIN — for that feature, Signal simply re-uses the device’s existing PIN, biometric, or other authentication scheme. An example of this response is from P37: “Protect application from opening from an unlocked phone.” Similar responses show this is a common misconception: “Pins are used to prevent unauthorized access to the app” from P227. Some individual participants also mentioned security as a general topic, without further describing it (2; 2%), or associated the PIN with inconvenience (1; 1%).

Why did participants set a PIN? In addition to knowing if participants understand the usage of the PIN, we also want to analyze how many actually set a PIN in their Signal app. In total, 202 or 86% of all 235 participants adopted a PIN. If we further divide those 202 participants based on their understanding, we see that more enthusiasts (116; 57%) than casuals (86; 43%) set a PIN.

To get a deeper understanding, **Q6a** asked participants to explain their decision. By far the most popular reason, equally distributed among enthusiasts and casuals, is *security*: 48 or 24% mentioned it in their answer. Once again, we find that enthusiasts display a detailed, in-depth understanding, exemplified by P14:

“I want to be able to use secondary identifier once it becomes available and not to lose my contacts that are not in my phone’s contacts list. I also want to be secure against SIM-swap attacks.”

This code is followed by participants mentioning that they were required to set a PIN (33; 16%). Among enthusiasts, we observed 25 that mentioned it was required (or 22%). P164 said “I had absolutely no choice if I wanted to continue to use Signal. Eventually, the box asking you to create a PIN kept you from opening any of your messages until you did what it wanted.”

This response may reflect the changing nature of the PIN requirement. Initially, it was required and then in a subsequent version, merely encouraged. The enthusiast-casual split here suggests perhaps more enthusiasts were early adopters of the Signal PIN. Another theme, of setting a PIN due to annoyance (12; 11%) may also reflect this changing communication strategy for Signal PINs. See for example Figure B.1, showing the initial prompt used by Signal to ask users to create a PIN; the prompt has subsequently been updated to B.2, current as of this writing. Observe the communication is also different when a user wishes to change their PIN as shown in Figure B.3, again current as of this writing.

Enthusiasts also regularly noted registration lock as a reason to set a PIN (14; 12%). P3 said “The PIN stop [sic] others from registering as me, and also protects access to my account details (profile, settings, contacts) if my device is misplaced.”

Casual Signal users noted *security* most frequently (26; 29%), but did so in a more general way as seen in this quote from P141: “for security and for reassurance if device gets stolen.” Additional codes include *don't know* (16; 18%) and *prompted* (13; 15%), suggesting that many casual users selected a PIN simply because they were prompted to do so and had no other underlying motivations. For example, P155 responded “I trusted the app and just did it when prompted.”

Why did participants not set a PIN? A total of 33 (14%) participants chose not to set a PIN (see Figure 5.3). A roughly equal number of enthusiasts and casual Signal users did not set a PIN: 16 enthusiasts (12%) did not set a PIN and 17 (17%) casual Signal users did not set a PIN. A χ^2 test revealed no significant differences between the groups.

When these $n = 33$ participants described why they did not set a PIN (Q6b), there were a number of differences. Both casual (3; 18%) and enthusiasts (4; 25%) described PINs as inconvenient, but casual users were more likely to note that either they do not need a Signal PIN (3; 18%) or that their phone lock provided security (4; 24%). For example P227 noted that their “. . . phone is always locked” and “Additional authentication seems unnecessary.”

Enthusiasts expressed distrust as a reason for not setting a PIN. Either this distrust is in the security of PINs for key derivation and management (3; 19%), or they distrust cloud storage (7; 44%). Distrust of cloud storage stems from privacy concerns with the SVR feature that backs up contacts and settings. P216, for example

stated, that they “had no desire to have any contact data uploaded,” and P207 said “i [sic] do not want to store personal information in the cloud.”

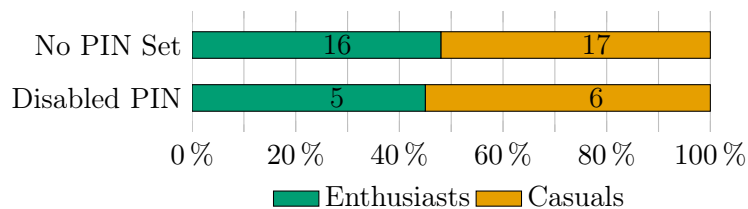


Figure 5.3.: Classification of the participants who disabled or did not set a Signal PIN.

Why do participants disable PINs? On top of the 33 users who declined to set a PIN, a total of 11 (5%) set a PIN and then later disabled it: 5 (45%) enthusiasts and 6 (55%) casual users, as shown in Figure 5.3. When asked to explain why they disabled their PIN (**Q8**), participants mentioned that the PINs were annoying (4; 36%) or inconvenient (2; 18%), which may be related to the periodic verification reminders. P212 explicitly mentioned the “verification overhead.” Anti-cloud hesitation to store data on Signal’s servers led (3; 27%) participants to disable their PIN: “Don’t want my data stored on their server” (P193). We also observed (2; 18%) participants who simply stated that they “do not need it” (P206).

RQ1 Results summary Signal users in our sample break down into two groups: enthusiasts who were aware of the features Signal PINs enabled, and more casual Signal users who were unable to describe how PINs are used within Signal. In both groups, though, setting a Signal PIN was highly prevalent. Only 33 of the 235 respondents chose not to set a PIN. Among enthusiasts, their choice to not set a PIN stemmed from either distrust in the key-derivation process or hesitancy to store information in the cloud generally. Casual users did not set a PIN because of inconvenience or a false belief that other authentication mechanisms, like locking their phone, provided adequate protection. When participants disabled their Signal PIN, inconvenience or annoyance were often cited, sometimes referring specifically to the periodic reminders.

5.5.2. RQ2: PIN Recall and Reminders

In this section, to address **RQ2**, we consider how participants remember their PINs and their reactions to the periodic PIN verification reminders. Throughout this section we consider the $n = 191$ participants who still have their PIN enabled, and not the 11 participants who since disabled their PIN.

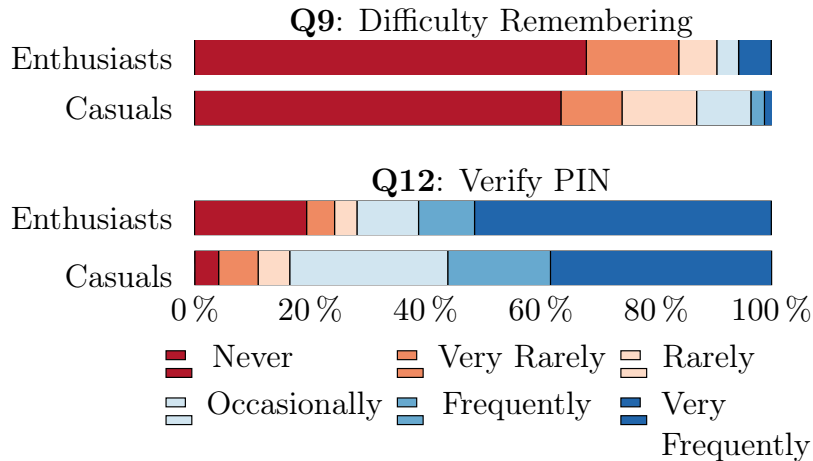


Figure 5.4.: PIN memorability and verification

Forgetting PINs We asked the ($n = 191$) participants who still use a Signal PIN in **Q9** if they encountered difficulty in remembering their PIN. Overwhelmingly, 89% of participants ($n = 170$) indicated that they *never*, *very rarely*, or *rarely* have difficulty remembering their PIN (see Figure 5.4; *top*). We compared the response to this question from enthusiasts ($n = 106$) and casual ($n = 85$) Signal users who still had their PIN enabled, and we found no statistical differences.

We asked participants in **Q10** what they would do if they forgot their Signal PIN. (Note that the PIN is not required to use Signal for messaging, and can be reset at any time in the settings menu.) Many enthusiasts noted that their PIN was stored in their password manager (45; 42%), and they would simply look it up. Fewer casual participants mentioned a password manager (12; 15%). A number of participants did not know what to do (27; 25% enthusiasts and 33; 40% casuals), while a few casuals suggested they would contact Signal (4; 5%) and two enthusiasts said they would reinstall the app (2; 2%). Others believed that their Signal account is now unrecoverable (2; 2% enthusiasts and 3; 4% casuals); some would create a new account (4; 4% enthusiasts and 5; 6% casuals). A handful (2; 2% enthusiasts and 4; 5% casuals) denied that they would forget stating “It is a PIN I use for my bank cards” (P145), for example. A small number of participants noted that they would wait (8; 7% enthusiasts and 4; 5% casuals), aware that the registration lock expires after 7 days of inactivity.

Periodic verification Perhaps recognizing that Signal PINs are only truly required when transferring a Signal account to a new device, Signal decided to employ *graduated interval recall* [165] (or, *spaced repetition*) that regularly prompted participants to verify their PIN when opening the Signal app. An example of such a reminder

is found in Figure 5.1. To our knowledge, Signal is the first mainstream app to implement such a feature.

We first asked participants if they were aware of the PIN verification reminders (Q11). Most participants ($n = 176$; 92%) indicated that they were aware, and a follow up question (Q13) asked if they have since disabled the reminders. Seventy-four percent ($n = 131$) of participants have the periodic PIN verification enabled, and many still verify their PIN when prompted. Seventy-six percent ($n = 135$) of participants either *occasionally*, *frequently*, or *very frequently* verify their PIN when prompted. When dividing this data by enthusiasts and casual Signal users (see Figure 5.4; *bottom*), we did not observe significant differences between frequency of PIN verification using a Mann-Whitney U test.

The remaining 23% ($n = 45$) disabled the PIN reminders. These 45 participants were asked why they disabled the reminders (Q14): (23; 51%) mentioned doing so because they use a password manager. P63 said “I don’t remember my PIN, it’s stored in my password manager, frankly, I don’t even want to remember it.” Ten (22%) said there was no need or their PIN was already memorized, and a further (11; 24%) found the reminders annoying. These figures suggest that the periodic reminders are generally viewed as beneficial, or at least not substantially invasive enough to warrant disabling them. As we rely on self-reported data, we do not independently verify PIN recall rates.

Password manager usage We found a large amount of password manager (PM) usage in our study. These reports were entirely unprompted as PMs were not mentioned in any survey material. Thirty-one percent ($n = 62$) indicated that they use a PM in response to questions regarding either what they would do if they forget their PIN Q10 or how they select their PIN Q20. As we did not explicitly ask about PM usage, the true number of PM users might be higher.

More striking is the combination of the classification of enthusiasts and casual participants combined with that of PMs: (52; 83%) of the 62 participants who said they use a PM were enthusiasts. Or, 50% of the 103 enthusiasts who have a PIN enabled use a PM. Only (10; 14%) of the 73 casual Signal users using a PIN mentioned PMs as a mechanism to either select or recall their PIN. Put another way, participants who mentioned a PM were overwhelmingly enthusiasts.

RQ2 Results summary Participants indicated that they have little difficulty remembering their PIN, many stating that this is a PIN they use all the time and thus would *never* forget it. A large number of participants, notably half of PIN-using enthusiasts, use password managers to both select and recall their Signal PIN, and are thus, not concerned with forgetting their PIN. Reactions to Signal’s periodic PIN verification requests were more mixed, but overwhelmingly participants verified

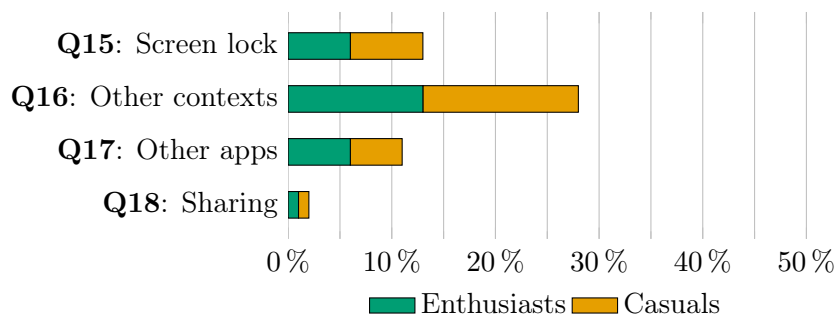


Figure 5.5.: Frequency of PIN reuse and sharing

their PIN when prompted. Roughly a quarter of participants disabled periodic PIN verification; most did so because they use a password manager. Others stated that the PIN was already memorized, so there was no need for the reminders, and some simply found the reminders annoying. Overall, since 76% of participants reported verifying their PIN when prompted, we conclude graduated interval recall used for Signal PIN verification is generally embraced by users, though the effectiveness of this intervention is obviously an area that deserves future work.

5.5.3. RQ3: PIN Re-use and Composition

In this section, we explore selection strategies of Signal PINs by asking participants if they re-use their Signal PIN in other contexts; the composition of their Signal PIN with respect to numbers, digits, and special symbols; and the perceived security of their Signal PIN in comparison to other PINs they use.

PIN re-use To explore the many ways in which PINs are reused, we adopted questions from Khan et al. [89] and Casimiro et al. [90] regarding PIN usage, more broadly. The responses of $n = 191$ participants using a Signal PIN are found in Figure 5.5, broken down by enthusiasts and casual users.

First, as a mobile application, we asked participants if they used their smartphone unlock PIN as their Signal PIN (**Q15**). Thirteen percent ($n = 26$) did so, composed of 12 enthusiasts and 14 casual users. In **Q16**, we asked if they used the Signal PIN in other contexts, ranging from ATM/Credit/Payment cards, to garage door codes, gaming consoles, and voice mail. (Refer to Appendix B.2 for the full list, derived from Khan et al. and Casimiro et al.) Twenty-eight percent ($n = 53$) of participants use their PIN in another context, consisting of (25; 43%) enthusiasts and (28; 53%) casual users. Among those who reused, casual users did so more often: 1.39 times on average, compared to enthusiasts who did so 1.24 times. The most common context of PIN re-use overall was for ATM/credit/payment cards where (17; 32%) of 53

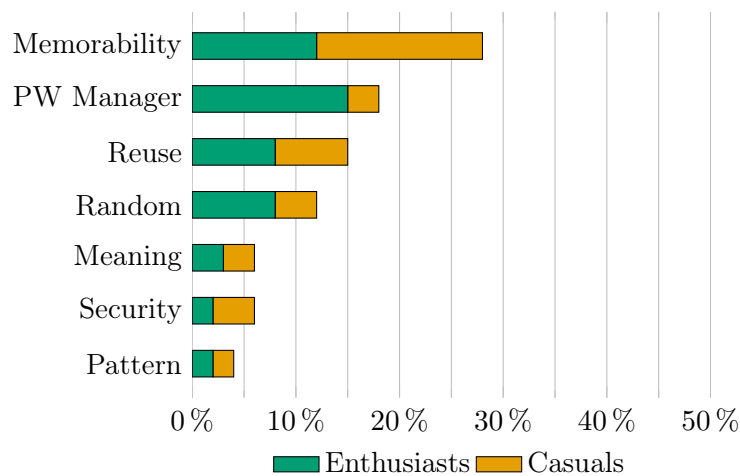


Figure 5.6.: Most popular codes assigned to the answers of **Q20**: What was your primary strategy in selecting your Signal PIN?

participants reused a PIN. Participants also mentioned laptop/PC authentication (13; 24%) and other online accounts (11; 21%).

We also asked if participants re-use PINs in other mobile applications (**Q17**): (21; 11%) reported they did, and of those, 12 were enthusiasts and 9 were casual users. Most commonly, the other app was WhatsApp ($n = 6$); WhatsApp implements the Signal Protocol. Other common mobile apps where this PIN was reused were banking apps ($n = 5$). In **Q25–Q28**, we asked participants if they use other messenger services, such as Facebook messenger, Telegram, and WhatsApp: (183; 95%) did. We also asked if they set a PIN in these services and found (49; 26%) did.

Finally, we asked if participants share their PIN with friends and family: this was rare. Only 3 participants did so, suggesting that PINs selected for Signal are not widely shared with others and are considered confidential.

PIN composition A participant’s understanding of the Signal PIN’s functionality had a large effect on the composition of their PIN. We asked participants what was their primary PIN selection strategy in **Q20**: code frequencies summarized in Figure 5.6 (with full details in Table B.8 in Appendix B.4).

Among enthusiasts, password managers (PM) were mentioned frequently (28; 26%). For example P100 noted that their “password safe generated it.” Some participants mentioned the name of their password manager explicitly, like KeePass or Bitwarden. Far fewer casual Signal users (6; 7%) mentioned a PM. The most-frequent code among casuals was *memorable*: (30; 36%), choosing a PIN easy to remember; among enthusiasts it was second-most frequent (23; 21%). For example, P7 noted their PIN was “Complicated enough but can still be remembered.” This result suggests that despite the prevalence of randomized password generation, most

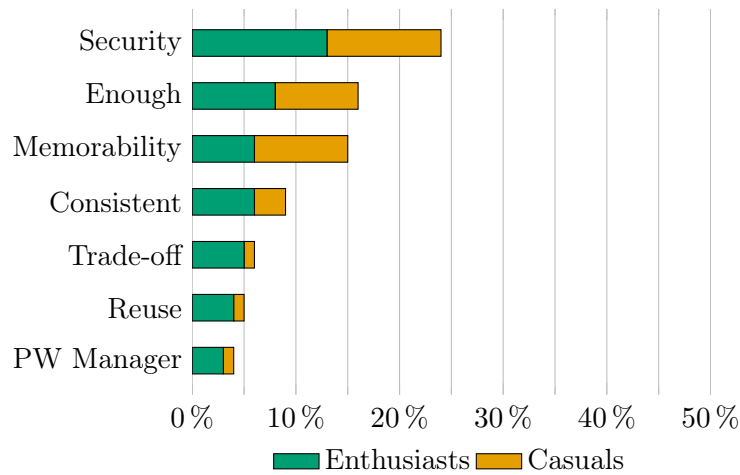


Figure 5.7.: Most popular codes assigned to the answers of **Q22**: Why did you choose a PIN with this security level?

participants want to select a PIN they can remember and recall easily, rather than having to look it up in a PM.

Interestingly, while the study of Markert et al. found dates to be the most-popular strategy for selecting a PIN, only 3 of our participants mentioned dates (2 enthusiasts and 1 casual) [186]. In the study of Markert et al. with ($n = 200$), *memorable* was the second-most frequent code (37; 19%).

We then asked participants why they select a PIN with the current “security level” (**Q22**). (Results summarized in Figure 5.7; full details in Table B.9 in Appendix B.4.) Among both enthusiasts (25; 23%) and casual Signal (20; 24%) users, many mentioned security; P44, an enthusiast, said “I am fairly security conscious.”

Casuals and enthusiasts roughly equally mentioned that they chose something that was simply *good enough*: (16; 15%) and (15; 18%) respectively. Slightly more casual users mentioned memorability: (12; 11%) enthusiasts and (18; 22%) casual users. A similar number of enthusiasts (11; 10%) and casuals (6; 7%) mentioned that they try to be consistent in their security choices around PINs (and authentication generally), for example “Because I always choose this security level” (P109).

Recall that while Signal refers to this secret as a PIN, it is not a traditional *personal identification number*, but rather has more of the properties of a password. We asked participants to provide metrics for how many numbers, characters, and special symbols they use in their Signal PIN (**Q24**). Participants were presented a slider for each class from 0 to 12. While it is of course possible that a participant might have more than 12 digits, as a practical matter more than this simply indicates the use of a PM, which we can see in our data. Results are shown in Table 5.2.

Enthusiasts on average chose PINs with an additional 1.3 digits, 3.0 letters, and 1.3 special characters, and length increased overall by 5.5 characters. Except for

Table 5.2.: PIN composition among different user groups of $n = 191$ participants who set a PIN and did not disable it. t -tests were performed between groups within categories; all p -values are displayed Bonferroni-corrected for 8 overlapping hypothesis tests.

Classification	Participants	Length	
		Mean (SD)	t -test
Enthusiast	106	12.7 (9.8)	$t = 4.65$
Casual	85	7.2 (5.7)	$p < 0.001^{**}$
PM User	62	17.3 (10.2)	$t = 9.42$
non-PM User	129	6.8 (5.3)	$p < 0.001^{**}$
Overall	191	10.3 (8.7)	–

Digits			Letters			Special Characters		
Mean (SD)		t -test	Mean (SD)		t -test	Mean (SD)		t -test
6.2 (3.3)		$t = 2.97$	4.4 (5.1)		$t = 4.57$	2.2 (4.1)		$t = 2.74$
4.9 (2.4)		$p = 0.026^*$	1.4 (3.3)		$p < 0.001^{**}$	0.9 (2.4)		$p = 0.05$
7.0 (3.7)		$t = 4.72$	6.7 (5.1)		$t = 8.79$	3.7 (4.7)		$t = 6.16$
4.9 (2.3)		$p < 0.001^{**}$	1.3 (3.2)		$p < 0.001^{**}$	0.6 (2.2)		$p < 0.001^{**}$
5.6 (3.0)		–	3.1 (4.7)		–	1.6 (3.5)		–

the number of special characters, we were able to observe significant differences between the enthusiasts and the casuals using a t -test with Bonferroni-correction (for 8 overlapping hypotheses).

When dividing the population by their use of PMs, the difference is even greater. (Note that more enthusiasts employed a PM.) PM users chose PINs with an additional 2.1 digits, 5.3 letters, and 3.1 special characters. Overall, they used PINs which are 10.5 characters longer on average. Using a t -test with Bonferroni-correction, we were able to observe significant differences for all those statistics.

RQ3 Results summary Many participants re-use Signal PINs in a number of ways. Roughly 15% indicated that they use their Signal PIN as their screen lock PIN, used to unlock their smartphone. Nearly 30% noted that the same PIN is used in other contexts, most commonly as an ATM/banking/payment card PIN. The Signal PIN is also reused in other mobile apps, such as a WhatsApp PIN, serving the same purpose as a Signal PIN for SVR and registration lock. When selecting a PIN, understanding of the purpose of Signal PINs led to much more diverse PINs, both in terms of the PIN length but also the presence of special characters and symbols. Among enthusiasts, the use of a password manager was particularly prominent when selecting a PIN, as compared to more casual users. But by far the largest factor in PIN selection overall is a desire for choosing a memorable PIN.

5.6. Discussion of Recommended Enhancements

Communicating about Signal PINs Our data show Signal's communication about the PIN feature has been effective for its traditional community of privacy enthusiasts. Without prompting, participants told us they learned about the PIN by reading blog posts, the Signal website, and tweets. Casual users, on the other hand, were much less likely to have exposure to these other sources. For this reason, in-app or in-the-moment resources nudging casual users in a more secure direction would almost certainly be of benefit.

As explained in Section 5.2, the case of Signal is especially challenging. While users are surely familiar with PINs as used in smartphone-unlock and payment-card scenarios, Signal PINs are actually used to *infrequently* derive encryption keys for SVR and *infrequently* act as a password for registration lock. Yet, despite the text in the Signal PIN enrollment prompt (see Figure B.2) saying "You won't need your PIN to open the app," many of the participants who did not set a PIN mentioned inconvenience as a reason for their decision.

When further exploring the cause for this, the name "PIN" itself, is likely causing confusion. The Signal PIN is fundamentally a countermeasure against account takeover and to offer recovery functionality. If for example, the Signal PIN were to be called the "Account Recovery Password," or perhaps "Restore/Recovery Password," that might better convey the usage pattern. Text could then inform the user of the ill consequences of a bad PIN choice. This end could be achieved with text like "This password protects you from account takeover." Re-framing the PIN in this way could break the users' mental association with device-unlock PINs while also inspiring dread of consequences. While our study does not directly measure the effectiveness of such an intervention, the themes we uncovered naturally point in this direction.

Encouraging password managers The Signal PIN ultimately is used to derive a symmetric key in SVR and to retrieve a copy of the encrypted profile backup. For this reason alone, it is worth encouraging users to generate and store their Signal PIN in a password manager (PM). Few users are willing to memorize long, random keys and a PM is much better at generation, storage, and recall of secrets. Importantly, the user interface of a PM is already designed to explain these concepts to a user. The longest and most diverse PINs observed in the data were selected by participants using a PM.

But to reach this goal, broader adoption of PMs is also needed: while half of the enthusiasts in our study are already using a PM to manage their Signal PIN, only 10 casual participants do (10%). For at least this group of users, this approach is preferable. The Signal app could reinforce this idea in the UI and encourage users

to adopt a PM if they have not yet — and if they have, to use it to manage their Signal PIN.

PIN security An account with a strong PIN is less likely to be taken over by an attacker on the network. Our data show large differences in how subgroups of participants select PINs. Although we did not ask participants for their Signal PIN, we asked for its composition among classes of characters: digits, letters, and special characters. Importantly Signal PIN security affects all users because account takeover can affect both the sender and receivers, especially in a group conversation. Even if a given user picks a strong PIN, if one of their messaging partners does not — that well-behaved user is at risk of mistakenly communicating sensitive data to an attacker who hijacked another account.

The current mechanisms of ensuring users select a strong PIN are minimal. Signal currently implements a very small blocklist of weak numeric PINs. These include the following:

- (a) not empty;
- (b) not sequential digits (e.g., 1234);
- (c) not all the same digit (e.g., 0000)

Note that this leaves other popular choices like recent years and dates as acceptable Signal PINs, which are often chosen by users [25, 186]. A targeted attack on an account where the victim’s birthdate, anniversary, etc. are known would likely greatly assist the attacker. The sequence check also only applies to numeric PINs — observe “abcd” and “aaaa” are both valid PINs. In addition, this approach fails to block popular passwords like “password.”

This situation could certainly be improved quite easily, for example implementing the blocklist as recommended by Markert et al. [186] and Bonneau et al. [25] for PINs and following recent guidance from the literature and from government agencies for passwords. NIST Special Publication 800-63B, recommends checking user password choices against lists of the most popular passwords [3]. PIN checks could easily occur locally on the user’s device; however full password checks would require additional features to protect the privacy of the user’s password.

PIN verification reminders To our knowledge, this is the highest-profile roll-out to date of PIN verification reminders (both on Signal and other messengers using the Signal protocol, like WhatsApp). While our study is based on user self-reported data, Figure 5.4 shows that participants do not generally feel they have a problem recalling their Signal PIN. This could be due to password manager use or that participants are using PINs they know well and use in other contexts. More than half of users say they frequently/very frequently verify their PIN when prompted, which points

to user acceptance of PIN reminders. Even though (45; 24%) of respondents turned off PIN reminders, many of those used a password manager; the remainder appear to be comfortable and appreciate periodic PIN verification.

5.7. Conclusion

We conducted an online study ($n = 235$) of Signal users recruited from Reddit, Signal Community Forum, snowballing, and Prolific about their understanding and choice of Signal PINs. In total, 86% of participants set a PIN, with 57% able to technically describe what Signal PINs are used for (enthusiasts) and 43% unable to accurately describe how Signal PINs are used (casuals). We also find that PIN composition followed similar lines: enthusiasts use significantly longer PINs with more complex compositions, and casual participants used more traditional, numeric PINs despite the fact that Signal allows PINs to be alphanumeric. This suggests that communication about the Signal PIN has been effective for part of the Signal population only and that new strategies will be needed to reach the remainder.

As an example of in-app authentication — an authentication mechanism that occurs within a mobile app setting — our investigation shows that in the case of Signal, in-app usage of PINs can be confusing for users who have grown accustomed to screen lock and website login. These authentication metaphors are used often enough that users can be reasonably expected to handle them without much explanation. Where some authentication machinery (a PIN, for example) is repurposed for symmetric-key derivation, only enthusiasts can be expected to read the blogs, documents, tweets, and online help text to gain a full understanding.

Thus, we conclude that communication needs to meet the understanding of the (possibly multiple) user communities. Outside of a core constituency, even something as simple as the name matters. Signal's choice of the term "PIN" can be seen as correct and well-understood by the developers and enthusiasts. However, Signal may be well served in renaming their PIN, e.g., to "Account Recovery Password," and other uses of in-app authentication will need to carefully choose names and messaging to match user expectations.

Though our study does not measure the effect of this intervention, we believe there is strong evidence that suggests renaming Signal PIN to better reflect its usage could be helpful. First, a number of participants described it as an authentication mechanism or message privacy mechanism or simply indicated they do not know. A more precise name, like "Account Recovery," would help users place the Signal PIN in context with other credentials they manage. Second, reusing the term "PIN" suggests to users that only digits are valid. Using the word "Password" or "Passcode" could elicit broader classes beyond digits and encourage more diverse composition.

5.8. Author Contribution

In this paper appearing in Symposium on Usable Privacy and Security, I was the first author and I personally contributed most elements. Given the previous work on PINs for smartphone unlock, I suggested the idea of performing a study on the Signal PIN. I contributed the research questions and the survey instrument. I contributed the analysis plan to interpret our results.

In addition, I was responsible for design and implementation of the online survey using the Qualtrics platform. I developed analysis scripts and implemented a data-analysis pipeline. I coded and analyzed qualitative data including the primary codebook. This effort led directly to my contribution of figures and tables in the final manuscript. I contributed our presentation and talk based on our work which appeared at the conference.

6

Analyzing How Untrained Attackers Guess PINs

Contents

6.1. Introduction	116
6.2. Related Work on Smartphone Threat Models	117
6.3. Methodology	118
6.3.1. Attacker Model	118
6.3.2. Survey Structure	119
6.3.3. Recruitment and Demographics	121
6.3.4. Ethical Considerations	121
6.3.5. Qualitative Analysis	122
6.3.6. Limitations	122
6.4. PIN Characteristics	124
6.4.1. PIN Features	124
6.4.2. Selection Strategies	125
6.4.3. PIN Re-use	125
6.4.4. Security and Usability Perception	126
6.4.5. Guessing Strategies	126
6.5. Novice Attackers' Performance	126
6.5.1. Individual Performance	127
6.5.2. Combined Performance	128
6.6. Context for Smartphone Access	130
6.6.1. Delegation and Emergency Access	130
6.6.2. Controlling Access and Guessing	132
6.6.3. Participant Misconceptions	133
6.7. Discussion and Conclusion	133
6.7.1. Novice Attackers' Guessing of PINs	133
6.7.2. Implications for System Designers	134
6.7.3. PIN Length	135
6.7.4. Users' Perceived Threat Models	135
6.8. Author Contribution	135

6.1. Introduction

Smartphones store sensitive and personal information, including emails, photos, videos, text messages, passwords, account numbers, among others [187]. Securing access to these devices is essential to protect this information. Since both Android and iOS allow smartphone unlocking using either a biometric or knowledge-based authenticator (PIN/password), device security requires a PIN that is hard to guess.

Prior work notes that PINs are the most widely used mechanism to secure access to smartphones, with about 60% of participants using PINs [12] to unlock their devices. At the same time, however, previous studies [12] have shown that many human-chosen 4-digit PINs can easily be guessed by adversaries in just a few attempts, and that upgrades to 6-digit PINs, unfortunately, do not meaningfully improve security [12, 83].

Previous research [49, 61, 83] on PIN-guessing assumes a well-informed attacker with access to commonly-used PIN datasets. Here, we investigate a different, more commonplace setting, where users do not have access to this information and simply try to guess a stranger’s PIN – a novice attacker.

Motivated by this unexplored attacker model and the threat models that real users face, we seek to address the following three research questions:

- RQ1** How do participants select PINs when primed to guess the PINs of other participants?
- RQ2** How well do novice attackers perform compared to the data-driven guessers used previously?
- RQ3** What PIN-based smartphone-unlock scenarios are participants most concerned about?

To answer these research questions, we followed a methodology used by Uellenbeck et al. [13] to study Android unlock patterns where participants were incentivized to guess patterns selected by others in the study, receiving candy if they successfully guessed another participant’s pattern. We adapt this methodology to study PINs through an online survey where participants ($n = 210$) first select a “secret” PIN they would use to protect their smartphone. Thereafter, participants entered five guesses for PINs selected by others in the study, receiving a bonus payment if they succeeded in guessing any other participant’s PIN. Each participant was assigned to one of two treatments, corresponding to 4- or 6-digit PIN selection and guessing. Each treatment had a total of 105 participants.

We find that PINs remain susceptible to guessing even by novices that have no information about the victim, with a total of fourteen 4-digit PINs and seven 6-digit PINs guessed by participants. There is also a benefit for 6-digit PINs as fewer of these

PINs were guessed. Previous work analyzing informed, data-driven attackers [12, 62, 83], found little to no benefit for 6-digit over 4-digit PINs.

To understand how novice attackers compare to data-driven attackers, we aggregated participants' guesses and measured their performance when guessing a large PIN dataset collected by Markert et al. [12] for 4- and 6-digit PINs. We find that in the throttled scenario of mobile authentication, where an attacker only has a limited number of guessing attempts, novice attackers perform similarly to data-driven attackers, in line with previous work [12, 49, 83]. When making up to 30 guesses, a novice attacker performs slightly better, guessing 8.1% of the 4-digit PINs collected by Markert et al., comparable to the 7.6% guessed by a data-driven attacker. Similarly, a novice guesser succeeds at guessing 9.6% of 6-digit PINs collected by Markert et al. after 30 guesses, again slightly more than 8.7% guessed by a data-driven attacker. Our results show that data-driven attackers indeed model real-world threats, and support the guessing approaches leveraged in previous studies [12, 49, 83] where the Amitay dataset has been used to guess 4-digit PINs, and the RockYou dataset to guess 6-digit PINs.

Participants additionally explained their expectations around PIN security and guessing. Thirty-seven percent of participants admitted attempting to access someone else's smartphone *without their knowledge*, while 45% of participants changed their PIN to *keep someone out*. About one-third (34%) of participants overall thought their PIN would be guessed. For those whose PIN was guessed, once again one-third (33%) thought it would be, suggesting participants' perception of the risk of PIN guessing is not related to their PIN choice. As ever, there is an opportunity for user education to help align threats and expectations.

Overall, we conduct the first user study exploring how users guess PINs selected by others as well as how they expect their PINs to be guessed. We find that novice attackers without any knowledge about the victim have surprising capacity to guess PINs, and can roughly guess one in eight PINs. These results suggest the need for more design interventions to nudge users towards more secure PIN choices. We additionally find that users express a need to delegate access to their smartphones for various reasons, suggesting a need for more user awareness and education about device-sharing options.

6.2. Related Work on Smartphone Threat Models

This study analyzes the security of both 4-digit and 6-digit PINs against a specific threat model in the form of real-world human-scale attackers. Hence, this section is two-fold, and presents work both on the security of mobile authentication as well as users' perceived threat models regarding access to their smartphones.

Prior work has analyzed the security of numerous knowledge-based authentication schemes used on mobile devices, e.g., alpha-numeric passwords [115, 121], LG Knock Codes [188], or Android unlock patterns [13, 58, 60, 189–191]. The study most closely related to our work is the one by Uellenbeck et al. [13]. In their study, participants created a “defensive” Android unlock pattern and were given 5 attempts, so called “offensive” patterns, to guess the defensive choices of other participants. We applied this method to PINs to model an average real-world attacker by asking participants to select their own PIN, and then provide 5 guesses for PINs they believe other participants selected. This is different from prior work on the security of PINs [12, 49, 61, 83] which considers a perfect knowledge attacker with full knowledge about the PIN distribution or a simulated attacker that has access to likely 4- or 6-digit PINs collected from experiments or extracted from password leaks.

Bonneau et al. [61] focused their analysis on human-chosen 4-digit PINs which are predominant for mobile devices but are also used in the banking sector for payment cards. Wang et al. [49], Markert et al. [12], and Munyendo et al. [83] on the other hand, also included 6-digit PINs which are predominant in Asia and the default on Apple devices since the roll out of iOS 9 in 2015 [192, 193]. Their analyses show that the security of 4- and 6-digit PINs is comparable and in certain cases, 4-digit PINs are more secure. This is particularly the case for an online attacker limited in the number of guesses they can make, with general knowledge of the distribution of PINs, but no targeted knowledge about the victim. Munyendo et al. specifically showed the limited security benefits of upgrading from 4- to 6-digit PINs.

6.3. Methodology

In this section, we describe the attacker model, datasets, structure of the survey, recruitment and demographics of participants, as well as the qualitative data analysis methods we used. We conclude with the ethical implications and limitations of our study.

6.3.1. Attacker Model

This work focuses on a surreptitious throttled attacker, relevant to Google’s Android and Apple’s iOS, where rate-limiting mechanisms are in place to slow or stop an attacker from simply trying every possible PIN. We will refer to “secret” PINs chosen by device owners to defend against attacks, and “guesses” or “guessing PINs” chosen by an attacker on offense. Our simulation considers the attacker successful if any of their 5 guesses matches the secret PIN of any other participant. We consider the results of each guesser entering the same guesses on 105 different devices. Given

the lifetime of unlock PINs, we assume each PIN will easily face many instances of unattended attack opportunities. We assume that:

(1) The attacker is performing an *online* or *UI-bound* attack, limited in the number of guesses. Based on the throttling implemented on iOS 15, the phone locks after 6 incorrect guesses, so an attacker can make 5 guesses without locking the phone and therefore being detectable by its owner. Similarly on Android, after 5 guesses, the phone locks for 30 seconds [194] (2) The attacker knows the secret PIN length, i.e., whether to guess a 4- or 6-digit PIN. This is the case for Apple devices where the GUI of the lock screen indicates the PIN length. (3) The attacker is *untargeted* and *novice*, i.e., has no personal information about the victim whatsoever and uses their intuition to guess PINs others may have chosen. As described in Section 6.2, it has been shown that an attacker who has knowledge about the victim, by knowing their birthday for example, can use this knowledge to increase their success rate. Our guessability results, therefore, provide a lower bound on an attacker’s capability.

6.3.2. Survey Structure

Below, we outline the overall survey structure. Please refer to Appendix C.1 for a detailed description of the survey instrument including the full wording of all questions.

To allow comparison with previous work, we use the same language and similar ordering of prompts and tasks, the same general appearance and functionality of the PIN pad, and survey questions from previous studies. For example, the practice, task description, and creation prompts match those from Markert et al. [12] for the selection of secret PINs. The overall aim of this approach is to minimize additional bias that might be introduced due to question presentation, phrasing, or a different PIN pad. All participants were required to complete the study using a smartphone: their user-agent string was recorded to ensure a smartphone was used.

Further, we assigned participants to one of two PIN treatments (4-digits or 6-digits). The study itself was identical for both groups, differing only when creating, recalling, or guessing a PIN, with the PIN-pad layout requesting a 4- or 6-digit PIN. After assignment to their treatment, each participant completed the following:

1. *Informed Consent*: On the landing page, participants were shown the consent form where we described the purpose and duration of the research project as well as any anticipated risks. We also informed participants that they can withdraw from the study at any time without penalty.
2. *Agenda*: This page briefly described the overall layout of the study including the three main tasks of the study: (1) creating a PIN, (2) making 5 attempts to guess the PIN of other participants, and (3) completing a short survey.

3. *Practice*: To ensure all participants were familiar with the PIN interface, we asked them to practice entering a PIN. The PIN length was set to 4- or 6-digit, depending on the treatment.
4. *Task Description*: Participants were told the context for which they would create a PIN, namely, to unlock their smartphone. Furthermore, we highlighted that the PIN should be remembered for the duration of the study without writing it down.
5. *Creation*: On the page shown in Figure 4.2, participants created their secret PIN; exactly the same as before, but now with the added word “secret” in “secret PIN.” To ensure that participants did not accidentally mistype it, they also had to confirm it. The layout of this page changed according to the assigned PIN length.
6. *Creation Strategy and Perception*: After creating a secret PIN, we asked participants about their creation strategy (**Q1**) and how they perceived it in terms of security, ease of entry, and memorability (**Q2–Q4**), following Markert et al [12]. Moreover, we asked participants if they reused one of their own PINs (**Q6**) and if they did, the context(s) for this (**Q7**), following Khan et al. and Casimiro et al [89,90]. In between these questions, we included an attention check (**Q5**).
7. *Task Description*: Followed by the questions about their own secret PIN, we framed the guessing task as shown in Figure 6.1. We highlighted that the 5 guesses must be unique. We further informed participants that more than 100 participants would take the study, and any number of correct guesses would earn the bonus payment of \$0.50.
8. *Guessing*: Now that participants were informed about the guessing task, they made their 5 guesses on the page shown in Figure 6.2. If participants provided the same PIN twice, we notified them that the guesses must be unique. Note, participants only guessed PINs of their assigned PIN length.
9. *Guessing Strategy and Threat Model*: After the guessing procedure, we showed the participants their 5 guesses and asked them about their overall strategy when making these guesses (**Q8**). Afterward, we asked participants about a scenario in which others may try to access their smartphone, including their reasons, and the strategies employed (**Q9**). We also asked participants if they considered this scenario when creating their secret PIN or making their guesses (**Q12** and **Q13**). With the next 5 questions (**Q14–Q18**), we intended to learn about participants’ perception and experience of someone accessing their smartphone.
10. *Recall*: We now asked participants to recall their secret PIN. If they could not recall their PIN within 3 attempts, they advanced to the next step, they were compensated normally and their data was included in the evaluation.

11. *Guessing Success*: We asked participants if and why they thought their secret PIN would be guessed as well as if and why they think they guessed the PIN of another participant.
12. *Demographics*: Questions **D1** to **D7** asked the age, gender, education and IT background of participants as well as their smartphone usage. To prevent interference of the participants' demographic background with the results, we asked these questions at the end, following survey best practices from Redmiles et al [125].
13. *Honesty*: We concluded the study by asking participants if they participated honestly. We highlighted that they would not be penalized in any sense if they indicated dishonesty. This approach has been taken with the prior work in this area in order to increase the quality of the data collected [12, 83, 195].

6.3.3. Recruitment and Demographics

We ran pilots with members of our institutions to ensure the clarity of the questions as well as correctness of the data collection process. As a result, we made some slight edits to the wording of some questions, leaving intact the “look and feel” of the PIN pads used on prior work. No data from the pilot studies was incorporated into the final results. For the main study, we recruited 226 participants through Prolific, restricting participation to those residing in the U.S. After excluding 16 participants who indicated dishonesty, we ended up with $n = 210$ participants, 105 for each PIN length. Participants were compensated \$3.50 for completing the study, taking on average 13 minutes for an hourly wage of \$16.15. In total, 179 participants correctly guessed at least one PIN and were compensated a \$0.50 bonus payment for a total compensation of \$4.00.

Table 6.1 depicts the demographics of our participants. The majority of participants were women (112; 53%) compared to men (90; 43%) or non-binary (8; 4%). As expected when using a crowdsourcing platform [196], participants tended to be younger, 73% below the age of 35, and more educated (59% had a Bachelor's degree or higher). The majority (143; 68%) reported that they did not have a technical background.

6.3.4. Ethical Considerations

To limit any negative implications resulting from our study and the data we collect through it, we took several steps. Foremost, our study and its design were approved by our Institutional Review Board (IRB). Further, we informed participants about the purpose of our study and required their consent to proceed. During the study, participants could opt out at any time without any consequences.

In regard to the data we collected, it may have happened that participants selected their actual PIN during the study. We also asked participants in **Q6** whether the secret PIN they selected is a PIN they actually use and a total of 58 participants (27%) affirmed. Although this supports the ecological validity of our data, it imposes the risk of harming the user. To mitigate this risk, we used the Prolific ID as the only identifier in our study and analyzed the PIN data separately from it. For **Q9**, we asked participants to describe a situation where someone might access their phone and if relevant include information about their relationship to this person. As this might pose a risk of participants including Personally Identifiable Information (PII), we explicitly told them not to include PII.

6.3.5. Qualitative Analysis

Several of our study questions involved open-ended answers. Each of these was reviewed by two independent coders in the following manner: A primary coder created a codebook and coded all responses. A secondary coder used the codebook to code all responses. Cohen's κ was calculated for all questions, ranging from 0.825 to 0.926, indicating that coding was reliable. To reach consensus, we observed that some disagreements were lapses in code assignment by researchers; other disagreements were resolved by disambiguating and combining some code descriptions.

6.3.6. Limitations

Our study has several limitations. Foremost, as this was an online study, we could not fully ensure that participants followed our instructions completely. To mitigate this, we included open-ended text based responses as well as an attention-check question. Additionally, participants could indicate if they did not participate honestly at the end of the study, without fearing any negative consequences. Through these questions, we identified 16 participants whose answers were excluded from the final analysis. As is typical for studies using Prolific or other crowdsourcing platforms, participants were younger and more educated. While the survey is not U.S. census-representative, Redmiles et al. [182] showed that crowdsourced samples used by researchers are generally an effective proxy for conditions in the real world, especially for U.S. participants aged 18-49 with at least some college education. While we believe that our results reflect common user attitudes, additional work is required to determine how well these results generalize, especially for more-diverse user populations. For example, it has been shown that populations from other locales such as China select PINs with different distributions [49]. Our study focused on participants from the U.S. only, and additional work is required to study PIN guessing in other countries and cultures.

Table 6.1.: Demographic information of participants

	Woman		Man		Non-Binary		Total	
	No.	%	No.	%	No.	%	No.	%
Age	112	53	90	43	8	4	210	100
18–24	46	22	20	10	5	2	71	34
25–34	41	20	40	19	1	0	82	39
35–44	17	8	21	10	2	1	40	19
45–54	7	3	7	3	0	0	14	7
55–64	1	0	1	0	0	0	2	1
65–74	0	0	0	0	0	0	0	0
75+	0	0	1	0	0	0	1	0
Prefer not to say	0	0	0	0	0	0	0	0
Education	112	53	90	43	8	4	210	100
Some High School	0	0	1	0	0	0	1	0
High School	0	0	1	0	0	0	1	0
Some College	20	10	11	5	2	1	33	16
Trade	24	11	14	7	3	1	41	20
Associate’s	2	1	2	1	0	0	4	2
Bachelor’s	11	5	3	1	0	0	14	7
Master’s	40	19	42	20	2	1	84	40
Professional	11	5	9	4	1	0	21	10
Doctorate	2	1	3	1	0	0	5	2
Prefer not to say	2	1	4	2	0	0	6	3
Background	112	53	90	43	8	4	210	100
Technical	24	11	35	17	1	0	60	29
Non-Technical	85	40	52	25	6	3	143	68
Prefer not to say	3	1	3	1	1	0	7	3

Similar to the study design used by Uellenbeck et al. [13], participants in our study knew from the agenda text seen in Appendix C.1 that other users would attempt to guess their PIN—which may have yielded more secure choices. Participants also knew they would have to use and remember the PIN only for the short duration of the study, which may also have encouraged the use of more secure PINs. On the other hand, previous studies that have used a similar method collected generalizable authentication data [12, 129]. Moreover, unlike in Uellenbeck et al.’s study, we highlighted the bonus payment for correctly guessing other participants’ PINs for the first time after participants had already created their secret PIN.

Some of our survey questions asked about behaviors that may not be seen as socially desirable. For example, participants were asked if they had ever tried to access someone else’s smartphone without their knowledge and if they had ever changed their PIN to prevent someone from accessing their smartphone. In both situations, a participant could be seen as admitting undesirable behavior: in the first case by exceeding granted permissions, or in the second case by the need to perhaps

Table 6.2.: Responses to Q6: PIN reuse

	4-digit		6-digit		Total	
	No.	%	No.	%	No.	%
Yes	34	32	24	23	58	28
No	68	65	73	70	141	67
Do not use PIN	3	3	6	6	9	4
Unsure	0	0	2	2	2	1

conceal something. Our study cannot determine if participants were untruthful, except to note that responses to both these questions were roughly the same, and were somewhat consistent with previous studies.

6.4. RQ1: PIN Characteristics

In this section, we discuss how participants selected their secret PINs and how they guessed other participants' PINs. We first describe features of both the secret and guessing PINs, and afterward, we discuss strategies used to create and guess PINs.

6.4.1. PIN Features

The features of secret and guessing PINs can be categorized into four groups: *date*, *repeat*, *sequential*, and *pattern*. Table 6.3 provides a detailed breakdown for each category. Note, some PINs such as 0101 can match multiple patterns, including *date* and *repeat*. Overall, *date* is the most popular pattern for secret and guessing PINs for both 4- and 6-digit PINs, as seen also in prior work [12, 61, 83]. Dates throughout this section correspond to four different types of sequences, including *yyyy* for “recent year,” defined as sequences 1940–2028, similar to the definition used by Wang et al [49]. Examining the 4-digit PINs shows that 11% of the secret PINs and 5% of the 4-digit guesses represent a recent year. Four-digit PINs of the format *mmyy* (beginning with digits 01-12) account for 26% of the secret PINs compared to 33% of the guesses. The format *mmdd* accounts for 20% and 12% of secret and guessing PINs respectively for 4-digit PINs. For the 6-digit PINs, we similarly observe that the secret PINs and guesses follow the same patterns, with 37% of both secret and guessing PINs following the format *mmyyyy*. PINs in the format *mmddy* account for 30% of secret PINs, but only for 16% of guessing PINs. Lastly, *yymmdd* accounts for 16% and 11% of secret and guessing PINs respectively. These results align with prior work [12, 61, 83] which find that dates represent a sizable percentage of 4- and 6-digit user-selected PINs.

The most popular sequential feature across secret PINs is an ascending order, e.g., 1234 for the case of 4-digit PINs. However, it only accounts for 2% of the total

4-digit secret PINs. For both PIN lengths, only about 5% of all secret PINs follow this pattern, despite its popularity among the guesses. Our results further reveal that one in three 6-digit guesses depicts a rectangular walk on the PIN pad, e.g., 139713, despite only 5% of the secret 6-digit PINs following this pattern. Overall, participants' guesses reflect the features imagined to be popular, which diverge from the actual secret PINs the participants selected.

6.4.2. Selection Strategies

When asked about the strategies used to create their secret 4- or 6-digit PINs, participants in both treatments often mentioned a *date* ($n = 33$; 31% for 4-digit and $n = 29$; 28% for 6-digit). This finding is in line with results from our PIN analysis as well as prior work [12, 61, 83]. However, the incidence of date-related responses was somewhat higher in the 4-digit PIN treatment compared to the 6-digit PIN treatment.

The second most frequent strategy in both treatments was *memorable*, i.e. choosing a PIN that is easy to remember ($n = 14$; 13% for 4-digit and $n = 19$; 18% for 6-digit), also in line with prior work [12, 83]. For instance, P3 mentioned that “it was just three 2 digit numbers i [*sic*] knew I’d remember.”

Selecting a PIN based on something that had a *meaning* to participants was the next most common strategy. This was more prevalent for 6-digit PIN participants compared to 4-digit PIN participants ($n = 8$; 8% for 4-digit and $n = 14$; 13% for 6-digit). For example, P40 mentioned “using numbers that hold personal meaning to me but don’t have to do with birthdays or anniversaries.”

Participants in both treatments also indicated using a *pattern* for various reasons, including convenience. For instance, P104 stated:

“The shape or movement of my thumbnail, I drew a rocket ship with the numbers so I could use muscle memory to sign in and not worry as much about the numbers.”

The use of *random* numbers and *reuse* of PINs were also frequently mentioned by participants, as well as creating a PIN that is *simple*. Other less frequently cited strategies included using *subsets* of phone numbers, *ZIP codes* and *words*.

6.4.3. PIN Re-use

Reuse of credentials remains a challenge in online safety and therefore, we asked participants whether they re-use their 4- or 6-digit PIN on any other accounts. Across both treatments, more than half of the participants indicated they do not re-use their PINs. However, *reuse* was more prevalent for 4-digit PINs at 32% compared to only 23% of 6-digit PINs, as shown in Table 6.2. While this may suggest a possible benefit

of 6-digit PINs, 4-digit PINs are more commonly used, and are thus more likely to be re-used.

6.4.4. Security and Usability Perception

When asked about the security perception of their chosen secret PIN, 83% of participants in both 4- and 6-digit PIN treatments felt their secret PIN was “secure” or “somewhat secure.” In reality, 90% of PINs were unguessed, so participants were roughly correct in our threat model. For memorability, however, 97% of participants in the 4-digit PIN treatment perceived their PIN to be “memorable” or “somewhat memorable” compared to 90% in the 6-digit PIN treatment. Overall, while participants perceive both their 4- and 6-digit PINs to be secure, they find 4-digit PINs to be slightly more usable compared to 6-digit PINs. Therefore, system designers should consider the additional usability burdens of 6-digit PINs as well as their limited security improvement over 4-digit PINs before increasing PIN lengths.

6.4.5. Guessing Strategies

In contrast to the selection strategies for secret PINs, which have a large dependency on PIN length, guessing strategies for 4-digit and 6-digit PINs were mostly similar. While most participants selected their secret PINs using *dates* to make them *memorable*, they often went for *simple* PINs when guessing other participants’ PINs ($n = 39$; 37% for 4-digit and $n = 34$; 32% for 6-digit). Rather than speculate on what dates or what sequences might be memorable to other users, most participants focused on simple PINs—even though participants themselves most often used dates to select their secret PINs. For instance, P2 noted:

“I knew ii [sic] wouldn’t be able to guess ones that were chosen because they were meaningful for some reason so I picked ones that are easy to type in.”

6.5. RQ2: Novice Attackers’ Performance

In this section, we analyze how novice attackers perform in guessing PINs compared to the data-driven guessers that have mostly been used in previous studies [12,49,83].

Datasets We collected new datasets by priming participants to select a secret PIN and five guessing PINs, either of four or six digits in length. In Section 6.4, we analyzed the features of these PINs and how they are selected. Now, we utilize them to study the guessing resistance of PINs against novice attackers who *do not* utilize knowledge of the general distribution of PINs. Following the way participants were

primed and their assigned PIN length during the study (see Section 6.3), we refer to the datasets as *Secret-4*, *Secret-6*, *Guess-4*, and *Guess-6*.

In addition to these four datasets collected, we also use datasets used in previous studies. Since PINs are usually stored and validated on individual devices instead of web servers, no large-scale leaks of PINs have yet appeared. Therefore, previous studies have relied on a bricolage of datasets that were collected in a variety of ways. Probably the most realistic, although not collected within the bounds of a controlled experiment, is composed of 204 508 4-digit PINs. These PINs were gathered by the iOS application “Big Brother Camera Security” from Daniel Amitay [51]; we refer to this dataset as *Amit-4*. Since Amitay collected only 4-digit PINs, a similarly-sized 6-digit dataset is not available. Other datasets have been derived from leaked alphanumeric passwords. For instance, 4- and 6-digit subsequences were extracted from the RockYou password leak by Bonneau et al. [61] and Wang et al [49]. We refer to these datasets as *Rock-4* and *Rock-6*. Two datasets have also been collected within the bounds of a controlled experiment; Markert et al. collected both 4- and 6-digit PINs from participants primed to choose secret PINs [12]. We refer to these as *Markert-4* and *Markert-6*, respectively.

6.5.1. Individual Performance

We now examine participants' individual guessing performance, or the smartphone-unlock guessing threat posed by surreptitious untrained attackers. We focus on their performance in successfully guessing the secret PINs collected as part of this study but also datasets used in prior work. We additionally highlight trends that emerged from the guessability analysis for both 4- and 6-digit PINs.

The 1050 guesses we collect (525 for each PIN length) comprise a total of 415 distinct PINs. In the 4-digit case, we observed 177 different PINs (1.77% of all possible PINs), and 238 different 6-digit PINs (0.0238% of all possible PINs). Of these 415 PINs, 91% (378) were guesses of three or fewer participants, and only eight (2%) were guesses of more than 20 participants. Interestingly, these eight very popular guesses split evenly between the two lengths and follow similar patterns: the 4-digit PINs were 0000, 1111, 1234, and 2580, the 6-digit PINs were 000000, 111111, 123456, and 987654.

In terms of the guessing resilience of participants' secret PINs, the PINs of 21 participants (10%) were guessed (amounting to 16 unique PINs). Fourteen (13%) of these were 4-digit, 7 (7%) a 6-digit PIN. Regarding the variety of the selections, the 21 participants selected 16 different secret PINs, 10 of them being 4-digit PINs (0000, 1234, 1478, 1990, 1995, 1997, 2000, 2468, 2580, 6666) and six 6-digit PINs (121212, 123456, 134679, 135790, 159753, 654321). From the guessers' perspective, 179 participants (85%) guessed at least one secret PIN: 95 (91%) in the 4-digit and

84 (80%) in the 6-digit treatment. In prior work [12, 49, 83], the insecurity of 6-digit PINs arose from a stark selection bias, particularly with 123456 tending to be overwhelmingly popular. Among the Secret-6 PINs, we did not observe this shift. The reasons for this shift are unclear; while our study carefully reproduced the “look and feel” of prompts used in prior work, our priming of users may have caused more secure PIN choices.

The individual guessing success, i.e., the success of the five guesses of each participant against datasets used in prior work can be seen in Figure 6.3. The violin plots show the range of individuals’ PIN-guessing performance on a guess-by-guess basis when guessing the Markert-4 and Markert-6 dataset. For example, after one guess, the median, shown as a vertical line, for both 4- and 6-digit is equal to the maximum proportion guessed: 2% for 4-digit and 4% for 6-digit PINs. The median gains in individual performance after the first guess are marginal for both 4- and 6-digit PINs. The highest increase happens after the third guess for 6-digit PINs and the fourth guess for 4-digit PINs. For 4-digit PINs, the proportion guessed increases from 4% (3 guesses) to 5% (4 guesses), for 6-digit, from 5% (2 guesses) to 8% (3 guesses). Overall, novice guessers perform better at guessing 6-digit than 4-digit PINs in Markert et al.’s dataset. This is in line with findings from prior work about the success rate of data-driven attacks against 4- and 6-digit PINs chosen without security focus [12, 49].

For comparison purposes, Figure 6.3 additionally depicts the simulated guessing attacks constructed from other previously published datasets which were analyzed by prior work. From these plots, we can see that the median performance of our individual novice guessers closely matches the performance of data-driven guessers built from Amit-4 (★) for 4-digit and Rock-6 (●) for 6-digit PINs. This supports the ecological validity of previous studies’ use of these datasets to evaluate the guessing resilience of PINs [12, 49, 83].

6.5.2. Combined Performance

To assess the performance of participants’ guesses in aggregate, we combined all guesses to carry out a simulated attack against different 4- and 6-digit PIN datasets, following the approach in the previous work of Markert et al., Munyendo et al. [12, 83], and others. This aggregated novice attacker is created based on the guessing order we derive from merging the five guesses of all participants into a single dataset. If two or more PINs share the same frequency in this dataset, we first try to rank them based on the order in which the participants guessed them. For example, if two PINs both occurred once but one of them was the third guess and other the fifth, we guess the third guess first because it had a higher “priority” for the participant.

An alternative approach sometimes seen in the guessing literature is to use the guesses from a study as a training set for a Probabilistic Context-Free Grammar (PCFG)- or Markov-based approach, following Wang et al. [197] that would generate many more guesses and therefore reach a strength estimation for the remaining (unguessed) PINs. A PCFG approach can be helpful when guessing variable-length passwords that follow a *system*, a generative structure like “name followed by date.” According to our participants, less than 5% used a system. But the contribution of our work is precisely to understand the actual guesses for 4- and 6-digit PINs directly observed in our study. Moreover, we aim for surreptitious guessers who are strictly limited to a small handful of guess attempts. Developing a guessing order for the remaining unguessed PINs only applies to *unthrottled* attackers who could make hundreds or thousands of guesses. As this attacker could merely try all possible 4-digit or 6-digit PINs, they are outside our threat model.

Figure 6.4a shows the performance of participants' 4-digit guesses, while Figure 6.4b shows it for the 6-digit guesses. When considering 4-digit PINs, participants' guesses have a comparable effectiveness when guessing the *Rock-4* (●) and Markert-4 (+) datasets, but perform slightly better on other participants' *Secret-4* PINs (●). Interestingly, the success rate against *Amit-4* (★) is noticeably better, with over 15% of PINs in this dataset guessed after 20 guesses. This difference between the guessing performance of the *Amit-4* dataset compared to the remaining sets could be attributed to users not being specifically primed for security during the selection of PINs in the Amitay app. This can be investigated in future research.

For 6-digit PINs, the success rate of participants' guesses is lower in guessing other participants' secret PINs (●) compared to PINs from the Markert-6 (+) dataset. However, the difference is even greater when guessing *Rock-6* PINs (●). This contrast can be attributed to the popularity of the PIN 123456 in *Rock-6*, leading to a substantial portion of PINs being guessed with just one guess. After this first guess, the rate appears to be more consistent with the guessing performance against the other two datasets.

Overall, our results from Figure 6.4 indicate that participants' guesses perform similarly or better on previously published datasets compared to the secret PINs selected by participants in our study, particularly for 6-digit PINs. As previously discussed in Section 6.5.1, there were only six different 6-digit secret PINs that were successfully guessed. This likely suggests that encouraging users to select secret PINs that cannot be easily guessed is a promising way to make them select more secure PINs. However, additional work is needed to specifically explore how this can be implemented.

Figure 6.5 shows how well participants' guesses, secret PINs, and the Amitay and RockYou datasets perform when guessing PINs from Markert-4 (Figure 6.5a) and

Markert-6 (Figure 6.5b). This serves as a benchmark to assess the effectiveness of participants' selections when applied to other datasets previously collected from the literature. Overall, participants' secret PINs (●) perform the worst of the datasets at guessing Markert-4 and by an even wider margin in the case of Markert-6. In the prior case, *Guess-4* (▲) performs similarly to *Amit-4* (*), while *Secret-4* performs comparably to *Rock-4* (●). However, in the case of Markert-6, the performance is more varied, with *Secret-6* (●) performing poorly, guessing only one PIN correctly. Conversely, *Guess-6* (▲) performs similarly to *Rock-6* (●) when guessing the Markert-6 dataset.

Overall, we find that the *Amit-4* dataset performs comparably to participants' *Guess-4* PINs, while the success rate of the *Rock-6* based attacker is similar to the *Guess-6* PINs of participants. This shows that on aggregate, novice attackers perform similarly to data-driven attackers, and further supports the ecological validity of previous studies [12, 49, 83] that have used the Amitay dataset to guess 4-digit PINs, and the RockYou dataset to guess 6-digit PINs.

6.6. RQ3: Context for Smartphone Access

This section reports on concerns expressed by participants from **RQ3**: “What smartphone-unlock scenarios are participants concerned about?” We report on how participants conceptualize situations of smartphone access by others with a focus on the critical context of *who* (Q9), *why* (Q10), and *how* (Q11), similar to Marques et al. [41]. Figure 6.6 summarizes these results.

6.6.1. Delegation and Emergency Access

Previous work [198, 199] has shown that sharing behaviors are common with regard to smartphones, and that users have a desire to grant others limited, temporary access to their devices. In this section, we explore how these sharing and delegation behaviors align with the threat models envisioned by participants.

In Q9, we asked participants to “describe a situation where someone is most likely to unlock your smartphone.” Answers overwhelmingly mention a close social contact, rather than a stranger or thief. The combined categories of *partner*, *friend*, and *family* account for 178 out of the total 210 responses (85%). In contrast, only seven participants (3%) mention a *thief* or *stranger*. Considering *partner*, the most-frequent code (81; 39%), P25 stated:

“I can imagine my wife needing my PIN to access my phone to make a payment in a store and when my hands were full.”

In spite of that, individuals in the partner category are not without risk. The prevalence of this category in our data combined with the work of Tseng et al. [200] on

intimate partner surveillance points to a need to further understand the nuances of PIN security in controlling types of unauthorized access by insiders. Relationships within the *family* (50; 24%) or with a *friend* (47; 22%) have already been shown to pose similar risks [201, 202].

Subsequently, **Q10** asked *why* the individual would access their phone. Most often, participants describe some form of *delegation* (88; 42%), with P37 saying:

“They would just be looking at a text or notification to let me know what it said while I was presumably otherwise occupied.”

Finding some form of information, not necessarily with a bad intent, was the second most popular answer (38; 18%). For example, P175 said the person might “check something like the weather or bus schedule.” A similar use case, specifically in the form of someone borrowing the phone, was described by 21 participants (10%).

Many participants also identified a special category of delegation: they indicated that someone would access their smartphone in *dire circumstances* (24; 11%), e.g., when the owner is physically incapacitated. P112 said “in case of an emergency he might be unlocking my phone to get help.” While unlocking the phone to call for help is a valid scenario, sharing one’s PIN for this purpose is not necessary as both Android and iOS allow anyone to make emergency calls even when the device is locked. System developers have also been improving features that can further assist with delegation and emergency access. For example, Android 13 allows user profiles to be switched on from the lock screen and guest users granted access to installed apps [203] and Android 14 may allow “cloned apps” to permit multiple installations of a given app to use different user profiles [204]. This development suggests a broader recognition in the industry of this user concern and would give added prominence to the feature of user profiles. In our study, *no* participants mentioned user profiles despite the preponderance of scenarios where profiles would be of benefit.

Overall, there appears to be a prevalence of benign motivations from close social contacts when it comes to accessing participants’ phones (see Figure 6.6). A reason for this might be the social-desirability bias which might have made participants hesitate to be forthcoming about malicious activity in response to this question. On the other hand, more than 37% admitted to “trying to access someone else’s smartphone without their knowledge” in **Q17**, which is more than those who mentioned *mal-intent* (16; 8%) in **Q10**. Moreover, this finding is comparable to the 39% found by Marques et al [40].

Participants also provided responses to **Q11** asking about “the strategy the individual would use to gain access to your smartphone.” Although guessing was the most prominent answer, responses continued with the general theme of benign access by social contacts. More than half of participant responses indicated that the person

would simply *ask* (62; 30%) or already know the PIN because it is *shared* (59; 28%). For example, P15 said “I would tell them the code,” while P79 added:

“He knows my PIN... but my trick is to say ‘it’s my birthday’ which shows me who knows when my birthday is!”

This willingness to allow others access was further confirmed by the responses to **Q18**. A large majority (184; 88%) said they have granted someone else access before, while (23; 11%) had not.

6.6.2. Controlling Access and Guessing

Our survey found that participants had some difficulty controlling access, with nearly half (95; 45.2 %) reporting that they had changed their PIN specifically to prevent someone’s access in **Q16**. As guessing is the main focus of our study, we additionally reviewed all responses to **Q11** that mention guessing. By this more-inclusive measure, 78 responses were included in this additional analysis.

A total of 51 responses mentioned the use of personal knowledge in formulating guesses. As an example, P86 said “she would use my date of birth, or she would use the ages of my grandkids.” This finding suggests an opportunity for future work focusing on guessers who incorporate personal knowledge and is echoed by Munyendo et al. [83] who found a personal hint can be effective in helping predict someone’s PIN. In contrast, comparatively fewer participants (12; 15 %) reported shoulder-surfing in their scenario. Previous work such as from Aviv et al. [63] suggest that only about 11 % of shoulder-surfing attacks on 6-digit PINs are successful.

For comparison, of the 51 responses who mentioned personal hints, there were (33; 65 %) occurrences of *birthday*, *day*, or *date*. This confirms prior work [61, 83] that has similarly shown that personal hints including birthdays can reveal users’ PINs.

Q14 and **Q15** specifically asked about participants’ level of concern about their phone being accessed *without* their consent. About half of participants were “somewhat concerned” (67; 32%) or “concerned” (26; 12%), with most of them saying that they keep important information on their phone. Except for 8% who were indecisive, all others indicated they were “somewhat unconcerned” (53; 25%) or “unconcerned” (47; 22%). Most of these participants do not believe their phone can be unlocked by someone else while others simply trust their surroundings or believe they have nothing to hide.

To investigate participants’ perception about the security of their secret PIN, **Q19** asked if they think their secret PIN will be guessed by other participants. Most participants said no (138; 66%), and only five of them were wrong. The PINs they picked were years (1990, 2000) or common schemes (2580, 6666, 121212). In contrast, of those who believed their PIN would be guessed (71; 34%), 16 were

right. In other words, there was a tendency across participants to underestimate the security of their PIN.

Finally, **Q20** asked “Do you think you guessed someone else’s secret PIN? Why or why not?” Most participants said yes or maybe (164; 78%), with a majority (134; 82%) of them right. On the other hand, (43; 20%) said they would not guess correctly: of these, only (14; 33%) were right. Participants who thought they did not guess the PIN of another participant mostly mentioned that the PIN would be of *personal importance* to a stranger or *random*. Participants who thought they did guess correctly indicated that other participants artificially *picked easy PINs*, e.g., P112 was right in saying:

“I think I guessed someones PIN. Someone definitely used 0000.”

6.6.3. Participant Misconceptions

We were particularly interested in the survey responses of the 21 participants whose PINs were guessed by another participant. Nine mentioned having important information or valuing their privacy while only four mentioned some variation on trusting others or having nothing to hide. When asked how someone would access their phone, only four mentioned guessing compared to 10 who said the person would ask. Five participants chose a PIN that was *simple*, while another five chose a PIN that was *memorable*.

Of the 21, only 7 thought their secret PIN would be guessed by another participant, suggesting that most participants were not purposefully setting weak PINs. The 14 participants who were overconfident highlight an opportunity for user education, as 3 said their PIN was reused while another 3 said their PIN was random.

6.7. Discussion and Conclusion

In this paper, we analyze the guessability and threat models of human-chosen 4- and 6-digit PINs for smartphone unlock by considering more commonplace novice attackers. Overall, we find that novice guessers perform comparably to the data-driven attackers employed in prior work [12, 49, 83], and that most people would like to delegate access to their smartphones in some way. In the rest of this section, we discuss these findings further and offer recommendations to system designers that can improve the security of human-chosen PINs.

6.7.1. Novice Attackers’ Guessing of PINs

The exact probability distribution of PINs in various contexts of smartphone device unlock is not known. Previous studies have considered an attacker whose guessing

order is informed by proxies, with the Amitay dataset used to guess 4-digit PINs, PINs extracted from digit sequences in RockYou to guess 6-digit PINs alongside additional PINs from participants primed to choose a new secret PIN [12, 49, 61]. To our knowledge, we are the first to gather PINs from participants ($n = 210$) specifically primed for guessing, using an approach modified from Uellenbeck et al. [13] to study Android unlock patterns. We find that in the aggregate, our individual novice attackers perform very closely to the simulated attacking routines built from the proxies that have been used in previous studies. The direct comparison is shown in the median performance in Figure 6.3. Our work therefore supports the approaches employed in previous work. We additionally show that human-chosen PINs remain susceptible to guessing, even by novice attackers that have no personal information about the victim. Therefore, more design interventions as well as user education are required to guide users towards more secure PINs.

6.7.2. Implications for System Designers

The previous work has shown how to build optimized blocklists [12] from data-driven attackers. Here, we find a justification for a small additional blocklist consisting of those PINs guessed in our study. These PINs that were guessed tended to be guessed by many participants. Therefore, systems could introduce additional friction for the user based on our results. Messaging to users could take on additional urgency, such as “This PIN would be guessed by *many* people.” In addition, though the iOS blocklist currently allows a user to select “Use Anyway,” the risk is so great with the guessed PINs that these should be completely disallowed. There is, as ever, an opportunity for user education shown by the fact that only one-third of participants whose PIN was guessed believed this would be true. Our work agrees with Marques et al. [41] who suggested adding some realistic details when explaining attacks to users, emphasizing the aspects of unattended access by non-strangers. Where Marques et al. dub this the *shower-time attack*, as the phrase emphasizes threats from non-strangers, low technical and temporal barriers to entry, combined with a re-evaluation of what may be heretofore considered safe physical spaces. This period of unattended access by insiders has been previously called the *lunchtime* or *midnight* attack, noted by Naor and Yung to be “folklore” as early as 1990 [205]. Future work could test which of these phrases leads to changes in user behavior. Finally, users could be notified when unsuccessful guessing occurs on their device. It is by now common for websites to send email when the site is accessed by a new device or in a new location. The device could send an email to the owner when a PIN was guessed incorrectly. This approach would make this attack less surreptitious and perhaps discourage attackers.

6.7.3. PIN Length

In our attack setting where novice individuals provide five guesses applied to 105 simulated smartphones, participants performed better against 4-digit PINs than 6-digit PINs; nearly twice as many 4-digit PINs (13%) compared to 7% of 6-digit PINs were guessed. However, when using our participants' offensive PINs to guess a larger dataset such as the one collected by Markert et al. [12], we find that 6-digit PINs are more easily guessed compared to 4-digit PINs, matching prior work [12, 49, 83]. In addition, we find 32% of our 4-digit participants compared to only 23% of 6-digit participants reused their secret PIN. While this may suggest a possible benefit of 6-digit PINs, 4-digit PINs are more commonly used and are therefore more likely to be re-used. Nonetheless, encouraging users to select secure PINs, of either 4- or 6-digit, might have better security outcomes than simply asking users to upgrade to 6-digit PINs, as noted by Munyendo et al. [83]. The exact messaging and design can be explored further as part of future research.

6.7.4. Users' Perceived Threat Models

When asked about scenarios in which someone would actually access their smartphone, participants overwhelmingly mentioned close social contacts. Most common was a desire to *delegate* some forms of access to the device for example when driving. Others in the literature such as Karlson et al. [199] have similarly observed that users express a desire for delegated or guest accounts. Our survey also found that participants had some difficulty controlling this access, with nearly half ($n = 95$; 45.2 %) reporting that they had changed their PIN specifically to prevent someone's access. Android 13 revamps the user-profile feature making it easier to delegate access. Given that no participants mentioned user profiles, there is a need for more user awareness and education on this improved feature, as well as how users can use it. Further, participants mostly do not consider (and report being generally unconcerned about) scenarios of smartphone unlock by strangers when selecting either their "secret PIN" or their guesses, suggesting they value an easily-remembered PIN over guessing resistance. More than half of respondents who mentioned guessing in their scenario said the person would use a personal hint. More than half of those mentioned "day," "date," or "birthday" suggesting an opportunity for future work to better understand the resistance of PINs to guessing in the presence of hints or other personal or additional information.

6.8. Author Contribution

In this paper to appear in European Symposium on Usable Security in Oct. 2023, I was the first author and I personally contributed most elements. Given the previous

work on PINs for smartphone unlock, and my advisor's previous study on Android Patterns, it was natural to develop an idea like this. I contributed the research questions and the survey instrument. I contributed the formal analysis plan to interpret our results.

I developed some of the analysis scripts, which in some cases were based on our earlier study on smartphone PIN unlock. I coded and analyzed qualitative data including the primary codebook. This effort led directly to my contribution of figures and tables in the final manuscript, along with being the lead author. I am currently developing our presentation and talk based on our work which will appear at the conference.

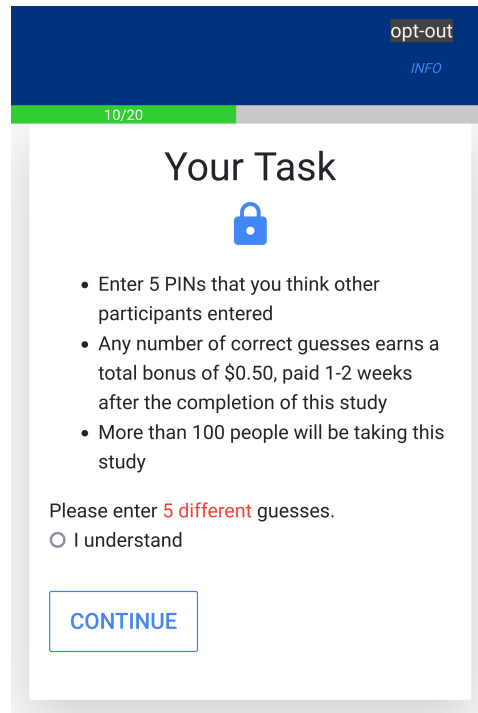


Figure 6.1.: The instructions provided before the participants were asked to guess others' secret PIN

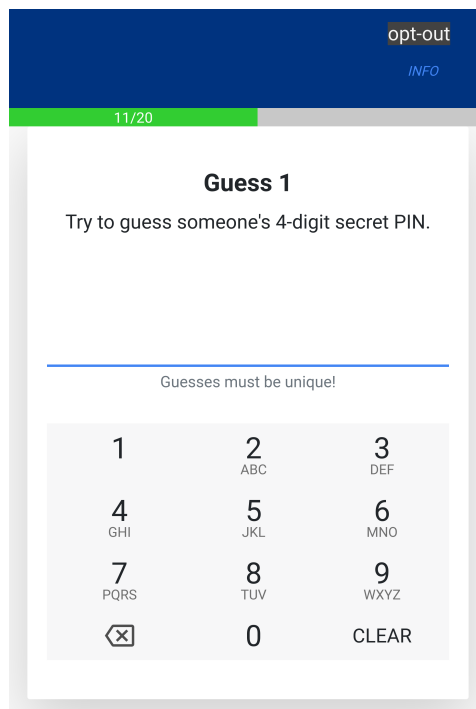


Figure 6.2.: The page on which we asked participants to guess other participants' secret PIN

Table 6.3.: Features of the Secret PINs and Guesses

Secret PINs						Guesses					
4-digit			6-digit			4-digit			6-digit		
Date											
mmyy	27	26%	mmddy	31	30%	mmyy	172	33%	mmddy	83	16%
mmdd	21	20%	yymmdd	16	15%	mmdd	64	12%	yymmdd	55	11%
yyyy	11	11%	ddmmyy	14	13%	ddmm	51	10%	ddmmyy	46	9%
ddmm	9	9%	mmyyyy	4	4%	yyyy	24	5%	mmyyyy	9	2%
total	41	39%	total	34	32%	total	198	38%	total	86	16%
Repeat											
couplet	12	11%	couplet	7	7%	abab	141	27%	ababab	97	19%
triplet	4	4%	ababab	6	6%	couplet	126	24%	couplet	92	18%
abba	4	4%	aabbcc	2	2%	triplet	121	23%	triplet	85	16%
abab	3	3%	triplet	1	1%	abba	121	23%	aabbcc	78	15%
aaaa	2	2%			aabb	120	23%	abcabc	77	15%	
aabb	2	2%			aaaa	119	23%	aaaaax	77	15%	
								xaaaaa	76	15%	
								aaaaaa	74	14%	
total	15	14%	total	13	12%	total	150	29%	total	120	23%
Sequential											
asc	2	2%	asc	2	2%	asc	100	19%	asc	95	18%
asc even	1	1%	asc odd	1	1%	desc	36	7%	desc	40	8%
asc odd	1	1%	desc	1	1%	asc even	6	1%	asc odd	8	2%
			double-asc	1	1%	asc odd	5	1%	asc even	7	1%
									double-asc	4	1%
total	4	4%	total	5	5%	total	147	28%	total	154	29%
Walk											
vertical	8	8%	rectangle	5	5%	vertical	47	9%	rectangle	173	33%
diamond	2	2%	vertical	1	1%	corners	27	5%	horizontal	8	2%
diagonal	2	2%			diamond	12	2%	vertical	5	1%	
corners	2	2%			rectangle	7	1%	corners	3	1%	
rectangle	1	1%			box	6	1%	box	1	1%	
horizontal	1	1%			diagonal	5	1%				
total	16	15%	total	6	6%	total	104	20%	total	190	36%
Total											
total	105		total	105		total	525		total	525	

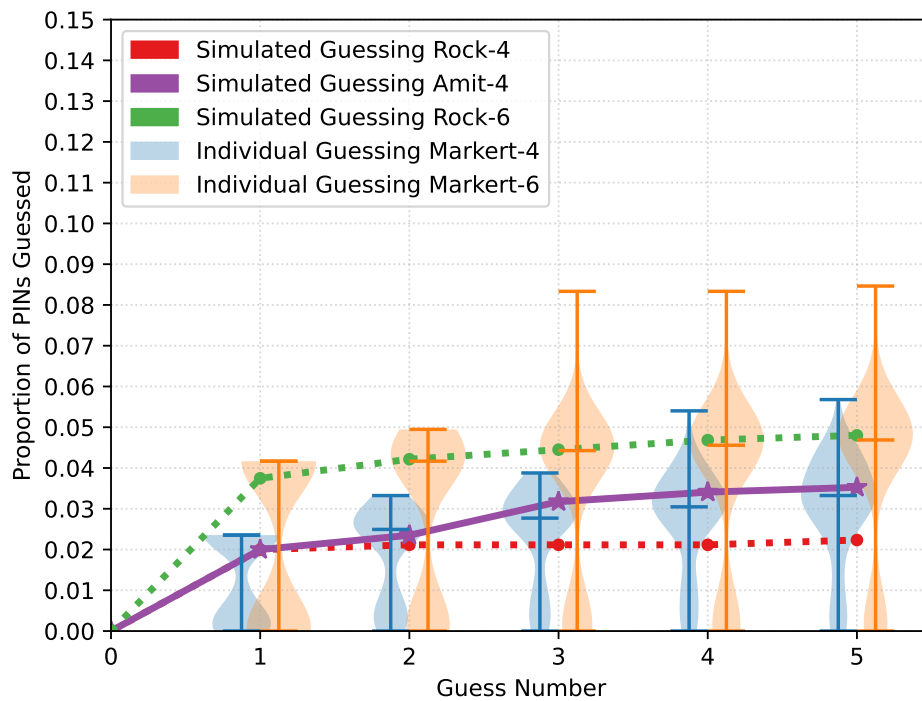
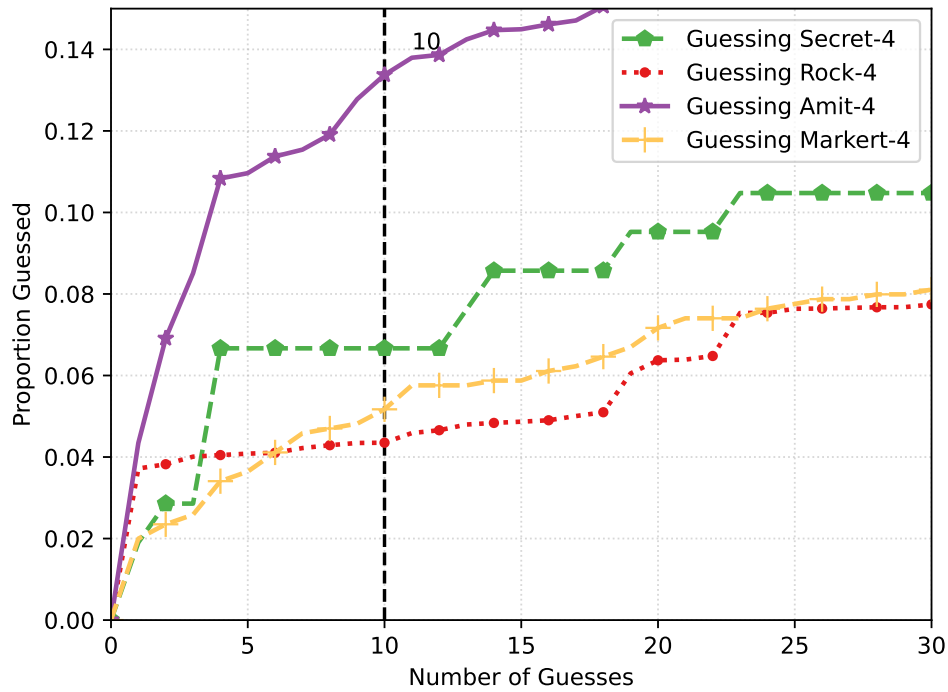
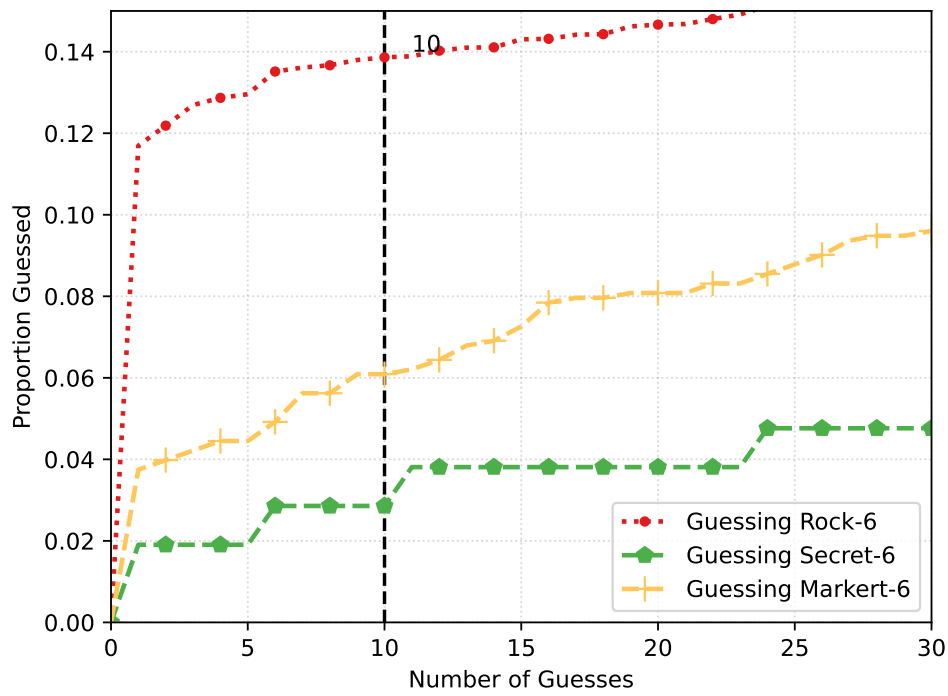


Figure 6.3.: Individual performance of participants' guesses and simulated guessing performance of other datasets when guessing Markert-4 and Markert-6.

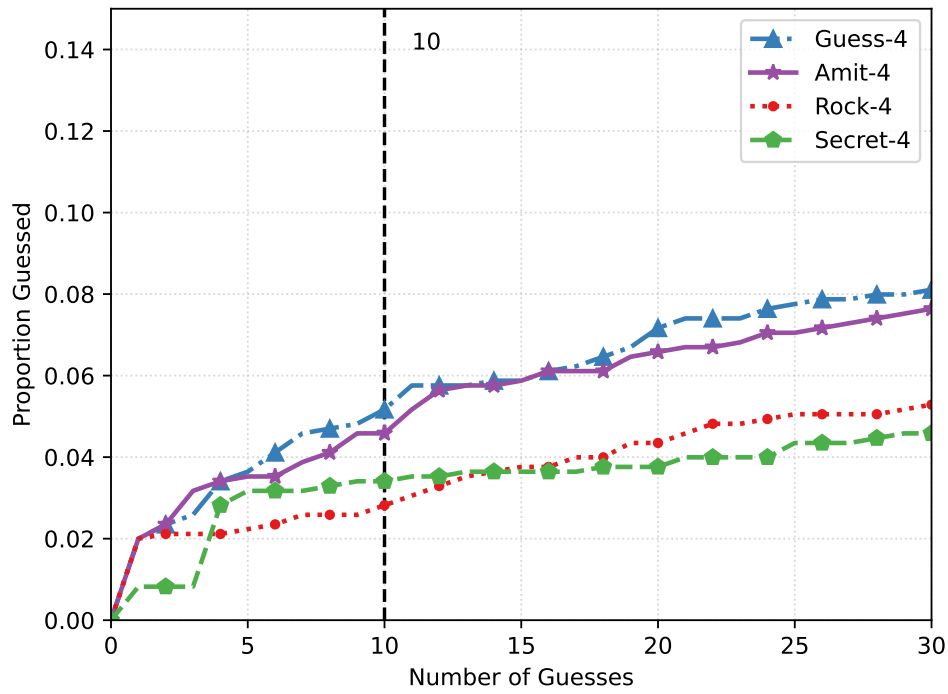


(a) Guessing 4-digit PIN datasets

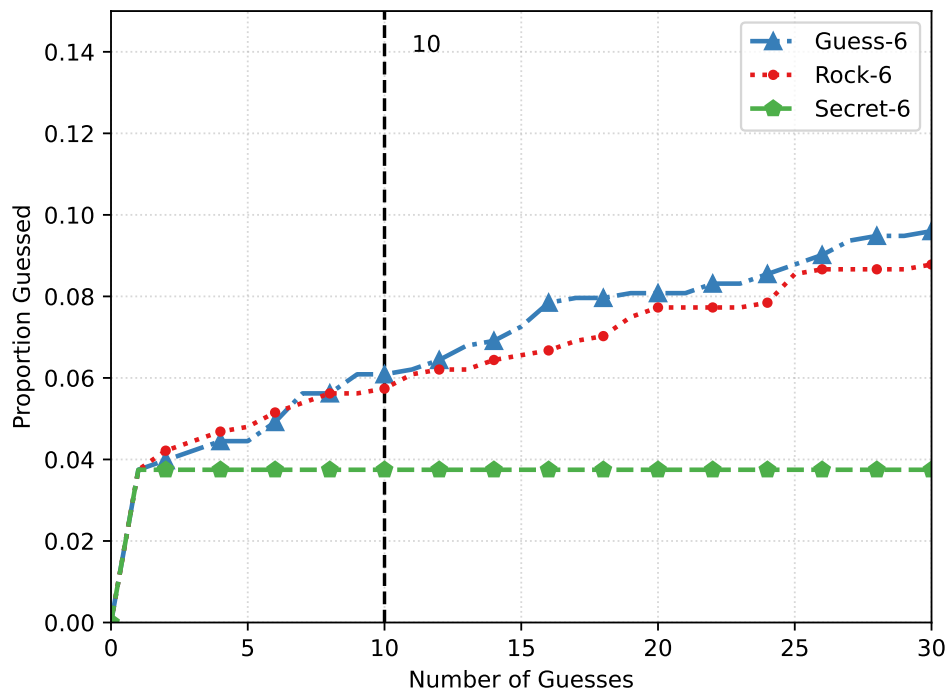


(b) Guessing 6-digit PIN datasets

Figure 6.4.: Using participants' guessing PINs to guess 4- and 6-digit PIN datasets



(a) Guessing Markert-4 dataset



(b) Guessing Markert-6 dataset

Figure 6.5.: Using indicated datasets to guess 4- and 6-digit PIN datasets from Markert

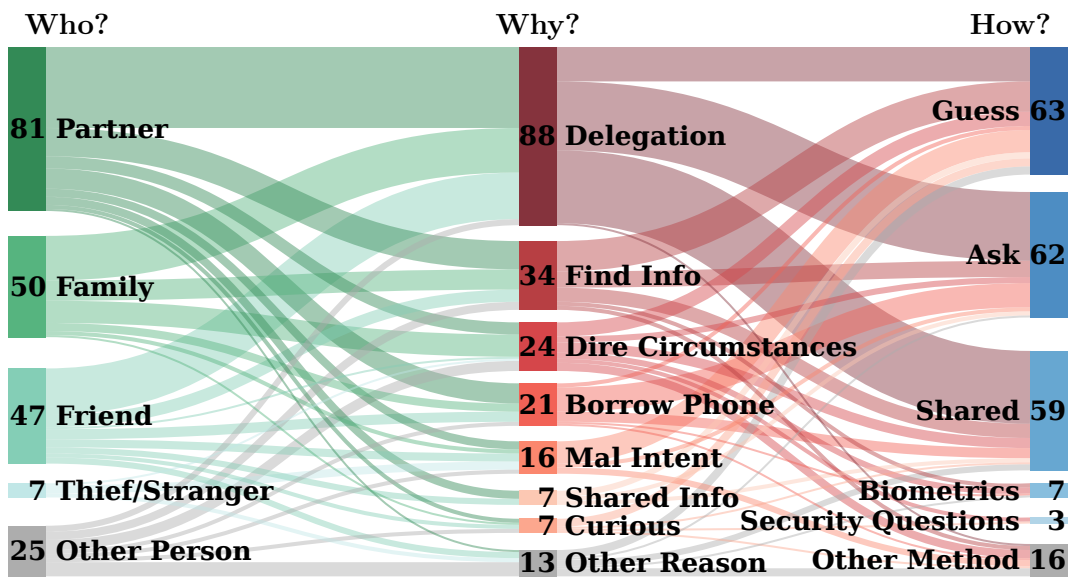


Figure 6.6.: Participants' answers to **Q9** asking *who* would try to access their smartphone, **Q10** asking *why* this person would try to access it, and **Q11** asking *how* this person would try to gain access.

7

Summary and Future Work

Contents

7.1. Summary and Key Results	144
7.2. Thesis Statement Evaluation	145
7.3. Outlook and Future Work	146

7.1. Summary and Key Results

In this dissertation, we examined several aspects around use of KBAs, especially PINs used for mobile device unlock. These necessarily have grown out of newly collected and unique datasets. As explained above, studying the use, re-use, and composition of KBAs is a challenge: users are generally told not to reveal their KBA, and users might not have full awareness of the number of times a KBA is re-used. When it first appeared, Chapter 3 represented a fresh look at the problem of password reuse, drawing on data collected from users' everyday activities by the Zeus banking trojan. We found that password complexity for financial sites was indeed higher than for social-networking and shopping sites, a finding that has been confirmed by subsequent work.

Chapter 4 focused on mobile device PINs, to our knowledge gathering the first new datasets in a rigorous study specifically priming participants for mobile device unlock. Among our findings is that, at least for the crucial first guesses before a device throttles or locks, 6-digit PINs offer no additional benefit over their 4-digit counterparts. Six-digit PINs are of benefit once large guess counts are considered, but alas, an attacker with unthrottled access can easily perform exhaustive search.

We additionally looked at the practice of blocklisting, especially as currently deployed on Apple iOS. Their 274-member 4-digit PIN blocklist, for example, is about as effective as a much smaller blocklist. On the other hand, for a blocklist to be significantly more effective, it would need to be quite large: a 2740-member list in the throttled setting would lead to only 1% of PINs guessed by a data-driven attacker within 100 guesses. Ultimately the purpose of blocklisting is to provide a meaningful intervention that can educate users on the strength of their PINs. This hopeful sign is a thread explored throughout this thesis. Users have unfortunate tendencies toward picking bad KBAs, but education of the right kind can help address the problem.

Chapter 5 examined use of the PIN feature in the Signal secure messaging app. What we found is a fairly large disparity in PIN complexity between two groups of users: those who understand (and can explain) the purpose of the Signal PIN and those who cannot. This observation reinforces one of our central points: if educated users are picking more-diverse PINs, then education can point the way to better outcomes in KBA selection. We also found that password managers are far more prevalent in the group that understands the PIN. As password managers are often used to generate better KBAs, there is an opportunity to educate users on their use.

Chapter 6 is the first to our knowledge to actually engage participants to try and guess one another's PINs. This dataset provides experimental justification that helps validate the PIN-guessing attacker modeled in preceding chapters. Our data showed that 31% of participants attempted to access someone else's smartphone

in the previous year. We examine how successful these novice attackers might be, showing that given 5 guesses, a novice can unlock 1 in 8 smartphones.

Considering the sheer number of smartphones in circulation, this is quite a high number, suggesting the need for more design interventions and user education to encourage users to select more secure PINs.

7.2. Thesis Statement Evaluation

Here we re-state our thesis and examine it in light of the foregoing evidence: The widespread use of knowledge-based authentication means a shared responsibility for security in the face of usability challenges. Adapting authentication to the way users comprehend, select, use, and re-use their passwords and PINs along with studying the way attackers guess online can lead to improved outcomes for all stakeholders.

Our unique datasets have shown once again the usability challenges of KBAs, such as re-use and users selecting easy-to-guess PINs/passwords. But the picture is more nuanced than that. The risk is spread unequally. User choices are not universally bad, but instead our work shows users are influenced by a number of factors.

- Account value (and password composition policy) affects the re-use rate
- PIN selection process — especially blocklist usage — affect guessability
- User comprehension of the purpose of a PIN affects PIN diversity
- The user’s mental model of the attacker’s capability affects PIN lifetime
- A desire to delegate or share access with close social connections affects users’ methods of PIN selection

In each of these cases, we can identify steps to better balance security and usability. For example, in the case of blocklisting, we showed that for iOS, a much smaller blocklist is just as effective in shifting the probability mass of user PIN selections — despite being only 10% of the size of the existing blocklist. A smaller blocklist would be encountered less often and therefore be less of a nuisance for users. Additionally, we showed that contrary to intuition, 6-digit PINs are not necessarily harder to guess than 4-digit PINs. Given that a smartphone unlock PIN will be entered many times over the course of a day, having fewer digits to enter should enhance usability.

Better communication around the use of PINs, especially for local authentication as in Signal, should lead to increased comprehension. From our study, it is evident that for part of the user base, Signal’s messaging is working as intended. For the other part who may not be privacy technology early adopters, the communication is less effective and security behaviors such as password manager usage is not as

widespread. Additional communication could help users understand why the Signal PIN is important.

Our study analyzing the guessability and threat models of human-chosen 4- and 6-digit PINs shows that novice guessers perform comparably to the data-driven attackers employed in previous chapters of this thesis [12, 49, 83], lending credence to our overall approach. More importantly, these novice guessers have the most opportunities to actually attempt smartphone unlock and therefore we argue this may be the most relevant attack scenario to address. Additionally, 45 % of participants changed their PIN specifically to keep someone out. Explaining PIN selection in light of these actual user concerns may lead to more-secure PIN choices overall.

Participants in our study on novice guessing attackers place a premium on allowing friends and social contacts to access their device. Moreover, our data show that 31% of participants in our study have attempted to access someone else’s smartphone without their knowledge in the previous year. This use case suggests a need to support forms of limited delegated access that don’t involve PIN sharing or guessing. Android 13 revamps the user-profile feature making it easier to delegate access. Given that no participants mentioned user profiles, there is a need for more user awareness and education on this improved feature, as well as how users can use it.

7.3. Outlook and Future Work

System designers continue to implement more and more features that rely on a KBA. Take for example Apple; as of this writing it has just rolled out Advanced Data Protection (iOS 16.3, 13 Dec 2022) worldwide which relies on the device passcode (PIN or password). In this way, a user’s iCloud backups can be secured so that Apple does not maintain the decryption keys. Observe this “end-to-end encryption” concept is also used in Signal. As these different uses of the KBA increase, the number of threat models a user must understand increases. How will users respond? Future work is needed to determine what educational interventions may be effective in protecting PINs from guessing.

KBA replacement As always, industry and academia look to potential replacements for KBAs. WebAuthN shows a great deal of promise, as does Apple’s new (as of this writing) support for detachable hardware security modules. Though they are sometimes called “key fobs” or “security keys,” the net effect is the same: a user can prevent unlocking of their smartphone unless a hardware security module is attached to the phone. Future work is needed to understand why users elect to use them and what they believe they are achieving. Certainly this is a security feature with a big usability tradeoff: the user must insert/otherwise make available the hardware security module only when they want to use the phone. If the security key is always

present/stored with the smartphone, an attacker can simply carry on as before. If the security key is kept in a secure location, the user must retrieve it in order to use their smartphone. Will the users of this feature diligently manage the security key-smartphone relationship? Or will users disable this feature in frustration? Closer to our work, many of the hardware security modules themselves require a PIN to unlock it. Do users select better PINs for this use case? Or do they fall back to old habits selecting and re-using guessable PINs? All these topics merit further investigation.

KBA re-use Users have an ever-growing collection of passwords to choose, remember, and enter. A common coping strategy is to reuse a handful of passwords across a number of sites. Still, research efforts have been hampered by lack of data. Innovative methodologies could lead to better understanding of re-use as it applies in new settings such as local (in-app) authentication. For example, we have explored how often do people re-use their mobile device unlock KBA as their Signal PIN? What exactly is the unaddressed risk and what could be done about it? We also explored the impact of password policies on re-use, finding that financial-site passwords were more likely to be re-used. Since that initial experiment, prevailing guidance on password composition rules in NIST SP 800-63B [3] has changed. Password rotation and complexity rules involving specific characters, such as requiring a number or symbol, are no longer recommended. As the research community delves deeper into this area, guidance is certain to change again. What effect do these new rules have on re-use for accounts of different value? Are there targeted interventions that could assist users in selecting better passwords?

Local (In-app) authentication As with the deployment of the Signal PIN, many apps — especially financial — require secondary authentication beyond the device being unlocked to use the app. It remains to be seen how widespread this practice will become. From our work, we know that people routinely share access to their device with friends and family members who may be trusted to play games or change the music, but not trusted to transfer funds for example. Will more apps require this secondary authentication? Will users struggle to comprehend the purpose of local authentication, as with the Signal PIN? Will users cope with the demand for yet another KBA by re-using PINs that are easily guessed?

Targeted guessing Our work shows that most participants envision friends and family as most likely to unlock their device. Our experimental results show something of a lower bound: the guessing performance of a novice stranger. But we know that users often rely on numbers from their personal history (birthdays, anniversaries, etc.) when selecting a KBA. Friends and family could take advantage of this fact by modifying the guessing order: prioritizing known special numbers. At this point, we

can surmise these hints *should* help the attacker, but it is not yet known how much of an advantage is conferred.

Concluding remarks Despite their known shortcomings, KBAs are here to stay. Rather than their eventual phase-out, which has been predicted many times before, we seem to have more KBAs than ever. In this dissertation we have studied the case of 4- and 6-digit PINs for mobile unlock authentication in detail. Yet, it is clear from our summary of future work that there is much more to be done. This thesis is not the last word on PIN guessability by any means. Researchers have much more to contribute to lead to better outcomes for all stakeholders.

List of Figures

2.1. iOS lock screen shows the number of digits in a user's PIN	19
3.1. Fraction of passwords successfully guessed	40
3.2. Fraction of passwords successfully guessed (zoomed in)	41
3.3. Re-use of passwords for variable levels of similarity	47
4.1. Priming information provided to the participants	59
4.2. PIN creation interface	59
4.3. Blocklist warning with the ability to "click through"	60
4.4. Blocklist warning without the ability to "click through"	60
4.5. Guessing performance against mobile authentication systems	69
4.6. Guessing performance by dataset based on the required time	70
4.7. PIN creation times among the different treatments	73
4.8. 4-digits PINs: Ideal blocklist size	77
4.9. 6-digits PINs: Ideal blocklist size	78
4.10. Participants' PIN selection and changing strategies	81
4.11. Participants' perceptions of their PIN	82
4.12. Participants' sentiments classified using EmoLex	83
5.1. Prompt used by Signal to occasionally ask users to verify their PIN	96
5.2. Classification of the participants (enthusiasts and casuals)	100
5.3. Classification of participants who disabled/did not set a PIN	103
5.4. PIN memorability and verification	104
5.5. Frequency of PIN reuse and sharing	106
5.6. Primary Signal PIN selection strategy	107
5.7. Why a PIN with this security level?	108
6.1. Instructions shown before guessing others' secret PIN	137
6.2. PIN guessing interface	137
6.3. Individual performance of participants' guesses	139
6.4. Participants' guessing performance	140
6.5. Guessing performance against Markert-4 and Markert-6	141
6.6. Participants' answers to <i>who</i> would try to access their phone	142
B.1. First prompt to ask Signal users to create a PIN	167
B.2. Updated prompt to ask Signal users to create a PIN	167
B.3. Prompt used when Signal users wish to change their PIN	168

List of Tables

3.1. Overview of the password lists	34
3.2. Accuracy of experimental partial guessing entropy	40
3.3. Experimental partial guessing entropy using John the Ripper	42
3.4. Comparing our results on password re-use with previous work	45
4.1. Rate limiting on mobile operating systems	55
4.2. Datasets for strength estimations and comparisons	56
4.3. Overview of studied treatments	61
4.4. Usage of mobile unlock authentication schemes	64
4.5. Guessing difficulty for a perfect knowledge attacker	67
4.6. Overlap of the First-4 PINs with the First-6 PINs	72
4.7. Security metrics and usage times for PINs	73
4.8. Attacker’s gain from blocklist knowledge	75
4.9. Changes in participants’ PIN selection strategies across treatments	80
5.1. Demographics of participants divided by subgroups	98
5.2. PIN composition among different user groups	109
6.1. Demographic information of participants	123
6.2. Responses to the PIN reuse question	124
6.3. Features of the Secret PINs and Guesses	138
A.1. Participant demographics	158
A.2. Participant device usage	159
A.3. Top ten selection strategies	160
A.4. Strategies of participants that changed their PIN	160
A.5. Participant reactions to blocklisting	161
B.1. Enthusiasts: How PINs are used by Signal	169
B.2. Casuals: How PINs are used by Signal	170
B.3. Why set a Signal PIN?	171
B.4. Why not set a Signal PIN?	172
B.5. Why disable the Signal PIN?	173
B.6. What participants would do, if they forgot their Signal PIN	173
B.7. Why disable Signal PIN reminders?	174
B.8. Primary Signal PIN selection strategy	174
B.9. Why a PIN with this security level?	175
B.10. Why a PIN in other messengers?	176
B.11. Signal PIN reuse in another messenger	177



PIN Blocklists

Appendices for Chapter 4, based on the publication:

Philipp Markert, Daniel V. Bailey, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv, “On the Security of Smartphone Unlock PINs,” in *ACM Transactions on Privacy and Security (ACM TOPS)*. ACM, Nov. 2021.

Includes:

- Survey Instrument
- Demographics
- Device Usage
- PIN Selection and Changing Strategies
- Feelings and Sentiments

A.1. Survey Instrument

Questions for participants who **hit** the blocklist.

We noticed that you received the following warning while choosing your PIN:

[A screenshot of the same warning message that the participant saw during the study.]

People use different strategies for choosing their PINs. Below, we will ask about your strategy.

1. Prior to seeing the warning above, what was your strategy for choosing your PIN?

Answer: _____

2. After receiving the warning message, please describe how or if your strategy changed when choosing your PIN.

Answer: _____

The “Extra” question was only asked if the participant had the option to ignore the warning and did so by clicking “Use Anyway.”

- (Extra) You selected “Use Anyway” when choosing your final PIN. Please describe why you did not change your final PIN after seeing this warning message.

Answer: _____

3. Please describe three general feelings or reactions that you had after you received this warning message.

Feeling 1: _____ Feeling 2: _____ Feeling 3: _____

Please select the answer choice that most closely matches how you feel about the following statements:

4. My initial PIN creation strategy caused the display of this warning.
 - Strongly agree ○ Agree ○ Neutral ○ Disagree ○ Strongly Disagree

Questions for participants who did **not hit** the blocklist.

People use different strategies for choosing their PINs. Below, we will ask about your strategy.

1. What was your strategy for choosing your PIN?

Answer: _____

Imagine you received the following warning message after choosing your PIN:

[A screenshot of the warning message as in Figure 4.3 or Figure 4.4.]

2. Please describe how or if your strategy would change as a result of the message.

Answer: _____

3. Please describe three general feelings or reactions that you would have had after you received this warning message.

Feeling 1: _____ Feeling 2: _____ Feeling 3: _____

Please select the answer choice that most closely matches how you feel about the following statements:

4. My PIN creation strategy would cause this warning message to appear.
 Strongly agree Agree Neutral Disagree Strongly Disagree

From now on all participants saw the same questions.

5. It is appropriate for smartphones to display warning messages about PIN security.
 Strongly agree Agree Neutral Disagree Strongly Disagree

Please select the answer choice that most closely matches how you feel about the following statements referring to the final PIN you chose:

The order of questions 6, 7, and 9 was chosen randomly for each participant. The attention check question was always the 8th question.

6. I feel the PIN I chose is:
 - Secure Somewhat secure Neither secure nor insecure Somewhat insecure Insecure
7. I feel the PIN I chose is:
 - Easy to remember Somewhat easy to remember Neither easy nor hard to remember
 - Somewhat hard to remember Difficult to remember
8. What is the shape of a red ball?
 - Red Blue Square Round
9. I feel the PIN I chose is:
 - Easy to enter Somewhat easy to enter Neither easy nor hard to enter Somewhat hard to enter Difficult to enter
10. What is your age range?
 - 18-24 25-34 35-44 45-54 55-64 65-74 75 or older Prefer not to say
11. With what gender do you identify?
 - Male Female Non-Binary Other Prefer not to say
12. What is the highest degree or level of school you have completed?
 - Some high school High school Some college Trade, technical, or vocational training
 - Associate's Degree Bachelor's Degree Master's Degree Professional Degree Doctorate
 - Prefer not to say
13. Do you use any of the following biometrics to unlock your primary smartphone? (Select all that apply)
 - Fingerprint Face Iris Other biometric I do not use a biometric I do not use a smartphone Prefer not to say

If the participant stated they use a biometric in question 13:

- 14A) How do you unlock your smartphone, if your biometric fails or when you reboot your primary smartphone?
 - None Pattern 4-digit PIN 6-digit PIN PIN of other length Alphanumeric password
 - I use an unlock method not listed here I do not use a smartphone Prefer not to say

If the participant stated they do not use a biometric in question 13:

- 14B) What screen lock do you use to unlock your primary smartphone?
 - None Pattern 4-digit PIN 6-digit PIN PIN of other length Alphanumeric password
 - I use an unlock method not listed here I do not use a smartphone Prefer not to say
15. What is the operating system of your primary smartphone?
 - Android iOS (iPhone) Other I do not use a smartphone Prefer not to say

16. Which of the following best describes your educational background or job field?
 - I have an education in, or work in, the field of computer science, computer engineering or IT.
 - I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.
 - Prefer not to say to say

17. Please indicate if you have honestly participated in this survey and followed instructions completely. You will not be penalized/rejected for indicating 'No' but your data may not be included in the analysis:
 - Yes ○ No

18. Please feel free to provide any final feedback you may have in the field below.
Answer: _____

A.2. Demographics

Table A.1.: Overall demographics of the participants. For the sake of clarity, we grouped answers for *Non-Binary*, *Other*, and *Prefer not to say* under *Other*.

	Male		Female		Other		Total	
	No.	%	No.	%	No.	%	No.	%
What is your age range?	923	54%	768	45%	14	1%	1705	100%
18-24	125	7%	87	5%	5	0%	217	13%
25-34	461	27%	350	21%	5	0%	816	48%
35-44	231	14%	195	11%	2	0%	428	25%
45-54	72	4%	84	5%	0	0%	156	9%
55-64	24	1%	47	3%	0	0%	71	4%
65-74	10	1%	5	0%	0	0%	15	1%
Prefer not to say	0	0%	0	0%	2	0%	2	0%
What is the highest degree or level of school you have completed?	923	54%	768	45%	14	1%	1705	100%
Some High School	3	0%	4	0%	0	0%	7	0%
High School	95	6%	66	4%	3	0%	164	10%
Some College	208	12%	142	9%	5	0%	355	21%
Training	33	2%	28	2%	0	0%	61	4%
Associates	85	5%	104	6%	2	0%	191	11%
Bachelor's	389	23%	321	19%	2	0%	712	42%
Master's	82	5%	86	5%	0	0%	168	10%
Professional	13	1%	8	0%	0	0%	21	1%
Doctorate	14	1%	9	0%	0	0%	23	1%
Prefer not to say	1	0%	0	0%	2	0%	3	0%
Which of the following best describes your educational background or job field?	923	54%	768	45%	14	1%	1705	100%
Tech	360	21%	109	7%	3	0%	472	28%
No Tech	534	31%	638	37%	8	0%	1180	69%
Prefer not to say	29	2%	21	1%	3	0%	53	3%

A.3. Device Usage

Table A.2.: Answers of participants regarding their device usage. Note, for the biometrics question, participants selected all that apply. For the sake of clarity, we grouped answers for *Non-Binary*, *Other*, and *Prefer not to say* under *Other*.

	Male		Female		Other		Total	
	No.	%	No.	%	No.	%	No.	%
Do you use any of the following biometrics to unlock your primary smartphone?	923	54%	768	45%	14	1%	1705	100%
Fingerprint	504	30%	395	23%	7	0%	906	53%
Face	161	9%	102	6%	0	0%	263	15%
Iris	41	3%	17	1%	0	0%	58	4%
Other Biometric	19	1%	26	2%	0	0%	45	3%
No Biometric	299	18%	266	16%	5	0%	570	34%
No Smartphone	2	0%	0	0%	0	0%	2	0%
Prefer not to say	28	2%	28	2%	2	0%	58	4%
How do you unlock your smartphone, if your biometric fails or when you reboot your primary smartphone?	594	55%	474	44%	7	1%	1075	100%
None	2	0%	5	0%	0	0%	7	1%
Pattern	93	9%	55	5%	0	0%	148	14%
4-digit PIN	262	24%	245	23%	3	0%	510	47%
6-digit PIN	177	16%	141	14%	4	0%	322	30%
PIN of other length	20	2%	12	1%	0	0%	32	3%
Alphanumeric	30	3%	12	1%	0	0%	42	4%
Other method	6	1%	2	0%	0	0%	8	1%
No smartphone	1	0%	0	0%	0	0%	1	0%
Prefer not to say	3	0%	2	0%	0	0%	5	0%
What screen lock do you use to unlock your primary smartphone?	329	52%	294	47%	7	1%	630	100%
None	85	13%	104	17%	0	0%	189	30%
Pattern	54	8%	32	5%	2	0%	88	13%
4-digit PIN	115	18%	101	16%	2	0%	218	36%
6-digit PIN	32	4%	27	4%	0	0%	59	8%
PIN of other length	8	1%	3	0%	0	0%	11	2%
Alphanumeric	8	1%	7	1%	0	0%	15	3%
Other method	10	2%	4	1%	0	0%	14	2%
No smartphone	0	0%	1	0%	0	0%	1	0%
Prefer not to say	17	3%	15	2%	3	0%	35	6%
What is the operating system of your primary smartphone?	923	54%	768	45%	14	1%	1705	100%
Android	592	35%	408	24%	8	0%	1008	59%
iOS	323	19%	349	21%	4	0%	676	40%
Other	2	0%	4	0%	0	0%	6	0%
No smartphone	0	0%	0	0%	0	0%	0	0%
Prefer not to say	6	0%	7	0%	2	0%	15	1%

A.4. PIN Selection and Changing Strategies

Table A.3.: We coded and analyzed a sample of 314 PIN selection strategies. Below, we list the top 10 selection strategies. Two coders independently coded the data. The level of agreement among the coders, measured by Cohen’s kappa, was $\kappa = 0.90$. Question: “*People use different strategies for choosing their PINs. Below, we will ask about your strategy. What was your strategy for choosing your PIN?*”

Code Name	Frequency	Description	Example PIN	Sample from the Study
Memorable	77	Memorability was the main concern	2827 / 777888	“A number easy to remember.”
Date	65	Special date like anniversary, birthday, graduation day	1987 / 112518	“A date I won’t forget.”
Pattern	37	Visualized a pattern on the PIN pad	2580 / 137955	“The numbers on how they appeared on the PIN pad.”
Random	33	Randomly chosen digits	4619 / 568421	“Random numbers that do not repeat.”
Meaning	27	Personal meaning; Familiar or significant number	6767 / 769339	“I chose my favorite numbers and used them repeatedly.”
Reuse	18	Reused PIN from a different device/service	0596 / 260771	“The one I normally use.”
Simple	16	Simplistic, comfortable, easy	0000 / 123987	“To just chose an easy PIN.”
Word	12	Textonyms; Converted a word to a number	2539 / 567326	“Dog name.”
System	10	User’s established systematic strategy	0433 / 041512	“I used the numbers from the current time 04:33 PM.”
Phone	7	(Partial) phone number	1601 / 407437	“I used the first four digits of a friend’s phone number.”

Table A.4.: We coded and analyzed a sample of 183 PIN changing strategies of participants that encountered a blocklist and in response changed their PIN. Below we list and explain our codes. Two coders independently coded the data. The level of agreement among the coders, measured by Cohen’s kappa was $\kappa = 0.92$. Question: “*After receiving the warning message, please describe how or if your strategy changed when choosing your PIN.*”

Code Name	Frequency	Description	Use Case	Strategy	Sample from the Study
Same	37	Same strategy for both	Selection	Date	“Birthday of relative.”
			Change	Date	“Chose another birthday.”
Minor	51	Slight modification of strategy	Selection	Meaning	“It’s one I remember, a number with personal significance.”
			Change	Meaning++	“I changed one number in the sequence to get the app to accept it.”
New	95	New strategy that is different	Selection	Date	“I used my girlfriend’s birthday.”
			Change	Phone	“I changed my strategy to a memorable phone number’s last 4 digits.”

A.5. Feelings and Sentiments

Table A.5.: As part of our questionnaire, we asked participants for 3 feelings about the blocklist warning. We coded and analyzed these feelings from a sample of 182 participants that encountered a blocklist. We also included 21 participants that only imagined hitting a blocklist. Below, we list the top 20 reported feelings. Two coders independently coded the data and the level of agreement between the coders, measured by Cohen’s kappa was $\kappa = 0.98$. Question: “Please describe three general feelings or reactions that you had after you received this warning message.” or “Please describe three general feelings or reactions that you would have had after you received this warning message.”

Code Name	Frequency	Sample from the Study	Sentiment
Annoyance	125	“Annoyed by this message.”	Negative
Worried	81	“I am worried about my PIN’s security.”	Negative
Frustrated	56	“This message frustrates me.”	Negative
Surprised	53	“Surprised to see this message.”	Neutral
Indifference	48	“Don’t care about this message.”	Negative
Thinking	47	“Thinking about my PIN’s security.”	Neutral
Acceptance	46	“I agree with this message.”	Positive
Fear	43	“Afraid of attackers.”	Negative
Compelling	41	“Motivated to change my PIN.”	Positive
Doubt	39	“I distrust the veracity of this message.”	Negative
Confusion	35	“This message is confusing.”	Negative
Angry	32	“Angry this message appeared.”	Negative
Cautious	30	“Cautious about my PIN.”	Positive
Happy	24	“Happy my PIN will be stronger.”	Positive
Curiosity	19	“I wonder why this message appeared.”	Positive
Shame	19	“Ashamed my PIN wasn’t strong.”	Negative
Remember	17	“I might forget my PIN.”	Neutral
Alert	15	“I’m now more aware.”	Neutral
Disappointed	14	“Disappointed seeing this warning.”	Negative
Safe	13	“Confident this PIN will be safe.”	Positive

B

Signal PINs

Appendices for Chapter 5, based on the publication:

Daniel V. Bailey, Philipp Markert, and Adam J. Aviv, “I have no idea what they’re trying to accomplish:” Enthusiastic and Casual Signal Users’ Understanding of Signal PINs, in *Seventeenth Symposium on Usable Privacy and Security (SOUPS ’21)*. Virtual Conference: USENIX, Aug. 2021.

Includes:

- Additional Pre-Screening Study
- Survey Instrument of the Main Study
- Additional Figures
- Codebooks

B.1. Additional Pre-Screening Study

The following question was asked in an additional pre-screening study on Prolific to be able to recruit more Signal users for our main study:

- P1** Which instant messaging apps do you use? (Select all that apply)
 WhatsApp Facebook Messenger Signal Telegram iMessage WeChat QQ Other, please specify: _____

B.2. Survey Instrument of the Main Study



- Q1** Signal Private Messenger is a cross-platform encrypted messaging service. Do you use Signal?
 Yes No

[Participants who indicate No are screened out of the survey at this point, and only Signal users move forward]

- Q2** I use Signal primarily on:
 Android Apple iPhone
 Other, please specify: _____

- Q3** I also use Signal on: (Select all that apply)
 Desktop Tablet None of these

- Q4** PINs are a new feature provided by Signal. In your own words, please explain how PINs are used by Signal.
 Answer: _____

- Q5** Did you set a Signal PIN?
 Yes No

[Participants who indicate Yes to Q5]

- Q6a** Why did you choose to set a PIN?
 Answer: _____

[Participants who indicate No to Q5]

- Q6b** Why did you choose not to set a PIN?
 Answer: _____

[Participants who indicate No to Q5 skip ahead to Q25]

- Q7** Since setting your Signal PIN, are you still using it, or have you since disabled it?
 My Signal PIN is currently enabled My Signal PIN is currently disabled

[Participants who indicated that their PIN is disabled in Q7]

- Q8** Why did you disable your Signal PIN?
 Answer: _____

[Participants who indicated that their PIN is disabled in Q7 skip ahead to Q25]

[Participants who indicated that their PIN is enabled in Q7]

- Q9** How frequently do you have difficulty remembering your Signal PIN?
 ○ Very frequently ○ Frequently ○ Occasionally ○ Rarely ○ Very rarely ○ Never

[Participants who indicated that their PIN is enabled in Q7]

- Q10** If you were to forget your Signal PIN, what would you do?

Answer: _____

[A screenshot of the Verify PIN prompt (see Figure 5.1, Page 96)]

- Q11** Have you seen this dialog in Signal?

○ Yes ○ No

[Participants who indicated that they have seen the dialog in Q11]

- Q12** When prompted, how frequently do you verify your Signal PIN?

○ Very frequently ○ Frequently ○ Occasionally ○ Rarely ○ Very Rarely ○ Never

[Participants who indicated that they have seen the dialog in Q11]

- Q13** Have you disabled Signal PIN reminders?

○ Yes ○ No

[Participants who indicated that they have seen the dialog in Q11 and that they have disabled reminders in Q13:]

- Q14** Why did you disable Signal PIN reminders?

Answer: _____

- Q15** Many smartphone users also unlock their phone using a PIN or passcode. Is your Signal PIN the same one you use to unlock your smartphone?

○ Yes ○ No ○ Unsure ○ I do not lock my smartphone with a PIN or passcode

- Q16** Do you use your Signal PIN in other contexts besides unlocking your smartphone? (Select all that apply)

ATM/Credit/Payment Card Laptop/PC Online Accounts

Electronic Door Lock Home Security System/Safe Garage Door Opener Car/Truck/SUV

Bike/Gym lock Voicemail Gaming Console

Smartwatch Other, please specify: _____

- Q17** Do you use your Signal PIN in any other mobile applications?

○ Yes, please specify: _____ ○ No

- Q18** Do you share your Signal PIN with friends or family?

○ Yes ○ No

- Q19** How long is your Signal PIN?

Answer: _____

- Q20** What was your primary strategy in selecting your Signal PIN?

Answer: _____

- Q21** Compared to other PINs you use, did you try to pick a Signal PIN that was:

○ The most secure PIN you use ○ About the same security as other PINs you use ○ Less secure than other PINs you use

- Q22** Why did you choose a PIN with this security level?

Answer: _____

- Q23** What is the shape of a red ball?

○ Red ○ Round ○ Blue ○ Square

[For each category, this question uses sliders so the user can choose a value between 0 and 12, or check the category's box for "Not applicable:"]

Q24 My Signal PIN contains:

Digits: _____

Letters: _____

Special characters: _____

Q25 Do you use other messenger services like: (Select all that apply)

Facebook messenger Skype Telegram WeChat WhatsApp

Other, please specify: _____

[For the services above, place them in order of how often you use them:]

Q26 Besides Signal, did you set a PIN in one or more other messengers?

Yes No

[Participants who indicate Yes to Q26]

Q27a Why did you set a PIN in the other messenger(s)?

Answer: _____

[Participants who indicate No to Q26]

Q27b Why didn't you set a PIN in the other messenger(s)?

Answer: _____

[Participants who indicate No to Q26 skip ahead to D1]

[Participants who indicate Yes to Q26]

Q28 In which other messenger(s) did you set a PIN? (Select all that apply)

Facebook Messenger Skype Telegram WeChat WhatsApp

Other, please specify: _____

[Participants who indicate Yes to Q26]

Q29 Did you re-use the same PIN with any of these other messengers?

Yes No

[Participants who indicate Yes to Q29]

Q30a Why did you re-use the same PIN in another messenger?

Answer: _____

[Participants who indicate No to Q29]

Q30b Why didn't you re-use the same PIN in another messenger?

Answer: _____

D1 What is your age range?

18-24 25-34 35-44 45-54 55-64 65-74 75 or older Prefer not to say

D2 With what gender do you identify?

Male Female Non-Binary Other Prefer not to say

D3 What is the highest degree or level of school you have completed?

Some high school High school Some college Trade, technical, or vocational training Associate's Degree Bachelor's Degree Master's Degree Professional Degree Doctorate Prefer not to say

D4 What is your country of residence?

[Drop-down all countries]

D5 Does your educational background or job field involve IT?

Yes No Prefer not to say

B.3. Additional Figures

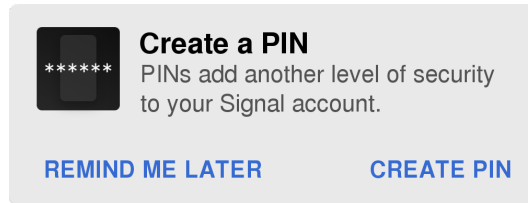


Figure B.1.: First prompt to ask Signal users to create a PIN

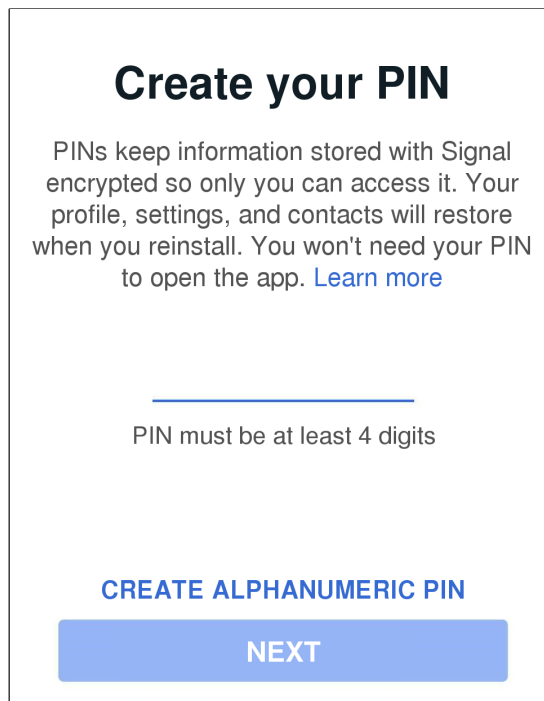
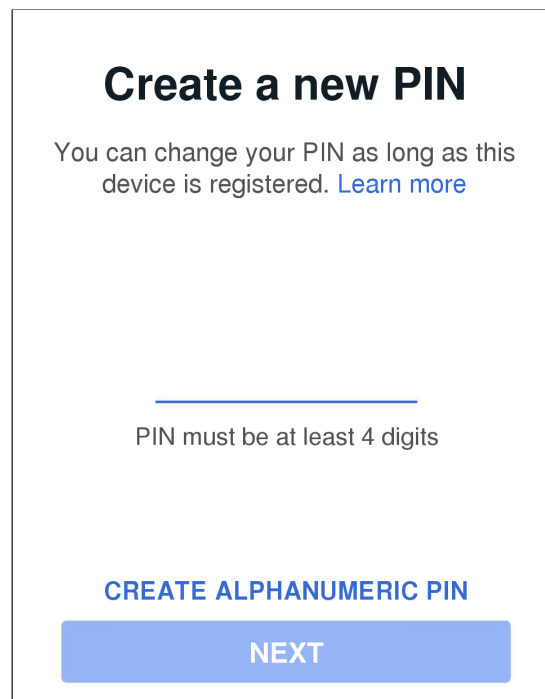


Figure B.2.: Updated prompt to ask Signal users to create a PIN



Create a new PIN

You can change your PIN as long as this device is registered. [Learn more](#)

PIN must be at least 4 digits

CREATE ALPHANUMERIC PIN

NEXT

Figure B.3.: Prompt used when Signal users wish to change their PIN

B.4. Codebooks

We have 10 open-ended questions in our study for which two coders independently coded all answers we received. The two coders compared and combined codes until they agreed. For each question, n depicts the number of responses. As a single response might receive multiple codes, the number of codes does not sum to n . All codes of participant responses are shown below.

Table B.1.: **Q4:** “PINs are a new feature provided by Signal. In your own words, please describe how PINs are used by Signal.” ($n = 235$) Based on the answer to **Q4**, 132 participants were classified as *enthusiasts*.

Code Name	No.	%	Description	Sample from the Study
Backup	65	49%	Participant mentions secure backup of settings and contacts but not messages	“The PIN enables storing a backup of the user’s settings on the signal servers in an encrypted form.” (P13)
Encryption	45	34%	Participant mentions encryption based on the PIN	“deriving a key to encrypt data stored on signals servers” (P82)
Contacts	31	24%	Participant mentions the backup of contact data	“To secure contacts data saved on signal server with your own pin” (P7)
Registration	23	17%	Participant mentions the registration lock	“to prevent reregistration of an account for the same mobile phone number for a given amount of time” (P91)
Settings	8	6%	Participant mentions the backup of settings	“For encrypted backups - on cloud storage - for the user settings and profile. Not the messages themselves.” (P127)
Keying	7	5%	Participant mentions the keying of the PIN	“They say it’s part of a keying mechanism providing a non-phone-number value that allows secure storage and retrieval of contacts and social graph info across devices.” (P2)
Phone number	6	5%	Participant mentions the intention of Signal to move away from the phone number as an identifier	“I think for backup purposes and to later fade out the phone number as identifier.” (P106)
Profile	4	3%	Participant mentions the backup of profile information	“PINs are used for recovery of settings and profile information after re-installation of Signal app.” (P54)
Groups	3	2%	Participant mentions the backup of group memberships	“They are used to secure private information such as group membership and store it on the Signal server” (P35)
Anti-Cloud	2	2%	Participant expresses negative sentiment about the data being stored by Signal	“they are used to secure data in acloud service that is beeing forced on users” (P208)
SVR	1	1%	Participant mentions Secure Value Recovery (SVR)	“Secure Value Recovery” (P222)

Table B.2.: **Q4:** “PINs are a new feature provided by Signal. In your own words, please describe how PINs are used by Signal.” ($n = 235$) Based on the answer to **Q4**, 103 participants were classified as *casuals*.

Code Name	No.	%	Description	Sample from the Study
Don't Know	57	55 %	Participant does not mention any terms that may indicate an understanding	“I don't understand their purpose very well. I thought that they might be using the PIN system to verify the identity of the person using signal (if for instance someone unauthorized gained access to the phone), but the way that pin entry is optionally offered every few weeks doesn't align with such a purpose. as such, I have no idea what they're trying to accomplish.” (P178)
Messages	21	20 %	Participant mentions the backup of messages	“Secure backup of messages” (P23)
Unlock	21	20 %	Participant mentions that the PIN is used to protect access to the app	“Protect application from opening from an unlocked phone” (P37)
Security	2	2 %	Participant mentions security	“Security somehow...” (P7)
Inconvenient	1	1 %	Participant mentions inconvenience	“I have not tried it considering that it'd pop up for additional verification through the pin.” (P212)

Table B.3.: Codes assigned to the answers of the participants for (Q6a) on adopting a PIN.
Q6a: “Why did you choose to set a PIN?” ($n = 202$).

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Security	7	5%	26	25%	Participant mentions security	“I wanted some extra security” (P50)
Required	25	19%	6	6%	Participant mentions that there was no other choice	“I did not see an option to not set one” (P121)
Prompted	8	6%	13	13%	Participant mentions that Signal showed a prompt that suggested it	“cause signal asked me to do so” (P78)
Don’t Know	4	3%	16	16%	Participant does not mention any of the terms that indicate an understanding	“So that people that get a hold of my phone would have greater difficulty accessing my messages.” (P159)
Annoying	12	9%	6	6%	Participant mentions the feature was annoying	“Because it kept hassling you with a pop up screen” (P154)
Registration	14	11%	2	2%	Participant mentions the registration lock	“I chose to set a PIN to both set registration lock and to backup my contacts.” (P51)
Features	8	6%	3	3%	Participant mentions features without further defining them	“To be able to use the features that depend on a PIN” (P111)
No harm	8	6%	3	3%	Participant describes there being no drawbacks	“No disadvantage doing so” (P20)
Trust	4	3%	2	2%	Participant expresses trust in Signal	“I trusted the app and just did it when prompted.” (P155)
Privacy	2	2%	3	3%	Participant mentions valuing privacy	“Because privacy is important to me and it’s an added layer of it” (P162)
Contacts	3	3%	0	0%	Participant mentions the backup of contact data	“I want to be able to access contact data saved on signal server if I somehow can’t access my current phone” (P7)
Comfort	2	2%	0	0%	Participant mentions feeling comfortable	“Because it I felt comfortable with the trade-off. Picked a long passphrase rather than a four digit PIN.” (P127)
Encryption	1	1%	1	1%	Participant mentions encryption based on the PIN	“For me it’s okay to encrypt and store data on Signal’s servers as I have no high threat model.” (P87)
Lock	1	1%	0	0%	Participant mentions locking apart from registration lock	“basically to lock and to avoid sim hijacking” (P19)

Table B.4.: Codes assigned to the answers of the participants for (Q6b) on adopting a PIN.
Q6b: “Why did you choose not to set a PIN?” ($n = 33$).

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Inconvenient	4	25 %	3	18 %	Participant mentions inconvenience	“I want to access my apps as seamless and fast as possible.” (P212)
Anti-Cloud	7	44 %	0	0 %	Participant expresses negative sentiment about the data being stored by Signal	“had no desire to have any contact data uploaded” (P216)
Key management	3	19 %	0	0 %	Participant described the use of the PIN in key derivation	“I don’t trust Signal’s encryption strategy involving SGX. It’s my belief that SGX is likely to be compromised by nation-state actors, and cannot be used securely. If any of my private information must be stored persistently in a cloud service, it is unacceptable to use anything other than an encryption key that I personally control.” (P203)
Lock	0	0 %	4	24 %	Participant falsely links the phone lock to the PIN	“My phone is always locked. Additional authentication seems unnecessary” (P227)
No need	0	0 %	3	18 %	Participant mentions seeing no need	“it’s not necessary for me” (P233)
Memorability	1	6 %	1	6 %	Participant described memorability issues	“I didn’t want to be bothered with remembering another code.” (P224)
No awareness	1	6 %	0	0 %	Participant did not know Signal had a PIN	“I didn’t know it existed.” (P232)
Not prompted	0	0 %	1	6 %	Participant said they were not prompted to set a PIN	“was not asked.” (P218)
Rarely use	0	0 %	1	6 %	Participant described using Signal only rarely	“I dont use signal much, its not for sensitive messages so dont need the extra security” (P223)
Unsupported	0	0 %	1	6 %	Participant described using an unsupported client	“Not possible because of using a unsupported native client for SailfishOS” (P220)

Table B.5.: Q8: “Why did you disable your Signal PIN?” (n = 11)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Annoying	3	60 %	1	17 %	Participant mentions being annoyed	“It was annoying.” (P188)
Anti-Cloud	2	40 %	1	17 %	Participant expresses negative sentiment about the data being stored by Signal	“Don’t want my data stored on their server” (P193)
Inconvenient	1	20 %	1	17 %	Participant mentions inconvenience	“Verification overhead” (P212)
No backup	1	20 %	0	0 %	Participant describes not needing a backup	“It’s annoying to re-enter the PIN and I don’t need backup for signal since there’s no important conversation” (P231)
No need	1	20 %	1	17 %	Participant sees no necessity	“I do not need it” (P206)

Table B.6.: Q10: “If you were to forget your Signal PIN, what would you do?” (n = 191)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Don’t know	27	25 %	33	40 %	Participant does not know what to do	“Honestly don’t know” (P68)
PW Manager	45	42 %	12	15 %	Participant has the PIN stored in a password manager	“I’ve stored my Signal PIN in my PW manager” (P74)
Reset	0	0 %	12	15 %	Participant describes resetting the account	“Check the help page for how to reset” (P158)
Wait	8	7 %	4	5 %	Participant is aware that the PIN expires and would wait	“wait for pin expiration” (P161)
New PIN	4	5 %	7	6 %	Participant would set a new PIN	“as long as I have access to my Signal account I can set a new PIN at any time” (P18)
New account	4	4 %	5	6 %	Participant would create a new account	“I would make another account” (P181)
Reused	2	2 %	4	5 %	Participant reuses the PIN and does not expect to forget it	“It is a PIN I use for my bank cards, so I would not forget it.” (P145)
Unrecoverable	2	2 %	3	4 %	Participant accepts that there is not way to recover	“Signal said there is no way to recover it. All chats constants block list will be lost.” (P79)
Contact	0	0 %	4	5 %	Participant would contact Signal directly	“Contact the signal team” (P137)
Guess	0	0 %	3	4 %	Participant would try to guess the PIN	“try a lot of PINs i use” (P98)
Reinstall	2	2 %	0	0 %	Participant would reinstall Signal	“delete the app and reinstall it” (P106)
Written	1	1 %	1	1 %	Participant mentions that the PIN has been written down	“I would check the PIN on my journal, I wrote it down with all the passwords and the login info.” (P143)

Table B.7.: Q14: “Why did you disable Signal PIN reminders?” ($n = 45$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
PW Manager	22	67%	1	8%	Participant has the PIN stored in a password manager	“Because I have a password safe and do not need to remember the pin” (P49)
Annoyed	6	18%	5	42%	Participant describes being annoyed	“Because it asked my pin to often” (P70)
No need	5	15%	4	33%	Participant describes not needing them	“I don’t think I need them” (P160)
Memorized	0	0%	1	8%	Participant does not expect to forget the PIN	“Thought I’d be able to remember it” (P157)
Effective	0	0%	1	9%	Participant mentions the effectiveness of the reminders	“After a few reminders I was sure not to forget the PIN” (P87)

Table B.8.: Q20: “What was your primary strategy in selecting your Signal PIN?” ($n = 191$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Memorable	23	21%	30	36%	Participant mentions memorability	“My ability to remember it.” (P116)
PW Manager	28	26%	6	7%	Participant describes using a password manager	“My password safe generated it.” (P100)
Reuse	16	15%	13	16%	Participant describes reusing a PIN	“I used my PIN that I often use.” (P176)
Random	15	14%	7	8%	Participant describes choosing a random PIN	“random number generator” (P63)
Meaning	6	6%	6	7%	Participant describes choosing a meaningful PIN	“Something meaningful to me” (P77)
Security	3	3%	8	10%	Participant describes selecting a secure PIN	“just something safe and long” (P200)
Pattern	3	3%	4	5%	Participant describes choosing a PIN that depicts a pattern	“Thinking of a pattern that memorable to me” (P142)
None	2	2%	3	4%	Participant describes not having a strategy	“no strategy” (P115)
Word	2	2%	3	4%	Participant describes converting a word to a PIN (textonyms)	“Words to numbers” (P115)
Date	2	2%	1	1%	Participant describes using a date	“It’s a date that is relevant but nobody knows” (P154)
System	2	2%	1	1%	Participant describes having a certain system	“My preferred format” (P138)
Typable	0	0%	1	1%	Participant mentions a PIN that is easy to enter	“Strong alphanumeric password that is secure enough but fairly easy to type on the phone, even if I couldn’t paste it from password manager for some reason.” (P54)
Simple	0	0%	1	1%	Participant mentions simplicity	“Something simple” (P154)
Phone	0	0%	1	1%	Participant mentions a phone number	“Old phone number I can remember” (P108)

Table B.9.: Q22: “Why did you choose a PIN with this security level?” ($n = 191$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Memorability	12	11 %	18	22 %	Participant mentions memorability	“Because I wanted it to be easy to remember.” (P5)
Enough	16	15 %	15	18 %	Participant describes the security level being sufficient	“I think that’s enough” (P179)
Security	25	23 %	20	24 %	Participant mentions security	“I am fairly security conscious” (P44)
Consistent	11	10 %	6	7 %	Participant describes this level being the standard	“Because I always choose this security level.” (P109)
Trade-off	9	8 %	2	2 %	Participant describes some form of trade-off	“trade-off between remembering and security” (P60)
Reuse	7	7 %	2	2 %	Participant describes reusing a PIN	“The same as the iPhone passcode.” (P144)
PW manager	6	6 %	2	2 %	Participant describes using a password manager	“why not, if i can use a pw manager” (P76)
Don’t know	1	1 %	6	7 %	Participant cannot remember the strategy	“I don’t remember” (P84)
None	2	2 %	4	5 %	Participant describes not having a strategy	“no strategy” (P88)
Convenience	3	3 %	1	1 %	Participant mentions convenience	“Convience over security” (P113)
Privacy	2	2 %	1	1 %	Participant mentions privacy	“The chats and contacts in Signal have a relatively high level of privacy, so it should be properly protected. Yet the pin is not as good as for example my computers encryption password but as good as my android encryption phrase.” (P94)
Low-threat	0	0 %	2	2 %	Participant sees little need for data security	“The info isn’t super important” (P168)
Indifference	2	2 %	0	0 %	Participant says the PIN is unimportant	“Dont think that the pin is too important” (P120)
Rarely use	2	2 %	0	0 %	Participant described using the PIN only rarely	“Unlike my smartphone unlock pin for example, I don’t have to enter my Signal PIN frequently (never really, unless I set up a new smartphone) and thus had no problem with selecting a long and complicated PIN” (P20)
Minimum	1	1 %	0	0 %	Participant mentions a Signal requirement	“Initially 6 digits were required.” (P132)

Table B.10.: Codes assigned to the answers of the participants for (Q27a) and (Q27b) on setting a PIN in other messengers.

(a) Q27a: “Why did you set a PIN in other messenger(s)?” ($n = 49$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Security	24	67%	8	62%	Participant mentions security	“For more security, 2FA” (P95)
Prompted	4	11%	1	8%	Participant mentions being prompted by the application	“Prompted to do so, and I understand the reasons why it is a good idea.” (P196)
Required	4	11%	1	8%	Participant mentions that there was no other choice	“Forced to set” (P93)
Feature	2	6%	1	8%	Participant mentions being given the option to	“Because I could” (P117)
Don’t know	1	3%	2	16%	Participant doesn’t address the question	“Telegram” (P48)
Reuse	1	3%	0	0%	Participant mentions reusing a PIN when possible	“Since I already has a pin memorized, why not use it in other messengers” (P50)

(b) Q27b: “Why didn’t you set a PIN in other messenger(s)?” ($n = 131$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
No feature	20	35%	24	33%	Participant mentions not being able to set a PIN	“They don’t have that option” (P35)
No need	18	32%	12	17%	Participant describe that there is no necessity	“Not required” (P156)
Not asked	10	17%	20	28%	Participant describes not being asked to	“Was not asked to” (P67)
Use rarely	3	5%	4	6%	Participant describes only using them rarely	“I don’t use them often, if at all.” (P51)
Screen lock	3	5%	2	3%	Participant describes that the phone lock is sufficient	“The phone in itself has a pin” (P194)
Annoyed	2	3%	1	2%	Participant describes being annoyed	“They are inconvenient, do not know how, and I do not use them for secure messaging. My Signal is already password protected so a pin seems redundant.” (P55)
Insecure	0	0%	2	3%	Participant describes that they don’t use them for secure communication	“Not intended for secure communication.” (P62)
Comfort	1	2%	0	0%	Participant mentions feeling comfortable	“comfort” (P102)

Table B.11.: Codes assigned to the answers of the participants for (Q30a) and (Q30b) on reusing the Signal PIN in another messenger.

(a) Q30a: “Why did you re-use the same PIN in another messenger?” ($n = 10$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Memorability	5	63 %	2	100 %	Participant mentions memorability	“I was too lazy to memorize a new one... not good I know” (P50)
Messenger PIN	2	25 %	0	0 %	Participant mentions using a PIN for messengers	“Because I have one pin for messengers.” (P5)
Convenience	1	12 %	0	0 %	Participant mentions convenience	“Convivence, but it was probably a poor decision, as WhatsApp is more vulnerable to a secret warrant.” (P196)

(b) Q30b: “Why didn’t you re-use the same PIN in another messenger?” ($n = 37$)

Code Name	Enthusiasts		Casuals		Description	Sample from the Study
	No.	%	No.	%		
Security	17	65 %	6	60 %	Participant mentions security	“Reusing PINs is a bad practice.” (P54)
PW Manager	8	31 %	2	20 %	Participant describes using a password manager	“Why would i? Thats what passwordmanagers are for.d” (P23)
Other options	3	12 %	0	0 %	Participant describes having other options	“Some of them gave me the option of using my thumbprint.” (P3)
Don’t know	1	4 %	1	10 %	Participant cannot explain the reason	“I didn’t really think about it, it just happened” (P179)
Required	0	0 %	1	10 %	Participant mentions different requirements	“different lengths” (P381)



Novice Attackers

Appendices for Chapter 6, based on the publication:

Daniel V. Bailey, Collins W. Munyendo, Hunter Dyer, Philipp Markert, Miles Grant, and Adam J. Aviv, “‘Someone Definitely Used 0000’: Analyzing How Novice Attackers Guess Unlock PINs,” to appear at European European Symposium on Usable Security, Oct. 2023.

Includes:

- Survey Instrument
- Qualitative Codes

C.1. Survey Instrument

Agenda

PINs or passcodes are often used to unlock mobile devices. You will be asked to choose a PIN just like you would to protect your smartphone. Afterward, you will complete a short survey and then try to guess the PINs of other participants. You will enter 5 guesses. Finally, you will answer some questions about your experience.

You contribute to research, so please answer correctly and as detailed as possible. Next, you will practice the PIN selection.

Practice entering a [4/6]-digit PIN.

PIN pad as shown in Figure 4.2

Your Task

You will be asked to choose a digit PIN you would use to unlock your smartphone. You will need to remember your secret PIN for the duration of the study. Please DO NOT write down your secret PIN.

Create a [4/6]-digit secret PIN

A secret PIN protects your data and is used to unlock your smartphone.

PIN pad as shown in Figure 4.2

Questionnaire

People use different strategies for choosing their PINs. Below, we will ask about your strategy.

Q1 What was your strategy for choosing your secret PIN?

Answer: _____

Please select the answer choice that most closely matches how you feel about the following statements:

Q2 I feel the secret PIN I chose is:

- Secure Somewhat secure Neither secure nor insecure
- Somewhat insecure Insecure

Q3 I feel the secret PIN I chose is:

- Easy to enter Somewhat easy to enter Neither easy nor hard to enter Somewhat hard to enter Hard to enter

Q4 I feel the secret PIN I chose is:

- Easy to remember Somewhat easy to remember
- Neither easy nor hard to remember Somewhat hard to remember
- Hard to remember

Q5 What is the shape of a red ball?

- Red Blue Square Round

Q6 Was the secret PIN that you entered a PIN that you use on your smartphone or other personal devices?

- Yes No Unsure I do not lock my smartphone with a PIN

If participants indicated reuse of their PIN in Q6:

Q7 Did you choose a secret PIN you use in other contexts besides unlocking your smartphone? (Select all that apply)

- ATM/Credit/Payment Card Laptop/PC Online Accounts Bike/Gym lock Electronic Door Lock Home Security System/Safe Garage Door Opener Car/Truck/SUV
- Voicemail Gaming Console Smartwatch
- Other, please specify:
- No, I did not choose a PIN from other contexts.

Your Task

- Enter 5 PINs that you think other participants entered

- Any number of correct guesses earns a bonus \$0.50, paid 1-2 weeks after the completion of this study
- More than 100 people will be taking this study

Please enter 5 different guesses.

The following page appeared 5 times

Guess x

Try to guess someone's [4/6]-digit secret PIN. Guesses must be unique!

PIN pad as shown in Figure 6.2

About Your Guesses

Previously, you made the following guesses:

Guess 1: [pin]

Guess 2: [pin]

Guess 3: [pin]

Guess 4: [pin]

Guess 5: [pin]

Q8 In two to three sentences, please describe your overall strategy you used when guessing other participants' secret PINs.

Answer: _____

Q9 Please describe a situation where someone is most likely to unlock your smartphone. If relevant, indicate your relationship to this person but do NOT include Personally Identifiable Information (PII).

Answer: _____

Q10 In this situation, why would the individual be accessing your smartphone?

Answer: _____

Q11 In this situation, describe the strategy the individual would use to gain access to your smartphone.

Answer: _____

Q12 Did you consider this situation when you chose your secret PIN?

Yes No Unsure

Q13 Did you consider this situation when you were guessing the PINs of other participants?

Yes No Unsure

Q14 My level of concern for someone accessing my phone without consent is:

Unconcerned Somewhat unconcerned
 Neither concerned nor unconcerned Somewhat concerned Concerned

Q15 Why do you feel this level of concern?

Answer: _____

Q16 Have you ever changed your PIN to keep someone from accessing your smartphone?

Yes No Unsure

Q17 Have you ever tried to access someone else's smartphone without their knowledge?

Yes No Unsure

Q18 Have you ever granted someone else access to your smartphone?

Yes No Unsure

Re-enter your [4/6]-digit PIN

PIN pad as shown in Figure 4.2

About Your Guesses

Q19 Do you think the secret PIN you entered at the start of this survey will be guessed by other participants in this study? Why or why not?

Answer: _____

Q20 Do you think you guessed someone else's secret PIN? Why or why not?

Answer: _____

Enter Demographic Information

D1 What is your age range?

- 18–24
- 25–34
- 35–44
- 45–54
- 55–64
- 65–74
- 75 or older
- Prefer not to say

D2 What is your gender?

- Woman
- Man
- Non-binary
- Prefer to self-describe
- Prefer not to say

D3 What is the highest degree or level of school you have completed?

- Some high school
- High school
- Some college
- Trade, technical, or vocational training
- Associate's Degree
- Bachelor's Degree
- Master's Degree
- Professional Degree
- Doctorate
- Prefer not to say

D4 Do you use any of the following biometrics to unlock your primary smartphone? (Select all that apply)

- Fingerprint
- Face
- Iris
- Other biometric
- I do not use a biometric
- Prefer not to say

If participants indicated to use a biometric in D4:

D5a How do you unlock your smartphone, if your biometric fails or when you reboot your primary smartphone?

- None
- Pattern
- 4-digit PIN
- 6-digit PIN
- PIN of other length
- Alphanumeric password
- I use an unlock method not listed here
- Prefer not to say

*If participants indicated **not** to use a biometric in D4:*

D5b What screen lock do you use to unlock your primary smartphone?

- None
- Pattern
- 4-digit PIN
- 6-digit PIN
- PIN of other length
- Alphanumeric password
- I use an unlock method not listed here
- Prefer not to say

D6 What is the operating system of your primary smartphone?

- Android
- iOS (iPhone)
- Other
- Prefer not to say

D7 Which of the following best describes your educational background or job field?

- I have an education in, or work in, the field of computer science, computer engineering, or IT.
- I do not have an education in, nor do I work in, the field of computer science, computer engineering, or IT.
- Prefer not to say

One More Thing

Please indicate if you've honestly participated in this survey and followed instructions completely. You will not be penalized/rejected for indicating 'No' but your data may not be included in the analysis:

- Yes
- No

C.2. Qualitative Codes

- **no (183)**
- **yes (134)**
- **pattern (127)**
physical (2), word (2), sequential (1), odd-numbers (1)
- **simple (106)**
- **maybe (101)**
- **delegation (97)**
entertainment (24), respond (11), find-information (7), shopping (4), location (3), take-photo (2)
- **partner (96)**
- **memorable (94)**
- **random (90)**
guess (1)
- **date (89)**
- **guess (79)**
personal-hint (51), shoulder-surf (12), smudges (1)
- **ask (63)**
- **shared (61)**
- **important-information (61)**
financial (13), personal (13), sensitive (8), private (6), social-media (3), messages (2), email (2), confidential (2), apps (2), photo (1), health (1), photos (1), other-passwords (1), relationships (1), mature-content (1), encrypted (1), files (1), communication (1), browsing-history (1), contacts (1)
- **friend (58)**
- **family-member (58)**
- **personal-importance (56)**
birthday (13), zip-code (3), date (2), partner (2), favorite-number (2), months (1), year (1), date (1), name (1), acquaintance (1), lucky-number (1)
- **phone-locked (43)**
pin (20), biometric (6), password (4), access (1), pattern (1)
- **find-information (41)**
location (2)
- **physical-control (36)**
- **meaning (31)**
- **users-lazy (28)**
unaware (1)
- **picked-common (28)**
- **nothing-to-hide (28)**
- **borrow-phone (26)**
contact (12), find-information (5), entertainment (4)
- **dire-circumstances (25)**
- **reuse (24)**
- **hard-to-guess (23)**
nobody-made-easy (3), needed-more-guesses (1), generation (1), not-enough-information (1), tricky (1)
- **picked-easy (23)**
- **privacy (22)**
- **mal-intent (21)**

- **picked-pattern (21)**
- **other-method (17)**
- **uncommon (17)**
- **picked-sequence (16)**
- **easy-to-guess (15)**
simple (3), obvious (1)
- **unique (13)**
- **word (13)**
- **statistically-unlikely (12)**
- **picked-memorable (11)**
- **phone (10)**
- **picked-simple (10)**
- **shared-info (10)**
financial (2), files (2), contacts (2)
- **other-pins-more-secure-than-irl (9)**
- **family-trusted (9)**
partner (2)
- **undescribed (9)**
- **no-one (8)**
- **picked-easy-enter (8)**
- **trust-others (8)**
acquaintances (2), familiar (1)
- **system (8)**
- **biometrics (8)**
- **curious (8)**
- **picked-random (8)**
- **laziness (8)**
- **unspecified (7)**
- **none (6)**
- **picked-repetition (6)**
- **guessed-by-luck (6)**
- **no-important-information (6)**
financial (1)
- **malice (6)**
financial (2), scam (1), manipulate-documents (1)
- **hard (6)**
no-personal-info (1), no-feedback (1), others-mindset (1)
- **distress (5)**
- **thief (5)**
- **app-locked (5)**
- **picked-date (5)**
year (3)
- **common (5)**
numbers (1)
- **shared-passcode (4)**
partner (1), family (1)
- **not-obvious (4)**
- **need-prior-knowledge (4)**

- **phone-insecure (4)**
 - pin (3)*
- **financial (4)**
- **number-users (4)**
 - will-guess (1)*
- **picked-other (3)**
 - zip (1), no-strategy (1), diverse (1)*
- **unconcern (3)**
- **security-questions (3)**
- **question-of-legality (3)**
- **not-pattern (3)**
- **zipcode (3)**
- **similar-thought-process (3)**
- **stranger (3)**
- **no-motive (2)**
- **resigned (2)**
- **police (2)**
- **luck (2)**
- **similar-reasoning (2)**
- **insecure (2)**
- **used-friends-pin (2)**
- **picked-insecure (2)**
- **brick-phone (2)**
- **lost-phone (1)**
- **smartphone-pin (1)**
- **not-complex (1)**
- **easily-change-password (1)**
- **not-difficult (1)**
- **multiple-attempts (1)**
- **boss (1)**
- **not-random (1)**
- **not-a-target (1)**
- **appearance (1)**
- **manufacturer-trust (1)**
- **no-knowledge (1)**
- **other-pins-more-simple-than-irl (1)**
- **cryptocurrency (1)**
- **no-pattern (1)**
- **convenient (1)**
- **search-warrant (1)**
- **phone-unlocked (1)**
- **number-participants (1)**
- **app-security (1)**
- **typical (1)**
- **need-consent (1)**

Bibliography

- [1] Google, Inc. Google Transparency Report, July 2022. <https://transparencyreport.google.com/https/overview?hl=en/>, as of July 16, 2023.
- [2] Alma Whitten and J. Doug Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *USENIX Security Symposium*, SSYM '99, pages 169–183, Washington, District of Columbia, USA, August 1999. USENIX.
- [3] Paul A. Grassi, James L. Fenton, and William E. Burr. Digital Identity Guidelines – Authentication and Lifecycle Management: NIST Special Publication 800-63B, June 2017.
- [4] Philip G. Inglesant and M. Angela Sasse. The True Cost of Unusable Password Policies: Password Use in the Wild. In *ACM Conference on Human Factors in Computing Systems*, CHI '10, pages 383–392, Atlanta, Georgia, USA, April 2010. ACM.
- [5] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. Of Passwords and People: Measuring the Effect of Password-Composition Policies. In *ACM Conference on Human Factors in Computing Systems*, CHI '11, pages 2595–2604, Vancouver, British Columbia, Canada, May 2011. ACM.
- [6] Sonia Chiasson and Paul C. Van Oorschot. Quantifying the Security Advantage of Password Expiration Policies. *Designs, Codes and Cryptography*, 77(2–3):401–408, December 2015.
- [7] William E. Burr, Donna F. Dodson, and W. Timothy Polk. Electronic Authentication Guideline: NIST Special Publication 800-63, June 2004.
- [8] Peter Mayer, Jan Kirchner, and Melanie Volkamer. A Second Look at Password Composition Policies in the Wild: Comparing Samples from 2010 and 2016. In *Symposium on Usable Privacy and Security*, SOUPS '17, pages 13–28, Santa Clara, California, USA, July 2017. USENIX.
- [9] PCI Security Standards Council. Req. and Security Assessment Proc., Version 3.2.1, September 2022.
- [10] PCI Security Standards Council. Req. and Security Assessment Proc., Version 4.0, March 2022.
- [11] Joseph Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *IEEE Symposium on Security and Privacy*, SP '12, pages 538–552, San Jose, California, USA, May 2012. IEEE.
- [12] Philipp Markert, Daniel V. Bailey, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. On the Security of Smartphone Unlock PINs. *ACM Transactions on Privacy and Security*, 24(4):30:1–30:36, September 2021.
- [13] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. Quantifying the Security of Graphical Passwords: The Case of Android Unlock Patterns. In *ACM Conference on Computer and Communications Security*, CCS '13, pages 161–172, Berlin, Germany, November 2013. ACM.
- [14] Tony Freeth, Yanis Bitsakis, Xenophon Moussas, John H Seiradakis, A Tselikas, H Mangou, M Zafeiropoulou, R Hadland, D Bate, A Ramsey, et al. Decoding the Ancient Greek Astronomical Calculator Known as the Antikythera Mechanism. *Nature*, 444(7119):587–591, 2006.

- [15] Wolfram Hoepfner. Ein kombinationsschloss auf dem kerameikos. *Archäologischer Anzeiger*, pages 210–213, 1970.
- [16] Muhammad ibn Hamid al-Isfahani al Asturlabi. *Combination Lock Box*, chapter 55.1113. Museum of Fine Arts, Boston, Massachusetts, USA, 1197. <https://collections.mfa.org/objects/21956>.
- [17] Ibn al-Razzaz al Jaziri. The book of ingenious mechanical devices. ed. and trans. *Donald Hill. Dordrecht: D. Reidel*, 1974.
- [18] Benjamin Blayney, editor. *Holy Bible: King James Version, 1611 Edition*. Clarendon Press, 1769.
- [19] Steven M. Bellovin. Permissive Action Links, Nuclear Weapons, and the History of Public Key Cryptography. In *Symposium on Usable Privacy and Security*, Boston, MA, May 2006. USENIX.
- [20] Eric Schlosser. *Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety*. Penguin, 2013.
- [21] Anne Adams and Martina Angela Sasse. Users Are Not the Enemy. *Communications of the ACM*, 42(12):40–46, December 1999.
- [22] Emil Lundberg, James Jones, Akshay Kumar, Dirk Balfanz, Alexei Czeskis, Angelo Huakai Liao, Michael B. Jones, Jeff Hodges, and Rolf Lindemann. Web Authentication: An API for Accessing Public Key Credentials – Level 1, March 2019. <https://www.w3.org/TR/2019/REC-webauthn-1-20190304/>, as of July 16, 2023.
- [23] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE Symposium on Security and Privacy*, SP '12, pages 553–567, San Jose, California, USA, May 2012. IEEE.
- [24] Robert Morris and Ken Thompson. Password security: a case history. *Commun. ACM*, 22(11):594–597, 1979.
- [25] Joseph Bonneau, Sören Preibusch, and Ross Anderson. A Birthday Present Every Eleven Wallets? The Security of Customer-Chosen Banking PINs. In *Financial Cryptography and Data Security*, FC '12, pages 25–40, Kralendijk, Bonaire, February 2012. Springer.
- [26] Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. Pushing on String: The “Don’t Care” Region of Password Strength. *Communications of the ACM*, 59(11):66–74, October 2016.
- [27] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It’s a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Symposium on Usable Privacy and Security*, SOUPS '14, pages 213–230, Menlo Park, California, USA, July 2014. USENIX.
- [28] Apple, Inc. iOS Security: iOS 12.1, November 2018. https://web.archive.org/web/20190228085414/https://www.apple.com/business/site/docs/iOS_Security_Guide.pdf, as of July 16, 2023.
- [29] Android Open Source Project. Full-Disk Encryption – Storing the Encrypted Key, August 2018. https://source.android.com/security/encryption/full-disk#storing_the_encrypted_key, as of July 16, 2023.
- [30] Lily Hay Newman. Google’s Making it Easier to Encrypt Even Cheap Android Phones, February 2019. <https://www.wired.com/story/android-encryption-cheap-smartphones/>, as of July 16, 2023.

- [31] Apple Inc. Apple Advances User Security with Powerful New Data Protections, Dec 2022. <https://www.apple.com/newsroom/2022/12/apple-advances-user-security-with-powerful-new-data-protections/>, as of July 16, 2023.
- [32] Ahmed Mahfouzab, Ildar Muslukhova, and Konstantin Beznosova. Android Users in the Wild: Their Authentication and Usage Behavior. *Pervasive and Mobile Computing*, 32:50–61, October 2016.
- [33] Lydia Kraus, Tobias Fiebig, Viktor Miruchna, Sebastian Möller, and Asaf Shabtai. Analyzing End-Users’ Knowledge and Feelings Surrounding Smartphone Security and Privacy. In *Mobile Security Technologies*, MoST ’15, San Jose, California, USA, May 2015.
- [34] Matthews, Tara and Liao, Kerwell and Turner, Anna and Berkovich, Marianne and Reeder, Rob and Consolvo, Sunny. “She’ll just grab any device that’s closer”: A Study of Everyday Device & Account Sharing in Households. In *ACM Conference on Human Factors in Computing Systems*, CHI ’16, pages 5921–5932, San Jose, California, USA, May 2016. ACM.
- [35] Yusuf Albayram, Mohammad Maifi Hasan Khan, Theodore Jensen, and Nhan Nguyen. “...better to use a lock screen than to worry about saving a few seconds of time”: Effect of Fear Appeal in the Context of Smartphone Locking Behavior. In *Symposium on Usable Privacy and Security*, SOUPS ’17, pages 49–63, Santa Clara, California, USA, July 2017. USENIX.
- [36] Elham Al Qahtani, Mohamed Shehab, and Abrar Aljohani. The Effectiveness of Fear Appeals in Increasing Smartphone Locking Behavior Among Saudi Arabians. In *Symposium on Usable Privacy and Security*, SOUPS ’18, pages 31–46, Baltimore, Maryland, USA, August 2018. USENIX.
- [37] Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow. Implicit Authentication Through Learning User Behavior. In *International Conference on Information Security*, ISC ’10, pages 99–113, Boca Raton, Florida, USA, April 2010. Springer.
- [38] Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. Know Your Enemy: The Risk of Unauthorized Access in Smartphones by Insiders. In *Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI ’13, pages 271–280, Munich, Germany, August 2013. ACM.
- [39] Karen Levy and Bruce Schneier. Privacy Threats in Intimate Relationships. *Journal of Cybersecurity*, 6(1):1–13, May 2020.
- [40] Diogo Marques, Ildar Muslukhov, Tiago Guerreiro, Luís Carrico, and Konstantin Beznosov. Snooping on Mobile Phones: Prevalence and Trends. In *Symposium on Usable Privacy and Security*, SOUPS ’16, pages 159–174, Denver, Colorado, USA, July 2016. USENIX.
- [41] Diogo Marques, Tiago Guerreiro, Luis Carrico, Ivan Beschastnikh, and Konstantin Beznosov. Vulnerability & Blame: Making Sense of Unauthorized Access to Smartphones. In *ACM Conference on Human Factors in Computing Systems*, CHI ’19, pages 589:1–589:13, Glasgow, Scotland, United Kingdom, April 2019. ACM.
- [42] Daniel V. Bailey, Markus Dürmuth, and Christof Paar. Statistics on Password Re-use and Adaptive Strength for Financial Accounts. In *Security and Cryptography for Networks*, SCN ’14, pages 218–235, Amalfi, Italy, September 2014. Springer.
- [43] Claude E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.
- [44] J.L. Massey. Guessing and entropy. In *IEEE International Symposium on Information Theory*, page 204, 1994.

- [45] Joseph Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *IEEE Symposium on Security and Privacy*, SP '12, pages 538–552, San Jose, California, USA, May 2012. IEEE.
- [46] Boris Köpf and Geoffrey Smith. Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks. In *2010 23rd IEEE Computer Security Foundations Symposium*, 2010.
- [47] Christian Cachin. *Entropy Measures and Unconditional Security in Cryptography*. PhD thesis, ETH Zürich, 1997.
- [48] Joseph Bonneau. *Guessing human-chosen secrets*. PhD thesis, University of Cambridge, May 2012.
- [49] Ding Wang, Qianchen Gu, Xinyi Huang, and Ping Wang. Understanding Human-Chosen PINs: Characteristics, Distribution and Security. In *ACM Asia Conference on Computer and Communications Security*, ASIA CCS '17, pages 372–385, Abu Dhabi, United Arab Emirates, April 2017. ACM.
- [50] Hyounghick Kim and Jun Ho Huh. PIN Selection Policies: Are They Really Effective? *Computers & Security*, 31(4):484–496, June 2012.
- [51] Daniel Amitay. Most Common iPhone Passcodes, June 2011. <http://danielamitay.com/blog/2011/6/13/most-common-iphone-passcodes>, as of July 16, 2023.
- [52] David Schültz. Accidental \$70k Google Pixel Lock Screen Bypass, November 2022. <https://bugs.xdavidhu.me/google/2022/11/10/accidental-70k-google-pixel-lock-screen-bypass/>, as of July 16, 2023.
- [53] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. Stay Cool! Understanding Thermal Attacks on Mobile-Based User Authentication. In *ACM Conference on Human Factors in Computing Systems*, CHI '17, pages 3751–3763, Denver, Colorado, USA, May 2017. ACM.
- [54] Almos Zarandy, Ilia Shumailov, and Ross Anderson. Hey Alexa What Did I Just Type? Decoding Smartphone Sounds With a Voice Assistant. *CoRR*, abs/2012.00687:1–18, December 2020.
- [55] Oleksiy Lisovets, David Knichel, Thorben Moos, and Amir Moradi. Let's Take it Offline: Boosting Brute-Force Attacks on iPhone's User Authentication through SCA. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2021(3):496–519, July 2021.
- [56] Gregor Haas, Seetal Potluri, and Aydin Aysu. iTimed: Cache Attacks on the Apple A10 Fusion SoC. *2021 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, pages 80–90, 2021.
- [57] Adam J. Aviv, Devon Budzitowski, and Ravi Kuber. Is Bigger Better? Comparing User-Generated Passwords on 3x3 vs. 4x4 Grid Sizes for Android's Pattern Unlock. In *Annual Computer Security Applications Conference*, ACSAC '15, pages 301–310, Los Angeles, California, USA, December 2015. ACM.
- [58] Marte Løge, Markus Dürmuth, and Lillian Røstad. On User Choice for Android Unlock Patterns. In *European Workshop on Usable Security*, EuroUSEC '16, Darmstadt, Germany, July 2016. ISOC.
- [59] Emanuel von Zezschwitz, Malin Eiband, Daniel Buschek, Sascha Oberhuber, Alexander De Luca, Florian Alt, and Heinrich Hussmann. On Quantifying the Effective Password Space of Grid-Based Unlock Gestures. In *Conference on Mobile and Ubiquitous Multimedia*, MUM '16, pages 201–212, Rovaniemi, Finland, December 2016. ACM.

- [60] Collins W. Munyendo, Miles Grant, Philipp Markert, Timothy J. Forman, and Adam J. Aviv. Using a Blocklist to Improve the Security of User Selection of Android Patterns. In *Symposium on Usable Privacy and Security*, SOUPS '21, pages 37–56, Virtual Conference, August 2021. USENIX.
- [61] Joseph Bonneau, Sören Preibusch, and Ross Anderson. A Birthday Present Every Eleven Wallets? The Security of Customer-Chosen Banking PINs. In *Financial Cryptography and Data Security*, FC '12, pages 25–40, Kralendijk, Bonaire, February 2012. Springer.
- [62] Ding Wang, Zijian Zhang, Ping Wang, Jeff Yan, and Xinyi Huang. Targeted Online Password Guessing: An Underestimated Threat. In *ACM Conference on Computer and Communications Security*, CCS '16, pages 1242–1254, Vienna, Austria, October 2016. ACM.
- [63] Adam J. Aviv, John T. Davin, Flynn Wolf, and Ravi Kuber. Towards Baselines for Shoulder Surfing on Mobile Authentication. In *Annual Conference on Computer Security Applications*, ACSAC '17, pages 486–498, Orlando, Florida, USA, December 2017. ACM.
- [64] Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. Now You See Me, Now You Don't: Protecting Smartphone Authentication from Shoulder Surfers. In *ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2937–2946, Toronto, Ontario, Canada, April 2014. ACM.
- [65] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *ACM Conference on Human Factors in Computing Systems*, CHI '17, pages 4254–4265, Denver, Colorado, USA, May 2017. ACM.
- [66] Adam J. Aviv, Flynn Wolf, and Ravi Kuber. Comparing Video Based Shoulder Surfing with Live Simulation and Towards Baselines for Shoulder Surfing on Mobile Authentication. In *Annual Conference on Computer Security Applications*, ACSAC '18, pages 453–466, San Juan, Puerto Rico, USA, December 2018. ACM.
- [67] Farid Binbeshr, Miss Laiha Mat Kiah, Lip Yee Por, and A. A. Zaidan. A Systematic Review of PIN-Entry Methods Resistant to Shoulder-Surfing Attacks. *Computers & Security*, 101, February 2021.
- [68] Emanuel von Zezschwitz, Alexander De Luca, Philipp Janssen, and Heinrich Hussmann. Easy to Draw, but Hard to Trace?: On the Observability of Grid-based (Un)Lock Patterns. In *ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2339–2342, Seoul, Republic of Korea, April 2015. ACM.
- [69] Oliver Wiese and Volker Roth. See You Next Time: A Model for Modern Shoulder Surfers. In *Conference on Human-Computer Interaction with Mobile Devices and Services*, Mobile-HCI '16, pages 453–464, Florence, Italy, September 2016. ACM.
- [70] Adam J. Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. Smudge Attacks on Smartphone Touch Screens. In *USENIX Workshop on Offensive Technologies*, WOOT '10, pages 1–7, Washington, District of Columbia, USA, August 2010. USENIX.
- [71] Seunghun Cha, Sungsu Kwag, Hyoungshick Kim, and Jun Ho Huh. Boosting the Guessing Attack Performance on Android Lock Patterns with Smudge Attacks. In *ACM Asia Conference on Computer and Communications Security*, ASIA CCS '17, pages 313–326, Abu Dhabi, United Arab Emirates, April 2017. ACM.
- [72] Alexander Peslyak (“Solar Designer”) and Community. John the Ripper, July 1996. <http://www.openwall.com/john/>, as of July 16, 2023.

- [73] Jens Steube (“atom”) and Community. Hashcat, June 2016. <https://hashcat.net/hashcat/>, as of July 16, 2023.
- [74] Arvind Narayanan and Vitaly Shmatikov. Fast dictionary attacks on passwords using time-space tradeoff. In *Proc. 12th ACM conference on Computer and communications security (CCS)*, pages 364–372. ACM, 2005.
- [75] Matt Weir, Sudhir Aggarwal, Breno de Medeiros, and Bill Glodek. Password cracking using probabilistic context-free grammars. In *Proc. IEEE Symposium on Security and Privacy*, pages 391–405. IEEE Computer Society, 2009.
- [76] Blake Ives, Kenneth R. Walsh, and Helmut Schneider. The domino effect of password reuse. *Communications of the ACM*, 47(4):75, April 2004.
- [77] D. Florêncio and C. Herley. A Large-Scale Study of Web Password Habits. In *Proc. 16th international conference on World Wide Web (WWW '07)*, pages 657–666. ACM, 2007.
- [78] Shirley Gaw and Edward W. Felten. Password management strategies for online accounts. In *Proc. Symposium On Usable Privacy and Security (SOUPS '06)*, 2006.
- [79] Joseph Bonneau. Measuring Password Re-Use Empirically, February 2011. <http://www.lightbluetouchpaper.org/2011/02/09/measuring-password-re-use-empirically/>.
- [80] Trusteer, Inc. Reused login credentials. Security Advisory, online at <https://libertadzero.files.wordpress.com/2010/02/cross-logins-advisory.pdf>, 2010.
- [81] Sena Sahin, Suood Al Roomi, Tara Poteat, and Frank Li. Investigating the password policy practices of website administrators. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 1437–1454. IEEE Computer Society, 2023.
- [82] Yinqian Zhang, Fabian Monrose, and Michael K. Reiter. The Security of Modern Password Expiration: an Algorithmic Framework and Empirical Analysis. In *Proc. ACM Conference on Computer and Communications Security (CCS)*, pages 176–186, 2010.
- [83] Collins W. Munyendo, Philipp Markert, Alexandra Nisenoff, Miles Grant, Elena Korkes, Blase Ur, and Adam J. Aviv. “The Same PIN, Just Longer”: On the (In)Security of Upgrading PINs from 4 to 6 Digits. In *USENIX Security Symposium, SSYM '22*, Boston, Massachusetts, USA, August 2022. USENIX.
- [84] Troy Hunt. *Have I Been Pwned?* – Check If Your Email Has Been Compromised in a Data Breach, December 2013. <https://haveibeenpwned.com>, as of July 16, 2023.
- [85] Troy Hunt. *Have I Been Pwned?* – Pwned Websites, August 2022. <https://haveibeenpwned.com/PwnedWebsites>, as of July 16, 2023.
- [86] Kurt Thomas, Jennifer Pullman, Kevin Yeo, Ananth Raghunathan, Patrick Gage Kelley, Luca Invernizzi, Borbala Benko, Tadek Pietraszek, Sarvar Patel, Dan Boneh, and Elie Bursztein. Protecting Accounts From Credential Stuffing With Password Breach Alerting. In *USENIX Security Symposium, SSYM '19*, pages 1556–1571, Santa Clara, California, USA, August 2019. USENIX.
- [87] Rick Wash, Emilee Radar, Ruthie Berman, and Zac Wellmer. Understanding Password Choices: How Frequently Entered Passwords are Re-used Across Websites. In *Symposium on Usable Privacy and Security, SOUPS '16*, pages 175–188, Denver, Colorado, USA, July 2016. USENIX.
- [88] Chun Wang, Steve T.K. Jan, Hang Hu, Douglas Bossart, and Gang Wang. The Next Domino to Fall: Empirical Analysis of User Passwords across Online Services. In *ACM Conference on Data and Application Security and Privacy, CODASPY '18*, pages 196–203, Tempe, Arizona, USA, March 2018. ACM.

- [89] Hassan Khan, Jason Ceci, Jonah Stegman, Adam J. Aviv, Rozita Dara, and Ravi Kuber. Widely Reused and Shared, Infrequently Updated, and Sometimes Inherited: A Holistic View of PIN Authentication in Digital Lives and Beyond. In *Annual Computer Security Applications Conference, ACSAC '20*, pages 249–262, Austin, Texas, USA, December 2020. ACM.
- [90] Maria Casimiro, Joe Segel, Lewei Li, Yigeng Wang, and Lorrie Faith Cranor. A Quest for Inspiration: How Users Create and Reuse PINs. In *Who Are You?! Adventures in Authentication Workshop, WAY '20*, pages 1–7, Virtual Conference, August 2020.
- [91] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa Redmiles, and Blase Ur. “What was that site doing with my Facebook password?” Designing Password-Reuse Notifications. In *ACM Conference on Computer and Communications Security, CCS '18*, pages 1549–1566, Toronto, Ontario, Canada, October 2018. ACM.
- [92] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Michael Backes, and Sven Bugiel. “Better managed than memorized?” Studying the Impact of Managers on Password Strength and Reuse. In *USENIX Security Symposium, SSYM '18*, pages 203–220, Baltimore, Maryland, USA, August 2018. USENIX.
- [93] Pete Lepage. New in Chrome 69, 2018. <https://developer.chrome.com/blog/new-in-chrome-69/>, as of July 16, 2023.
- [94] Dan Goodin. Hack of Cloud-Based LastPass Exposes Hashed Master Passwords. *Ars Technica*, June 2015. <https://arstechnica.com/information-technology/2015/06/hack-of-cloud-based-lastpass-exposes-encrypted-master-passwords/>, as of July 16, 2023.
- [95] LastPass. Notice of Recent Security Incident. <https://blog.lastpass.com/2022/12/notice-of-recent-security-incident/>, December 2022. Accessed on March 8, 2023.
- [96] Daniel R. Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R. Beresford. Ethical Issues in Research Using Datasets of Illicit Origin. In *Internet Measurement Conference, IMC '17*, pages 445–462, London, United Kingdom, November 2017. ACM.
- [97] U.S. Department of Homeland Security. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research, August 2012. https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/, as of July 16, 2023.
- [98] Marcello Ienca and Effy Vayena. Ethical Requirements for Responsible Research with Hacked Data. *Nature Machine Intelligence*, 3(9):744–748, September 2021.
- [99] Maximilian Golla and Markus Dürmuth. On the Accuracy of Password Strength Meters. In *ACM Conference on Computer and Communications Security, CCS '18*, pages 1567–1582, Toronto, Ontario, Canada, October 2018. ACM.
- [100] S. M. Taiabul Haque, Matthew Wright, and Shannon Scielzo. A study of user password strategy for multiple accounts. In *Proc. 3rd ACM Conference on Data and Application Security and Privacy (CODASPY)*, pages 173–176, 2013.
- [101] C. Herley, P. van Oorschot, and A. S. Patrick. Passwords: If we’re so smart, why are we still using them? In *Proc. 13th International Conference on Financial Cryptography and Data Security (FC 2009)*, 2009.
- [102] Jason RC Nurse, Sadie Creese, Michael Goldsmith, and Koen Lamberts. Trustworthy and Effective Communication of Cybersecurity Risks: a Review. In *Proc. Workshop on Socio-Technical Aspects in Security and Trust (STAST)*, pages 60–68. IEEE, 2011.
- [103] D. Florêncio and C. Herley. Where do Security Policies Come From? In *Symposium on Usable Privacy and Security (SOUPS '10)*, 2010.

- [104] Marc Jacob. Trusteer detects rapid spread of new polymorphic version of zeus online banking trojan. Security Advisory, online at <https://www.globalsecuritymag.fr/Trusteer-Detects-Rapid-Spread-of>, 20100421, 17206.html, 2010.
- [105] Konstantinos P Grammatikakis, Ioannis Koufos, Nicholas Kolokotronis, Costas Vassilakis, and Stavros Shiaeles. Understanding and mitigating banking trojans: From zeus to emotet. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 121–128. IEEE, 2021.
- [106] Stat Counter. Microsoft market dominance. Online at <https://gs.statcounter.com/os-market-share/all/worldwide/2013>, 2013.
- [107] StatCounter. Desktop operating system market share. <https://gs.statcounter.com/os-market-share/desktop/worldwide/2022>, as of July 16, 2023.
- [108] Jason Mick. Inside the Mega-Hack of Bitcoin: the Full Story, June 2011. <https://www.scribd.com/document/166287990/DailyTech-Inside-the-Mega-Hack-of-Bitcoin-the-Full-Story-pdf>, as of July 16, 2023.
- [109] Brian Krebs. Fraud Bazaar Carders.cc Hacked, May 2010. <http://krebsonsecurity.com/2010/05/fraud-bazaar-carders-cc-hacked/>.
- [110] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. Of passwords and people: Measuring the effect of password-composition policies. In *Proc. Conference on Human Factors in Computing Systems (CHI 2011)*, 2011.
- [111] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proc. 17th ACM conference on Computer and communications security (CCS 2010)*, pages 162–175. ACM, 2010.
- [112] R. Dhamija and A. Perrig. Deja vu: A user study using images for authentication. In *Proc. 9th USENIX Security Symposium*, 2000.
- [113] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas. Generating and remembering passwords. *Applied Cognitive Psychology*, 18(6):641–651, 2004.
- [114] OWASP. Testing Local Authentication. <https://mas.owasp.org/MASTG/iOS/0x06f-Testing-Local-Authentication/>, 2021. Accessed on March 21, 2023.
- [115] William Melicher, Darya Kurilova, Sean M. Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Michelle L. Mazurek. Usability and Security of Text Passwords on Mobile Devices. In *ACM Conference on Human Factors in Computing Systems, CHI '16*, pages 527–539, San Jose, California, USA, May 2016. ACM.
- [116] Ivan Cherapau, Ildar Muslukhov, Nalin Asanka, and Konstantin Beznosov. On the Impact of Touch ID on iPhone Passcodes. In *Symposium on Usable Privacy and Security, SOUPS '15*, pages 257–276, Ottawa, Canada, July 2015. USENIX.
- [117] Patrick Kelley, Saranga Kom, Michelle L. Mazurek, Rich Shay, Tim Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio López. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In *IEEE Symposium on Security and Privacy, SP '12*, pages 523–537, San Jose, California, USA, May 2012. IEEE.
- [118] Richard Shay, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Alain Forget, Saranga Komanduri, Michelle L. Mazurek, William Melicher, Sean M. Segreti, and Blase Ur. A Spoonful of Sugar?: The Impact of Guidance and Feedback on Password-Creation Behavior. In *ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 2903–2912, Seoul, Republic of Korea, April 2015. ACM.

- [119] Apple, Inc. Apple Platform Security, May 2021. https://manuals.info.apple.com/MANUALS/1000/MA1902/en_US/apple-platform-security-guide.pdf, as of July 16, 2023.
- [120] Android Open Source Project. Android 9 – “Pie”: GateKeeper – ComputeRetryTimeout Function, February 2018. <https://android.googlesource.com/platform/system/gatekeeper/+pie-release/gatekeeper.cpp#253>, as of July 16, 2023.
- [121] Florian Schaub, Ruben Deyhle, and Michael Weber. Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms. In *International Conference on Mobile and Ubiquitous Multimedia*, MUM ’12, pages 13:1–13:10, Ulm, Germany, December 2012. ACM.
- [122] Maximilian Golla, Jan Rimkus, Adam J. Aviv, and Markus Dürmuth. Work in Progress: On the In-Accuracy and Influence of Android Pattern Strength Meters. In *Workshop on Usable Security and Privacy*, USEC ’19, San Diego, California, USA, February 2019. ISOC.
- [123] Jeremi M. Gosney (“epixoip”). How LinkedIn’s Password Sloppiness Hurts Us All, June 2016. <https://arstechnica.com/information-technology/2016/06/how-linkedins-password-sloppiness-hurts-us-all/>, as of July 16, 2023.
- [124] Troy Hunt. I’ve Just Launched “Pwned Passwords” V2 With Half a Billion Passwords for Download, February 2018. <https://www.troyhunt.com/ive-just-launched-pwned-passwords-version-2/>, as of July 16, 2023.
- [125] Elissa M. Redmiles, Yasemin Acar, Sascha Fahl, and Michelle L. Mazurek. A Summary of Survey Methodology Best Practices for Security and Privacy Researchers. Technical Report CS-TR-5055, UM Computer Science Department, May 2017.
- [126] Lina Qiu, Alexander De Luca, Ildar Muslukhov, and Konstantin Beznosov. Towards Understanding the Link Between Age and Smartphone Authentication. In *ACM Conference on Human Factors in Computing Systems*, CHI ’19, pages 163:1–163:10, Glasgow, Scotland, United Kingdom, May 2019. ACM.
- [127] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. “I Added ‘!’ at the End to Make It Secure”: Observing Password Creation in the Lab. In *Symposium on Usable Privacy and Security*, SOUPS ’15, pages 123–140, Ottawa, Ontario, Canada, July 2015. USENIX.
- [128] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. Design and Evaluation of a Data-Driven Password Meter. In *ACM Conference on Human Factors in Computing Systems*, CHI ’17, pages 3775–3786, Denver, Colorado, USA, May 2017. ACM.
- [129] Blase Ur, Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Saranga Komanduri, Darya Kurilova, Michelle L. Mazurek, William Melicher, and Richard Shay. Measuring Real-World Accuracies and Biases in Modeling Password Guessability. In *USENIX Security Symposium*, SSYM ’15, pages 463–481, Washington, District of Columbia, USA, August 2015. USENIX.
- [130] Joshua Sunshine, Serge Egelman, Hazim Almuhtedi, Neha Atri, and Lorrie Faith Cranor. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *USENIX Security Symposium*, SSYM ’09, pages 399–416, San Diego, California, USA, June 2009. USENIX.
- [131] Devdatta Akhawe and Adrienne Porter Felt. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *USENIX Security Symposium*, SSYM ’13, pages 257–272, Washington, District of Columbia, USA, July 2013. USENIX.

- [132] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettis, Helen Harris, and Jeff Grimes. Improving SSL Warnings: Comprehension and Adherence. In *ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 2893–2902, Seoul, Republic of Korea, April 2015. ACM.
- [133] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa Redmiles, and Blase Ur. “What was that site doing with my Facebook password?” Designing Password-Reuse Notification. In *ACM Conference on Computer and Communications Security, CCS '18*, pages 1549–1566, Toronto, Canada, November 2018. ACM.
- [134] Thomas Reed. GrayKey iPhone Unlocker Poses Serious Security Concerns, March 2018. <https://blog.malwarebytes.com/security-world/2018/03/graykey-iphone-unlocker-poses-serious-security-concerns/>, as of July 16, 2023.
- [135] Oleg Afonin. iPhone 5 and 5c Passcode Unlock with iOS Forensic Toolkit, August 2020. <https://blog.elcomsoft.com/2020/08/iphone-5-and-5c-passcode-unlock-with-ios-forensic-toolkit/>, as of July 16, 2023.
- [136] Emily Stark. The URLephant. In *USENIX Enigma Conference*, Enigma '19, Burlingame, California, USA, January 2019. USENIX.
- [137] Sonia Secher Wichmann. Self-Determination Theory: The Importance of Autonomy to Well-Being Across Cultures. *Journal of Humanistic Counseling*, 50(1):16–26, March 2011.
- [138] Karen Renaud and Melanie Volkamer. Exploring Mental Models Underlying PIN Management Strategies. In *World Congress on Internet Security, WorldCIS '15*, pages 19–21, Dublin, United Kingdom, October 2015. IEEE.
- [139] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3):436–465, August 2013.
- [140] Nathanael Andrews. Can I Get Your Digits: Illegal Acquisition of Wireless Phone Numbers for SIM-Swap Attacks and Wireless Provider Liability. *Northwestern Journal of Technology and Intellectual Property*, 16(2):79–106, November 2018.
- [141] Roger Piqueras Jover. Security Analysis of SMS as a Second Factor of Authentication: The Challenges of Multifactor Authentication Based on SMS, Including Cellular Security Deficiencies, SS7 Exploits, and SIM Swapping. *ACM Queue*, 18(4):37–60, July 2020.
- [142] Kevin Lee, Benjamin Kaiser, Jonathan Mayer, and Arvind Narayanan. An Empirical Study of Wireless Carrier Authentication for SIM Swaps. In *Symposium on Usable Privacy and Security, SOUPS '20*, pages 61–79, Virtual Conference, August 2020. USENIX.
- [143] Joshua Lund. Technology Preview for Secure Value Recovery, December 2019. <https://signal.org/blog/secure-value-recovery>, as of July 16, 2023.
- [144] Jim Oleary. Improving Registration Lock with Secure Value Recovery, January 2020. <https://signal.org/blog/improving-registration-lock>, as of July 16, 2023.
- [145] Paul Rösler, Christian Mainka, and Jörg Schwenk. More is Less: On the End-to-End Security of Group Chats in Signal, WhatsApp, and Threema. In *European Symposium on Security and Privacy, EuroSP '18*, pages 415–429, London, United Kingdom, April 2018. IEEE.
- [146] Katriel Cohn-Gordon, Cas Cremers, Benjamin Dowling, Luke Garratt, and Douglas Stebila. A Formal Security Analysis of the Signal Messaging Protocol. *Journal of Cryptology*, 33(4):1914–1983, December 2020.
- [147] Moxie Marlinspike. WhatsApp’s Signal Protocol Integration Is Now Complete, April 2016. <https://signal.org/blog/whatsapp-complete>, as of July 16, 2023.

- [148] Moxie Marlinspike. Facebook Messenger Deploys Signal Protocol for End-to-End Encryption, July 2016. <https://signal.org/blog/facebook-messenger>, as of July 16, 2023.
- [149] WhatsApp. Answering Your Questions About WhatsApp’s Privacy Policy, February 2021. <https://faq.whatsapp.com/general/security-and-privacy/answering-your-questions-about-whatsapps-privacy-policy>, as of July 16, 2023.
- [150] Manish Singh. Signal’s Brian Acton Talks About Exploding Growth, Monetization, and WhatsApp Data Sharing Outrage, January 2021. <https://tcrn.ch/38BHusb>, as of July 16, 2023.
- [151] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the Adoption of Secure Communication Tools. In *IEEE Symposium on Security and Privacy*, SP ’17, pages 137–153, San Jose, California, USA, May 2017. IEEE.
- [152] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. Expert and Non-Expert Attitudes Towards (Secure) Instant Messaging. In *Symposium on Usable Privacy and Security*, SOUPS ’16, pages 147–157, Denver, Colorado, USA, July 2016. USENIX.
- [153] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A. Dabbish, and Jason I. Hong. The Effect of Social Influence on Security Sensitivity. In *Symposium on Usable Privacy and Security*, SOUPS ’14, pages 143–157, Menlo Park, California, USA, July 2014. USENIX.
- [154] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In *ACM Conference on Computer and Communications Security*, CCS ’14, pages 739–749, Scottsdale, Arizona, USA, November 2014. ACM.
- [155] Sean Oesch, Ruba Abu-Salma, Oumar Diallo, Juliane Krämer, James Simmons, Justin Wu, and Scott Ruoti. Understanding User Perceptions of Security and Privacy for Group Chat: A Survey of Users in the US and UK. In *Annual Conference on Computer Security Applications*, ACSAC ’20, pages 234–248, Virtual Conference, December 2020.
- [156] Nik Unger, Sergej Dechand, Joseph Bonneau, Sascha Fahl, Henning Perl, Ian Goldberg, and Matthew Smith. SoK: Secure Messaging. In *IEEE Symposium on Security and Privacy*, SP ’15, pages 232–249, San Jose, California, USA, 2015.
- [157] Elham Vaziripour, Justin Wu, Mark O’Neill, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala. Is That You, Alice? A Usability Study of the Authentication Ceremony of Secure Messaging Applications. In *Symposium on Usable Privacy and Security*, SOUPS ’17, pages 29–47, Santa Clara, California, USA, July 2017. USENIX.
- [158] Elham Vaziripour, Justin Wu, Mark O’Neill, Daniel Metro, Josh Cockrell, Timothy Moffett, Jordan Whitehead, Nick Bonner, Kent Seamons, and Daniel Zappala. Action Needed! Helping Users Find and Complete the Authentication Ceremony in Signal. In *Symposium on Usable Privacy and Security*, SOUPS ’18, pages 47–62, Baltimore, Maryland, USA, August 2018. USENIX.
- [159] Justin Wu, Cyrus Gattrell, Devon Howard, Jake Tyler, Elham Vaziripour, Daniel Zappala, and Kent Seamons. “Something Isn’t Secure, but I’m Not Sure How That Translates Into a Problem”: Promoting Autonomy by Designing for Understanding in Signal. In *Symposium on Usable Privacy and Security*, SOUPS ’19, pages 137–156, Santa Clara, California, USA, August 2019. USENIX.

- [160] Elham Vaziripour, Devon Howard, Jake Tyler, Mark O'Neill, Justin Wu, Kent Seamons, and Daniel Zappala. I Don't Even Have to Bother Them! Using Social Media to Automate the Authentication Ceremony in Secure Messaging. In *ACM Conference on Human Factors in Computing Systems*, CHI '19, pages 934:1–93:12, Glasgow, Scotland, United Kingdom, May 2019. ACM.
- [161] Scott Ruoti, Nathan Kim, Ben Burgon, Timothy van der Horst, and Kent Seamons. Confused Johnny: When Automatic Encryption Leads to Confusion and Mistakes. In *Symposium on Usable Privacy and Security*, SOUPS '13, pages 5:1–5:12, Newcastle, United Kingdom, July 2013. ACM.
- [162] Sunyoung Seiler-Hwang, Patricia Arias-Cabarcos, Andrés Marín, Florina Almenares, Daniel Díaz-Sánchez, and Christian Becker. "I Don't See Why I Would Ever Want to Use It" Analyzing the Usability of Popular Smartphone Password Managers. In *ACM Conference on Computer and Communications Security*, CCS '19, pages 1937–1953, London, United Kingdom, November 2019. ACM.
- [163] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Why People (Don't) Use Password Managers Effectively. In *Symposium on Usable Privacy and Security*, SOUPS '19, pages 319–338, Santa Clara, California, USA, August 2019. USENIX.
- [164] Herbert F. Spitzer. Studies in Retention. *Journal of Educational Psychology*, 30(9):641–656, December 1939.
- [165] Paul Pimsleur. A Memory Schedule. *The Modern Language Journal*, 51:73–75, February 1967.
- [166] Arthur W. Melton. The Situation With Respect to the Spacing of Repetitions and Memory. *Journal of Verbal Learning and Verbal Behavior*, 9(5):596–606, October 1970.
- [167] Thomas K. Landauer and Robert A. Bjork. Optimum Rehearsal Patterns and Name Learning. In *International Conference on Practical Aspects of Memory*, PAM '78, pages 625–632, Cardiff, United Kingdom, September 1978. Academic Pres.
- [168] Ann-Marie Horcher and Gurvirender P. Tejay. Building A Better Password: The Role of Cognitive Load in Information Security Training. In *IEEE International Conference on Intelligence and Security Informatics*, ISI '09, pages 113–118, Richardson, Texas, USA, June 2009. ACM.
- [169] Steven Mujye and Yair Levy. Complex Passwords: How Far Is Too Far? the Role of Cognitive Load on Employee Productivity. *Online Journal of Applied Knowledge Management*, 1(1):122–132, June 2013.
- [170] Joseph Bonneau and Stuart Schechter. Towards Reliable Storage of 56-bit Secrets in Human Memory. In *USENIX Security Symposium*, SSYM '14, pages 607–623, San Diego, California, USA, August 2014. USENIX.
- [171] Jeremiah Blocki, Saranga Komanduri, Lorrie Faith Cranor, and Anupam Datta. Spaced Repetition and Mnemonics Enable Recall of Multiple Strong Passwords. In *Symposium on Network and Distributed System Security*, NDSS '15, San Diego, California, USA, February 2015. ISOC.
- [172] Christopher Novak, Jim Blythe, Ross Koppel, Vijay Kothari, and Sean Smith. Modeling Aggregate Security With User Agents That Employ Password Memorization Techniques. In *Who Are You?! Adventures in Authentication Workshop*, WAY '17, Santa Clara, California, USA, July 2017. USENIX.

- [173] Stuart Schechter and Joseph Bonneau. Learning Assigned Secrets for Unlocking Mobile Devices. In *Symposium on Usable Privacy and Security*, SOUPS '15, pages 277–295, Ottawa, Ontario, Canada, July 2015. USENIX.
- [174] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *ACM Conference on Human Factors in Computing Systems*, CHI '08, pages 1065–1074, Florence, Italy, April 2008. ACM.
- [175] Robert Biddle, P. C. van Oorschot, Andrew S. Patrick, Jennifer Sobey, and Tara Whalen. Browser Interfaces and Extended Validation SSL Certificates: An Empirical Study. In *ACM Workshop on Cloud Computing Security*, CCSW '09, pages 19–30, Chicago, Illinois, USA, 2009.
- [176] Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhammedi, and Sunny Consolvo. Experimenting at Scale with Google Chrome's SSL Warning. In *ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2667–2670, Toronto, Ontario, Canada, April 2014. ACM.
- [177] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *ACM Conference on Human Factors in Computing Systems*, CHI '18, pages 512:1–512:13, Montreal, Quebec, Canada, April 2018. ACM.
- [178] Baruch Fischhoff, Donna Riley, Daniel C. Kovacs, and Mitchell Small. What Information Belongs in a Warning? *Psychology & Marketing*, 15(7):663–686, 1998.
- [179] Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie S. Downs, and Saranga Komanduri. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *IEEE Security & Privacy*, 9(2):18–26, March 2011.
- [180] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. Your Attention Please: Designing Security-Decision UIs to Make Genuine Risks Harder to Ignore. In *Symposium on Usable Privacy and Security*, SOUPS '13, pages 6:1–6:12, Newcastle, United Kingdom, 2013.
- [181] Cristian Bravo-Lillo, Lorrie Faith Cranor, Saranga Komanduri, Stuart Schechter, and Manya Sleeper. Harder to Ignore? Revisiting Pop-Up Fatigue and Approaches to Prevent It. In *Symposium on Usable Privacy and Security*, SOUPS '14, pages 105–111, Menlo Park, California, USA, July 2014. USENIX.
- [182] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *IEEE Symposium on Security and Privacy*, SP '19, pages 227–244, San Francisco, California, USA, May 2019. IEEE.
- [183] Eleanor E. Maccoby and Nathan Maccoby. *The Interview: A Tool of Social Science*, volume 1, pages 449–487. John Wiley & Sons, New York, USA, 1954.
- [184] Robert J. Fisher. Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2):303–315, September 1993.
- [185] Prolific Team. Deciding on a Reward, September 2018. <https://researcher-help.prolific.co/hc/en-gb/articles/360009500733-Deciding-on-a-Reward>, as of July 16, 2023.
- [186] Philipp Markert, Daniel V. Bailey, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. This PIN Can Be Easily Guessed: Analyzing the Security of Smartphone Unlock PINs. In *IEEE Symposium on Security and Privacy*, SP '20, pages 286–303, San Francisco, California, USA, May 2020. IEEE.

- [187] Federal Trade Commission. How to Protect Your Phone from Hackers, Aug 2022.
- [188] Raina Samuel, Philipp Markert, Adam J. Aviv, and Iulian Neamtiu. Knock, Knock. Who's There? On the Security of LG's Knock Codes. In *Symposium on Usable Privacy and Security*, SOUPS '20, pages 37–59, Virtual Conference, August 2020. ACM.
- [189] Panagiotis Andriotis, Theo Tryfonas, George Oikonomou, and Can Yildiz. A Pilot Study on the Security of Pattern Screen-Lock Methods and Soft Side Channel Attacks. In *ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '13, pages 1–6, Budapest, Hungary, April 2013. ACM.
- [190] Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. Complexity Metrics and User Strength Perceptions of the Pattern-Lock Graphical Authentication Method. In *Conference on Human Aspects of Information Security, Privacy and Trust*, HAS '14, pages 115–126, Heraklion, Crete, Greece, June 2014. Springer.
- [191] Adam J. Aviv, Devon Budzitowski, and Ravi Kuber. Is Bigger Better? Comparing User-Generated Passwords on 3x3 vs. 4x4 Grid Sizes for Android's Pattern Unlock. In *Annual Computer Security Applications Conference*, ACSAC '15, pages 301–310, Los Angeles, California, USA, December 2015. ACM.
- [192] Apple. iOS 9 Preview, June 2015. <https://web.archive.org/web/20150608223846/http://www.apple.com/ios/ios9-preview/>, as of June 8, 2015.
- [193] Cyrus Farivar. Apple to Require 6-digit Passcodes on Newer iPhones, iPads Under iOS 9, June 2015. https://arstechnica.com/?post_type=post&p=679147, as of July 16, 2023.
- [194] Android Open Source Project. Android 12: GateKeeper – ComputeRetryTimeout Function, January 2021. <https://android.googlesource.com/platform/system/gatekeeper/+refs/heads/android12-release/gatekeeper.cpp/#282>, as of July 16, 2023.
- [195] Daniel V. Bailey, Philipp Markert, and Adam J. Aviv. “I have no idea what they're trying to accomplish:” Enthusiastic and Casual Signal Users' Understanding of Signal PINs. In *Symposium on Usable Privacy and Security*, SOUPS '21, pages 417–436, Virtual Conference, August 2021. USENIX.
- [196] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research. *Journal of Experimental Social Psychology*, 70(5):153–163, May 2017.
- [197] Ding Wang, Ping Wang, Debiao He, and Yuan Tian. Birthday, Name and Bifacial-Security: Understanding Passwords of Chinese Web Users. In *USENIX Security Symposium*, SSYM '19, pages 1537–1555, Santa Clara, California, USA, August 2019. USENIX.
- [198] Frank Stajano. One User, Many Hats; and, Sometimes, No Hat: Towards a Secure Yet Usable PDA. In *International Conference on Security Protocols*, SP '04, pages 51–64, Cambridge, United Kingdom, April 2004. Springer.
- [199] Amy K. Karlson, A.J. Bernheim Brush, and Stuart Schechter. Can i borrow your phone? understanding concerns when sharing mobile phones. In *ACM Conference on Human Factors in Computing Systems*, CHI '09, pages 1647–1650, Boston, Massachusetts, USA, April 2009. ACM.
- [200] Emily Tseng, Rosanna Bellini, Nora McDonald, Matan Danos, Rachel Greenstadt, Damon McCoy, Nicola Dell, and Thomas Ristenpart. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. In *USENIX Security Symposium*, SSYM '20, pages 1893–1909, Virtual Conference, August 2020. USENIX.

- [201] Lorrie Faith Cranor, Adam L. Durity, Abigail Marsh, and Blase Ur. Parents' and Teens' Perspectives on Privacy In a Technology-Filled World. In *Symposium on Usable Privacy and Security*, SOUPS '14, pages 19–35, Menlo Park, California, USA, July 2014. USENIX.
- [202] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill. “Privacy is not for me, it’s for those rich women”: Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Symposium on Usable Privacy and Security*, SOUPS '18, pages 127–142, Baltimore, Maryland, USA, August 2018. USENIX.
- [203] Google Developers. Android 13 Developer Preview, March 2022. <https://developer.android.com/about/versions/13>, as of July 16, 2023.
- [204] Mishaal Rahman. Android 14 Could Let You Clone Apps, February 2023. <https://www.xda-developers.com/android-14-app-cloning/>, as of July 16, 2023.
- [205] Moni Naor and Moti Yung. Public-key Cryptosystems Provably Secure against Chosen Ciphertext Attacks. In Harriet Ortiz, editor, *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*, pages 427–437. ACM, 1990.