2018-06-01

Evaluating the Usability of Two-Factor Authentication

Kendall Ray Reese
Brigham Young University

Evaluating the Usability of Two-Factor Authentication

Kendall Ray Reese

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Kent Seamons, Chair
Daniel Zappala
Dan Ventura

Department of Computer Science

Brigham Young University

ABSTRACT

Evaluating the Usability of Two-Factor Authentication

Kendall Ray Reese
Department of Computer Science, BYU
Master of Science

Passwords are the dominant form of authentication on the web today. However, many users choose weak passwords and reuse the same password on multiple sites, thus increasing their vulnerability to having their credentials leaked or stolen. Two-factor authentication strengthens existing password authentication schemes against impersonation attacks and makes it more difficult for attackers to reuse stolen credentials on other websites. Despite the added security benefits of two-factor authentication, there are still many open questions about its usability. Many two-factor authentication systems in widespread usage today have not yet been subjected to adequate usability testing. Previous comparative studies have demonstrated significant differences in usability between various single-factor authentication systems.

The main contributions of this work are as follows. First, we developed a novel user behavior model that describes four phases of interaction between a user and an authentication system. This model is designed to inform the design of future usability studies and will enable researchers and those implementing authentication systems to have a more nuanced understanding of authentication system usability. Second, we conducted a comparative usability study of some of the most common two-factor authentication systems. In contrast to previous authentication usability studies, we had participants use the system for a period of two weeks and collected both timing data and SUS metrics on the systems under test. From these studies, we make several conclusions about the state of usability and acceptance of two-factor authentication, finding that many users want more security for their sensitive online accounts and are open to using multiple forms of two-factor authentication. We also suggest that security researchers draw upon risk communication theory to better help users make informed security decisions.

Keywords: two-factor authentication, usable security, two step verification

ACKNOWLEDGMENTS

I would like to acknowledge my family for their continued love and support during my graduate education, particularly my brother's statistical advice. Additional thanks to my advisor, Kent Seamons, for his insightful feedback and mentorship, and to my committee for their efforts on my behalf. Finally, I would like to thank the many graduate and undergraduate research assistants in the Internet Security Research Lab of which I had the privilege of working; it was only with their assistance that I was able to conduct a longitudinal user study with a considerable number of participants.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

Passwords are the most widespread form of user authentication on the web today [6]. Although a plethora of password-replacement schemes have been proposed, none of them fully measure up to the deployability and usability attributes of passwords [5]. More recently, large service providers including Google, Facebook, and Microsoft, have deployed an optional two-factor authentication layer as part of their authentication processes to defend against widespread impersonation attacks. A two-factor authentication scheme requires users to present two of the following types of authentication factors:

1. Something they *know* (traditionally a password)
2. Something they *have* (such as a phone or hardware token)
3. Something they *are* (referring to biometrics, such as a fingerprint)

Two-factor authentication provides a strong defense against remote impersonation attacks. For example, if an attacker were able to steal or guess a user's password, the attacker would still need to compromise the user's phone or steal a physical token in order to gain access to the account. Thus, it is significantly more difficult for a remote attacker to conduct a successful impersonation attack on a user whose account is secured with a second-factor.

Many forms of two-factor authentication have been proposed. Systems such as SMS, TOTP (time-based one-time password), and hardware code generators (such as the RSA SecurID) require the user to enter a 6-digit single-use code in addition to their

password. These codes are either sent to the user via a separate channel or are generated on the fly by the user's device. In commercial and government settings, smart cards are a commonly used second-factor, requiring the user to insert an ID badge to a card reader attached to their computer. Online banking systems, particularly in the UK, frequently use variants of hardware code generators and card readers in their two-factor authentication implementations. USB hardware tokens, such as the YubiKey, have been adopted by Google for their employees [18].

The need to properly secure account credentials is further underscored by a number of recent password database leaks [11]. Because users tend to reuse the same username and password across multiple sites [9], password leaks from a single site can lead to a chain-reaction of account compromises as attackers access other accounts with the same credentials [15]. Two-factor authentication helps prevent these types of attacks, since even an attacker with knowledge of the user's password would still be unable to compromise the account without access to the second factor.

Despite the attractive security benefits of two-factor authentication, its impact on the user experience remains unclear. A number of previous studies on two-factor authentication systems have produced results which may appear contradictory. While one set of studies [13][16][17][29] concludes that two-factor authentication is completely unusable, others [12][18] have drawn very different conclusions, finding that some two-factor authentication systems are actually very usable.

The most obvious difficulty in trying to draw conclusions from these results as a whole is that the usability studies and surveys that led to these conclusions were performed on different sets of users under very different test conditions. Importantly, many of the studies did not test the same two-factor authentication systems, making it intractable to determine how the different systems compare in terms of usability. To resolve these problems, we defined the following set of objectives in our research:

1. **Identify adoption hurdles**— According to a 2015 estimate performed by Petsas et al. [23], only 6.4% of Google accounts have two-factor authentication enabled and other online service providers have an even lower rate of adoption. Even in a commercial setting, two-factor authentication adoption was estimated at only 17% by Humphries et al. [14]. We wish to identify any usability hurdles that may discourage two-factor authentication adoption.

2. **Better understand user attitudes about two-factor authentication**— In addition to studying the two-factor authentication systems themselves, we conducted in-person interviews with participants to better understand how they felt about using two-factor authentication. This type of qualitative data is essential in understanding how to design secure systems that integrate well with the needs of everyday users.

3. **Study two-factor authentication in daily lives**— Many previous studies focused heavily on participants' experiences using two-factor authentication in a laboratory setting. Although such studies are helpful as a way to gain initial focus in identifying obvious usability concerns, one weakness of such studies is that they do not allow the user enough time to become appropriately acclimated to using the system itself, thus potentially skewing the usability results downward. In contrast, we allowed users in our study to experience a two-factor authentication system an average of 10 times over a period of 14 days before interviewing them. We also quantified the learnability of each system by measuring the authentication time each day the participant used the system within the study period.

Chapter 2

Related Work

In this section, we discuss previous research published in the area of authentication usability and two-factor authentication usability.

2.1 Summary of Previous Work

Bonneau et al. [5] analyzed a number of different authentication systems and rated their usability, security, and deployability. They also proposed an evaluation framework for measuring the viability of newly proposed authentication systems. Importantly, this work highlights the incomplete and overly optimistic views that authors of authentication schemes often award their own systems. In contrast to our proposed work, this work did not perform any studies with end-users.

Braz and Robert [7] demonstrated the conflict between traditional user interface design strategies and the security goals of authentication systems. Although this paper did not perform any user studies, it did highlight several flaws with some existing authentication systems and performed a comparative security and usability analysis of 14 authentication systems. System usability was evaluated using an ad hoc subjective rating scale.

In 1993, Wood and Banks [31] identified human error as being a significant issue in computer security, and in 2005 Schultz [27] posited that the computer security problems were primarily people problems. In 2009, Liginlal et al. [19] found that the number of privacy breach incidents due to human error were increasing. Although this paper argued

for more effective organizational policies as an antidote to this increase in privacy breaches, it also acknowledged the role that poorly designed human-computer interaction played in these breaches. Norman [22] provides a comprehensive guide to designing usable systems in general. In this guide, he argues, "in my experience, human error usually is a result of poor design: it should be called system error." Usable design is not just a way to make users "feel good" about using a system; usability is foundational to the security of the system. Thus, understanding existing usability hurdles in two-factor authentication systems is a necessary step to improving security on the Internet.

Just and Aspinall [16] surveyed the two-factor authentication systems used by 10 UK banking websites and identified a number of common implementation patterns. They evaluated both the security and usability of the system, but did not interact with any end-users of these banks. In 2015, Krol et al. [17] conducted interviews with 21 individuals who used two-factor authentication as part of the login process for several UK banks. Participants used a variety of two-factor systems, including card readers, hardware code generators, SMS, phone calls, and smartphone apps that generated single-use codes. Hardware code generators were particularly disliked by participants; in fact, a few individuals changed banks because of the difficulty of using the tokens.

Gunson et al. [13] studied two-factor authentication in automated telephone banking, and found that users reported lower usability of the two-factor systems. However, users also perceived a higher level of security with two-factor authentication. Participants in the study were given a hardware code generator and were asked to authenticate with a simulated telephone banking system. A 7-point Likert-type scale was used to assess usability via 22 randomly ordered questions. As demonstrated by other studies, users disliked having to carry a dedicated code generator device around. Furthermore, users were unsure of how the code generator provided better security to their account as compared to traditional knowledge-based questions.

In 2014, De Cristofaro et al. [12] conducted a Mechanical Turk survey of online users already using two-factor authentication. In this work, they specifically studied hardware code generators, one-time codes via SMS and email, and smartphone code generator apps. They found that email or SMS messages were the most commonly used second-factor for financial or personal sites, but that hardware tokens were most common for work. Interestingly, this study reported SUS (System Usability Scale) scores in the 'A' range for all two-factor systems studied. We validated this work by conducting a controlled study of users over a period of two weeks.

Weir et al. [29] compared the usability of three variants of hardware code generators being evaluated by a bank in the UK. The first system would generate a code with the push of a button. The second and third systems both required the user to insert their bank card into the code generating device. In addition to inserting the bank card, the the third system also required the user to enter a PIN using a scroll wheel mechanism. Participants used each of the three systems, then were interviewed and asked to fill out a short usability questionnaire. Finally, participants were asked to authenticate once more using their favorite system. The study found that users would almost always choose as their favorite the system that they perceived to be the most usable, not the system with the highest perceived security. Push button hardware tokens were perceived to be more usable, and total authentication time was significantly less than PIN-secured tokens. In a similar study, Weir et al. [30] conducted an in-lab study of three authentication systems, including SMS and hardware code generator based two-factor systems. They found that participants were most successful using the SMS-based system.

Lang et al. [18] report on Google's internal deployment of YubiKeys to their employees. Although this work does not report SUS scores or a similar metric, it does report a long-term reduction in the number of authentication-related support tickets after deploying the hardware keys. Further, they demonstrate that overall authentication time was significantly reduced as compared to other one-time code based systems. Google now

allows consumers to secure their accounts using YubiKeys as well, though it has only been until quite recently that such applications have been academically studied.

Ruoti et al. [25] conducted a comparative user study of seven single-factor web authentication systems using a novel tournament structure. Participants used a small number of authentication systems to log into a mock forum website and mock banking website in a laboratory setting. In contrast to many previous works, this study reported SUS metrics for each of the systems tested. Their results show that users prefer single sign on systems, but that users still have some qualms about trusting a single sign-on system. Similarly, we also conducted a comparative study of authentication systems (described later), though our study differs in that it specifically studied two-factor authentication and took place over a two-week time period instead of 45 minutes in a laboratory.

Das et al. [10] performed two studies measuring both the usability and the acceptability of the YubiKey on a Google account. In their initial study, the researchers had participants attempt to set up the YubiKey on their Google account. Employing a think-aloud protocol, they made a number of recommendations to Yubico based on common points of confusion. After one year, they repeated the study with a second group of users, and found that although many of the previous usability concerns had been addressed, many users still did not see much benefit in using the YubiKey. Das et al. postulated that this lack of acceptability was due partly to the lack of awareness of the benefits and risks of using the YubiKey. Also, because the YubiKey worked in addition to a password instead of replacing it, they point out that there is a net increase in the cognitive load. One of the key takeaways of this work was to underscore the importance of clear communication with a user—it is not enough to reduce the number of usability concerns and assume that the users will follow. Users need at least a basic understanding of the risks they face and the security benefits of adopting a particular behavior (such as using a YubiKey for two-factor authentication).

Concurrent to our study, Colnago et al. [8] conducted a large-scale survey of faculty and students at Carnegie Mellon University during a campus-wide deployment of the Duo two-factor authentication system. Duo is a commercial two-factor authentication product and supports second factor authentication using a smartphone, phone calls, U2F, and several others. Colnago et al. found that many participants in the survey recognized the security benefits of using two-factor authentication. They also identified a number of usability issues with the deployment of Duo. One interesting takeaway is that the differences in perceived usability between users that voluntarily adopted two-factor authentication and those that were required to adopt two-factor authentication was fairly small; many participants that were required to use two-factor authentication reported it to be easier to use than they expected.

2.2 Analysis of Related Work

It is difficult to draw any certain conclusions about the usability of two-factor authentication in general from the works summarized above. A number of studies demonstrated the poor usability of several two-factor systems, particularly in the realm of online banking in the UK, but we caution that these results may not generalize to two-factor authentication as a whole. Interestingly, results from both Lang et al. [18] and De Cristofaro et al. [12] indicate that users find two-factor authentication to be much more usable once they overcome the initial adoption phase.

Largely missing from previous studies is any mention of smartphone-based authentication, which involves the user either using a time-based code-generator app or receiving a push notification sent to their smartphone. While much previous work studied the hardware code generators common in commercial settings, these hardware code generators are not generally supported by consumer-level service providers such as Google and Facebook. USB tokens such as the YubiKey are used in both commercial and consumer

settings, but the only published work of the usability of these keys was in a commercial setting [18].

Attempts to draw conclusions about the overall usability of two-factor authentication are perilous at best—there are simply too many possible systems that would need to be tested to make an accurate statement about their usability. But even drawing conclusions about the usability of a single system may be flawed. Reynolds et al. [24] describes two usability studies of YubiKeys (a type of FIDO U2F compliant hardware token) recently conducted by our research team. Preliminary results from these studies indicate that participants found the setup and initialization process of using the YubiKey to be extraordinarily difficult. In a follow-up study however, participants were guided through the setup process by a coordinator and asked to use the YubiKey as a second-factor to authenticate with their Google, Facebook, and Windows accounts for a four week period. These participants reported significantly higher SUS scores, suggesting that there is a significant difference in usability between setup process of the YubiKey and the day-to-day use aspect of using a YubiKey. These results, though still preliminary, indicate that we need a finer-grained model of user behavior to inform the design of authentication usability studies.

Part II

Groundwork

Chapter 3

User Behavior Model

We now describe a four-phase model of user behavior that describes the way that users think about and interact with authentication systems. Our behavior model is specifically intended to improve the quality and specificity of results that can be derived from authentication user studies, though we believe that this model will be helpful in an even wider range of product and user study design tasks.

3.1 Buy-in

The buy-in phase is a group of precursor events that happen before the user decides to adopt a particular system. Although the user traditionally must know about the system in order to use it, there are examples of zero-interaction authentication systems that may not require the user to actually know that the system is working or available to them (e.g., IP-geolocation authentication systems).

If the user is aware of the existence of the system, they must make a decision whether or not to begin using it. We model this decision in the context of risk communication theory, which is traditionally used in the public-health and disaster preparedness sectors. However, researchers such as Blythe et al. [4] have also used risk communication principles in designing usable security interfaces. We summarize such a risk analysis with three questions:

1. **Awareness**—Does the user perceive a risk of compromise for their online account?

2. **Evaluation**—What is the significance and likelihood of that perceived risk of compromise?
3. **Efficacy**—Does the user believe they can avoid this risk? (And at what cost?)

Studying the usability of a system during the buy-in phase is challenging because of the large number of variables at play. For example, a user may perceive a system to be highly usable based on advertising, social connections, or previous experience using other similar systems. In the context of account security, users may not properly understand the threats that the system would defend them against or may wrongly estimate their true vulnerability to attack. Finally, the user may perceive the cost of adopting the system to be too high (i.e., using the system would require too much time or money). Each of these factors are examples of the costs and benefits that users may consider when choosing to use an authentication system. Ultimately, the user will make a decision to either use or not use the authentication system. If the user chooses to use the system, they will enter into the adoption phase.

Interviews with users and sentiment surveys are helpful in discerning patterns of user buy-in for different authentication systems. Usability studies in this stage are mostly concerned with understanding users' typical online behavior and can assist with better discerning how users value security and privacy.

3.2 Adoption

If the user decides to adopt the authentication system during the buy-in phase, then they will take action on that decision. This action could involve making a purchase (such as purchasing a hardware token) or could involve entering personal information, such as entering a cell phone number into a form on a webpage. The boundary between the buy-in phase and the adoption phase is defined by whether the user has taken a definitive action with intent to use the system.

Financial and temporal investments made by the user during the adoption phase lead to an escalation of commitment when using the system. For instance, if the user spends $40 purchasing a hardware authentication token, they will be more committed to spending additional time to setup the token on their accounts and will be more likely to continue using the system on a daily basis. Because the adoption phase does not usually require a large amount of time, laboratory user studies are well-suited to understanding most installation-phase usability concerns of authentication systems.

3.3 Day-to-day Use

When the adoption phase is complete, the user begins the day-to-day use phase. This phase includes all events that happen during the user's regular interactions with the authentication system. In a typical password-based authentication system for example, this would include the user needing to recall their username and password pair and correctly input these credentials into the verification system.

Online surveys, such as the survey conducted by De Cristofaro et al. [12], of individuals already using the system may be helpful for determining its day-to-day usability. The advantage of such an approach is that surveys are usually able to reach a much wider and diverse population of users. However, it is difficult to control confounding factors in such surveys, such as how often and how long the user has actually been using the system. Furthermore, this approach does not allow testing of unreleased or minimally-deployed authentication systems. As an alternative to surveys, controlled longitudinal studies provide a way to capture the day-to-day usage aspect of a system, while better controlling experiment variables. Longitudinal studies may provide the opportunity to collect more qualitative results through in-person interviews with study participants.

3.4 Revocation, and Recovery

The revocation and recovery phase includes events outside of the typical authentication experience. If the user is unable to authenticate using the normal means (due to a lost hardware token or forgotten password), they will need a way to recover their account. If the account credentials have been lost or stolen, then they should be revoked in order to prevent misuse. Presumably, these events will be rare, though more research needs to be done to determine exactly how often users do need to recover or revoke their account credentials.

This class of events is difficult to study scientifically, since events requiring revocation or recovery typically happen sporadically over a long time period. Thus, recruiting and retaining users to participate in user studies could prove difficult. Possibly, study participants could be required to correctly recover their account at the end of a longitudinal authentication study. However, this approach could risk unduly swaying participants' view of the day-to-day usability of the authentication systems under study.

3.5 Implications for Usability Studies

We developed this model specifically to inform the way that usability research in the authentication space is performed. In particular we felt that existing research could be improved by being more clear about how results obtained in a research setting, especially laboratory settings, would apply to real-life situations. We believe that a study designed to study specific phases of usability will generate more concrete, actionable results. Furthermore, our behavior model highlights the importance of studying areas of authentication that have previously not received as much as attention, such as the set up phase and the revocation/recovery phase.

Chapter 4

YubiKey Usability

On the basis of our previously described model of user behavior, our research team designed and executed two user studies to gain insight into the usability of U2F Security Keys. Both studies are described in more detail in Reynolds et al. [24]. We include a brief overview of the design and results of these studies because of their critical role in validating the efficacy of our behavior model and of studying phases of usability separately.

The purpose of these studies was to explore the usability of the setup and day-to-day phases of participants' experience in using Security Keys. Previously, Lang et al. [18] had conducted an internal usability study of Security Keys being used by Google employees. As compared to other two-factor authentication systems currently in use at Google, the Security Key performed well, reducing both authentication time and the long-term number of IT support tickets.

However, participants in the Lang et al. study were not necessarily representative of everyday consumers wishing to further protect their accounts. Employees that are required to use two-factor authentication to protect their accounts have a different system of incentives as compared to consumers protecting a personal account, and may have access to additional employer-provided resources (such as an IT support desk). Employees that feel they have nothing to protect personally, and thus reject two-factor authentication for their personal accounts, may feel more responsible to protect sensitive company information. Further complicating the usability dynamic of two-factor authentication in a

corporate setting, employees that are mandated to use two-factor authentication on their company accounts may feel resentful or irritated about additional difficulties they may experience while logging in. By contrast, consumers that voluntarily secure their personal accounts with two-factor authentication may perceive the two-factor authentication to be more usable, or at least more worth the trade off as compared to corporate workers. Given these differences, we designed two studies targeting consumers as opposed to corporate employees.

4.1 Measuring the Usability of YubiKey Setup

Our first study examined the user experience of setting up the YubiKey on Google, Facebook, and for a local Windows 10 account. Each of these systems were selected because they are already widely used and all support authentication with a YubiKey. By sheer volume of users, Google is thought to have one of the largest deployments of two-factor authentication, despite having less than 6.4% of its total users that have enabled two-factor authentication. Facebook similarly boasts a vast set of users, although it is unclear how many of these users actually use two-factor authentication to protect their account. Windows 10 supports multiple methods of authenticating with the YubiKey and has a wide installation base. We felt it important to test the YubiKey in not only a web setting (Google and Facebook), but also in a local operating system context (Windows 10). Each of these contexts has a differing set of potential threats that the YubiKey could defend against.

4.1.1 Study Design

At the commencement of each study, participants were given a YubiKey in its original shipping envelope. We then directed the participant to spend 5 minutes on the Internet learning about the YubiKey. The intent of this time was to help the participant familiarize themselves with what a YubiKey was (most participants had not heard of the YubiKey

before) and to simulate the process of a consumer researching a product online before deciding to purchase. Beyond telling the participant to, "use the Internet," we did not offer any further guidance to the participant during this 5 minute time period.

Following the brief familiarization phase, we asked the participant to set up the YubiKey on each of the three accounts in turn. We varied the order these tasks were assigned such that each of the six possible orderings to configure the three systems were covered an equal number of times. For this study, participants did not use their personal accounts, instead using account credentials provided by the study coordinator. Although only Google Chrome and Opera supported the YubiKey at this time, we included desktop shortcuts for all major browsers supported on Windows 10 at the time: Google Chrome, Opera, Mozilla Firefox, Microsoft Edge, and Internet Explorer.

Participants did not receive any assistance from the study coordinator in configuring the YubiKey for any of the accounts. In the event that the participant did ask the coordinator for help, the coordinator would reassure the participant and tell them to simply do their best. Participants were instructed to tell the study coordinator when they had completed each task. The coordinator would note whether the participant had correctly set up the YubiKey for the given account and allow the participant to move on. Although participants were encouraged to give a good effort to setting up the YubiKey on each system, we did allow them to abandon a task and move on if they decided that they would not be able to be successful at configuring the YubiKey within a reasonable time frame. To ensure that the participant would have enough time to experience each of the three systems, we limited each task to approximately 20 minutes.

At the conclusion of the three tasks, participants took a brief questionnaire, including several short answer questions and a standard SUS (System Usability Scale) survey. We used only a single SUS survey at the end of the study instead of three SUS surveys after each task because we wanted to understand the overall usability of the YubiKey itself, not necessarily the particular usability metric of each system; additionally,

we felt that multiple questionnaires would have made the study overly fatiguing for the participant.

4.1.2 Setup Results

Participants identified a number of usability problems when attempting to set up the YubiKey. Problems included outdated documentation, lack of success indicators, and accidental account lockout on Windows 10. Participants were by far the most successful in setting up the YubiKey for the Google account, exceeding an 80% success rate. We believe that Google's guided "wizard" approach was one reason that participants were so successful. By contrast, many participants were unable to configure the YubiKey for Facebook. Many participants mistakenly believed they had been successful at configuring the YubiKey for two-factor authentication on Facebook due to a misleading dialog box message telling the user that the YubiKey was ready to be used with Facebook. In reality, additional steps were needed to complete the two-factor authentication registration process. Windows 10 faired no better than Facebook with barely over 40% of participants managing to correctly locate and follow a dense 17-page PDF containing some out of date instructions. According to notes taken by the study coordinators, participants were overall much more frustrated by the process of setting up the YubiKey for Windows 10. Furthermore, in addition to being unsuccessful at configuring the YubiKey, nearly 20% of the participants also locked themselves out of the machine due to a user interface flaw. Depending on the exact chain of events, this lockout situation requires booting into safe-mode or reinstalling the operating system to undo the damage.

4.2 Day-to-day Usability of YubiKeys

Having uncovered several concerns with the setup phase of using the YubiKey, we next turned our attention to the day-to-day use aspect of using the YubiKey. In particular, given the somewhat dismal usability findings from the first part of our study, we endeavoured

to determine how users would then react to the device when using it as part of their daily lives. We also wanted to test how the form-factor of the YubiKey affected its usability—the YubiKey comes in both a full sized form-factor and a significantly more compact Nano form-factor.

4.2.1 Study Desgin

Participants began the study by meeting with a study coordinator to receive their YubiKey. We assisted each participant in configuring the YubiKey on their personal Google, Facebook, and Windows 10 accounts. To minimize the risk of participants accidentally locking themselves out of their account, we also configured other forms of two-factor authentication (SMS and printed backup codes) for each account. We asked participants to use the YubiKey whenever prompted, and to use the backup method only if they were unable to access their YubiKey. Although the YubiKey supports two-factor authentication on some mobile devices, this support is not consistent. Thus, when logging into Gmail or Facebook on a mobile device, participants used an alternate form of two-factor authentication, such as receiving a verification code via a text message. Participants used the YubiKey on their personal accounts for the period of four weeks, after which they reported back to be interviewed. The interviewer followed a semi-structured pattern, following a set list of questions, but deviating at their discretion with follow-up questions to explore any particularly salient points made by the participant. Three researchers coded each interview with an agreed-upon codebook.

4.2.2 Day-to-day Results

In contrast to the setup phase, the YubiKey performed surprisingly well in terms of usability during the day-to-day study. Generally, participants felt that the YubiKey was not overly intrusive, and several mentioned that the key was just as usable (if not more

so) than SMS two-factor authentication. Nearly all participants (93%) believed that the YubiKey helped make their account more secure.

In terms of acceptability however, the YubiKey was a mixed bag. Some participants were enamored with the security benefits of the key and mentioned wanting to purchase one. Other participants mentioned previous experience with having their (or a friend's) online accounts broken into. These participants were much more likely to view the YubiKey as being useful in protecting themselves online. At the same time, many others felt that the security benefits were superfluous, claiming, for instance, they had "nothing to protect". Similar sentiments were echoed by participants in the concurrent study performed by Das et al. [10].

These types of negative sentiments suggest that user adoption of the YubiKey, and more generally of two-factor authentication, will not be driven so much by a user-friendly interface design (although this is helpful), but by demonstrating how small investments in better account security can offer longer-term payoffs against lost productivity due to account compromise. Perhaps the greatest challenge will not be in merely improving the usability of the YubiKey, but in demonstrating its actual utility in the lives of regular users.

4.3 Application to Behavior Model

The YubiKey studies were the first application of our novel user behavior described in Chapter 3. From these studies, we were able to directly identify specific usability problems (and successes) that were unique to either the set up or day-to-day use phases of the user experience. These findings would not have been as evident without our novel study design of studying each phase in isolation. The success of the YubiKey studies demonstrates the applicability of our model to authentication usability research and bridged the gap between the way we theorized users would behave in our model and how they actually behaved in the real world.

Part III

Two Weeks of Two-factor

Chapter 5

Background and Objectives

In the Reynolds et al. [24] YubiKey study previously described, we wanted to test the efficacy of applying the user behavior model described in Chapter 3. This study allowed us to gain valuable insights into some of the differences in the user experience of setting up the YubiKey and using the YubiKey on a day-to-day basis. In particular, we were impressed by the good day-to-day usability results from using the YubiKey as well as the strong security guarantees made by the U2F protocol. What remained unclear was whether the YubiKey would hold up to other two-factor authentication systems in terms of usability.

We found it intractable to compare our YubiKey usability results directly with results from previous studies on other two-factor authentication systems for several reasons. Although we had collected numerical SUS scores in both the short and long-term study, these scores were intrinsically tied not only to the YubiKey, but also to the Google, Facebook, and Yubico application interfaces seen by the users. The complexity of isolating the usability of the YubiKey itself caused us some reticence in drawing any firm conclusions about how the YubiKey would compare to any other two-factor authentication system, particularly, since the research surrounding those systems had used very different methodologies.

The lack of comparable results between different two-factor authentication systems is somewhat systemic simply due to the extreme variability in the test conditions and overall goals of the published research. Furthermore, many of these studies [18][29] either

focused on corporate authentication or compared two-factor authentication systems that are not supported by actual online service providers like Google and Facebook. We were most interested in testing the usability of two-factor authentication systems that a consumer would be able to enable for their personal accounts in the real world.

5.1 Objectives

Our overarching research objective was to better understand the user experience of using different two-factor authentication systems available to consumers today. Each of the following objectives represents a component of this goal.

1. **Quantify usability**— An important objective was to compare numeric usability metrics for each two-factor authentication system. To support this object, we gathered both timing data and SUS (System Usability Score) data for each system under test.
2. **Quantify learnability**— We wanted to determine the effect of time and additional experience on user's performance in using two-factor authentication. We hypothesized that users would become faster at using two-factor authentication as they became more familiar with using the system. Supporting our underlying research goal, we also wanted to compare whether certain second-factor systems were more learnable than others.
3. **Qualitative user-experience analysis**— To provide background and context to our research, we wanted to conduct interviews with individuals to gain more insight into how they felt about two-factor authentication. We also asked participants more generally about their online security posture, such as whether they were concerned about any of their online accounts being broken into, and what steps they had taken to secure those accounts (such as enabling two-factor authentication). We coded this data to better understand in aggregate the sentiments expressed by participants.

5.2 Description of Systems Under Test

Our study compared five prevalent two-factor authentication systems available from many large online service providers today. A brief description of each system follows below.

1. **SMS**—The user is sent a six (sometimes seven) digit verification code through a text message to their mobile phone. Partly because most consumers already own a mobile phone capable of receiving text messages—99% of Americans according to a recent Pew study [20]—this two-factor authentication method is one of the most widely deployed. Potential usability problems may include delayed delivery, lack of cellular service (such as in a foreign country or remote location), and miscopying the code from phone to computer.

2. **TOTP**—This is an acronym for Time-Based One Time Password. To set up this two-factor authentication method, the user first synchronizes a secret key generated by the provider to their smartphone, usually by scanning a QR code. In order to generate a verification code, the app combines the secret with a truncated timestamp, hashes the result, and truncates to derive a verification code (as with SMS, usually 6 or 7 digits long). The server verifies the user-supplied code using the same method. The full specification is described by M'Raihi et al. in RFC 6238 [21]. The advantage of using a TOTP code generator app is that once the secret has been synced, the user does not need to rely on a cellular provider to deliver the one time codes—eliminating both a potential attack surface and a problem in usability. However, if the TOTP secret is stolen from the server or the phone, then the user could be subjected to an impersonation attack.

   Each code is only valid for a set time interval, usually 30 seconds, after which a new code must be generated. Crucially, this means that users will typically have *less* than 30 seconds to actually enter the code because codes can be generated in the middle of the 30 second interval. However, it is unclear what the exact usability

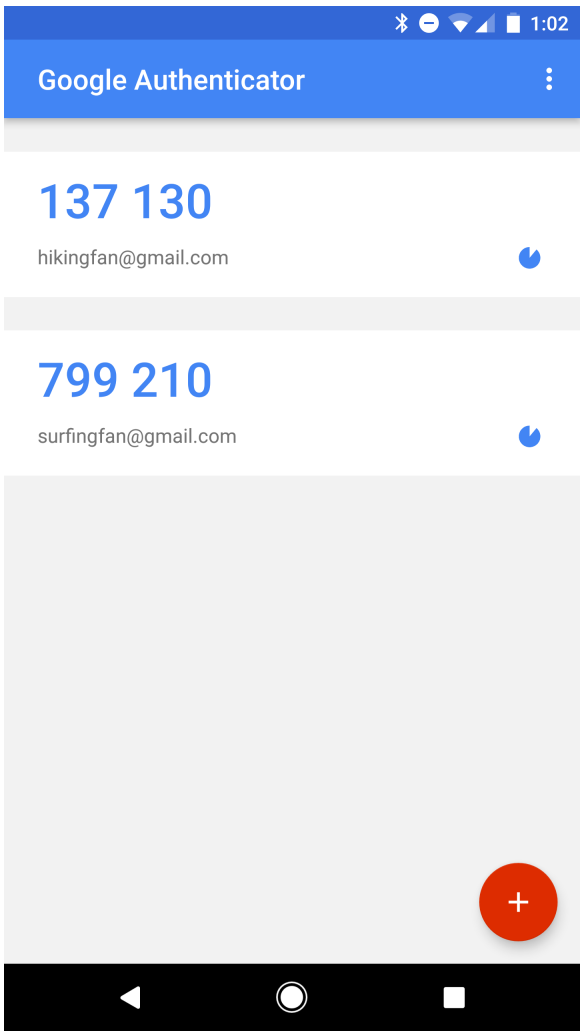

Figure 5.1: Example of TOTP authentication through the Google Authenticator interface

implications of this time interval are. Additionally, the smartphone and the server must both have a clock that is reasonably in sync. As with SMS, the verification codes still must be manually keyed in by the user, leaving additional room for user error. According to the same Pew study as above, only 77% of Americans own a smartphone, meaning that TOTP is not as broadly deployable to all customer-bases as SMS is. An example of TOTP as seen in the Google Authenticator app is displayed in Figure 5.1.

3. **Pre-generated codes**—This form of two-factor authentication is usually used in conjunction with other two-factor authentication methods as a backup authentication method in case the user is unable to access their primary two-factor authentication method. Implementation is straightforward: the service provider simply generates a list of verification codes and has the user print or write the codes down. The length of the list itself is variable and the codes are usually around 8 digits long. The codes may be used in any order and must be kept secure against theft both on the server-side and on the user's end. Because the pre-generated codes are usually longer than the verification sent through SMS or generated with TOTP, there may be additional room for user error when entering the codes. Furthermore, the user must be careful not to lose the medium on which they recorded the codes.

4. **Push**—In this two-factor authentication method, the user receives a push notification on their smartphone that allows the user to either "Approve" or "Deny" a login attempt. This technique is supported by Google (through their "Google prompt") and is available as a commercial service through Authy OneTouch and DUO Mobile. The advantage of this system is that there is less chance of user error, since there are no numbers that must be correctly copied off a phone screen. Push authentication does require an active Internet connection in order to work, though this requirement is *likely* to be fulfilled by virtue that the user is already trying to login to an online service provider. We hypothesize that not having to type in numbers, as required by other two-factor authentication systems, will be both faster and perceived as more usable by participants. An example of an authentication approval using the Authy app is shown in Figure 5.2.

5. **U2F Security Keys**—Originally developed through a collaboration with Google and Yubico, and now sponsored by the FIDO (Fast IDentity Online) Alliance, Universal 2nd Factor (U2F) is an open standard allows users to use a USB hardware device to authenticate online. In contrast to the other four two-factor authentication

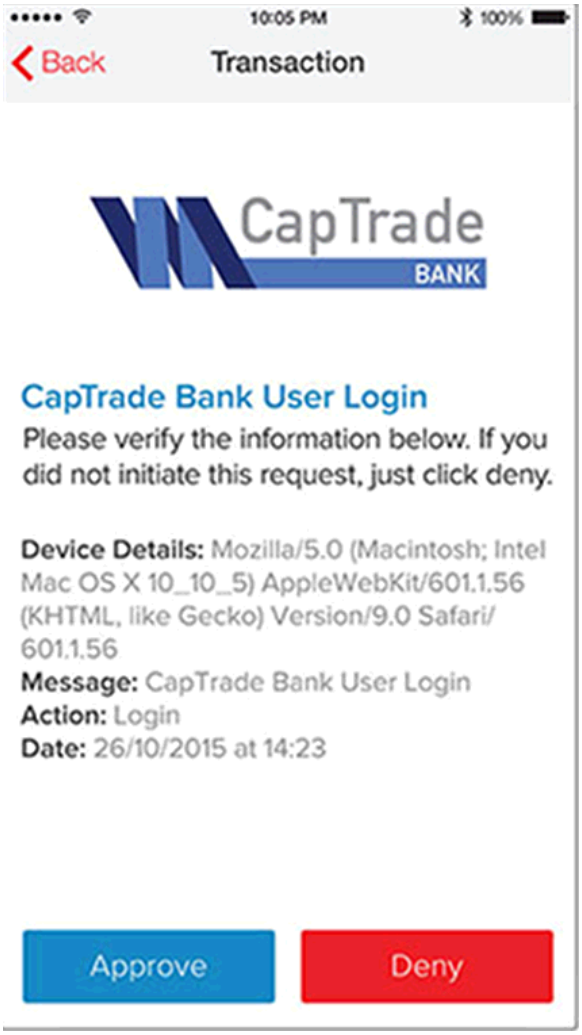

Figure 5.2: Example of push-based authentication through Authy OneTouch

Figure 5.3: Representation of the YubiKey NEO used by our participants

methods described above, the U2F standard itself is designed to be unphishable and provide more security and privacy protections than other forms of two-factor authentication.

In order to authenticate with a Security Key, the user must connect the device to their computer and activate the device when prompted by the website. We used the YubiKey NEO (pictured in Figure 5.3) in this study, which is a particular instantiation of the U2F Security Key.

Chapter 6

Methodology

We conducted a 72-person comparative longitudinal study of two-factor authentication, approved by our institution's ethics review board. Our study was designed to achieve our research objectives of being able to compare the usability and learnability of the five systems described in Section 5.2: SMS, TOTP, pre-generated-codes, push, and U2F Security Keys.

6.1 Study Design

Participants were divided into 6 groups of 12 participants each. Five of the groups were assigned to a specific two-factor authentication scheme and the final group was a control group that used only passwords with no two-factor authentication at all. Each participant initially met with a study coordinator in order to create an account on the study website. During this meeting, the participant was given a list of 12 tasks to complete on the study website over the next two week period (with no more than one task per day). As part of completing each task, the participant would need to login to the study website each day using their assigned authentication mechanism. At the conclusion of the two week participation period, participants reported back to participate in an exit interview with a study coordinator. Using a combination of authentication event timing data, survey responses, and qualitative data gathered from the exit interviews, we compared the usability of the various authentication systems under test and made a number of key observations and recommendations on the basis of this data.

6.2 Banking Website

Our test scenario was that of a participant needing to log into an online banking interface and complete a task, such as transferring money between accounts or paying a bill online. To support this scenario, we built a simple online banking interface, pictured in Figure 6.1. The banking interface supported authentication through either a password alone or a password in tandem with one of the five two-factor authentication systems described previously.

We automatically recorded and a number of events in the system, including those surrounding the authentication process. Events were triggered at the beginning and end of the password login phase and at the beginning and end of the two-factor authentication segment. Each beginning and completion event were associated through a unique identifier that allowed us to correlate distinct login attempts for each user. By computing the difference in the timestamp for a beginning event and the timestamp for its conjugate end event, we were able to determine the amount of time taken by each event. We used this data to determine the median authentication time for each two-factor authentication system as well perform a repeated measures correlation test for each user to determine the correlation between time to authenticate with two-factor authentication and time elapsed since they began the study; that is, whether the participant became faster (or slower) at logging in with more experience.

The frontend of the website was built as a single-page application using the React.js library. The backend banking and event-tracking systems were built using Sanic,[1] an asynchronous Python 3 HTTP server framework built on libuv. All account information and logged event data was stored in MongoDB. In order to protect participant privacy, all passwords were transmitted to our server using TLS 1.2 and were stored at rest using the Argon2id [3] password hashing algorithm.

[1] `https://github.com/channelcat/sanic`

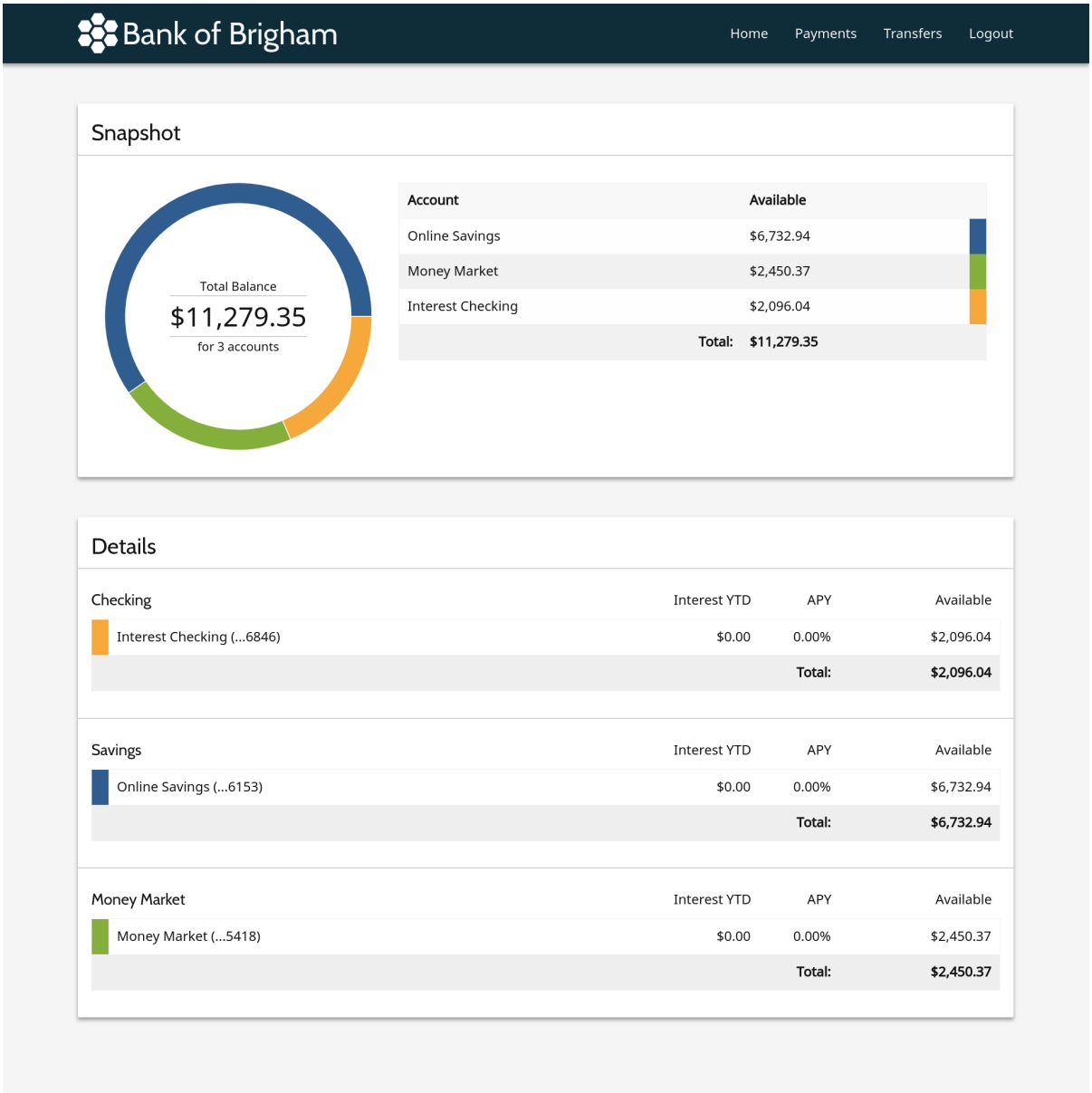

Figure 6.1: Example of the banking interface we constructed for our study

6.3 Recruitment

We recruited 72 participants using flyers posted throughout a university campus. Prospective subjects were told they would need daily access to an Internet-connected computer with Google Chrome. In order to be considered for the study, potential participants filled out a short survey to see if they owned an Android or iOS smartphone, or if they owned a phone able to receive text messages. Participants were then assigned to a particular study group for which they would be eligible (two study groups required a smartphone, for instance).

6.4 Demographics

We had a slightly higher number of female participants (38; 55%) as compared to male participants (31; 44%) in our study. Participants were largely young adults: 18–19 years (3; 4%), 20–29 (61; 88%), and 30–39 (5; 7%). Over two-thirds of the participants (49; 71%) had completed some college but had not yet completed a degree. One participant had completed only high school, with the remainder having completed an associate's degree (8; 11%), bachelor's degree (9; 13%), or master's degree (2; 2%). Participants self-reported their level of computer expertise: far above average (13; 18%), somewhat above average (28; 40%), average (25; 36%), and somewhat below average (3; 4%).

6.5 Setup and Initial Meeting

Participants scheduled an initial appointment to meet with a study coordinator. During the initial meeting, the study coordinator assisted them in setting up an account on the online banking interface. We allowed participants to choose their own username and password, with the only restriction being that the password had to be at least eight characters long. If the participant was part of one of the study groups using a second-factor scheme, the coordinator would also help them configure the two-factor authentication

on their account for the study website. Depending on the study group, this included helping the participant install any necessary apps, verifying their phone number, issuing the participant a YubiKey NEO, or printing the backup codes.

Participants were issued a list of 12 tasks that they would need to complete during the study period. The order of these tasks was generated randomly and the tasks were designed such that no permutation of their ordering would ever cause any of the participant's accounts on the banking interface to overdraw. The study coordinator assisted the participant in completing the first listed task during the intial meeting, leaving the participant with 11 tasks to complete on their own.

To avoid confusion and at request of our ethics review board, participants were expressly told that the bank was only a simulation and that they would not be able to withdraw or make deposits to any real bank. Additionally, we asked participants not to use their actual banking credentials for the study.

6.6 Two-week Task Completion Period

Over the next two weeks, participants were asked to complete no more than one task per day in the order given on their task list. At their discretion, we allowed participants to skip completion of a task for 1–2 days during the study period.

To complete each task, the participant would need to visit our online banking website and login with their previously selected username and password. With the exception of the control group using only a username-password pair, the participant would also authenticate using their assigned second-factor system for each login. Our event system recorded all failed and successful login attempts, including timing data for each attempted login. After logging in, the participant would go to either the "Payments" or "Transfers" page and complete the banking component of the task.

The purpose of having participants complete the banking-related task after logging in (as opposed to simply having the individual login and do nothing) was to encourage

the user to act more naturally during the login process and make the simulation more realistic—most real-world users do not authenticate for amusement; rather authentication is a means to an end and not an end itself. Simply, we did not want the authentication step to become the end goal in the mind of the participant, but rather daily completion of a banking task, as would be the case using a real banking website.

6.7 Exit Interview

Participants reported back for an exit interview with a study coordinator at the conclusion of the two-week period. The coordinator would first have the participant take a brief survey to gather a small amount of demographic data. Participants also completed a SUS (System Usability Scale) assessment of the website as a whole and for the authentication system they had used for the study. Following this, the coordinator would conduct a semi-structured interview with the participant to gather additional information about how the participant felt about the website overall as well as the login process. In particular, we asked participants questions about their overall online security posture to better understand their background and feelings about online security. With consent of each participant, we recorded audio of each interview. Two coders working together then listened to the recordings and coded each interview. We selected the codes on the basis of the questions that the coordinator asked in each interview.

6.8 Compensation

Participants were compensated a maximum of 25 USD at the conclusion of their participation in the study according to a tiered compensation structure they agreed to before beginning the study. In order to incentivize participants to login to the website and complete a task, compensation was based on the total number of tasks completed. We asked participants to self-report the number of tasks they completed in the exit interview and remunerated them accordingly.

Chapter 7

Results

We collected both quantitative and qualitative data in this study. Quantitative data that we collected included timing data for each second factor authentication step as well as results from a SUS (System Usability Scale) assessment completed by each participant. We also collected qualitative data by conducting a semi-structured interview at the conclusion of the study with each participant. We begin with a report of our timing data and SUS scores, including a statistical analysis of these results. Following this, we will provide more detail about the results of our qualitative analysis.

7.1 Timing Data

7.1.1 Individual Learnability

We computed the correlation between the amount of time an individual had been in the study and the amount of time it took them to authenticate. We used the repeated measures correlation (rmcorr) technique described by Bakdash and Marusich [1] to estimate the common regression slope for each two-factor authentication system being tested. Our hypothesis was that participants would get faster over time as they became more familiar with the two-factor authentication system, that is, there would be a negative relationship between the amount of time elapsed since beginning the study and authentication time. Table 7.1 summarizes the repeated measures correlation results for each two-factor authentication system being tested. We found statistically significant

Table 7.1: Repeated measures correlation (rmcorr) between amount of time participating in study versus amount of time to authenticate.

2FA System	p-value	r	df	95% confidence interval
SMS	0.2797	-0.0970	124	(-0.2688, 0.0807)
TOTP	0.5857	-0.0494	122	(-0.2251, 0.1294)
Push (Authy)	0.0288	-0.2038	113	(-0.3744, -0.0198)
U2F (YubiKey)	<0.003	-0.2690	118	(-0.4289, -0.0927)
Printed Codes	0.4255	-0.0760	110	(-0.2595, 0.1128)

($p < 0.05$) support for this hypothesis for both push notifications and U2F Security Keys, but not for the other systems.

7.1.2 Comparison of 2FA Authentication Times

We applied a Kruskal-Wallis one-way analysis of variance and found there was a significant difference ($p < 0.001$, $\alpha = 0.05$) in the median authentication time between the systems. We did not include the time that it took the user to enter their password; the observed authentication times reported here include only the time to get through the second-factor authentication step. The Security Key (U2F) devices had the fastest median authentication time, followed by push notifications. These timing results are summarized in Table 7.2 and Figure 7.1.

Table 7.2: Authentication Time (in seconds), Summary Statistics

Authentication System	Q1	Median	Mean	Q3
Printed Codes	11.340	17.230	28.010	25.370
Push (Authy)	8.437	11.840	16.130	17.580
SMS	12.950	16.610	18.460	22.090
TOTP (Google Authenticator)	10.650	15.050	23.890	23.340
U2F (YubiKey NEO)	4.482	9.092	13.010	16.250

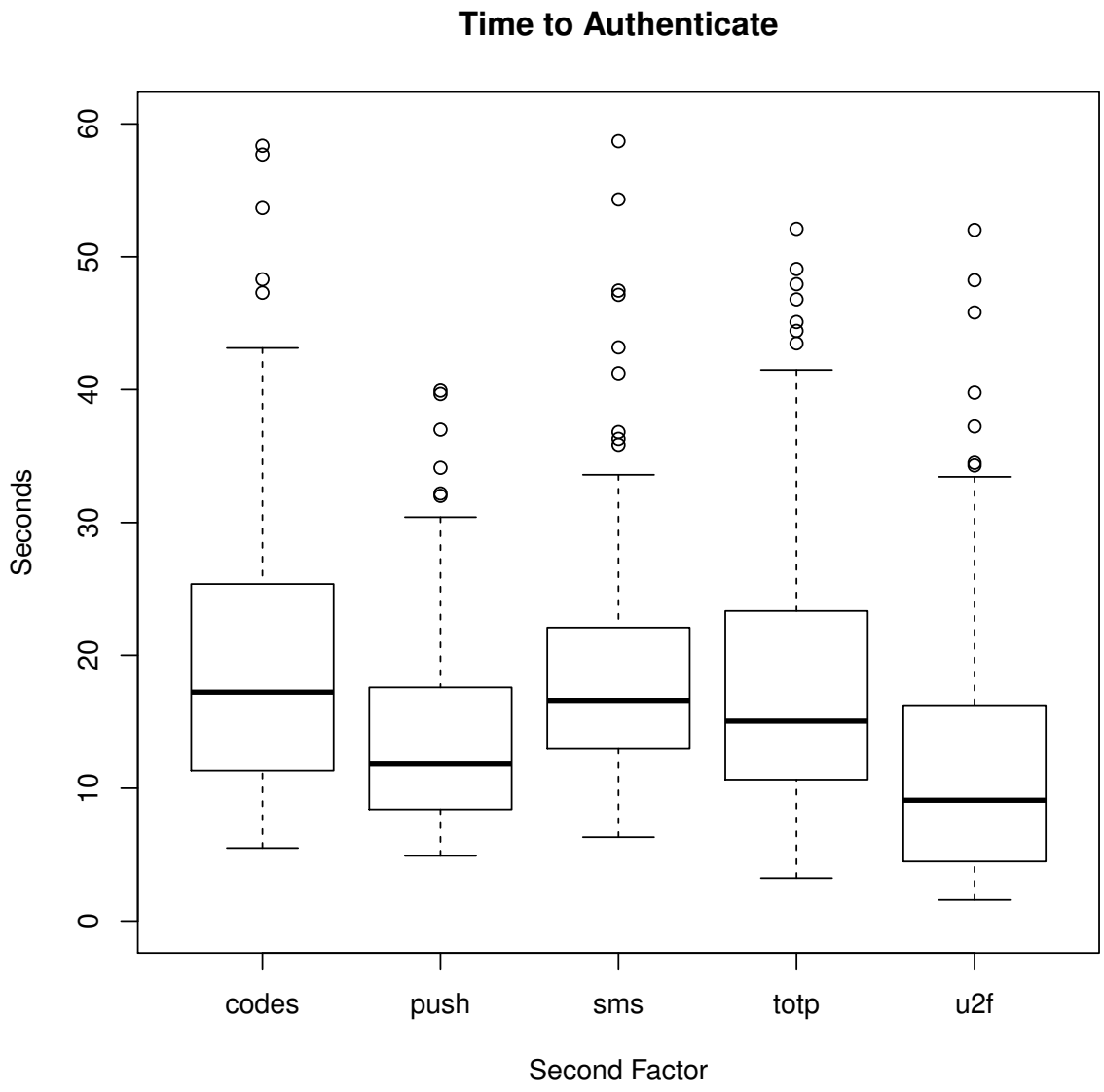

Figure 7.1: Time to authenticate for each second-factor authentication system being tested.

7.2 Usability Survey Rankings

We administered two SUS (System Usability Scale) surveys to participants at the beginning of each exit interview session. The first survey considered the usability of the banking website as a whole and the second had participants consider only the usability of the login system. The purpose of administering both SUS surveys was to determine how

large of an impact the banking website itself had on the participants' feelings about the authentication system. Additionally, we felt that participants would be more accurate with their opinions about the login system if we had first given them opportunity to both consider and express their feelings about the system as a whole; had we only given a SUS survey on authentication system, we felt participants would be more likely to (incorrectly) report their feelings about unrelated website features. The SUS results for the overall website (grouped by the authentication system) are shown in Figure 7.3. Similarly, the results for each authentication system is shown in Figure 7.2.

We performed a Kruskal-Wallis one-way analysis of variance and determined that the authentication system used was a statistically significant ($p = 0.02579$, $\alpha = 0.05$) predictor of the median SUS score for the two-factor authentication system. We also computed value of $\rho = 0.7576$ for Spearman's rank correlation coefficient and confirmed that there was a significant ($p < 0.001$) correlation between the overall website SUS scores and the SUS scores of the individual authentication systems. Summary statistics for each two-factor authentication system are shown in Table 7.3. Passwords with no second-factor had the highest median SUS score, with a median score of 95, followed by TOTP (via Google Authenticator) which had a median SUS score of 88.75.

Table 7.3: SUS Scores for each two-factor authentication system, Summary Statistics

Authentication System	Q1	Median	Mean	Q3
Password	87.5	95.0	92.5	98.75
Printed Codes	75.0	80.0	80.23	90.0
Push (Authy)	72.5	81.25	81.04	92.5
SMS	68.75	75.0	75.0	80.0
TOTP (Google Authenticator)	75.0	88.75	83.12	92.5
U2F (YubiKey NEO)	61.88	75.0	73.12	93.12

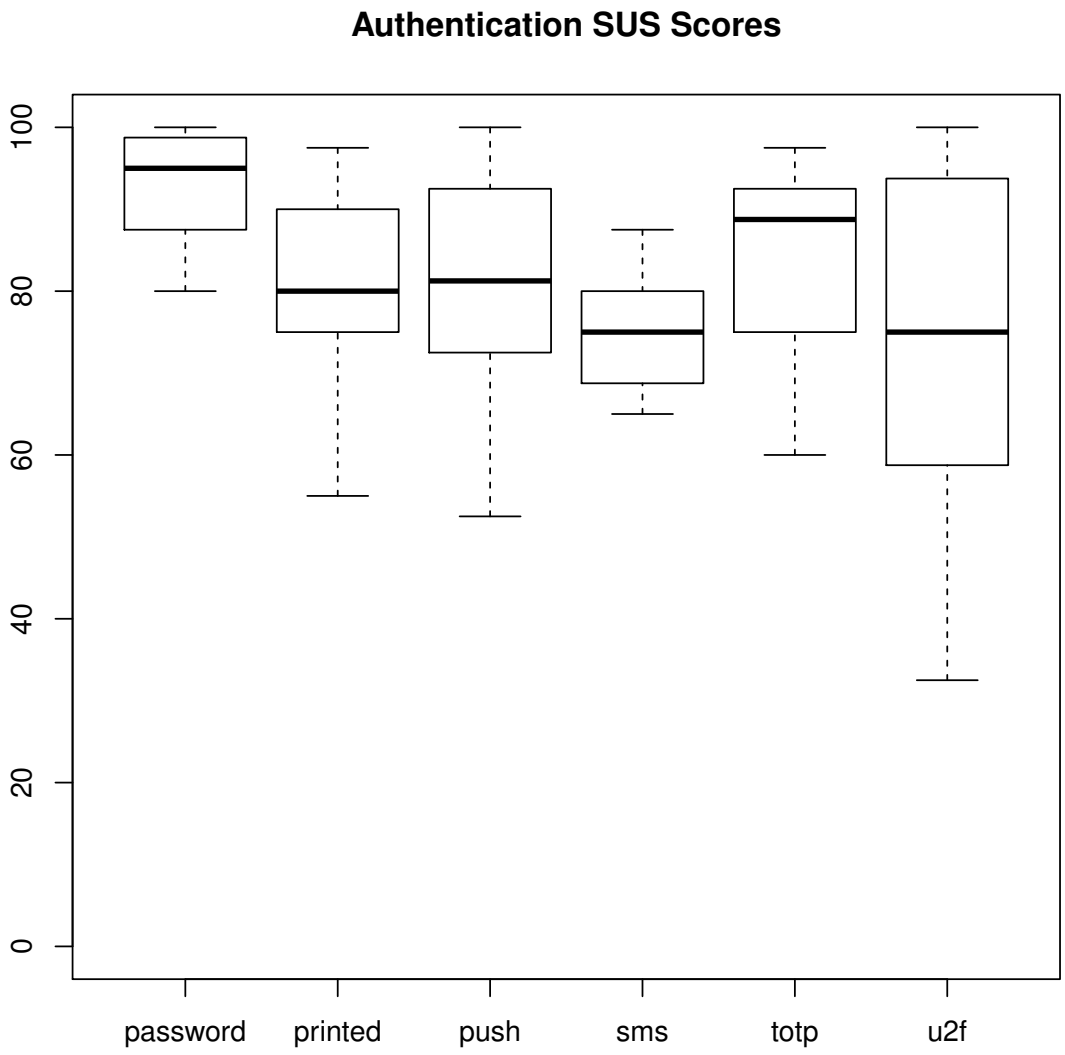

Figure 7.2: SUS scores for each authentication system being tested.

7.3 Qualitative Results

7.3.1 Previous Experiences with Account Compromise

A few individuals in the Reynolds et al. [24] study mentioned, unprompted, that one or more of their online accounts had previously been broken into. We followed up on this result by explicitly asking participants in this study whether any of their online

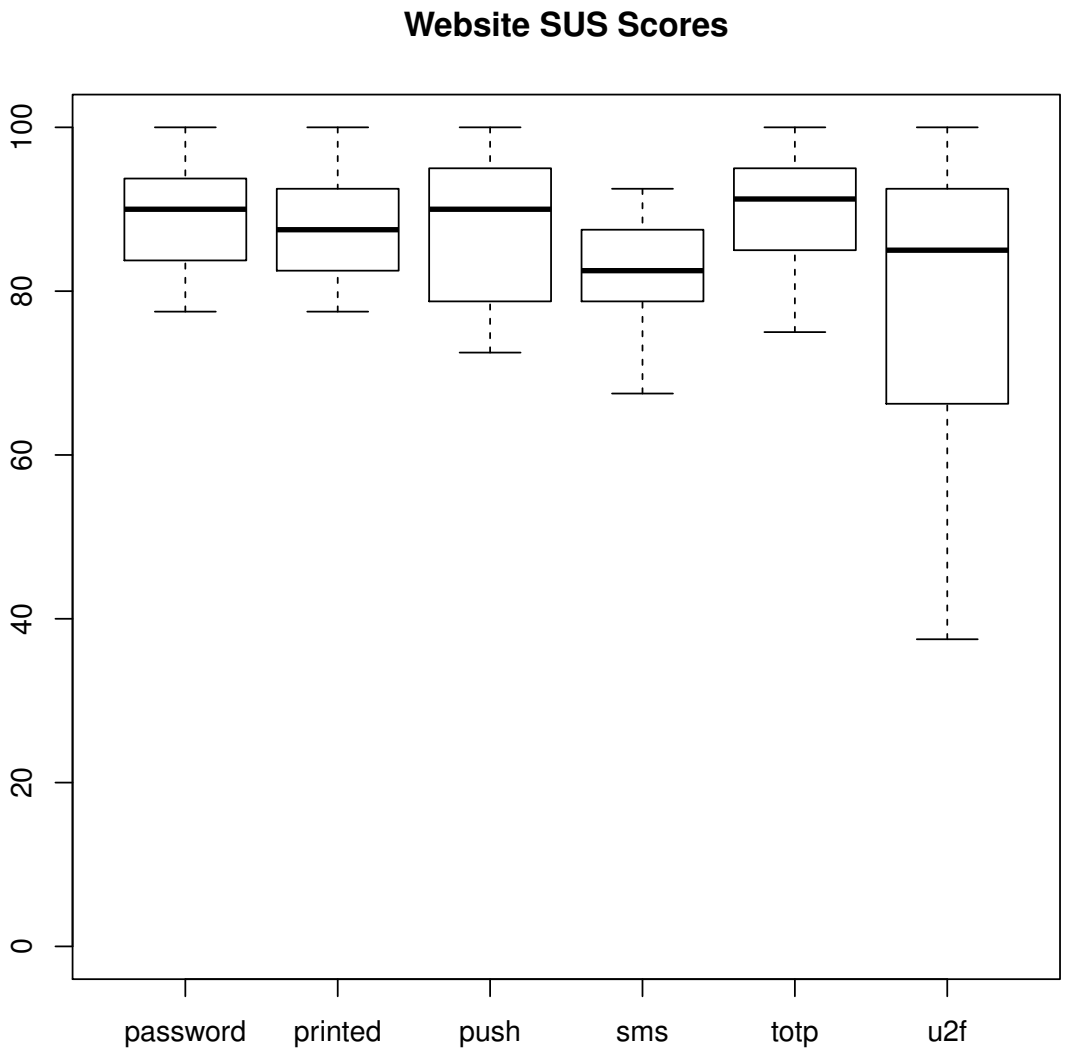

Figure 7.3: SUS scores for overall website, organized by authentication system being used.

accounts had ever been broken into. A number of participants (26; 37%) described experiences with remote attackers taking over their online accounts and a few people (7; 10%) mentioned that someone they know had had one of their online accounts hacked. Although not directly a form of online account compromise, a few participants also mentioned experiences with financial theft from having their credit or debit card number stolen or having their bank account credentials stolen. Others mentioned having their

personal information stolen as part of one or more data breach events, including the highly publicized Equifax compromise of millions of individuals' personally identifying information [2]. When asked how they noticed that the account had been compromised, most participants said they received an email indicating a new login from a suspicious location.

> ***P19:*** *"My Facebook account was broken into... I received a notification that someone had logged in from Africa, and I had not been to Africa. I changed my password, made it a little longer."*

> ***P23:*** *"My Apple account has been hacked before. I had to go on and change my password.... Apple sent me a [message] and said someone had logged in from Nevada or something."*

> ***P27:*** *"Something happened with my Facebook... I clicked on a link someone sent me, but it wasn't really from them. I changed my password right away after I figured it out. Over the summer I had someone spend a bunch of money from my bank account. I don't know how that happened—it was really scary; they spent a lot of money."*

7.3.2 Security and Inconvenience

We asked participants whether they felt that the second verification step while logging in made them feel more secure. Most participants did feel more secure, although 3 of 12 participants that used the printed backup codes did not feel like the codes added any additional security to the system.

> ***P6:*** *"I felt like the codes didn't accomplish anything, because that's just more passwords—anyone could guess them."*

We also asked participants if the additional security would be worth the additional login time or inconvenience they might face when using the second factor system. Several

people (20; 28%) said the extra security was definitely worth the trade off, and an additional group (25; 36%) said that they would be willing to use two-factor authentication depending on the importance of the account.

> ***P25:*** *"In my opinion it may be a little obsessive for everything, but for banking it's something that I actually do want some authentication. I almost wish that it was a requirement that the bank said, oh here set [two-factor authentication] up. Because now that I think about it, I don't know how to set up two-factor authentication with my bank. If it were an option I would definitely use two-factor authentication."*

> ***P33:*** *"It was pretty quick, so that was good; I didn't feel like I had to jump through a lot of hoops. I can imagine it being nice having an extra wall of security if it's your bank information, so that even if somebody else gets your password, it's not like they're going to be able to hack into your account because they don't have the [Security Key]."*

Some participants were particularly concerned about the centrality and importance of their email account, particularly considering the potentially large amount of sensitive data being stored there. For example, one participant reported they had already turned on two-factor authentication for their Gmail account to gain extra protection:

> ***P24:*** *"I use my email for everything, and so I thought it wouldn't hurt to have some extra security. The thought of someone hacking into [my account] and having everything vulnerable... better to be safe than sorry."*

Other participants (9; 13%) expressly stated that they would not be willing to use two-factor authentication in order to gain additional security because the inconvenience was too high.

P37: *"I don't know how much my level of convenience and my need for level of security would balance out because for me having something that is convenient and is at hand is almost more important than having something that is more secure... I know if people hack your credit cards, then the bank will take care of that and get the money back and so having that extra security makes me care less about having a second-factor."*

7.3.3 Experience with Compromise and Worth Inconvenience

We hypothesized that participants with previous experience having an account compromised would be more likely to feel that using a second-factor was worth any extra inconvenience. Using data extracted from coding the interviews, we used Pearson's chi-squared test with two degrees of freedom to test the dependence of these variables. Not all participants expressly talked about both of these variables, thus we analyzed only participants for which we had coded data for both variables.

Table 7.4: Account Compromise and Inconvenience

	Hacked	Not Hacked
2FA definitely worth inconvenience	11	9
2FA sometimes worth inconvenience	6	19
2FA not worth inconvenience	4	5

We observed no statistically significant relationship between a participant's previous history with account compromise and whether they felt that two-factor authentication was worth the inconvenience ($\chi = 4.6332$, $p = 0.0986$, $\alpha = 0.05$). One limitation of this analysis is that it does not consider the exact nature of the previous account compromise (such as whether financial loss was involved). However, we do note that numerous individuals independently stated that using two-factor authentication would be worth the inconvenience at least some of the time, particularly for financial accounts.

7.3.4 Perception of Likelihood for Account Compromise

Participants expressed a wide spectrum of views on how much value they placed on their online accounts. Some participants (9; 13%) felt that they had nothing to protect and would therefore not be a target of criminals.

> ***P5:*** *"I guess maybe because it's that I don't have anything to protect... I'm at a stage in my life where nothing I own is that valuable and none of my information is that wanted that it makes a difference."*

> ***P8:*** *"I mean, you hear a lot about stuff being broken into; I just don't think I have anything that people would want to take from me, so I think that's why I haven't been very worried about it."*

> ***P30:*** *"I don't have a lot of money in my accounts right now, so if someone stole my money, that would be bad, but its not enough that it would be the end of the world if I lost all my money— I don't feel like I'm a target for someone to steal my stuff. I can imagine in the future if I had a huge retirement fund or something then I would want that to be more secure."*

7.3.5 Availability of Second Factor Device

In order to login, each participant in the study in one of the two-factor authentication groups was required to use something external to their computer in order to login, whether it be the sheet of paper with printed codes, a YubiKey, or their phone. Many participants (24; 34%) mentioned not having their second-factor immediately available to them when they needed to login.

> ***P8:*** *"I don't always have my phone on me, and so if I'm doing something on the computer, I'm usually doing homework, so I actually try to keep my phone away from me."*

P42: *"Honestly, once I'm home I kind of just set my phone down and forget where I put it sometimes, so that was a little bit hard . . . I needed to go find my phone and pull up the app."*

7.3.6 TOTP Timeout

Although the participants using TOTP (via the Google Authenticator app) were overall very positive about their experience, 8 of 12 participants mentioned that they had problems entering the six digit verification code before it timed out.

P30: *"I have to type in these numbers so fast or else it's going to go away."*

7.4 Discussion of Results

In this section, we will further highlight some of the most interesting results of our study and discuss their meaning in context of usable two-factor authentication.

7.4.1 Relationship between Authentication Time and Usability

Although both push-based authentication and the U2F Security Keys had faster median authentication times, neither of these systems received the highest median SUS score. Conversely, TOTP was the highest scoring second-factor system we tested, but had a median authentication time that was slower than either push or U2F. From our exit interviews we identified some explanations for this result. First, some participants receiving push requests through Authy did not always receive the authentication request in their notification area and instead had to manually open the app and approve the request. It was unclear whether this was a bug in the Authy or the result of notification configuration on some participants' phones. Several U2F participants using both Windows and Mac operating systems reported a variety of minor troubles getting the YubiKey to work with their computers (possibly they plugged it in the wrong direction). However, other participants reported no problems using the YubiKey. Ultimately, participants using

TOTP reported liking the relative simplicity of the Google Authenticator app. The app functioned very similarly to SMS, a two-factor authentication system with which many participants were already familiar while not requiring them to always have cellular service.

We believe that the minor issues encountered by participants using the Authy app and the YubiKey likely explains most of the lower scores it received. That said, no authentication system we tested scored extremely poorly in terms of usability, suggesting that, although there is a noticeable impact on usability from requiring two-factor authentication, the presence of two-factor authentication itself does not doom the system as a whole to poor usability.

7.4.2 Remember Me?

In our study, we purposely did not provide a "Remember Me" option, thus requiring participants in the non-control groups to use their second-factor every day. We believe that some of the usability impacts of needing a second factor could be mitigated by only requiring the second-factor on new computers or after logging out. This would provide a similar level of protection against remote attackers while mostly allowing users unfettered access to their accounts. Some systems allow access for a limited amount of time (30 days for instance) without requiring a second-factor on the same machine. Participants with previous experience using such systems (typically for a university login system) made some remarks to the effect that they were never quite sure when the second-factor would be required. One solution to this problem would be to have a small count-down displayed to the user telling them how many days were left until they would need to again provide their second-factor to avoid the "ambush" effect described by Sasse et al. [26]. Further research needs to be done to determine the right balance of when to ask the user for the second-factor again when they have already been logged in previously on the same machine.

7.4.3 Acclimation and Likability

One unique design attribute of our study is that participants used their second-factor repeatedly over a period of two weeks instead of merely using it in a laboratory setting. Given the weak usability results of previous two-factor authentication studies, we expected an overall poor usability response. During the exit interviews, we were surprised at the number of participants that reported an overall positive experience using two-factor authentication. Many participants wanted to use two-factor authentication for some of their actual online accounts, but were either unaware it was an option or were unsure how to configure it. We believe that our participants were more willing to use two-factor authentication than previously has been reported because in our study they had an acclimation period to become adjusted to using two-factor authentication.

7.4.4 Differentiating Between High and Low-value Accounts

Although participants generally tended to care less about the security of their social media accounts, many expressed concern about the security of their banking and financial accounts. There were mixed feelings about frequently used accounts like email accounts, however, particularly in balancing whether it would be worth using two-factor authentication for such accounts. Participants generally agreed that they did not want to always have to use their second-factor to login to their email account when logging in from a known computer. Other participants felt they had no confidential information in their email, and that having a second-factor would not be worth the extra login step. In general, the higher the perceived value of the account, the more likely the participant was to be willing to use two-factor authentication for the account.

7.5 Limitations

Several limitations were inherent to our methodology. Because our participants were recruited from a university campus, they tended to be younger and more technically

savvy than the population as a whole would have been. A sample of university students would also be more likely to have fewer material assets to be concerned with as discussed in Sections 7.3.4 and 7.4.4. Additionally because we wanted to capture authentication timing data, we were unable to have participants use a real banking system or an existing online account; this may have altered their behavior. Participants also were required to use two-factor authentication for every authentication attempt, which may have caused them to acclimate to using two-factor authentication more quickly than would be seen if two-factor authentication had only been required on new machines. Finally, participants' discussions of the necessity of two-factor authentication and online security in general would have been different had we mocked our website as a social media site.

Part IV

Epilogue

Chapter 8

Future Work

Based on our behavior model described in Chapter 3, we would like to further explore several usability components of two-factor authentication. In this section, we discuss some ideas for potential future areas of research.

8.1 DUO Authentication

DUO offers a commercial two-factor authentication package deployed by many organizations. It supports multiple forms of two-factor authentication, including push, TOTP, and U2F. Brigham Young University recently required many of its students and faculty members to begin using DUO two-factor authentication to protect their accounts. Many current students, when asked about DUO as part of an informal pilot survey, felt that any extra security that DUO offered was not worth the inconvenience of always having to authenticate. Because many students at BYU use open access lab computers (such as those at the library), they are unable to effectively use the "Remember Me" option which might otherwise lessen the authentication burden. One student in the pilot study reported that he had purchased a personal laptop expressly to avoid having to go through DUO authentication on the library computers. Many participants in our two-factor authentication comparison study (described in Part III) mentioned DUO as being one of the two-factor authentication systems with which they had previous experience

We are planning to conduct a campus-wide survey about DUO usage to better understand whether users feel more secure using DUO and to identify actionable usability

problems. This survey will be sent to both BYU faculty members and current BYU students. We believe that faculty may perceive DUO to be less burdensome because they tend to use personal office computers as compared to students that use many different machines. A potential solution to mitigate some of the usability concerns of DUO would be to only require DUO authentication while off-campus. However, this approach may have some unintended security consequences that would need to be examined more closely. We also want to explore the use of U2F devices with DUO and analyze how this might change (for better or worse) the perceptions of DUO for both faculty members and students.

8.2 Comparative Setup Phase Study

Additional studies need to be done to understand the roadblocks in two-factor authentication adoption. Although certainly not the only roadblock, many users in the Reynolds et al. [24] study experienced problems with poor documentation while setting up the YubiKey. These issues are likely not limited to the YubiKey itself and poor user experience during the setup phase may be systemic to many other two-factor authentication systems. Therefore, we would like to conduct similar studies to the Reynolds study using different two-factor authentication methods, such as TOTP through the Google Authenticator app and compare these results to the original study.

Chapter 9

Conclusion

Two-factor authentication is arguably the most effective means of securing online accounts against compromise. It is clear moving forward that passwords alone will be insufficient to protect individuals from determined attackers. There are many concerns about the usability of two-factor authentication systems that are yet to be addressed. We believe that our work has made several important contributions to understanding the usability of two-factor authentication, including:

1. **First studies of YubiKey**—We were the first to study the viability of the YubiKey as a means of two-factor authentication. Although there were many concerns with the setup phase of the YubiKey, we were able to demonstrate fair success of novice users in setting the key up for Google accounts, suggesting that it is possible for this phase to be much improved on other service providers. Results from both long term studies indicate that users are surprisingly accepting of U2F Security Keys as a form of two-factor authentication.

2. **Model the authentication user experience in separate phases**—We are the first to explicitly view the user experience of authentication in separate phases. Of particular interest is our separation of the setup and day-to-day use phases. Previous studies have long been limited by confounding these two phases, leading to results that, on the surface, contradict each other. By studying each phase alone, it is easier to isolate specific usability concerns that can be mended while avoiding participant bias.

3. **Importance of long-term usability**—Although laboratory studies are helpful, we firmly believe that understanding how users' preferences and authentication performance changes over time is of paramount importance as well. Our second longitudinal study particularly demonstrated how users' performance using a second-factor authentication system can increase over time as they become more familiar with the system. Laboratory studies are insufficient to show such improvements, simply because users are not exposed to the system under test for a sufficient amount of time.

One approach, suggested by Unger et al. [28], is to focus first on building a usable system, and then build as much security in as possible. A secure system that is unusable may create more problems than it solves, and in fact may make a system *less* secure than before [26]. But ultimately, it is not enough to merely reduce the number of usability concerns in a security system to a (potentially arbitrary) "acceptable" level.

Although we have focused extensively in this work on identifying such usability concerns—and this is certainly a critical aspect of building secure computer systems—merely reducing barriers is not the driving force that will lead users to adopt more secure practices. Outside the electronic world, industry experts in a number of fields employ strategies in risk communication to help people prepare for the occurrence of potentially catastrophic events. The insurance industry has been hugely successful at least in part because it is effective at communicating risks and mitigation strategies (typically involving the purchase of a financial instrument from the company) for the most common everyday dangers faced by the public. Although in most cases these policies are tacitly meant to mitigate against obvious physical risks—fires, floods, earthquakes, car crashes—insurance is at its crux a protection against the potential negative financial ramifications of these events.

There are significant financial ramifications for a malicious account takeover even on accounts that are not immediately tied to a person's financial accounts. For instance,

losing access to an email account would cause significant lost work hours (meaning loss in pay) as well as potential loss of intellectual assets (meaning loss in future profits). Social media accounts may seem to have no financial component until one considers the potential social-capital loss stemming from malicious posts made on a user's account by an attacker, not to mention the value of any private information being stored in the account that would enable impersonation attacks. This is not to mention the potential domino effect of compromised accounts; information in one account reveals additional accounts to target, as well as additional information about the user to enable further compromise of their identity.

By no account is communicating these cyber-risks to users a trivial task, nor is there an exact mapping between the risks we face in the physical world and those we face in the cyber world. People frequently purchase insurance plans for high-value physical objects—such as a house or car—because it is more clear what is being protected. What is not as clear is how we can appropriately communicate the risk of account compromise for a user's email account—an intangible good that they paid nothing for.

At the same time, many users do care about security in general, and many are open to exploring two-factor authentication as an option for additional security. From our research, we find overall that users are not apathetic about account security; rather, they simply do not have enough information to be able to judge their amount of online-risk and take appropriate mitigating steps. Two-factor authentication is no silver bullet for improving the users' online security. However, it is critical that researchers evaluate existing systems in terms of usability and likewise consider usability as a component for new two-factor authentication schemes.

Our user behavior model forms the foundation for a new generation of authentication usability studies. We have demonstrated the viability of this model by conducting multiple user studies of various two-factor authentication systems. These studies have provided a valuable snapshot of the state of two-factor authentication usability and we

have been able to identify several usability successes and concerns from these studies. We look forward to many more authentication usability studies that will inform the design and improvement of two-factor authentication and protect millions of users from account compromise.

Bibliography

[1] J. Z. Bakdash and L. R. Marusich, "Repeated Measures Correlation", *Frontiers in Psychology*, vol. 8, p. 456, 2017.

[2] T. S. Bernard, T. Hsu, N. Perlroth, and R. Lieber. (Sep. 2017). Equifax Says Cyberattack May Have Affected 143 Million in the U.S., [Online]. Available: `https://www.nytimes.com/2017/09/07/business/equifax-cyberattack.html`.

[3] A. Biryukov, D. Dinu, and D. Khovratovich, "Argon2: New Generation of Memory-hard Functions for Password Hashing and Other Applications", in *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, IEEE, 2016, pp. 292–302.

[4] J. Blythe, J. Camp, and V. Garg, "Targeted Risk Communication for Computer Security", in *Proceedings of the 16th International Conference on Intelligent User Interfaces*, ACM, 2011, pp. 295–298.

[5] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes", in *2012 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2012, pp. 553–567.

[6] J. Bonneau and S. Preibusch, "The Password Thicket: Technical and Market Failures in Human Authentication on the Web", in *The Ninth Workshop on the Economics of Information Security (WEIS)*, 2010.

[7] C. Braz and J.-M. Robert, "Security and Usability: The Case of the User Authentication Methods", in *Proceedings of the 18th Conference on l'Interaction Homme-Machine*, ACM, 2006, pp. 199–203.

[8] J. Colnago, S. Devlin, M. Oates, C. Swoopes, L. Bauer, L. Cranor, and N. Christin, ""It's not actually that horrible": Exploring Adoption of Two-Factor Authentication at a University", in *2018 CHI Conference on Human Factors in Computing Systems*, ACM, 2018, p. 456.

[9] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang, "The Tangled Web of Password Reuse", in *Network and Distributed System Security (NDSS)*, vol. 14, 2014, pp. 23–26.

[10] S. Das, A. Dingman, and L. J. Camp, "Why Johnny Doesn't Use Two Factor: A Two-Phase Usability Study of the FIDO U2F Security Key", in *2018 International Conference on Financial Cryptography and Data Security (FC)*, 2018.

[11] (2017). Data Breach Investigations Report, [Online]. Available: `http://www.verizonenterprise.com/resources/reports/rp_DBIR_2017_Report_en_xg.pdf` (visited on 08/23/2017).

[12] E. De Cristofaro, H. Du, J. Freudiger, and G. Norcie, "A Comparative Usability Study of Two-Factor Authentication", in *Workshop on Usable Security (USEC)*, 2014.

[13] N. Gunson, D. Marshall, H. Morton, and M. Jack, "User Perceptions of Security and Usability of Single-factor and Two-factor Authentication in Automated Telephone Banking", *Computers & Security*, vol. 30, no. 4, pp. 208–220, 2011.

[14] D. Humphries. (Jan. 2015). Best Practices for Workplace Passwords, [Online]. Available: `https://www.softwareadvice.com/security/industryview/password-workplace-report-2015/`.

[15] B. Ives, K. R. Walsh, and H. Schneider, "The Domino Effect of Password Reuse", *Communications of the ACM*, vol. 47, no. 4, pp. 75–78, 2004.

[16] M. Just and D. Aspinall, "On the Security and Usability of Dual Credential Authentication in UK Online Banking", in *International Conference for Internet Technology And Secured Transactions (ICITST)*, IEEE, 2012, pp. 259–264.

[17] K. Krol, E. Philippou, E. De Cristofaro, and M. A. Sasse, "'They brought in the horrible key ring thing!' Analysing the Usability of Two-Factor Authentication in UK Online Banking", *arXiv preprint arXiv:1501.04434*, 2015.

[18] J. Lang, A. Czeskis, D. Balfanz, M. Schilder, and S. Srinivas, "Security Keys: Practical Cryptographic Second Factors for the Modern Web", in *International Conference on Financial Cryptography and Data Security (FC)*, Springer, 2016, pp. 422–440.

[19] D. Liginlal, I. Sim, and L. Khansa, "How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management", *Computers & Security*, vol. 28, no. 3, pp. 215–228, 2009.

[20] *Mobile Fact Sheet*, Jan. 2017. [Online]. Available: `http://www.pewinternet.org/fact-sheet/mobile/` (visited on 08/23/2017).

[21] D. M'Raihi, S. Machani, M. Pei, and J. Rydell, "TOTP: Time-Based One-Time Password Algorithm", RFC 6238, May 2011. [Online]. Available: `https://www.rfc-editor.org/rfc/rfc6238.txt`.

[22] D. Norman, *The Design of Everyday Things (revised and expanded edition)*. Basic Books (AZ), 2013.

[23] T. Petsas, G. Tsirantonakis, E. Athanasopoulos, and S. Ioannidis, "Two-factor Authentication: Is the World Ready? Quantifying 2FA Adoption", in *Eighth European Workshop on System Security (EuroSEC)*, ACM, 2015, p. 4.

[24] Reynolds, Joshua and Smith, Trevor and Reese, Ken and Dickinson, Luke and Ruoti, Scott and Seamons, Kent, "A Tale of Two Studies: The Best and Worst of YubiKey Usability", in *2018 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2018.

[25] S. Ruoti, B. Roberts, and K. Seamons, "Authentication Melee: A Usability Analysis of Seven Web Authentication Systems", in *Proceedings of the 24th International Conference on World Wide Web (WWW)*, 2015, pp. 916–926.

[26] M. A. Sasse, S. Brostoff, and D. Weirich, "Transforming the 'Weakest Link'—A Human/Computer Interaction Approach to Usable and Effective Security", *Bt technology journal*, vol. 19, no. 3, pp. 122–131, 2001.

[27] E. Schultz, "The Human Factor in Security", *Computers & Security*, vol. 24, no. 6, pp. 425–426, 2005.

[28] N. Unger, S. Dechand, J. Bonneau, S. Fahl, H. Perl, I. Goldberg, and M. Smith, "SoK: Secure Messaging", in *2015 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2015, pp. 232–249.

[29] C. S. Weir, G. Douglas, M. Carruthers, and M. Jack, "User Perceptions of Security, Convenience and Usability for Ebanking Authentication Tokens", *Computers & Security*, vol. 28, no. 1, pp. 47–62, 2009.

[30] C. S. Weir, G. Douglas, T. Richardson, and M. Jack, "Usable security: User Preferences for Authentication Methods in eBanking and the Effects of Experience", *Interacting with Computers*, vol. 22, no. 3, pp. 153–164, 2010.

[31] C. C. Wood and W. W. Banks, "Human Error: An Overlooked but Significant Information Security Problem", *Computers & Security*, vol. 12, no. 1, pp. 51–60, 1993.

Appendix A

Materials for Comparative Study

A.1 Exit Survey Questions

Question 1. Ask the study coordinator which group you are in.

- SMS
- TOTP
- Printed Codes
- U2F
- Push
- Password

Question 2. Please select your gender:

- Male
- Female
- Other

Question 3. Please select the range that includes your age:

- 18–19
- 20–29
- 30–39

Figure A.1: Flyer used to recruit participants in comparative authentication study (Part III)

- 40–49
- 50–59
- 60+

Question 4. Please select your highest level of education:

- Some High School
- High School Diploma or Equivalent
- Some College, No Degree
- Associate Degree
- Bachelor Degree
- Masters Degree
- Professional Degree
- Doctorate Degree

Question 5. How computer savvy do you consider yourself?

- Far above average
- Somewhat above average
- Average
- Somewhat below average
- Far below average

Question 6. In the following survey, the word "system" refers to the banking website you used. All questions must be answered. If you feel you cannot answer one of the items, mark the neutral on the scale. Please record your initial reaction after carefully reading each question. (Possible answers: Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree)

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very awkward to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Question 7. In the following survey, the word "system" refers to the authentication system you used. All questions must be answered. If you feel you cannot answer one of the items, mark the neutral on the scale. Please record your initial reaction after carefully reading each question. (Possible answers: Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree)

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very awkward to use.

9. I felt very confident using the system.

10. I needed to learn a lot of things before I could get going with this system.