Toward A Secure Account Recovery: Machine Learning Based User Modeling for protection of Account Recovery in a Managed Environment

Amos Alubala

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

Abstract

Toward A Secure Account Recovery: Machine Learning Based User Modeling for protection of Account Recovery in a Managed Environment

Amos Alubala

As a result of our heavy reliance on internet usage and running online transactions, authentication has become a routine part of our daily lives. So, what happens when we lose or cannot use our digital credentials? Can we securely recover our accounts? How do we ensure it is the genuine user that is attempting a recovery while at the same time not introducing too much friction for the user? In this dissertation, we present research results demonstrating that account recovery is a growing need for users as they increase their online activity and use different authentication factors. We highlight that the account recovery process is the weakest link in the authentication domain because it is vulnerable to account takeover attacks because of the less secure fallback authentication mechanisms usually used. To close this gap, we study user behavior-based machine learning (ML) modeling as a critical part of the account recovery process. The primary threat model for ML implementation in the context of authentication is poisoning and evasion attacks. Towards that end, we research randomized modeling techniques and present the most effective randomization strategy in the context of user behavioral biometrics modeling for account recovery authentication. We found that a randomization strategy that exclusively relied on the user's data, such as stochastically varying the features used to generate an ensemble of models, outperformed a design that incorporated external data, such as adding gaussian noise to outputs. This dissertation asserts that account recovery process security posture can be vastly improved by incorporating user behavior modeling to add

resiliency against account takeover attacks and nudging users towards voluntary adoption of more robust authentication factors.

Table of Contents

List of Figures

List of Tables

Abbreviations

Acknowledgments

My journey as a part-time doctoral student taught me how to balance life commitments. I leant how to juggle professional work, schoolwork and family. I was not a typical doctoral student and certainly I would not have completed this endeavor were it not for my advisors Prof. Salvatore Stolfo and Prof. Steven Bellovin. I would like to express my deepest gratitude for their invaluable advice and guidance. Sal and Steve helped me to truly learn the science behind Computer Science, how to synthesize new knowledge and how to apply research to solve real world problems. From the weekly meetings that we had for more than two years, I leant among other things, the importance of looking at solutions to complex problems from a practical perspective. One of the many lifelong lessons that I leant from Sal was that a little humor goes a long way. Sal and Steve certainly shaped my journey during my doctoral study. But I could not have undertaken this in the first place without Prof. Angelos Keromytis admitting me to the Computer Science department. I am deeply indebted to him for that opportunity. I would also like to extend my sincere thanks to my other defense committee members, Prof. Suman Jana and Prof. Asaf Cidon. Thank you for the time you put in and helping me to successfully complete my degree requirements. Lastly, to my parents, my wife, my children, my siblings and my family at large. You provided me the support and a foundation upon which all this was built. This is for you!

Dedication

To my children Jonah, Silla, and Noah, and my beloved wife, Afua

Introduction

Performing authentication is a routine part of our daily lives. It is even more so in the cyber world, where digital authentication is the precursor to most critical transactions that we perform online, such as financial-related transactions, healthcare, and social interactions. With this increased dependency on authentication to access services and accomplish routine activities online, users need to securely maintain access continuity even when they do not have or remember their nominal digital credentials. They can forget a knowledge credential such as a password; they can break or lose hardware with embedded credentials such as a security key or a phone or buy a new device and need that associated with their digital credentials. Credentials can also expire and require to be renewed. Account Recovery is the ability to re-establish authentication credentials using a fallback mechanism, thus allowing access continuity for the user. As a result of the many advances in the authentication domain, users have a wide variety of options to choose from to accomplish identification during authentication and an account recovery process. However, the proliferation of these different options, coupled with the fact that many systems implement account recovery using the weakest authentication options, such as security questions, has left the account recovery process vulnerable as a mechanism for account takeover attacks.

In this dissertation, our goal is to secure the account recovery process for legitimate users without adding friction to the user while maintaining their privacy. In the context of Account Recovery, the primary threat model is evasion attacks, making it possible to take over a victim's account. We thus leverage resilient Machine Learning based user behavioral modeling as a critical part of the account recovery process protection. We achieve this resiliency by using principles of randomization. We investigate the most optimal and efficient randomization

strategies that meet the constraints and requirements of an account recovery process. Our study is scoped to a managed environment since our approach leverages data collected on the end user device that is used as the input to the machine learning user modeling.

First, we start by reviewing the problem space. We present results from an empirical study against real-world authentication data covering several millions of user activity from an activity from a large US-based company. These results indicate that an increase in the adoption of multiple authentication factors exacerbates account recovery. We show that users do not voluntarily adopt the most secure means to protect their digital accounts and credentials. We also show that some secondary authentication mechanisms are less secure and often suffer usability challenges. All this points to the fact that user accounts are most vulnerable to account takeover attacks during the account recovery process. This is the weakest link in the authentication domain; thus, protecting this process goes a long way toward improving the overall security posture of the user accounts.

Second, explore the protection mechanisms for an account recovery process. We present a protection scheme that leverages a randomized machine learning user behavior-based modeling approach resilient to evasion attacks. Randomization principles have been used for security for a long time. Many studies have published results using randomization to add robustness against different security vulnerabilities. But there are many ways of achieving randomization. Our work determines and presents the optimal randomization strategy under account recovery conditions. We investigate data-independent-based versus data-dependent strategies to determine the most efficient and effective account recovery in a real-world setting. Some of the challenges in an account recovery setting include that a user is likely to use a different end device than their standard device during or after an account recovery process. Thus, the user model must function

just as efficiently and accurately on a new device. Moreover, unlike other user authentication activities, the user infrequently invokes the account recovery process. Our case study demonstrated that, on average, a user went through the account recovery process less than five times a year. The model, therefore, needs to stay current between uses. A user may not invoke this process for an extended period.

This dissertation presents research results demonstrating that account recovery is a growing need for users as they increase their online activity and use different authentication factors. We then discuss the current weaknesses of this process. We highlight that account recovery is the weakest link in the authentication domain and show an effective protection mechanism that can add resiliency against account takeover attacks. This work paves the way for closing this gap and improving the overall security posture around account recovery.

1.1 Thesis

Thesis Statement

Machine learning-based user behavioral biometrics can effectively strengthen the security posture of the Account Recovery process without adding friction to the user while maintaining their privacy.

1.2 Thesis contributions

This dissertation provides the following contributions:

1. Advancing voluntary adoption of cyber protective behavior: By analyzing millions of authentication events at a large US-based company, we study the additional authentication factors adopted by older adult users compared to younger users. In enterprise settings and organizations, older users outnumber younger users an, so the problem of Account Recovery is a more significant threat to these organizations and their users. The insights from this research provide solid foundations for understanding factors that influence the voluntary adoption of robust alternative authentication methods that can be used during Account Recovery.
2. Introduction of user behavior modeling to make Account Recovery more resilient to fraud.
3. Advancing User Behavior Modeling security posture by studying randomized modeling for adding resiliency against model theft.

1.3 Thesis organization

Chapters 1 and 2 provide a background on Account Recovery, Account Recovery methods, Account Recovery weaknesses, and current protections against these weaknesses. Chapter 3 presents research results of voluntary user adoption of multiple authentication factors. In Chapter 4, we offer research results for randomized modeling for adding resiliency against model theft used for user behavior modeling. We conclude by discussing how techniques presented in this dissertation can be combined to provide holistic proactive and reactive protection for the Account Recovery process.

Chapter 1:

Background

This chapter will provide the necessary background information on the concepts and topics covered in this thesis. We will first describe what Account Recovery is and what mechanisms are used for achieving account recovery today and then present an overview of why the account recovery process needs protection. We will review some of the weaknesses in the current account recovery process as a motivation to why we need a new way of protecting these. Lastly, we will cover user behavioral modeling concepts that we will apply as a new perspective on the protection of account recovery.

1.1 Account Recovery

1.1.1 What is Account Recovery?

Account Recovery is the process of restoring account access to a legitimate account owner when they cannot otherwise use their primary account credentials for access. This is usually accomplished using a secondary or a fallback authentication method or scheme. A fundamental property of Account Recovery is verifying the requestor's identity before granting them access to the account. In this context, identity verification is the authentication of the binding that asserts a name, or an identifier is pointing to the proper entity (the object or person). As explained in this article [2], there are two ways to achieve this; by the assertion of a trusted third party and through continuity, i.e., we presume that a binding between an identifier and the

proper entity is correct today because it was correct yesterday. This continuity can be represented as a timeline going back to the initial enrollment of the entity into the system when a credential is exchanged. Today, the predominant primary credential used is a password, and password forgetfulness is the leading cause of Account Recovery for this method [55, 56, 57, 58, 59, 64]. Account owners can also break or lose hardware with embedded credentials, such as a security key or a phone, or they can buy a new hardware and need that associated with their digital credentials. Credentials can also expire and require to be renewed.

1.1.2 Why does the Account Recovery process need protection?

Account Recovery process is currently the weakest link for many authentication schemes primarily because the fallback methods used during recovery are usually not as secure as the primary or nominal methods [141]. This weakness is enumerated under Common Weakness Enumeration CWE-640 [48], which describes it as a weakness resulting from security questions being too easy to guess or determine answers to, such as by looking at one's social media site. Other examples include an implementation weakness in the recovery mechanism code, which might be spoofed to send the recovery credentials to a threat actor. Moreover, an adversary can leverage the account recovery process to deny access to the legitimate user in a case where throttling is not done at the rate of the primary credential reset, thus leading to account lockout.

A weak Account Recovery implementation or process can compromise the overall security posture of the system regardless of how strong the primary authentication mechanism might be. Several reported Account Take Over breaches leveraged account recovery as the attack vector [45, 61, 62, 133,134,135,136,137].

1.2 Account Recovery methods

1.2.1 Personal Knowledge Questions

Personal knowledge questions, also referred to as Security questions is the most used method for account recovery today, despite its many usability issues and security weaknesses [48,49,51,52,54]. A large-scale study at Google [55] demonstrated unreliability of security questions primarily due to a poor memorability to secret answers. The study reported that about 40% of the users did not recall the answers to their security questions. The study reported that most of the memorability issues were because of users purposely setting incorrect answers to their security questions to make their questions harder for others to guess. But interestingly, this behavior had the opposite effect as the users “hardened” their answers in a predictable way, making it easier for adversaries to determine the correct answers. This Account Recovery method assumes a threat actor's difficulty in correctly guessing the answers.

1.2.2 Helpdesk

Using a helpdesk or service desk where users call in for assistance during account recovery is a standard method in enterprise settings [65]. The account recovery process typically involves the support staff unlocking the user’s account through an administrative interface after verifying the user. With an unlocked account, the user is provided a temporary credential, usually a password, to access their account, and is forced to reset the credential before continued use. The security assumption here is that the verification performed by the support staff is sufficient. This recovery method comes at a prohibitive cost [64]

1.2.3 Email

This is typically in the form of a credential reset link emailed to the user's email address on record. This link typically has a time-to-live period within which the reset activity can happen. The user clicks the link to be redirected to a page where they can go through a reset process, such as setting up a new password or enrolling a new authentication method. The security assumption behind this method is that the bad actor does not have access to the user's email, and the transmission of that email message is secure.

1.2.4 SMS

This recovery method enables users to receive a one-time code (OTC) on their registered mobile device sent by SMS. This natively works on the phone without requiring the user to install anything. On the one hand, SMS is very user-friendly, but on the other hand, it is vulnerable to several attacks, such as SIM-swapping and certain malware [141].

1.2.5 Voice call

This method requires a registered phone on record that the user can be called back on, and a temporary credential issued to. Possession of this phone is often used as a validation of the receiver as being the legitimate user. The voice call requires an interactive session with the user. The temporary credential is not left as a voicemail, instead it is delivered to the user interactively.

1.2.6 Magic links

This is a form of a non-password login that the user uses to access the account or the system, thus enabling them to recover their account. This can be sent through an email or an SMS message to the user's device. This can therefore be classified as a type of a bearer token.

1.2.7 Security Keys

These are small devices that securely store confidential information such as a private key. These hardware devices generate a One Time Code (OTC) for authentication.
Hardware security tokens are more secure and phishing resistant [63] but they have some operational challenges such as requiring the user to carry them around.

1.2.8 Trusted designated intermediaries

This involves a trusted intermediary vouching for the user trying to perform an account recovery that they are a genuine user. In some cases, the intermediary is trusted with the recovery codes that the user can retrieve and user to recover the account. Usage is limited as it requires the intermediary involved as part of the account recovery process.

1.2.9 Backup Account Recovery Codes

Backup account recovery codes are typically provided to the user ahead of time. They require to be stored securely. They are available for one-time use. This method scores higher on usability but its highly depended on the user securely storing and retrieving them when needed.

1.2.10 Temporary Access Pass

A temporary Access Pass is a passcode that can be issued for use for a limited time. These passcodes are usually issued by a system admin after the user has met a minimum set of verification requirements. Using the temporary access pass, a user can access their account and enroll or set up a new authentication method option that they can use subsequently. Temporary access pass differs from backup account recovery codes because these are usually generated on demand and have a limited time to live while, backup account recovery codes are generated ahead of time and provided to the user for future use.

1.3 Machine Learning User Behavioral Modeling in the context of Account Recovery

Today, the process of authenticating users is comprised of five distinct stages [140] as depicted in Figure 1.0.1 below:

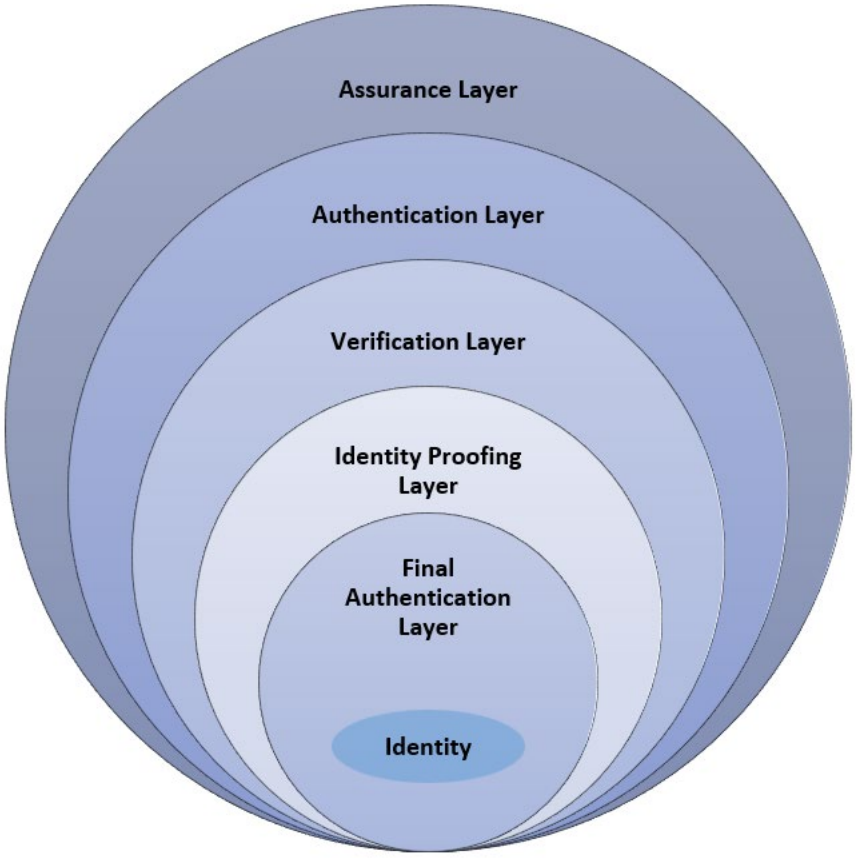

Figure 1.0.1: Five layers that comprise the process of authenticating a user used in industry today.

The outermost layer is used for providing Identity Assurance before the user provides any login credentials. Its purpose is to filter out bad actors by looking at an authentication request and comparing that to a profile of legitimate users. This profile is based on prior established declarative facts about the legitimate users' past access to the system, which typically includes known devices, known locations, known browser versions, and operating system versions. This is followed by the authentication layer, which leverages the user-supplied credentials such as passwords to establish the identity of the user. The verification layer follows, which challenges the user for an additional authentication factor beyond the one supplied in the prior layer. This typically involves a one-time passcode sent to the legitimate user's mobile device that

they have. Other factors, such as biometrics and push notifications, can be used to fulfill the challenge. Next is the Identity Proofing layer which sends the authorized user through a process that verifies their personally identifiable information or verification of a government-issued document such as driver's license that the user has. The final layer under this current model is a form of step-up authentication that can be performed on the user with an active session depending on the risk level assigned to a specific activity that the user initiates within the system.

User Behavioral Modeling adds another dimension to this process by introducing an ability to predict if the logged-on user is a legitimate user or an imposter, even when the input corresponds to an activity that the system has not seen before. This is what differentiates the current widely used approach based on a declarative user profile from one that leverages machine learning based user behavior modeling. This means that a system incorporating user behavior modeling is likely to detect a zero day type attack that a system based on classic user profiles will not since the latter's decisioning is primarily limited by historical information that the system has seen before and knows about.

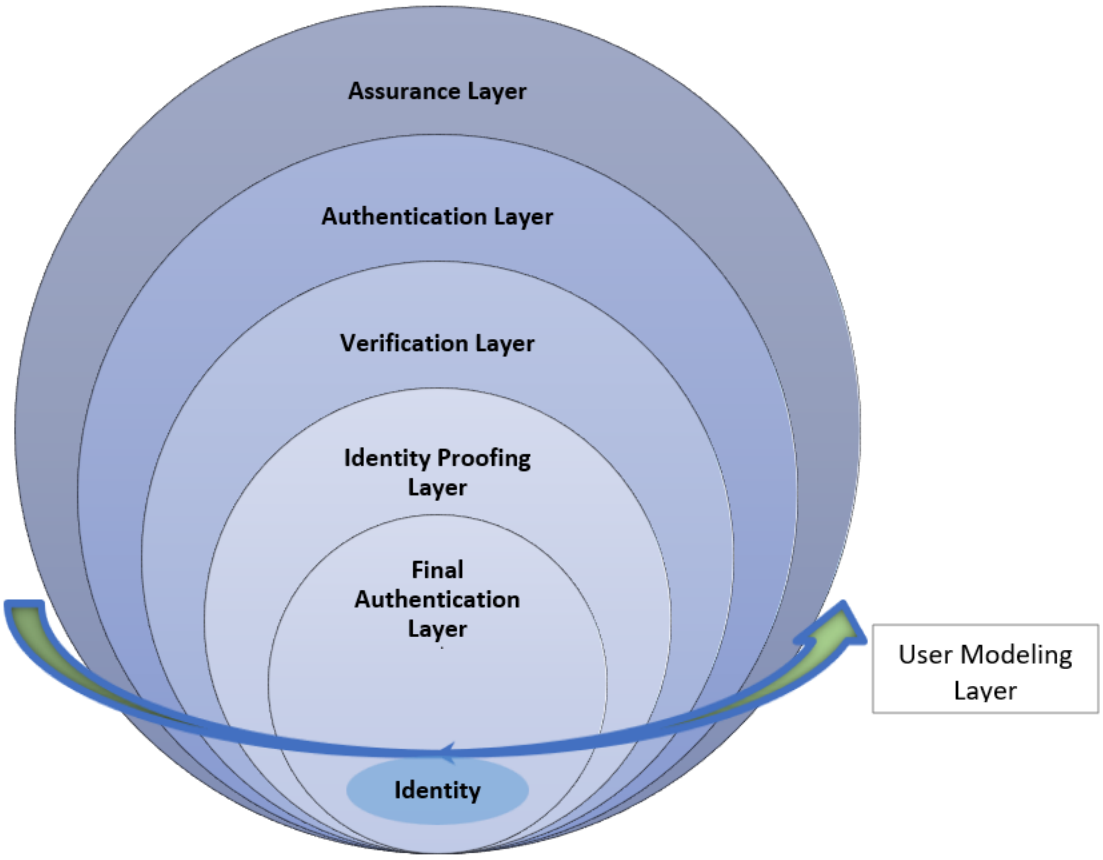

Figure 1.0.2: The 5 layers that make up the process of authenticating a user today plus the user modeling layer. There is a difference between the 'user model' introduced with user modeling and the 'user profile' used in the assurance layer

The input data used for training the user model typically comprises of events and activities that the user does in each period. This can involve the applications the user uses, the way the user interacts with a system, the duration of these interactions, a user's gait, among other events. Creation of this user model is done by a machine learning classifier that is trained to classify events corresponding to a legitimate user from an imposter. Formally, this classifier is initialized with n training input samples $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ drawn i.i.d. from a given distribution of user behaviors or events D, where $x_i \in R^d$ and $y_i \in \{+1, -1\}$. The objective is for the

classifier to make a prediction $\acute{y}_t$ given a new event x_t such that the prediction is within a given probability margin to that of the trained output.

Account Recovery typically involves a user establishing access to their account using an alternative authentication mechanism. This alternative mechanism might involve using a new device, such a new mobile device or initiating the request from a different location. As a result, the interaction with the system or the event performed by the user might be different from any of the historical information that the system knows about the user. In this scenario, the declarative profile-based approach where the system is simply performing pattern matching between an activity or event with the saved profile and detecting deviations from the historical profile to differentiate a legitimate user from an imposter does not work well. On the other hand, a well-trained Machine Learning based user model can still accurately predict whether the legitimate user is performing the new activity or the new event. This is one of the key reasons why this dissertation is calling out the gap in the current implementation of this process in the industry and proposing the use of ML-based user behavior modeling for the protection of the Account Recovery process, beyond just the declarative-based user profiles commonly in use today.

1.4 User Privacy in the context of Account Recovery

Today, the predominant method for Account Recovery is depended on knowledge-based questions and data derived from a user's Personally Identifiable Information (PII) [55]. As several studies have demonstrated, this is not secure since most of this information is easily discoverable from various sources such as social media sites. Moreover, use of a one's PII information often ends up exposing their confidential information. Enterprise service desk staff

often gets to know a user's confidential information when the user calls in and must provide this information as part of the verification step. The good news is that users are becoming more aware of the need to protect their private information. A study by Google revealed that about 32% of users provided false answers to their account recovery questions for privacy reasons [55]. They believed that falsifying the answers would ensure that their private information such as their date of birth was not revealed to others during the account recovery process. In another example, in early 2022, IRS had to abandon a facial recognition user verification system administered by a private company after users, privacy activists and congress criticized the agency for allowing a private third party company to collect private biometric information from users seeking agency services as part of the verification step [139].

Account Recovery feedback messages to users can potentially leak the account holder's private information. An example of this is the hints that some Account Recovery implementations provide to users. These include free text hints that the system lets users set to as reminders to answers for their secret questions or forgotten passwords. Moreover, some systems provide feedback to the user indicating a partial email address or phone number where the reset password is either emailed to or sent to. Some Account Recovery implementations show the last time a user successfully performed an account recovery. Whereas this feedback information might be helpful to the user, it can also potentially leak private information of the account holder.

Chapter 2:

Account Recovery - Current State

2.1 Why is the need for Account Recovery increasing?

2.1.1 Remote Work Style

COVID-19 directly contributed to an explosion of Remote Work style as described by a TIME report [66] and Pew Research [50]. Remote work requires a user to use several applications such as collaboration tools applications, remote time logging tools, and remote file sharing tools, among others. All these tools require some authentication which oftentimes means that the user will be required to maintain a set of different authentication credentials. A Dashlane report [53] supported this viewpoint through their survey of users and organizations, which showed that increased usage of password managers was the top change that organizations made because of remote work. The more credentials they must have, the more likely they will need to recover one or more.

2.1.2 Increased online activity

A study analyzing authentication-related log data from a large US-based company covering a period of one year (see Chapter 3) demonstrated that Account Recovery related issues increased with an increasing number of users active online.

2.1.3 Adoption of Passwordless

Adoption of FIDO credentials which usually means that passwords are not commonly used becomes a problem when one of those FIDO credentials must be recovered and the fallback method, which typically is a password must be relied upon. But in most cases, when FIDO credentials are used, users tend to forget their passwords simply because they rarely use them [63].

2.1.4 A variety of Authentication options

Single Factor Authentication using a secret such as a password is still the most prevalent means used for user authentication today even though there is a great deal of research demonstrating the weaknesses of this approach both from a security perspective and from a usability perspective [4]. Much published research has shown that adding a second factor during authentication tremendously mitigates the weaknesses evident when just a single factor is used [13]. 2FA raises the bar for the attacker since they not only have to compromise the victim's password, but they also must compromise the second factor used, such as the user's mobile device. Verizon Data Breach report approximates the number of security breaches involving a compromised password at 61% [21]. Microsoft reported that 99.9% of breaches are prevented by the introduction of 2FA [10]. In 2021 Google reported a 50% decrease in compromised within a few months after auto-enrolling more than 150 million users in 2FA [22]. However, an increase in authentication factors also implies an increase in methods that will need to be recovered. In Section 3.7, we demonstrate results from a study that highlighted this point. The study showed that account recovery reported issues increased with the increasing adoption of additional authentication factors.

2.2 Vulnerabilities and Weaknesses in current Account Recovery

Account Takeover happens when an adversary modifies a victim's account credentials such as something a victim knows or possesses to what an adversary knows or is in possession of. This results into the victim losing all access to their account and the attacker taking over full control of that account.

2.2.1 Account Takeover through MFA Bypass

Knowledge Based Authentication (KBA)

Security questions is the most common knowledge-based Account Recovery method but also it is the least secure. Most security questions are derived from personal data such as personally identifiable information (PII) such as one's mother's maiden name, one's father's middle name, one's date of birth, one's social security number, etc. These questions can also be contextualized to one's interaction with the system such as information about the bill that the user paid or items that a user purchased at a given time. One challenge with KBA's is that typically the information is something that is known by more than just the user. Usually, the user's family members or close friends will also know the answers to these questions. Moreover, users typically tend to expose this information on social media either intentionally or unintentionally, making it a major source of information harvesting by adversaries. The other challenge is that storage and validation of this information are not usually done securely. For example, a support staff that validates this information usually end up knowing the answers to the questions. This is unlike passwords which are typically stored as one-way hashes and are

usually changed upon a rest activity. Security questions based on personal information do not change even after the user has provided the answers to through a support staff.

SIM Swapping

In this attack, the adversary convinces the victim's service provider to transfer the target phone number to the attacker's SIM card. The result is the SMS One Time code indented for the victim being received by the attacker thus bypassing the MFA challenge.

Stolen One-Time Passcodes

This attack usually piggybacks on phishing attacks where the adversary lures the victim to a spoofed website. From there they can steal the victim's knowledge factors and use those to initiate an OTP request.

Phishing Emails

Phishing attacks involve tricking the victim into clicking on a link or downloading an attachment in an email. A click on the link usually presents the victim with a fake login experience that looks normal that captures the user's credentials or authentication codes. The attacker can then use those captured real user credentials to recover a user's account and modify it accordingly.

Forged recognized devices

To reduce friction for the end user, the application will not prompt for MFA on devices where the user has successfully logged on before. An adversary can use this as an attack vector by figuring out how the application remembers the device and thus forge the signature used for this remembrance. For example, if a particular cookie is used, an attacker can forge that cookie's value and add it to the user's request.

Session Hijacking:

This attack involves an attacker taking over a victim's web session without authenticating by stealing session cookies. This can be accomplished through tricking the user into clicking an attacker generated malicious link with a prepared session id. It can also be accomplished via pre-installed malware on the victim's device or through a man-in-the-middle (MiTM) attack. With the stolen session cookie, the attacker can access the user's account, modify recovery options, or even set up a new recovery option.

2.3 Account Recovery

The underlying threat model primarily drives the choice of an appropriate Account Recovery method. There are usually four concerns to be addressed; Ability for a user to continue access, the privacy of the user, security protection of the account and the cost.

2.3.1 Access Continuity, Security and Privacy

The risk underlying the service behind which the account is being used for might be low enough that both security and privacy concerns are of low importance to the ability for the user to maintain account continuity. An example of this might be a system that is open to the public for informational purposes only and thus an Account Recovery process for such a site will not require stringent user verification and authentication as part of the process. On the other hand, if the underlying risk is high enough, then the security and privacy requirements will outweigh the account continuity requirements thus delaying or discontinuing the access until the minimum security and privacy requirements are met. Google explains in this article [133] that they delay one's account recovery to notify the user that an account recovery attempt has been made on their account. This gives the user an opportunity to deny that request if an adversary initiated it thus securing the account.

2.3.2 Cost

The cost for implementing a particular Account Recovery mechanism directly influences the user's choice of that method. For example, some roaming authenticators such as hardware security keys come at a cost. Whereas platform authenticators such as mobile devices or laptops that the user already owns will not incur an additional cost

Chapter 3:

Rethinking Account Recovery

There has been previous research on account recovery and an enumeration of the weaknesses in account recovery [48]. The suggested mitigations against these weaknesses include several ways to strengthen the mechanism in which the fallback credential is transmitted to the user and the implementation of the recovery process. These include using multiple security questions, avoiding weak and easily guessable answers to the questions, ensuring all input supplied by the user to the recovery mechanism is validated and filtered, assigning a new temporary password to the user instead of revealing the original password, throttling on the number of incorrect answers to security questions, and providing the user with backup security keys or authentication codes.

The one commonality to all these mechanisms is that they rely heavily on the user complying with security policies or taking an action such as enrolling in an additional account recovery option or setting up security questions that are not guessable etc. Whereas this might work well in settings where users are required to enroll in multiple authentication mechanisms and security policies calling for that are enforced, it does not work well for most other cases where users are not forced to enroll or set up a recovery method. In addition, this will not work well for users that might not have devices capable of taking advantage of strong authentication schemes such as FIDO based methods. Moreover, past research has shown that a considerable number of breaches are caused by users who do not comply with security policies or circumvent these policies [124]. Given this heavy dependency on the human factor, we rethink Account

Recovery by studying how to make it more user-driven and more subjective to users' different preferences as dictated by varying user demographics. As defined earlier, Account Recovery requires use of an alternative authentication mechanism when the nominal authentication method is not available. As such, a user with multiple authentication methods will have that alternative mechanism to use for authentication and thus access their account and recover or re-establish the primary method. So, the question is "How do we ensure that users are voluntarily enrolling in multiple strong authentication factors for those services that make that option available"?

Unfortunately, voluntary 2FA adoption remains exceptionally low [9]. In 2021 Twitter reported that only 2.5% of their active users voluntarily signed up for 2FA [16]. In 2018, Google reported about 10% [17]. Other studies have shown that adoption remains low [7, 14, 15]. Besides the adoption challenge, there is also an increasing body of work on the 2FA usability issues [1, 2, 6, 7, 12, 13]. Our research corroborates these studies both on adoption as well as on usability. However, we contribute an entirely new perspective by focusing on the senior citizens user demographic. That is, US-based users aged sixty-five and above. Several studies have looked at adoption rates of other types of users such as university students, faculty, and staff [1, 15, 20] or a general population of users [16, 27] and a limited number of users [5]. This is the first study to our knowledge that has focused on 2FA adoption rates for older adults at scale in a real-world setting. The Pew Research Center's report on technology use among seniors [19] shows that over the last decade, the percentage of senior citizens using online services has almost tripled. This demographic of users represents a growing segment of internet users today. It is therefore imperative to understand any user perceptions towards technology adoption (in context of 2FA) that are specific to this segment of users. We demonstrate that older users' adoption rate is on average lower than that of younger users.

We seek to answer the research questions below with a twofold goal; First, to understand plausible reasons behind the low 2FA adoption rate among older adults and secondly, to provide suggestions that will help drive up the rate of 2FA adoption voluntarily and thus provide users that strong alternative Account recovery option.

a. **RQ1**: Is there a correlation between 2FA adoption and system usability?
b. **RQ2**: Were there any evident patterns between users' device types and the 2FA method adopted by the user?
c. **RQ3**: What associations can be gleaned between the login frequency and voluntary 2FA adoption?
d. **RQ4**: Were there any noticeable user behavior patterns right after they were notified of the upcoming changes requiring use of 2FA?

We addressed the above research questions by performing a detailed analysis of the authentication logs and help desk incidents reports. The log data contained authentication related transactions spanning a period of one year. A log transaction contained a user agent, a user type, 2FA enrolment status, user preference, a session id, a timestamp, and the type of transaction e.g., sign-in, logout, re-authentication etc. The help desk data contained the total number of authentication related incidents reported in each period.

A sure way to ensure adoption of 2FA is by making it mandatory. Our research corroborates this viewpoint. We demonstrate that the same users despite being aware of 2FA availability, did not enroll in it until when it was made mandatory. Only a small percentage of users opted to enroll voluntarily prior to the enforcement period. Several businesses, educational institutions and Government agencies have started mandating use of 2FA for their websites [17]. In 2021, the US Government issued an executive order requiring government agencies to adopt

use of 2FA [11]. Starting November of 2021, Google made it a requirement for approximately two million YouTube creator's accounts to have 2FA and auto enabled it by default for more than 150 million users [22]. Such steps are good measures in the right direction. However, this is not always possible or practical. Several businesses prioritize less friction for their users and view adding 2FA as a way disrupting their user's experience. Often, these businesses implement 2FA as an optional feature and leave it up to their users to adopt it should they choose to. Thus, the challenge of finding a way to encourage more users to voluntarily adopt protective behavior such as multiple authentication factors adoption is still an open research problem. Our research findings contribute towards solving this challenge.

We use an IRB approved log data with more than five hundred million records from a large US-based company to perform several empirical measurements towards answering this question. The log data covers authentication related events such as user registration, login methods used, device type and help desk reported incidents for hundreds of millions of unique authentication transactions. All the log data resided on the company's secured devices and within the company's network. No user personally identifying information was used as part of the analysis. The work was approved by both the company providing the data and the Columbia University IRB. The university IRB protocol number# IRB-AAAT9070 (Y01M00)

3.1 Methodology

This section covers a detailed description of the approach followed in studying the research questions below.

- **RQ1**: Is there a correlation between 2FA adoption and system usability?

This research question was aimed at determining the impact of 2FA adoption on the general usability of the authentication process. The International Organization for Standardization (ISO): ISO 9241-11 [18] defines usability as a measurement of the efficiency, effectiveness, and satisfaction of a product from the consumer's perspective. We used helpdesk reported authentication incident counts and performed an analysis of the numbers against 2FA enrollment and usage to quantify a usability score. For efficiency, we looked at the time it took for an authentication transaction to complete from start to end. For effectiveness, we looked at the ratio of failed logon sessions to successful sessions. Satisfaction was a little bit hard to quantify. We leveraged helpdesk call in numbers for login issues per number of login events. The higher rate indicates usability issues, which we directly mapped to a satisfaction score. This question was to address the concern some companies have about 2FA introducing friction on the user's authentication process thus contributing to their reluctance to enforce 2FA. We therefore wanted to understand to what extent that was true for older users when comparing measurement values before 2FA enforcement to after MFA enforcement.

- **RQ2**: Were there any evident patterns between users' device types and the 2FA method adopted by the user?

This research question was aimed at looking at the correlation between device type and the type of 2FA that a user selected. The goal for this was to determine if there was a strong correlation between the type of device preferred by the user with the type of 2FA that the user chose. Having this insight helps towards tailoring a notification regarding 2FA adoption to different users based on their preferred device types. Our log data contained user agent information from which we were able to determine the device type that the user was using. The 2FA type chosen by the user was also available in the log data. Combining these two pieces of information was sufficient to address this question.

- **RQ3**: What associations can be gleaned between the login frequency and voluntary 2FA adoption?

This research question was aimed at determining if there was a correlation between the frequency of logins and the total time spent within the website for a given session with the rate of 2FA adoption. The hypothesis was that users that logged in more frequently would also have a higher adoption rate. Additionally, users that spent more time per session would have a higher adoption rate. The login frequency was determined by counting the number of login sessions per user in each period. The total time spent using the web resource was determined by measuring the length of a unique session per user in each period.

- **RQ4**: Were there any noticeable user behavior patterns right after notification of the upcoming changes requiring use of 2FA?

The goal of this research question was to uncover any noticeable user behavior patterns within a three-month period between 2FA enforcement notification and just before the actual enforcement. These user behavior patterns provide insights into older users' mental models on 2FA or perceptions of 2FA. An understanding of a user's perceptions of 2FA is foundational to understanding the most effective way to nudge them towards voluntary adoption. User preferences within that 3-month period were analyzed and compared to other time periods to see any patterns that were distinct within this window.

Data Cleaning and Analysis

Since our analysis did not require any personally identifiable information, we redacted this information before running analytical processes on the data. This redaction was accomplished by filtering out fields containing the PII such as first name, last name, email address and username. Secondly, we removed any data that was not related to our study such as transactions within the application after the authentication process. Each log entry contained a source attribute corresponding to the module within the system where the log data was generated. We used this to define the sources of interest corresponding only to the authentication modules. Lastly, using a timestamp attribute that was available with each log entry, we restricted the data to cover just the period of interest. Any incomplete log entries were also discarded. A complete entry was one that had at the minimum all the three main attributes that we were interested in, i.e., a transaction id, a session id, and a timestamp.

This pre-processing resulted in 510 million records covering a period of 365 days. For determining associations between the different variable measurements, we used the Pearson Correlation Coefficient r (PCC) [43,44]. This value measures the strength of the linear relationship between a pair of variables. The value ranges from -1 to 1 with -1 being a perfect negative linear correlation, zero indicating no linear relationship and one indicating a perfect positive linear correlation. The analysis conducted in this work resulted in establishing the strength of linear associations between pairs of variables as well as produced different patterns and trends observed between these variables. This information was used to answer our research questions outlined above.

Limitations:

The log data did not contain detailed demographic information such as income levels, gender, and race of the users. Moreover, we did not have a detailed breakdown of the users' ages other than identifying whether they belonged to the 65+ years older group or younger than 65 years. Thus, we could not perform a more in-depth analysis based on these demographics. We did not survey users to get a better understanding of some of the findings in the log. For example, to understand from the user's perspective, the driver for some choices they made such as why they chose to use desktop computers, why they opted for certain 2FA options and some of the reasons behind the help desk calls that they made after enrolling in 2FA. In terms of 2FA options, our study was limited to Email and SMS as the options that were available to these users for receiving the one-time code used to fulfil a 2FA challenge.

Despite these limitations, our study contributes valuable insights towards voluntary Multiple Authentication Factors adoption considering the scale of the log data analyzed and the number of users covered under the study.

3.2 Voluntary Enrollments

Voluntary 2FA Adoption for the adult users was lower than that for the younger users. During the months of April, May and June, a total of 243,849 younger users voluntarily enrolled for 2FA compared to 84,865 older users. (Ratio: 0.35).

Table 3.0.1: Comparison of voluntary 2FA enrollments between older users (65+ years old). April, May, and June represent a period when 2FA enrollment notifications were sent out to users encouraging voluntary enrollment

Month	Older Users	Younger Users
June	51,642	143,450
May	30,132	91,309
April	3091	9090
December	1061	3791
March	439	1908
February	377	1018
January	333	950

The months of December through May accounted for 7,667 younger users compared to 2,210 older adults that voluntarily enrolled for 2FA (Ratio: 0.29). The main difference between these two periods of time was that April through June saw an active campaign notifying users about availability of 2FA and encouraging them to enroll for it while December through March, users were not actively notified to register for it. This accounts for the lower 2FA voluntary enrollment numbers during this period.

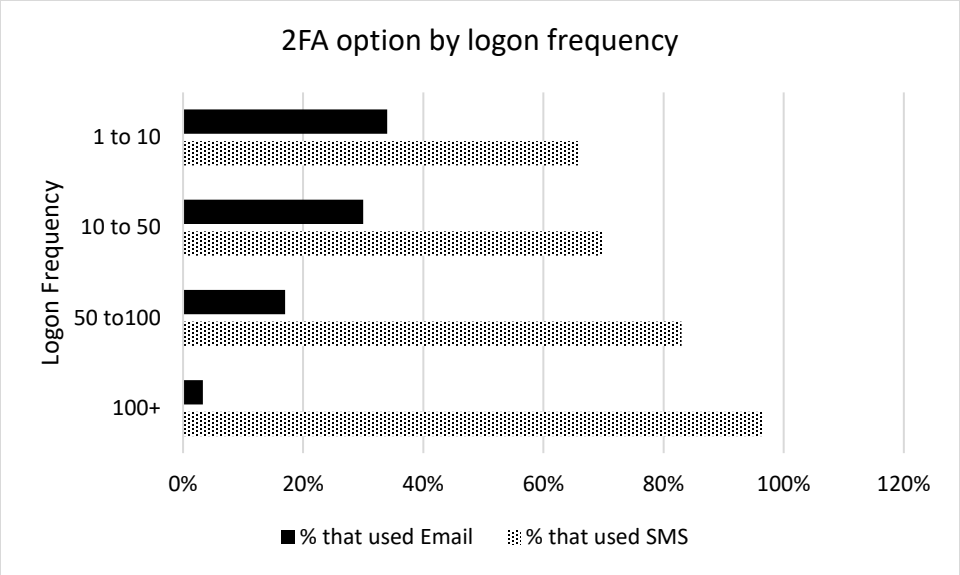

Figure 3.0.1: 2FA option by logon frequency, users that logged on at least one hundred times preferred SMS (97%) to email (3%) compared to those that logged on less than 10 times with 66% that preferred SMS to 34% for email

Our study shows a lower 2FA adoption among seniors and is consistent with prior literature [30, 31] findings on technology adoption rates. Among the reasons cited in these studies include a declined cognitive load as well as reduced physical abilities [19] for older users. A Pew Research study [19] reported that only 26% of senior citizens said that they felt very confident when using electronic devices online. 35% of these users said that they were a little bit confident (23%) or not confident at all (11%) using electronic devices which included computers, smartphones, and tablets in an online setting.

Table 3.0.2: Logon frequency per year showing that users that logged on most frequently (100+ in one year) chose SMS (97%) over email (3%) as a 2FA option.

Logon Frequency	Unique SMS Logons	Unique Email Logons	% Using SMS	% Using Email
100+	69912	2089	97%	3%
50 to 100	235395	48214	83%	17%
10 to 50	469101	201120	70%	30%
1 to 10	697801	358961	66%	34%

3.3 Logon Time

Table 3.0.3 shows the average logon time for different authentication activities and factors. For this study, the logon time represented the time it took for a user to initiate a logon session until when they completed that process and got redirected to the immediate resource following logon activity. Each logon session has a unique session id per user that was used for tracking the total time per user. The average is computed over a period of 365 days for all the users that logged on. In general, logon time increased (an average of 118 seconds) with 2FA adoption but decreased for sessions initiated with device remembrance enabled on the device in use (an average of 43 seconds). Email as a 2FA option took the most time (139 seconds) while SMS took 118 seconds on average. Sessions initiated before 2FA was enforced took an average of 96 seconds.

Table 3.0.3: Average logon time for all active users in the system over a period of one year as measured in seconds. Email as 2FA took the longest to complete (139 sec) With device remembrance the time is 43 sec. The time was measured from the moment the user loaded the sign-in page to when they completed all the required logon steps (submitting username/password/2FA challenge response etc.)

Activity	Average Time (Sec)
Email Login Time	139
2FA Logon Time	129
SMS Logon Time	118
Logon Time (Desktop)	103
General Logon Time	96
Logon Time (Mobile)	57
With Device Remembrance	43

These findings corroborate prior research that showed a general increase in logon time with 2FA adoption and a decrease in time when device remembrance is in effect [60]. From voluntary 2FA adoption perspective, it is imperative to encourage users to enable device remembrance where available as it drastically cuts down the time it takes perform a logon activity by not prompting the user for that second factor if they use the same device set to be remembered by the system. From usability perspective, this boosts the efficiency of logon process thus contributing to improvement of the system usability.

3.4 User Device Type

Users that primarily used a desktop device tended to opt for email as a way of receiving the one-time code for 2FA while most users that opted for mobile device tended to use SMS. This observation was consistent across both demographics of older and younger users. On average, 51% of older adult users used a mobile device compared to 63% of younger users.

AARP [32] report indicates that about 62% of older adults above 65 years old adopted a smartphone. We hypothesize that one of the reasons why there is a higher percentage of older users opting for desktop computers compared to younger users is because of a predominance of assistive technologies available on the desktop computers [40]. Desktop devices have larger screens and ergonomically designed keyboards to accommodate some of the challenges experienced by older generations.

In general, desktop devices can be easily upgraded and thus have a longer lifespan. Mobile devices have some drawbacks when it comes to these technologies, e.g., their smaller screens and limited disk storage space [42]. What this means from 2FA adoption perspective is that options optimized for desktop devices need to be made available by service providers to encourage voluntary adoption by this segment of users that use desktop devices. Figure 3.0.2 shows that most older users using desktop devices opted for email as their 2FA choice when compared to users in the same demographic that opted for mobile devices. On average, 42% of older desktop users chose Email as their 2FA option compared to just 25.3% of younger desktop users. This finding of a preference of email over SMS aligns with a Gallop study [41] that reported most adult Americans preferred communicating through phone calls (cellphone or landline) followed by email ahead of text messages. Within this demographic, they reported about 16% used email compared to 6% that read or sent a text message. The strong correlation between older desktop users and their preference of email as a 2FA option coupled with the correlation between mobile device users with their preference of SMS as the 2FA choice reinforces our conjecture that the type of device influenced the choice of 2FA option chosen by the users.

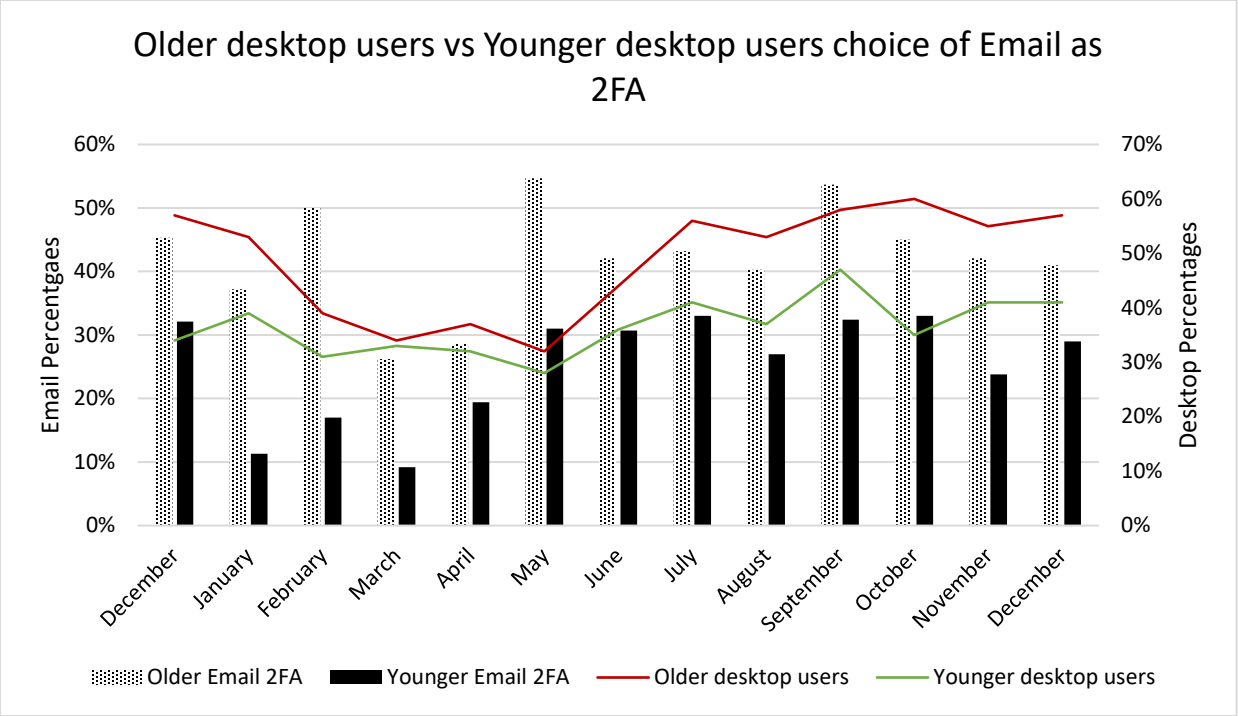

Figure 3.0.2: A comparison of the older desktop users to younger desktop users' choice of email as the option for receiving one time code used as part of 2FA. Throughout the year, a larger of percentage of desktop users were older users. Moreover, percentagewise more older users used email as a 2FA option compared to younger users

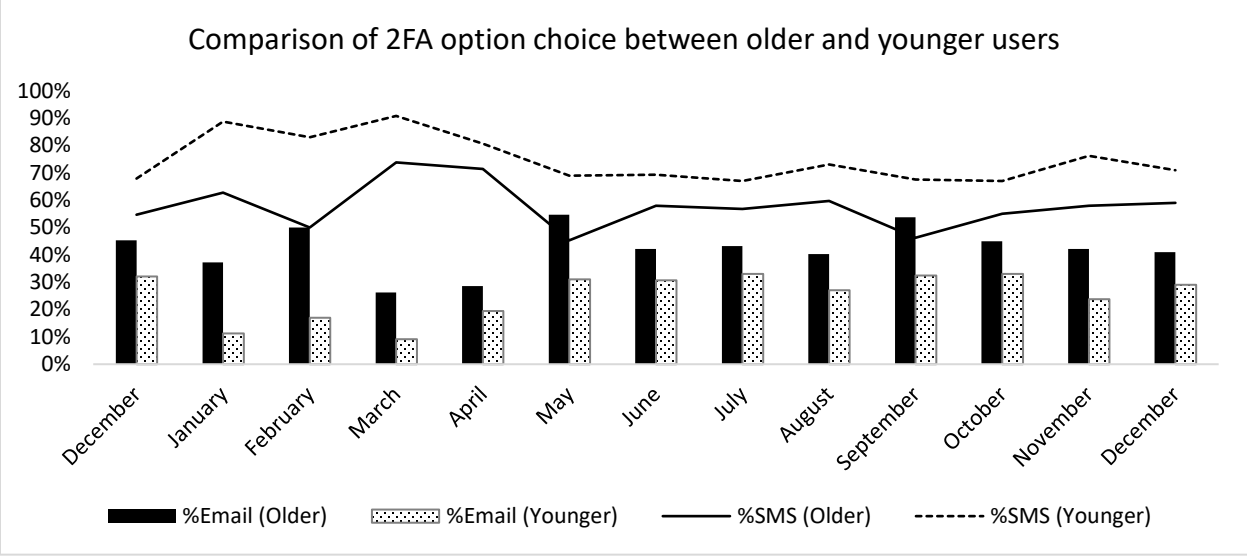

Figure 3.0.3: A comparison of device type between older users and younger users. On average, 63% of younger users used a mobile device compared to 51% of senior citizens while only 37% of younger users used a desktop device, compared to 49% of senior citizens

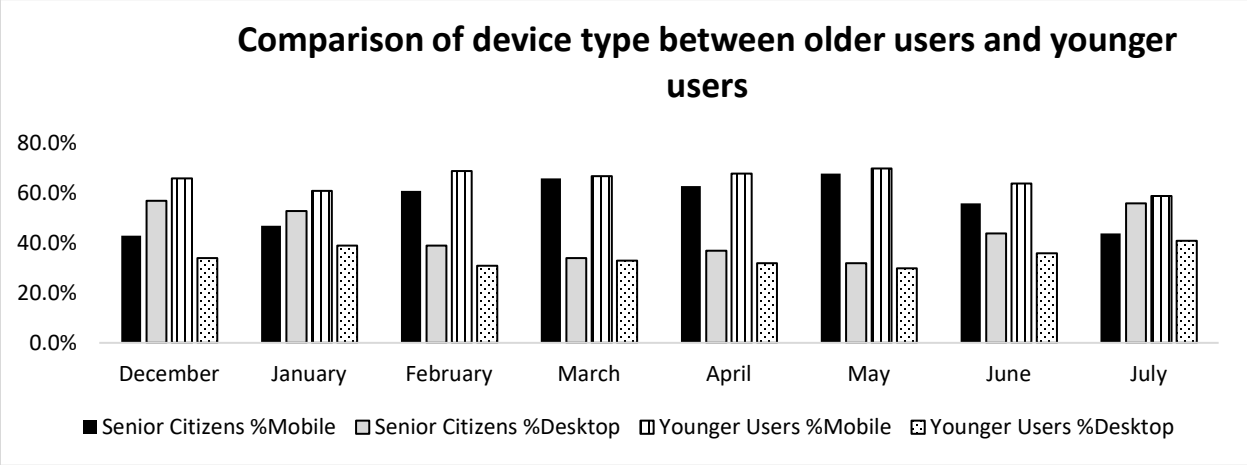

Figure 3.0.4: A comparison of 2FA option choice between younger users and older users. In general, a larger percentage of younger users opted for SMS compared to the percentage of older users that opted for the same. On the contrary, a larger percentage of older users opted for email as a 2FA option compared to the percentage of younger users accessing the same system during the same period

3.5 Logon Frequency

The logon frequency measurements in Table 3.0.1 and Figure 3.0.1 below show that users that logged on more frequently tended to use SMS as the 2FA option. 97% of all users that logged on at least one hundred times in one year had SMS set as their 2FA option compared to just 66% for those that logged in less than ten times during the same period.

The largest percentage of users (34%) that selected use of email as a 2FA option were those that logged in less than 10 times in a year (compared to just 3% for those users that logged in more than one hundred times a year). It is not surprising that SMS is the most used method and even more so for those users that frequently sign in. The 2021 Pew Mobile Fact Sheet report

[39] indicates that 92% of senior citizens use a cellphone and 61% use a smartphone. About 30% can only receive SMS messages or voice calls (i.e., cellphones that are not smartphones).

Table 3.0.4: Logon frequency per year showing that users that logged on most frequently (100+ in one year) chose SMS (97%) over email

Logon Frequency	Unique SMS Logons	Unique Email Logons	% Using SMS	% Using Email
100+	69912	2089	97%	3%
50 to 100	235395	48214	83%	17%
10 to 50	469101	201120	70%	30%
1 to 10	697801	358961	66%	34%

Besides the fact that SMS is the most common method, we believe that the other reason users that frequently signed-in chose SMS was because it is easier to enroll in and use. For those users that logged in multiple times in a year, they are likely to have opted for what they considered as an easier option to use. This viewpoint is supported by [60] in which the researchers demonstrated that besides pre-generated codes and push notification; SMS was the fasted 2FA method to set up. Our results show a strong correlation between login frequency and the type of 2FA option chosen by the user.

3.6 Noticeable User Behavior Patterns

There was a larger spike in the number of older adults opting for paper-based communication preference when notifications were sent out alerting them of the coming 2FA enforcement requirement after July of that year. These notifications started going out in April and that month the percentage of senior citizens that opted for paper-based communications jumped from 37% to 65%. This contrasts to younger users where the percentage went from

35.2% to 42%. There are several plausible reasons for this spike seen in older adults as compared to younger users. One conjecture is that the users opting for paper-based communication channel wanted a fall back means of receiving the service were they to get locked out of the online channel after the 2FA enforcement period.

A fear of account lockout was reported by authors in this study [26] as one of the barriers towards adoption of a new way of authenticating to a web resource. Majority of users that opted for paper-based communication preference did not sign up for 2FA until when it was made mandatory for them. Moreover, the lowest voluntary 2FA adoption rate among older users was for those that opted for a paper-based communication channel. A plausible explanation is that senior citizen users simply did not use the technology as much as younger users did.
This infrequent use of technology contributed to these older users seeing lower 2FA enrollment numbers when compared to the younger users. This viewpoint aligns with findings from several studies that demonstrated that older adults used technology less frequently when compared to younger users [33, 34, 35, 36].

Table 3.0.5: Pearson Correlation Coefficients r (PCC) for voluntary 2FA enrollment against paper-based communication preference, online communication preference, percent of seniors and younger users that opted for paper-based communication. There was a high positive correlation (r=0.863) between the percentage of seniors that opted for paper-based communication preference and 2FA enrollment, compared to that of younger users (r=0.304). This implies that during this period of 2FA enrollment, more seniors opted to use paper-based communication (i.e., both variables positively increased, with seniors increasing much more than was seen in younger users)

	2FA enrollments	Online preference	Paper-based Preference	% Seniors' paper-based	% Younger users' paper-based
2FA enrollments	1				
Online Preference	0.9972194	1			
Paper-based Preference	0.9991193	0.9932143	1		
% Seniors' paper-based	0.8634699	0.8758532	0.854385275	1	
% Younger users' paper-based	0.3041515	0.3232858	0.292633928	0.656115889	1

Table 3.0.1 shows the voluntary 2FA registrations between the months of December and June. During this period, an average of 34% of users that preferred a paper-based communication enrolled for 2FA compared to 66% that preferred only online-based communication. This points to a stronger association between a preference for paper-based communication channel and a lower adoption of 2FA. Table 3.0.5 shows a strong Pearson Correlation Coefficient (r=0.99) between paper-based communication preference and 2FA enrollment. This means that as 2FA enrollment increased so did paper-based communication preference. This rate of increase was mostly seen on the senior users (r=0.863) as compared to younger users (r=0.304). In general, there was not a high correlation between younger users' communication preferences (paper-based or online based) to the rate of 2FA adoption. This means that unlike the older users, no strong association could be drawn between these younger users' communication preferences and the awareness of the oncoming changes involving 2FA enforcement.

From voluntary adoption perspective, the implication for this is that the messaging targeted at older adults needs to build up their confidence in the system availability and reliability after a technological change. It also needs to address the fact that these older users infrequently access the web resources and thus it should simplify the enrolment process so that it can be completed with minimal steps and within a single session. It should also target a period when most users tend to use the system to maximum coverage.

3.7 User Reported Usability Issues

We analyzed help desk authentication related incidents reported by users during the same period as the log data analyzed in this study. These incidents were divided between generic logon

related incidents and account recovery related incidents. The graph in Figure 3.0.5 shows that the logon failure rate increased with the increasing 2FA adoption, meaning that on a percentage basis, the number of failed logons went up. During the period before 2FA enforcement, the rate of failed logons was an average of 1.6% compared to the period after 2FA enforcement (2.88%). This rate was calculated as the percentage of failed logon attempts out of all logons attempts per month. From the available log data, we could not determine all the causes of failed logon attempts but this involved either incorrect user credentials submitted, or abandoned logon attempts after the second factor prompt. Using this ratio as a measure of effectiveness, we can conclude that the effectiveness of the overall logon process diminished slightly thus contributing negatively towards overall system usability.

We also calculated the ratio representing the number of reported issues per month compared to the number of total logons in the same period. In this context, reported issues did not necessarily mean that the user was unable to logon, it was any issue reported regardless of the outcome of the logon. This could have included users reporting difficulty in logging in or a logon session taking longer than they anticipated etc. We loosely used this number of reported issues as a measure of satisfaction with the system. The rational being that users reporting an issue with the logon process were not satisfied with the logon experience in one way or another.

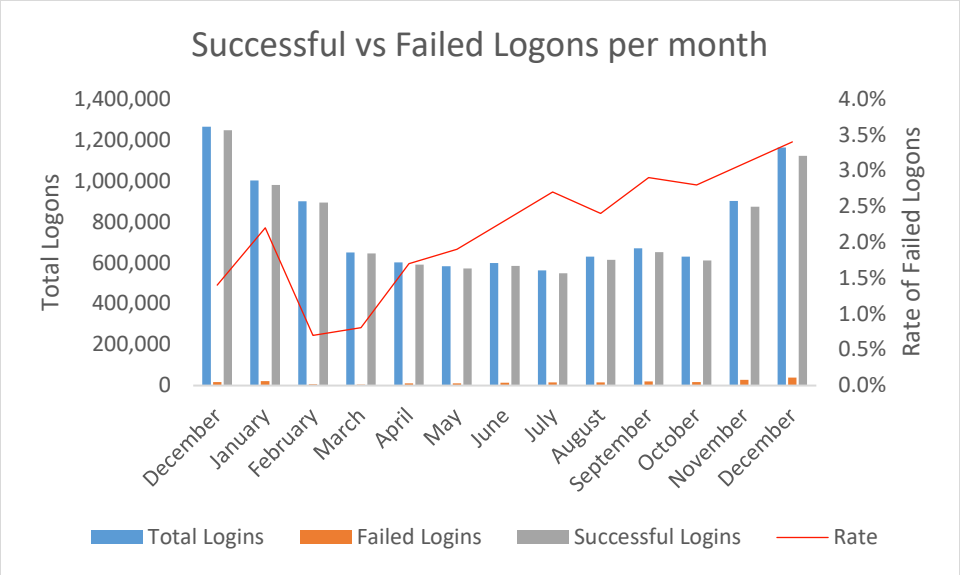

Figure 3.0.5: The rate of failed logons over a period of 1 year covering the period of voluntary 2FA enrollment as well as the period during mandatory 2FA enrollment. The failure rate was highest during the period of mandatory 2FA enrollment. The rate was lowest in February matching the period of least 2FA enrollment.

The graph in Figure 3.0.6 shows that between July and August, the rate of reported issues was highest (38%) compared to the lowest percentage in February (13%). This is the period when most users enrolled for 2FA. This means that during the period of most 2FA enrollments corresponded with a period of most dissatisfaction with the system and thus a lower usability score as per the ISO 9241-11 definition of a system usability [18].

Table 3.0.6 shows a correlation matrix between 2FA Adoption rate, 2FA account recovery incidents, 2FA login related incidents and the total help desk incidents. The lowest Pearson Correlation Coefficient (PCC) (r= -0.979) was between the percentage of incidents representing 2FA account recovery and the percentage representing general login incidents. This negative coefficient indicates that as one variable increased, the other one decreased at a high rate. It is not surprising that 2FA recovery related incidents will increase as more users adopt 2FA.

Table 3.0.6: A Pearson Correlation Coefficient matrix between the total help desk incidents, 2FA adoption rate, percentage login related incidents and percentage of incidents representing 2FA recovery related. The negative correlation (r= -0.979) between the percentage of incidents representing 2FA account recovery and the percentage of generic login related incidents indicates that as one increased, the other decreased. Conversely, the positive correlation (r=0.882) between 2FA adoption and the total helpdesk incidents indicates that they both increased positively.

	Total Helpdesk incidents	2FA Adoption rate	% Login incidents	% 2FA Recovery incidents
Total Helpdesk incidents	1			
2FA Adoption rate	0.882384022	1		
% Login incidents	-0.706106496	-0.782172379	1	
% 2FA Recovery incidents	0.644256946	0.720497216	-0.979694131	1

The importance of understanding this association between these 2 types of incidents is that the support staff needs to be well trained and equipped to handle more account recovery incidents over time as those tend to increase over time as users adopt 2FA compared to general logon related issues e.g., how to login using a second factor or how to enroll a second factor. It also underscores the importance of training users on self-remediation steps especially for those systems that enable use of more than one option for 2FA.

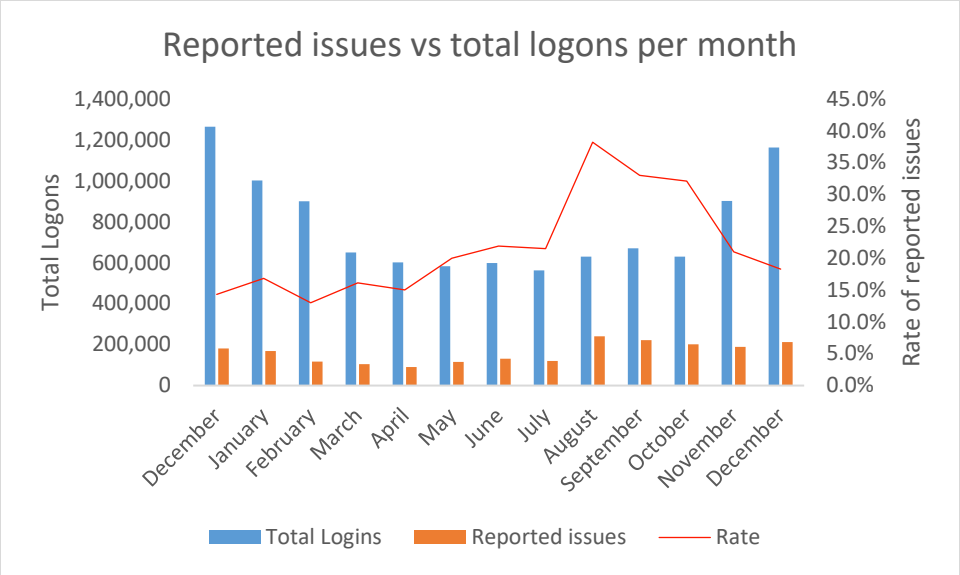

Figure 3.0.6: Reported issues versus total logons per month showing that there was a spike in the reported issues between July and August. This corresponds to the period when most users enrolled for MFA

Users become better at using the system over time, but the rate of account recovery related incidents does not necessarily reduce, that rate increases with time. The highest PCC (r= 0.882) was between 2FA adoption rate and the total help desk incidents. This means that the rate of increase of 2FA adoption closely matched that rate of increase in help desk incidents. This association between the 2FA adoption rate and the total number of help desk incidents was illustrated further by Figure 3.0.8 which shows an increase in the total number of help desk incidents with the increased 2FA adoption. This finding corroborates previous studies that showed an increase in helpdesk incidents with increasing 2FA adoption [138]. There is a higher positive correlation between the 2FA adoption rate, and 2FA account recovery incidents (r=0.816) compared to 2FA adoption rate and 2FA login related incidents (r=0.631).

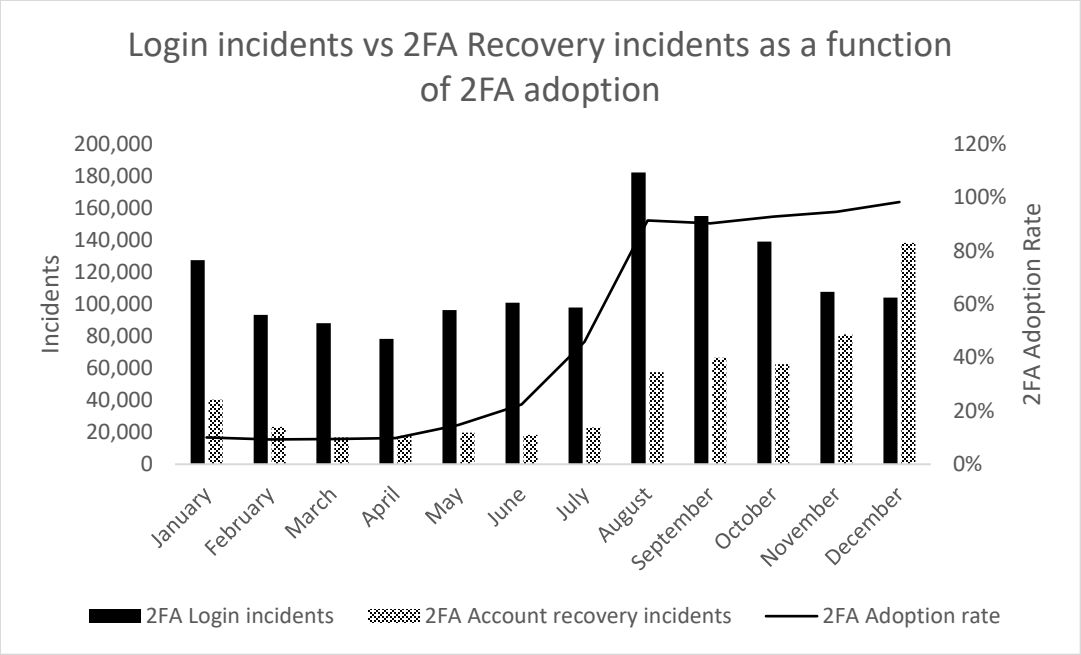

Figure 3.0.7: Generic logon related incidents compared to 2FA recovery incidents as a function of 2FA adoption. 2FA recovery issues increased over time while generic logon related issues tended downwards.

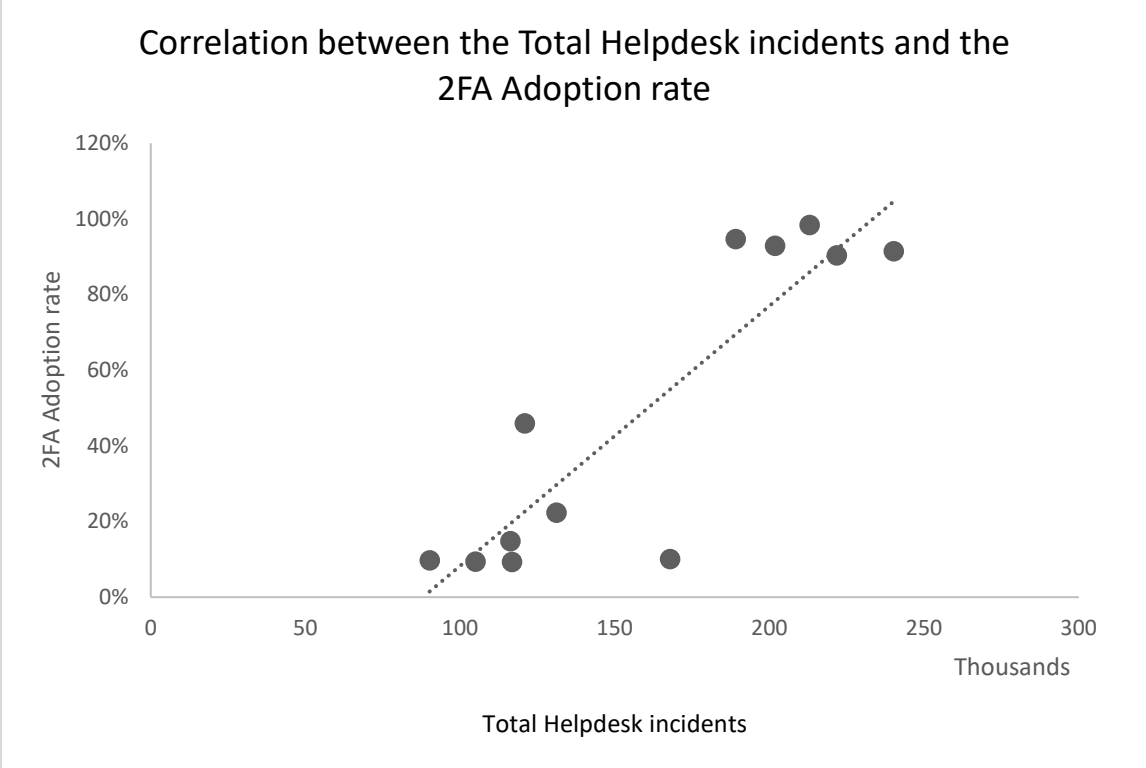

Figure 3.0.8: Illustration of a strong Pearson Correlation between total help desk incidents and 2FA adoption rate increase

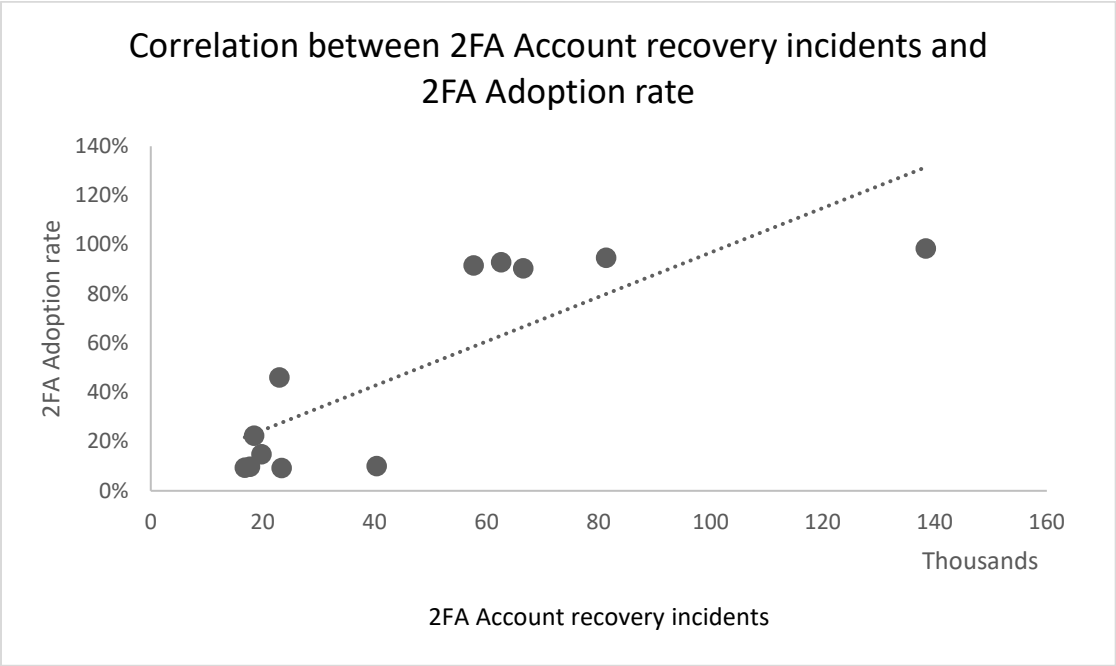

Figure 3.0.9: A positive Pearson Correlation between 2FA account recovery incidents and an increasing 2FA adoption

Intuitively, Account Recovery related incidents will increase with increasing 2FA adoption. However, this rate of increase is not uniform across the period. This is illustrated in Figure 3.0.10 below which shows that the largest percentage increase was observed in the month of December (65% of authentication related incidents were related to account recovery events compared to just 35% related to logon events) This contrasts to average of 17% for the months of March through July of the same year.

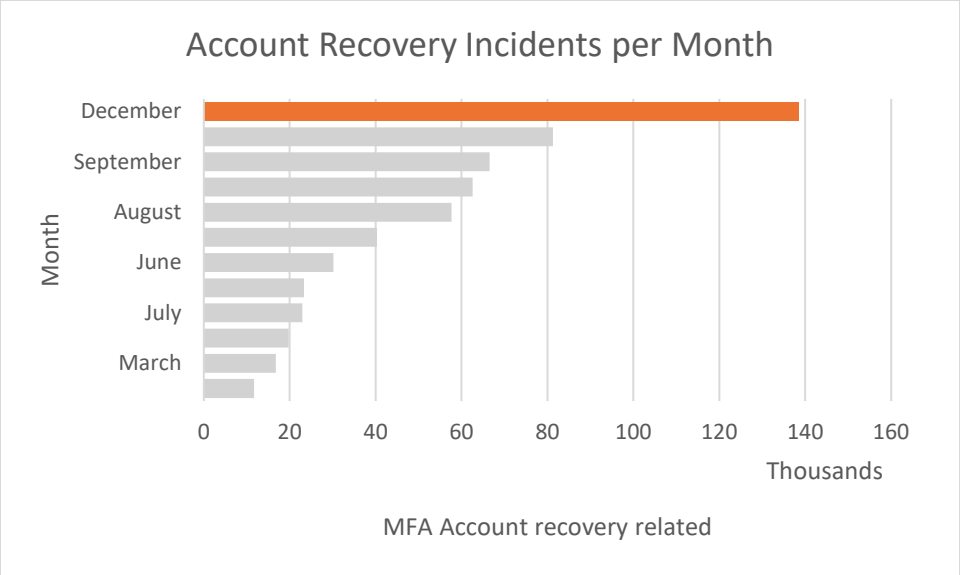

Figure 3.0.10: 2FA account recovery related incidents per month. December had the highest number of such incidents.

The considerable number of recovery related incidents in the months of November and December can be attributed to the likelihood of users getting new devices during the holiday season thus needing to update their 2FA phone number or devices. Moreover, there is an increased usage of the system during this time of the year as users complete the end of the year activities. This implies that the best time to target users for voluntary 2FA adoption is during this time when more users are likely to use the system as illustrated in Figures 3.0.5 and 3.0.6 that shows an increased total system logon during this period.

Conclusion

This chapter provides a new perspective and insights to plausible factors that influence user's voluntary adoption of protective cyber behavior. We use an IRB approved log data with millions of real-world authentication events from a large US-based company to perform several empirical measurements of different events as part of this study. The log data covered authentication transactions spanning a period of one year which totaled to more than five hundred million records. Using Pearson Correlation Coefficient (PCC), we showed associations between 2FA adoption and different variables such as helpdesk incident categories, user communication preferences and user demographics. We saw a high negative PCC (r= -0.979) between account recovery incidents and generic logon incidents which meant that with the adoption of 2FA and continued system usage, general logon incidents went down but account recovery related incidents increased. A high positive PCC between total helpdesk incidents and 2FA adoption meant that incidents increased as 2FA adoption went up. We saw a decrease in system efficiency (overall logon time went up by an average of 34%), a decrease in satisfaction as measured by the number of issues reported to the helpdesk during this period and a decrease in effectiveness as measured by an increase in the rate of failed logon attempts (from 1.6% to 2.9%). A combination of these three measures meant that the overall system usability decreased during this 2FA enablement period. We found a pattern of increased paper-based communications opt-in corresponding with a period just prior to a 2FA enforcement which indicated that users wanted an alternative means of accessing information in an event of a lockout from their online accounts. We also found that older desktop users that opted for paper-based communication preference also preferred use of Email as a 2FA option more than when

compared to younger desktop users accessing the same system during the same period. We corroborated results from previous studies confirming that voluntary 2FA adoption is still extremely low and that there is a general increase in helpdesk tickets with an increase in 2FA adoption. We conclude that driving 2FA adoption for older users requires a different approach that assures them of system availability and usable Account Recovery options. The most effective time to target voluntary 2FA enrollment for older adults is during the period when they use the system the most which in this study was towards the end of the year. Lastly, the correlations and patterns uncovered during this research between 2FA adoption and other related variables provide a rich foundation for guiding future work into understanding the causation behind the observed associations. These results also provide useful insights towards the planning and executing of a successful 2FA rollout strategy that can lead to a higher voluntary 2FA adoption outcome. Voluntary adoption of multiple authentication factors means that the user has stronger fallback authentication methods enrolled that they can use in an event that their primary authentication method is unavailable for their use.

Chapter 4:

Resilient User Behavioral Modeling for Account Recovery

4.1 Introduction

Account Recovery process involves a means to provide a user with an alternative authentication mechanism that enables them to regain their nominal authentication credentials. To avoid a legitimate user's account from being taken over by an imposter, it is critical that this process incorporates holistic protection from adversaries both proactively and reactively. In Chapter 3, we studied proactive protection mechanisms that investigated ways of getting users to voluntarily adopt stronger alternative authentication methods that they can use as secondary authentication during account recovery process. But what happens if an impostor still manages to take over a user's account? This chapter studies the reactive protection mechanism based on Machine Learning-based user behavioral modeling for active authentication of the user that can happen during and right after the Account Recovery process. Active Authentication implemented through user behavior modeling and validation seeks to mitigate some attacks resulting from account takeover by augmenting existing user authentication paradigms that traditionally authenticate a user only once, at the beginning of the session. These traditional modalities, regardless of how many factors are used, only validate the user at one time. Once the session is initiated the user is vulnerable to other attacks such as account take over by a masquerader or a malware on the system. The threat is even more profound in mobile devices which are vulnerable to being stolen or lost and compounded by their being widely used across the globe [108]. If

employed, Active authentication can mitigate these types of attacks by continuously validating the user via behavior analysis [74],[116],[114]. However, an adversary can evade this validation by compromising the machine learning implementation behind the active authentication. Moreover, in this context, a user's trained ML model is synonymous with their identity. We argue that stealing this trained ML model through techniques such as reverse engineering is another form of account takeover since that can be later used by an impostor to access an active authentication protected system conceptually like how an adversary can use a stolen traditional credential to access a username and password protected system. One potential way this can be accomplished is through the concept of model transferability [115]. A user's trained ML model should thus be deemed as confidential information. This study makes a novel contribution towards defending against model stealing by adversaries using principles of randomization in the context of user authentication.

4.2 Machine Learning Vulnerabilities in the context of Account Recovery

Model stealing through extraction involves an attacker obtaining a new model that can produce predictions equivalent to the original model, usually through formulating targeted queries and then using those to query the original model efficiently. ML Model stealing was demonstrated by [97] through the reverse-engineering technique of linear spam classification models more than a decade ago. The authors presented their model extraction attack by making membership queries via API calls. Over that period, several studies have been conducted and demonstrated that model stealing indeed poses a severe risk across numerous domains due to pervasive use of Machine learning in many real-world solutions. The authors in [98] presented

model stealing via prediction APIs using ML-as-a-service implementations such online services of BigML and Amazon Machine Learning as examples. Their research was focused on model extraction attacks where an adversary's goal was to reproduce the model's functionality without prior knowledge of the training data or parameters used in the ML model. To make this happen, they leveraged the confidence scores for class labels to get more accurate attacks. In [82], the researchers demonstrated how to infer useful information from ML classifiers using a meta-classifier designed to exploit other classifiers. In [102], the authors demonstrated stealing the model hyperparameters that are learned by a learner using Amazon Machine Learning platform. Their attacks affect several machine learning algorithms including support vector machines, neural networks and logistic regression. In this attack, the researchers assume that the attacker knows the training dataset and can obtain unknown model parameters using prediction APIs technique [98]. As demonstrated by [98], even protecting models using Secure Multi-Party Computation (SMPC) protocol [101] is not enough for thwarting model extraction attacks. Leveraging a MLaaS platform, the authors show Support Vector Machines (SVMs) and Support Vector Regression Machines (SVRs) based models can be efficiently stolen using a few hundred queries and for an extremely low cost. Model stealing has also been achieved in a setup where the adversary has no knowledge of the target's architecture or training data [85]. More recently, model extraction has been successfully demonstrated on Deep Neural Networks (DNN) [105, 106, 107] and based on features of Recurrent Neural Networks (RNN) [109]. Researchers have demonstrated an even more advanced technique for stealing ML models through side channels [77], [94].

All this research reinforces the fact that model stealing is indeed a realistic threat

especially in cases where the machine learning implementation is used in a critical security function such as access control. A stolen model can then be used to carry out other attacks like mimicry attacks [108] and evasion attacks [102] among others. In Active Authentication context, a user trained model is that user's one form of identity. Thus, a stolen model is indeed another form of that user's identity being stolen. Using a technique like model transferability [119], an extracted model from one user's device can be used to gain access to a different device This realization motivates our investigation of use of randomization at application time to add robustness to the ML implementation.

4.3 Threat Model

In our threat model, we assume an adversary in possession of a victim's device as well as their traditional credentials such as username and password (taken over through an account recovery process). The device is protected by a continuous authentication mechanism implemented by modeling user behavior through machine learning. The adversary's goal is to evade being detected by this system. To make the detection even more challenging, we assume that the adversary has some limited knowledge about the victim's behavior. The adversary starts off by recreating those known actions and observes how the user behavior model through the sensor responds after a given period. They build on those observations to determine the next set of actions to take. After a given period, the adversary uses the observed responses to build a profile that can be used later to mimic the victim. This activity entails model stealing through extraction [32]. The attacker's strategy is aimed at reproduction of predictive behavior that mimics the victim's behavior [43] to be used for fooling the original machine learning model. This setup is synonymous to the active learning [51] scenario where the learner draws unlabeled

samples and sends to an oracle via a query mechanism for a response in an interactive manner. In our case, the adversary is equivalent to the learner and the sensor that is implementing the continuous authentication on a system is the oracle. The adversary fires off certain events in each time epoch and observes responses in form of whether the sensor detects anomalous behavior or not. The goal being to steal (learn) the model, and once learnt (i.e., stolen), use that knowledge to successfully evade the sensor from detecting when the adversary is accessing the system.

Formally, the Adversary $\mathbf{A}$ queries the sensor implementing continuous authentication through a model $\mathbf{f}$ that is trained to the legitimate user $\mathbf{U}$ and safely stored by the sensor $\mathbf{S}$. The Adversary queries the sensor by triggering off several events $\mathbf{x} \in \mathbf{X}$ in a given time epoch. After the time epoch, the adversary's output can be represented as $\acute{\mathbf{f}}(\mathbf{x})$. This is considered successful if it is close to the real $\mathbf{f}(\mathbf{x})$ within a given error margin $\boldsymbol{\varepsilon}$. This margin is represented as an error function.

In summary, $\mathbf{A}$ sends $\mathbf{x} \in \mathbf{X}$ to $\mathbf{S}$ and receives $\mathbf{y} = \mathbf{f}(\mathbf{x})$. After the time epoch, $\mathbf{A}$ generates $\acute{\mathbf{f}}$. Using an error function, $\mathbf{A}$'s output is compared to the legitimate user's output $\mathbf{f}$. It is successful if it is within a predetermined error margin, i.e., if $\mathbf{Err}\left(\acute{\mathbf{f}}(\mathbf{x})\right) \leq \boldsymbol{\varepsilon}$ (that is with a high enough probability exceeding a predetermined threshold)

Putting this together, we can represent this as: $\mathbf{pr}\left[\mathbf{A\ outputs\ } \acute{\mathbf{f}} \mathbf{\ and\ Err}\left(\acute{\mathbf{f}}(\mathbf{x})\right) \leq \boldsymbol{\varepsilon}\right]$.

Given $\boldsymbol{\varepsilon},\ \boldsymbol{\delta} > \mathbf{0}$, the adversary seeks to

i) adaptively select events $\mathbf{x_1},\ \mathbf{x_2},\ \ldots \in \mathbf{X}$

ii) observe the responses from the sensor to the events in a given time epoch

iii) and generate an output $\acute{\mathbf{f}}$ of $\mathbf{f}(\mathbf{x})$ such that $|\acute{\mathbf{f}} - \mathbf{f}(\mathbf{x})|_2 < \boldsymbol{\varepsilon}$ with probability at least $\mathbf{1} - \boldsymbol{\delta}$.

There are several techniques for sampling the data formally developed in active learning discipline [112]. We will leverage these techniques for deriving the data used by our adversary in this study. We will use Random sampling, Uncertainty sampling and Query synthesis techniques [88]. Random sampling involves the adversary picking their next queries uniformly at random out of a given pool. Uncertainty sampling on the other hand involves the adversary selecting the samples from the pool that have the least confidence or with smallest margin. Query synthesis differs from random and uncertainty sampling techniques in that unlike these two which are based on existing events in each pool, it generates new samples or queries. Probabilistic clustering-based algorithms such as Gaussian Mixture Models used in this study are usually applied in unsupervised environments such as [1],[67] where there is not a tainted training dataset [69,77,78] that the attacker can leverage. So, for our threat model, this technique closely represents the way the adversary will generate the attack events (i.e., by adaptively synthesizing the events based on observations and other knowledge of the system). To implement uncertainty sampling technique, we propose using Silhouette measurements to determine events with the least confidence. Silhouette information evaluates clusters based on the comparison of a distance measure of each element in the cluster to a measure of the separation from the closest alternative cluster. In this work, we use a modified Silhouette measurement as defined in [92].

The idea being, rather than using distance measures, we build on the concept of intra-group similarity and inter-group dissimilarity as defined by the degree of confidence that we allocate to the cluster membership of the elements with high density points getting maximum confidence and the least dense getting assigned least confidence.

From the Gaussian Mixture Models equation (4), we have $\mathbf{x_i} \in \mathbf{X}$ drawn from a probability density function $\mathbf{g(x)}$ we can evaluate the posterior probability that it belongs to cluster $\mathbf{c_i}$, $\mathbf{i = 1}$, ... $\mathbf{N}$ as:

$$\tau_n(x_i) = \frac{\pi_n g_n(x_i)}{\sum_{n=1}^{N} \pi_n g_n(x_i)}$$

In this equation, the prior probability of cluster $\mathbf{c_n}$ is represented by $\boldsymbol{\pi_n}$ and $\mathbf{g_n}$ represents the probability density at element $\mathbf{x_i}$ which is obtained after creating the cluster with elements only from that specific cluster ($\mathbf{c_n}$). Given this, we can represent the density-based silhouette information (**dbs**) of element $\mathbf{x_i}$ as:

$$dbs_i = \frac{\text{loglog } \frac{\tau_{n_0}(x_i)}{\tau_{n_1}(x_i)}}{\max_{j=1,\ldots,m} \left| \text{loglog } \frac{\tau_{n_0}(x_i)}{\tau_{n_1}(x_i)} \right|}$$

The **dbs** information for each observation or element is thus proportional to the log ratio between the posterior probability that it belongs to the cluster it has been assigned to and the maximum posterior probability that it belongs to a different cluster. It follows that small **dbs** values correspond to low confidence in the clustering while large values indicate a high confidence level [92]. In our experiments, we will determine **dbs** values for each value and select all the events with least confidence and then use those as our input representing adversary's queries. We can formally represent this as an optimization problem:

$$\underset{(x)}{argmin}\ g(x)\ s.t.\ A\ \epsilon$$

$$\{\{a'_i\}_{i=1}^{m} \subset R^d \mid x_l \leq a'_i \leq x_u \text{ for } i = 1, \ldots, m\}$$

$$\text{where } g(x) \text{ is the } dbs_i$$

Where **A** is the vector of events that the adversary will be sending as queries in each time epoch. For the uncertainty sampling, these are bounded to some interval corresponding to the input that the adversary can control. Regardless of the technique used, the adversary's objective is to produce meaningful queries or events within this time epoch that will minimize the error i.e., $\mathbf{Err}\left(\hat{\mathbf{f}}(\mathbf{x})\right)$ as defined in this section.

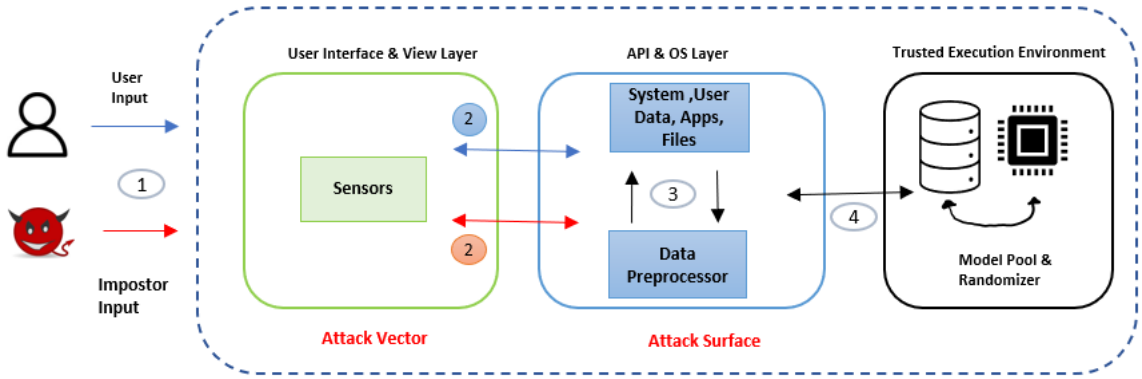

Figure 4.0.1: The threat model architecture showing the ML models stored in the device's TEE zone and the attack vector and surface that the adversary has access to.

Figure 4.0.1 shows a high-level architecture of the threat model. The device user interface represents the local attack vector through which the adversary (impostor who seeks to masquerade as the real user in this case) interacts with the device. For our study, the interaction involves activities performed on the device such accessing specific apps, reading specific data,

browsing the web, etc. An active authentication sensor tracks all those events including a system time stamp corresponding to each activity. The ML models and the decision processing sensor are protected in the device's isolated trusted execution environment (TEE). The adversary seeks to accomplish their goal by extracting the model and determining its decision boundary. The aggregated events and data fed to the ML model and decision engine is the attack surface that the impostor has control over. We assume that the adversary will not have ability to breach the device's TEE zone. Attacks where the adversary can compromise the device's TEE are beyond the scope of this study. The primary focus for this research is on evading the machine learning based security techniques. The decision processing sensor securely stores the model, represented as f that it uses to compute the prediction **f(x)** used to determine normal activity from abnormal ones. The conjecture is that if this sensor uses one fixed model that returns **f(x)** corresponding to a fixed threshold or boundary used to delineate normal from abnormal then it is easier for the adversary to compute events that satisfy this threshold compared to if the sensor uses multiple randomly generated models drawn from a pool $\mathbf{f} \in \mathbf{F}$ determined by some random routine and thus multiple thresholds used that depend on which model is used. The uncertainty introduced by the random routines employed makes it a more difficult problem for the adversary. Intuitively, we thus define robustness in this context as the ability of the active authentication system resisting ability for an adversary to determine the decision boundary used to delineate legitimate user from an impostor. We do this by varying this decision boundary for each time epoch, thus presenting a varying target that the adversary is challenged to determine during each time epoch. We validate this through extensive experiments.

4.4 Randomness as a defense

4.4.1 Defenses

A few techniques have been proposed for countering model stealing attacks. An interesting approach based on boundary differential privacy was proposed by [11]. Their approach protected against model stealing by obfuscating the prediction responses near the decision boundary, making it difficult for an adversary to learn predefined precision regardless of the number of queries sent to the prediction API. The downside of this approach however is the potential performance degradation it is likely to have on the original model. For model extraction attacks using prediction APIs that leverage confidence scores, the authors in [33] proposed rounding confidence values returned by machine learning service providers. This technique lessened the effect of the attack but did not completely mitigate it. Building on this, the researchers in [41] proposed a technique for monitoring extraction by observing the queries issued by multiple users of the machine learning service and provided a warning if a given threshold was crossed for the informed extracted by a given user or subset of users. This technique however posed a challenge to an ML service provider since the clients with more queries did not all translate to adversaries trying to extract models. These could be legitimate clients using the service extensively and in fact bringing in more revenue for the ML service provider. The authors in [45] proposed a defense based on embedding of digital watermarks during model training that the model owner could use afterwards for identifying their models in case they are stolen or misused. However, this defense is insufficient since it only detects a stolen model. It does not prevent the model from being stolen as demonstrated by [46]. There are defenses that aim to make the work of the adversary harder. For example, the authors in [42]

proposed a technique that aims to amplify the attackers error rate while minimally impacting the legitimate user.

4.4.2 Randomization

One promising defense strategy that researchers have studied for a long time is use of randomization techniques. For example, more than a decade ago the researchers in [5] leveraged randomization technique in their design of anagram, an anomaly detector that was based on n-gram analysis using binary-based modeling techniques. Barreno et al [27] in their work proposed a taxonomy of attacks and listed three general strategies: randomization, disinformation, and increased complexity of hypothesis space. The authors acknowledged that while randomization made it difficult for the adversary it also negatively affected the accuracy of the system and thus the key was to find a right balance between the system accuracy and the level or technique of randomization. The authors in [15] proposed a machine learning as a service implementation that used randomization technique to mitigate reverse engineering attacks. An important finding that they demonstrated was that randomization with large variance could be employed without the system incurring a loss in accuracy. Randomization techniques have also been applied at algorithm level. The researchers in [11] designed a perturbation algorithm that they referred to as randomized response. The objective was to make it difficult for the adversary to learn decision boundaries by predefined precisions.

Our study evaluates use of randomized modeling strategy as a mitigation against model stealing in continuous authentication implementation. Through extensive experiments, we show that randomization effectively mitigates these attacks by raising the bar for the attacker. We leverage continuous authentication work done by [1], [67]. We use their dataset and extend their

feature selection criteria in devising how to construct a randomized pool of models based on distinctive features.

4.5 Overview and Methodology

This study is designed to validate our claim that use of multiple randomized ML models from which one is selected randomly during run-time for a given time epoch offers robustness against model stealing in a continuous authentication implementation compared to when one fixed ML model is used. We also seek to determine which randomization technique is most effective. Precisely, we answer the question: *"which randomization approach is most effective: data-dependent randomization or data independent randomization?"* We define data dependent randomization as utilizing the input data (features in our case) in implementing the randomization scheme, while data independent as appending noise to the output label, i.e. $y = f(x) + n$ where n is the random noise added as part of the response. This appended noise is not part of the input data. In our study, this will mean that the sensor can modify the prediction from the model before acting (or responding to the user).

Dataset and Model

We use RUU [1] dataset in our experiments. Figure 4.0.2 shows a snippet of this dataset to illustrate the format which includes a time stamp, an event id, a category, an action and a description or detail information of the activity. The categories get converted into dataset features. Figure 4.0.3 shows a snippet of pre-processed dataset. This is accomplished by chunking the input into 5-minute time epochs from which the selected features are aggregated.

These features are represented as the columns in this figure. For example, the first row in Figure 4.0.3 indicates six unique activities of a particular feature in the first-time epoch. We choose activities from two different users as our baseline. Each has at least 150-time epochs where each epoch or time slice comprises of activity for 5 minutes, meaning that each sample has a total of more than 12 hours of activity or events. The features selected for use from the dataset are based on Fischer scores as determined in [1, 67]. This is a discriminant analysis technique that evaluates the value of each feature independently via a ratio of their inter-class and intra-class variance. So, a feature with both low within-class variance thus very stable and predictable and with a high between-class variance, i.e., appears vastly different for every class with a high score. Log-squashing technique was employed for smoothing out the measurements hence removing bias of a given feature dominating the others.

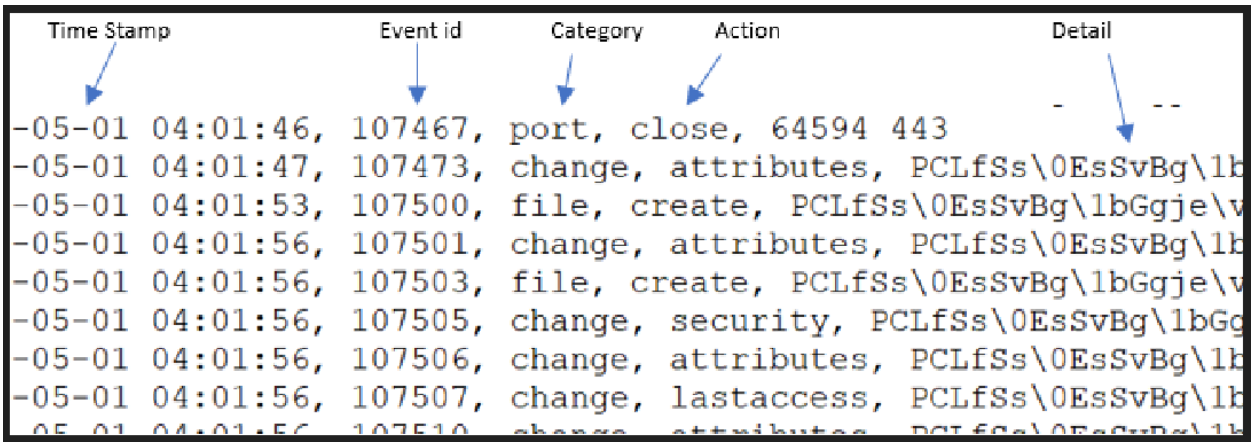

Figure 4.0.2: Dataset snippet showing input collected. This example shows the time stamp, event id, category, action, and the activity detail

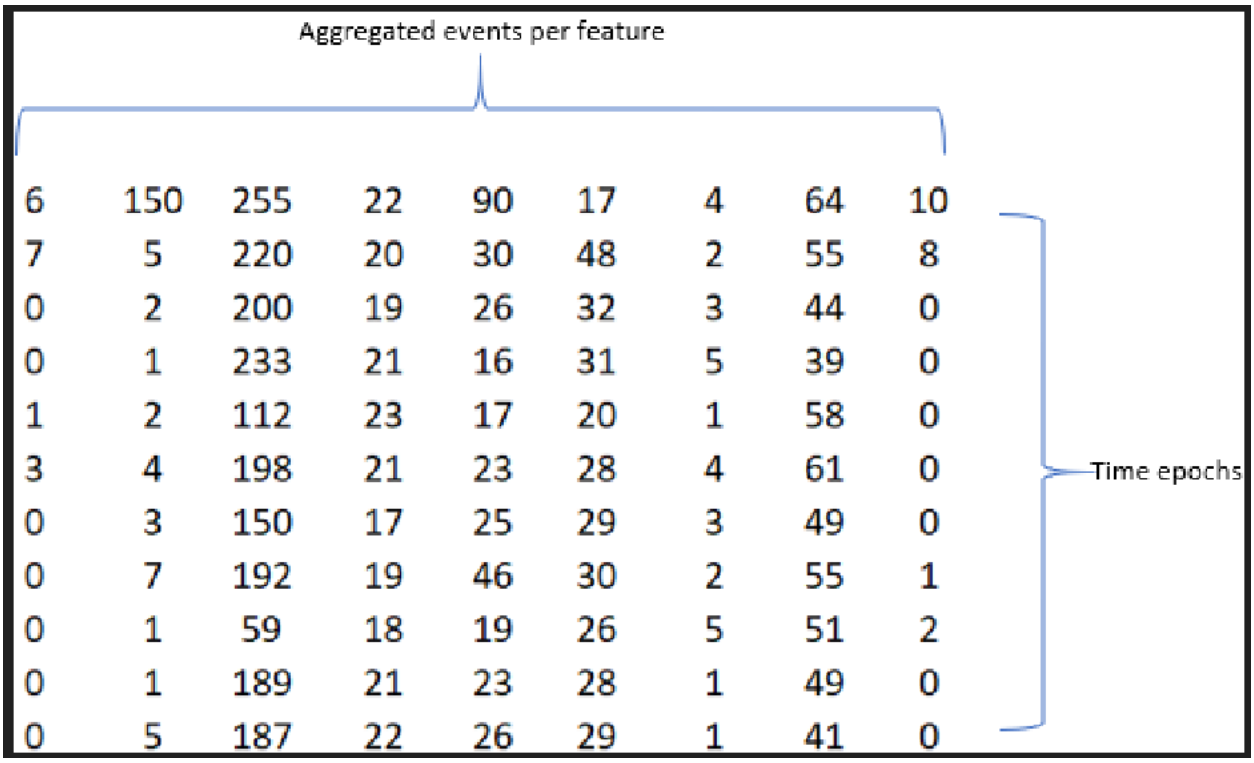

Figure 4.0.3: Snippet of aggregated events for each feature in each time epoch. The rows represent 5-minute time epochs, and the columns represent each feature

For the experiments below, we used clustering technique to represent the machine learning classifier for classifying events corresponding to a legitimate user from an imposter. This will be based on Gaussian Mixture Model (GMM) algorithm. GMM attempts to find a mixture of multi-dimensional Gaussian probability distributions that best model a given input dataset. It is a parametric probability density function which is represented as a weighted sum of Gaussian component densities. It uses an Expectation-Maximization (E-M) approach with two steps: The E-step looks at each point and finds weights encoding the probability of membership in each cluster. The M-step looks at each cluster, updates its location, normalizes, and shapes it based on all data points leveraging the weights which results into a smooth Gaussian model. It is worth noting that GMM is fundamentally a density estimation algorithm because the result of a GMM

fit to a given input data is a generative probabilistic model describing the distribution of that data.

More formally, given a feature space, $f \subset R^d$ a Gaussian Mixture Model $g: f \rightarrow R$ with n components is defined as:

$$g(x) = \sum_{i=1}^{n} w_i N_{\mu_i \sum i}(x)$$

$$N_{\mu_i \sum \ i}(x) = \frac{1}{\sqrt{(2\pi)^d \left| \sum \ i \right|}} e^{-\frac{1}{2}(x-\mu_i)\sum_i^{-1}(x-\mu_i)^T}$$

In the equations above, $\boldsymbol{\mu_i}$ is the center of the $\boldsymbol{i^{th}}$ GMM component, this is set to the mean of samples from the same cluster for that given iteration, i.e., $\boldsymbol{\mu_i = c_i}$. The covariance is set to covariance of the samples in the current cluster, i.e., $\sum \ _{\mathbf{i}}$ and $\mathbf{w_i}$ that is calculated based on the number of samples that are placed in each cluster. So, each component for GMM is created using the mean, the covariance, and the weights for each cluster from prior iteration.

Intuitively, similar distributions would imply similarity between the input data generated following similar patterns. In our RUU Active Authentication dataset [1], [138], this means that the activities or events indicating user behavior or habits from independent time slots will have a similarity in distributions and hence indicate the input belongs to the same user. A different user will be expected to produce a different distribution pattern. Input data corresponding to an outlier in either case would have those data points corresponding with smallest likelihood or probability values hence standing out as outliers. We are defining a boundary around normal events so that those can be distinguishable from abnormal events i.e., from the adversary. The adversary aims

to find out where this boundary lays. A fixed model with the same number of features implies a fixed boundary or threshold. While a pool of randomly generated models where one is randomly picked for a given time epoch implies a varying boundary thus a varying target that the adversary is challenged to determine during each time epoch. Intuitively, this makes the problem harder. Our experiments seek to validate this hypothesis.

4.6 Experiments and Results

Randomization Strategy

The first experiment that we ran was for determination of the randomization strategy. We sought to find out if using an input data-depended strategy was more effective than using a data independent approach. For the former approach, we used distinctive features to produce different trained models (see Algorithm 1 below). The testing dataset had about 17% of events from a different user. This was to act as our baseline for what we expected to be classified as outlier for the most accurate model. For each experiment, we only varied the number of features used, but kept everything else constant. The objective was to determine a methodology for varying the features without impacting accuracy. We approached this by first determining the top twenty Features with the highest Fisher scores. From these top features, we then ran experiments with different numbers of features varying the total features per model from 50% to 100% (of the top fisher scored features) and observed the performance scores. We defined our performance in terms of percentage of events that were categorized under an outlier cluster plus the highest density-based silhouette (dbs) coefficient measurement. Results summarized under Table 4.0.1 and Figure 4.0.4 show that at 80% features, we saw the highest median dbs score of 0.63. At this

same percentage, it had the second lowest percentage of events classified under outlier cluster at 25%. This was the closest number in terms of accuracy to expected given that the test data had 17% events from a different user. Based on this we determined that between 80% and 85% of the top features per model produced optimal results. So, our data-dependent randomization strategy simply selected 80% of the top features randomly out of the top scored features for each model instance. The data independent strategy was accomplished by selecting one of the model instances (as described above) and simply appending noise to the output. The same model was used for each experiment ran but only the output was slightly perturbed each time. The summarized results in Figure 4.0.5 and Figure 4.0.6 below show a comparison of these two approaches. For the estimation error, we used geometric error as the key metric for comparison. This estimation error was defined as $|\acute{f}- f(x)|_2$ where $f(x)$ is the probability output for the base model representing ground truth, while $\acute{f}$represented the output using the testing data with the specific randomization performed accordingly. The smaller the estimation error, the closer is the output to the ground truth. In other words, the nearer the adversary is to producing events that mimic the legitimate user.

ALGORITHM 1: Algorithm for generating randomized models

Input: Training dataset
Output: Pool of trained models
Initialization: Set appropriate X as total number of models in the pool
Select the top features using Fisher's criteria scores
for $i = X$ to 0 **do**
    Randomly pick 80% of the features from above
    Train one model using these features
    Add the trained model f_i to the pool: $F = F \cup f_i$
end for
return F

From the results in Figure 4.0.5 and Figure 4.0.6, we can see that with more testing rounds, the estimation error for the data independent approach reduced at a faster rate than for the data-depended approach. The estimation error in this case indicates how close the prediction is to the ground truth. Based on these results we conclude that data dependent randomization strategy is more effective. For the subsequent experiments below, we employ a data dependent randomization strategy.

Randomized Modeling:

Having determined our randomization strategy, the subsequent experiments involved comparing the performance of randomized models to fixed models over several queries. The first step was determining the methodology for generating the samples or queries for each round. We leverage active learning established query selection techniques [112]. Specifically, we use random sampling, uncertainty sampling, and query synthesis techniques. Random sampling involves selecting the queries uniformly at random from a given data source [85]. Uncertainty sampling involves drawing events with the least confidence scores from the classifier, which translates to regions closer to the classifier's decision boundary. In our case, this means events closer to the edge of a given cluster. Regarding DBS measures, these are activities with DBS scores at or close to zero (refer to the model theft threat model section). Query Synthesis [71,113] involves generating samples de novo, that is, samples that do not necessarily conform to a distribution or are not all part of a given pool. To simulate this technique, we start by using DBS measures to obtain events with the lowest confidence scores (as described in the model theft threat model). Then we employ a technique used by the authors in [113] to generate new

events that are similar or closer to these low-confidence events. For our study, we increment the user activity counts per time epoch by at least ten events or more. This will represent the new queries generated for the next round of adversary testing. In practice, the starting events represent the adversary's limited knowledge about the system and the victim. In our threat model, we assume the adversary has the user's mobile device and potentially has limited knowledge of the activities that the legitimate user performs on the device during specific times of the day. We summarized results after ten rounds of testing, where each round had an average of about one hundred tests. The three different probing techniques (random sampling, uncertainty sampling, and query synthesis) used for comparing the estimation errors between a fixed model and a pool of models where one was randomly selected for a given time epoch are each summarized in Figures 4.0.7, 4.0.8, and 4.0.9, respectively.

Query synthesis sampling technique best resembles the threat model where the adversary is in possession of the victim's device and their traditional credentials such as PIN, but potentially has no other additional information. The device in this case is protected by user behavior analysis based active authentication. The adversary can generate arbitrary queries that do not have to conform to any distribution in this scenario. This also allows us to relax any assumptions regarding the adversary's knowledge. Figure 4.0.9 summarizes the query synthesis-based results comparing when a fixed model is used to when a randomized pool of models' strategy is used. The trend line for both graphs shows a gradient of 0.0084 for the randomized models and 0.0194 for the fixed model. This implies that when a pool of randomized models is used with query synthesis, the rate of estimation error converging to the baseline is more than twice as fast when just one fixed model is used. What this means is that an adversary in possession of a victim's device would take more than twice as much time before they can

potentially extract the model. Uncertainty sampling trend lines (see Figure 4.0.8) are like query synthesis with gradients of 0.0183 and 0.0066 for fixed model approach and randomized model approach, respectively. Just like in query synthesis, the rate of convergence of the estimation error for randomized models is more than twice that of fixed models' approach. The largest difference is observed for the random sampling approach where the trend line gradients are 0.00007 when randomized models' approach is employed, and 0.0186 when a fixed model is used (see Figure 4.0.7). In practice, this sampling technique represents an adversary that has zero knowledge of the victim and device and are generating events randomly. Given that the adversary has no knowledge of the victim, it will take a much longer time or an extremely substantial number of queries before they can start seeing the estimation error start to converge. A realistic example of this is a case where an adversary recovers a victim's lost mobile device and have no prior knowledge of the victim or the device. On the other hand, uncertainty sampling technique would represent a case where an adversary steals a mobile device from a victim that they potentially have limited knowledge about.

Table 4.0.1: Median Density Based Silhouette (DBS) and Percentage of Events in the outlier cluster

% of top Fishers scrores	Median dbs	% outlier
0.5	0.47	0.38
0.55	0.49	0.36
0.6	0.48	0.37
0.65	0.53	0.33
0.7	0.57	0.29
0.75	0.59	0.26
0.8	0.63	0.25
0.85	0.63	0.24
0.9	0.61	0.25
0.95	0.59	0.27
100	0.62	0.27

Figure 4.0.9 represents results for all the three probing techniques on the same graph for comparison purposes. Overall, it is evident that the fixed model had a lower estimation error that decreased faster on average compared to the randomized pools approach regardless of the probing technique used. The lower the estimation error, the closer the adversary is to estimating the prediction (extracting the ground truth model). Similarly, the rate of change of the estimation error is related to the time it takes to estimate the ground truth model's prediction. So, a lower rate of change means that the adversary will take a longer time before extracting the model. These observations lead to our conclusion that randomized modeling increases robustness against model theft attacks through extraction. We attribute this to the fact that strategic varying of features (as employed in this study as a technique for generating the randomized models - see algorithm 1) leads to varying the mean and covariance of the GMM which means that the decision threshold will vary from model to model but still fit with the legitimate user’s profile that the model was trained on. This variance of the threshold raises the bar for the adversary that is trying to determine the threshold or the decision boundary. The adversary is challenged to determine which features are used for a given model at a given time. If a fixed model is used, the adversary has a fixed threshold target that they can compute through the probing queries by zeroing in on a fixed set of features.

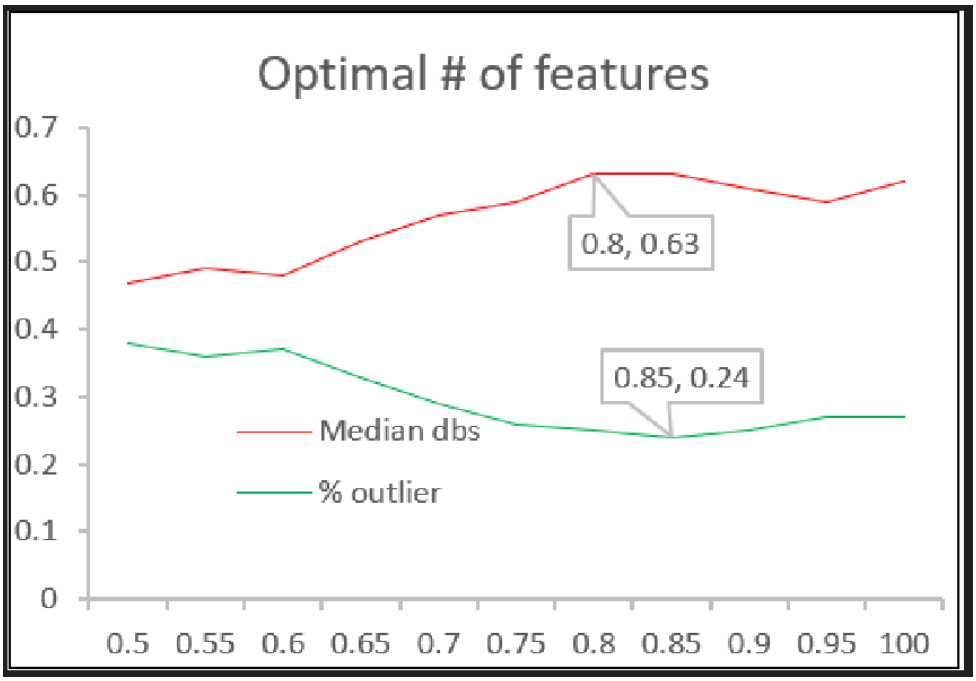

Figure 4.0.4: Median density-based silhouette (DBS) and % of events in outlier cluster values as a function of different numbers of features. Between 80% and 85% of the top fisher scored features produce optimal results.

Table 4.0.2: Results showing estimation errors when two different randomization strategies are used. Each testing round comprised of an average of 100 tests, so the reported error rate is the average value over those tests. On average, the data dependent strategy produced a higher estimation error compared to the data independent strategy meaning that data dependent strategy is more effective

Estimation Error for Randomized approaches		
	Data-Dependent	**Data-Independent**
	$y = f(x) \; : \; f \in F$	**$y = f(x) + n$**
Testing round (100 tests/round)	**(where f is picked randomly from a pool)**	**(where n is a data independent noise added)**
1	0.417	0.407
2	0.403	0.399
3	0.397	0.388
4	0.401	0.381
5	0.391	0.342
6	0.387	0.353
7	0.381	0.291
8	0.368	0.271
9	0.371	0.257
10	0.374	0.242

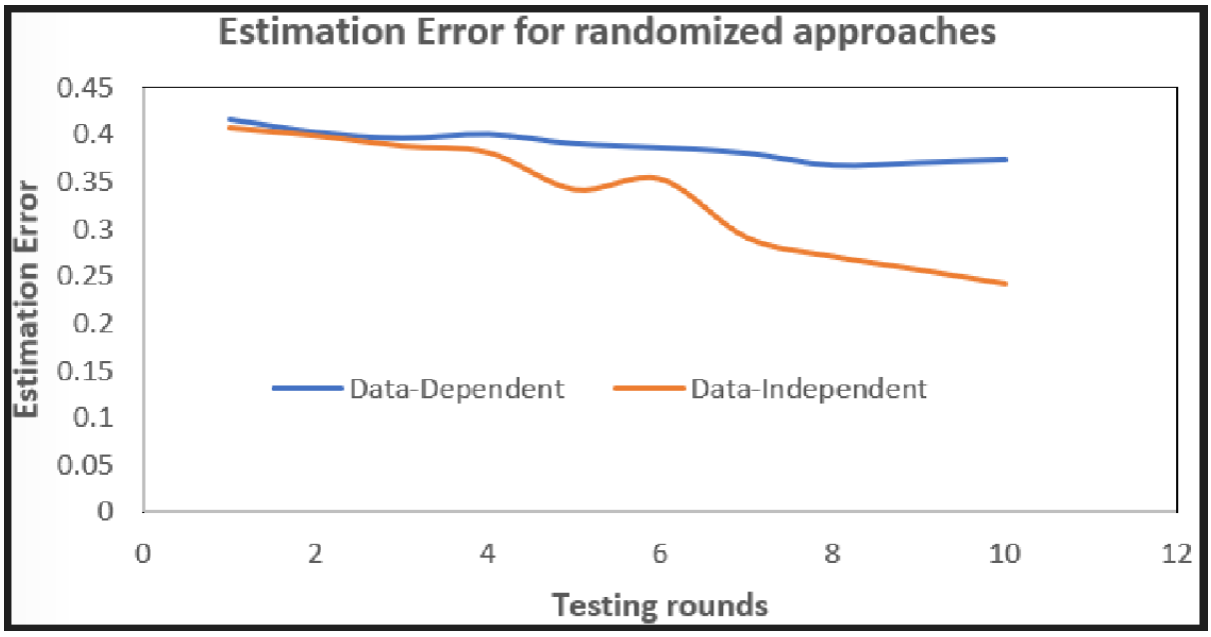

Figure 4.0.5: The estimation error against the probing rounds. Each round consists of an average of 100 tests. The data dependent strategy had a higher estimation rate compared to the data independent strategy. This translated into the data dependent strategy being more effective at mitigating the adversary extracting the model (i.e., synthesizing events leading to a closer estimate to the legitimate user's events)

Estimation Error (Random Sampling)		
Testing round	Fixed Model	Pool of Models
1	0.328	0.364
2	0.267	0.323
3	0.272	0.399
4	0.196	0.321
5	0.193	0.388
6	0.177	0.371
7	0.235	0.388
8	0.151	0.379
9	0.121	0.326
10	0.157	0.351

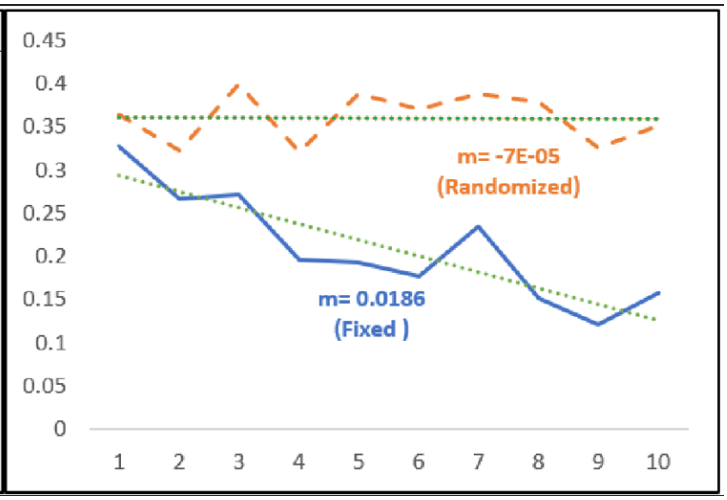

Figure 4.0.6: Estimation errors comparing a fixed model approach to a randomized model approach. Random sampling used to generate input data. Fixed models trend line has a gradient of 0.0186 compared to 0.00007 for randomized models. Implies that when using fixed model, normalized error converged at a much faster rate while randomized models the error rate was almost constant

Estimation Error (Uncertainty Sampling)		
Testing round	Fixed Model	Pool of Models
1	0.387	0.412
2	0.412	0.397
3	0.321	0.388
4	0.348	0.391
5	0.322	0.421
6	0.345	0.381
7	0.357	0.367
8	0.284	0.402
9	0.282	0.352
10	0.168	0.331

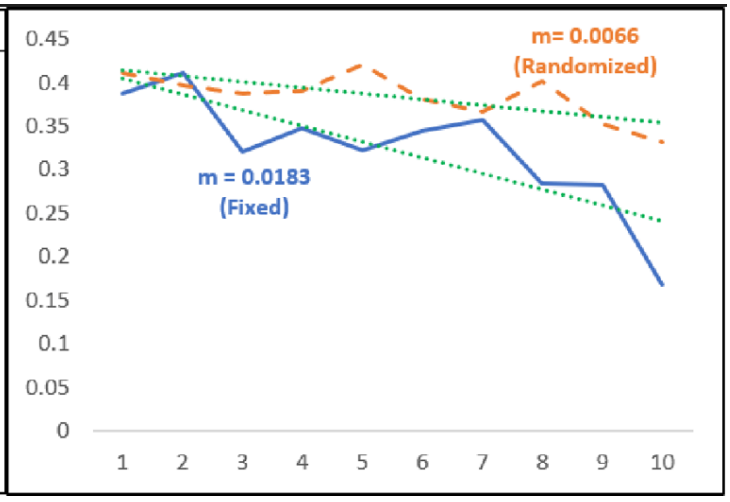

Figure 4.0.7: Estimation errors when uncertain querying technique is used to generate input data. The randomized modeling approach with a gradient of 0.0066 performed better than a fixed model approach with a gradient of 0.0183

Estimation Error (Query Synthesis)		
Testing round	Fixed Model	Pool of Models
1	0.393	0.401
2	0.361	0.407
3	0.284	0.391
4	0.279	0.365
5	0.323	0.387
6	0.241	0.317
7	0.201	0.354
8	0.269	0.381
9	0.199	0.345
10	0.207	0.312

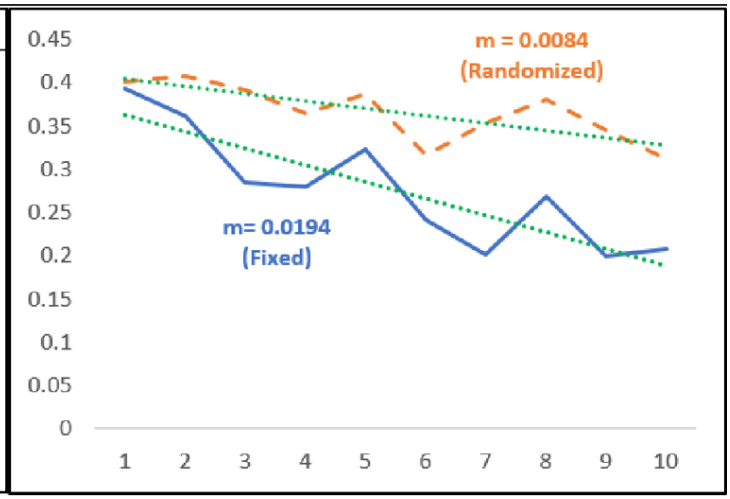

Figure 4.0.8: Estimation errors when query synthesis technique is used to generate input data. Fixed model approach had a gradient of 0.0194 compared to 0.0084 for a randomized approach

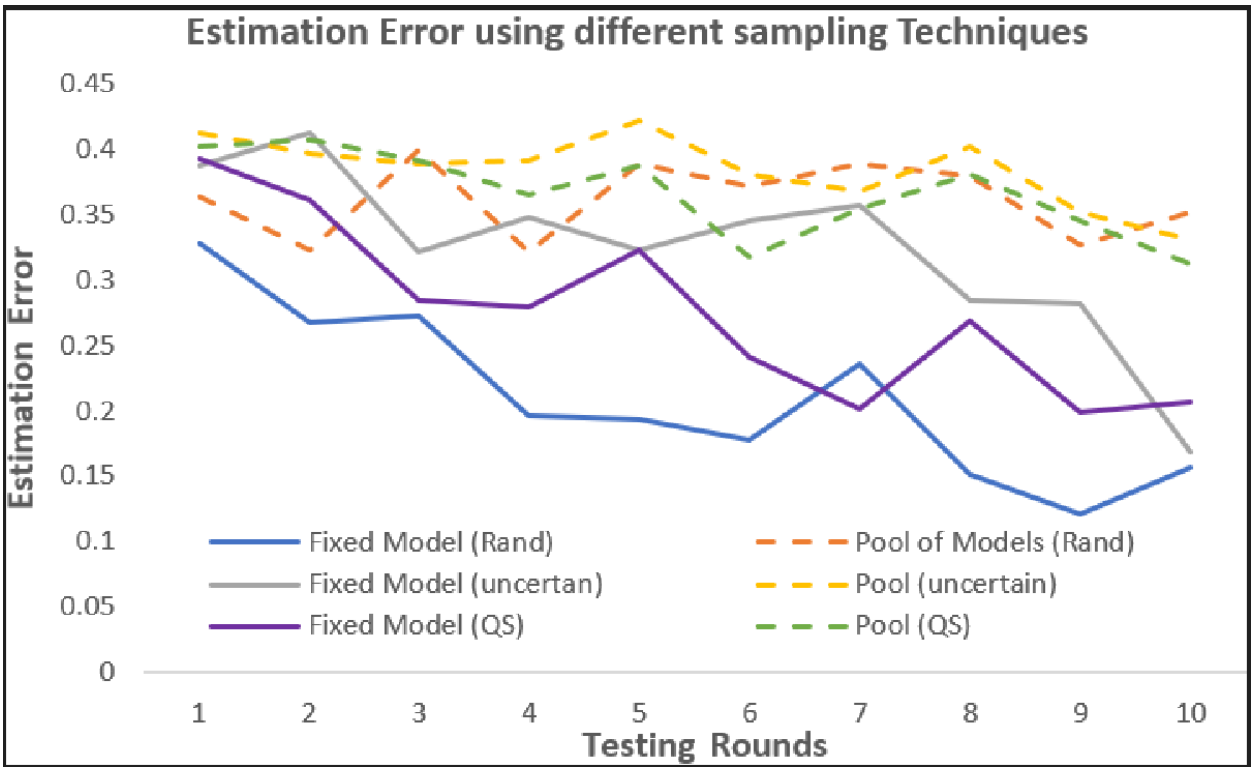

Figure 4.0.9: Results showing estimation errors when three different probing techniques are used i.e., random Sampling, Uncertainty sampling and Query synthesis. For each technique, we test the fixed model and the multiple models randomly selected. In all cases, the randomized approach leads to higher estimation error indicating that the adversary will take longer to extract the model compared to when the fixed model is used.

Conclusions

Protection of Account Recovery process requires a comprehensive approach that accounts for both proactive and reactive mechanisms. In Chapter 3, we covered several proactive mechanisms including insights derived from empirical measurements of voluntary multi-factor authentication adoption. This chapter covers a reactive mechanism that relies on user behavioral modeling to ensure that it is the legitimate user that performed the Account Recovery. In specific, the chapter demonstrated use of randomized modeling as an effective mitigation strategy against ML model stealing in context of user behavior based active authentication. We linked ML model stealing in this context to an evasion attack for achieving account takeover. To implement our randomized modeling, we started by determining an effective randomization strategy by comparing performance of a data dependent randomization approach versus a data independent approach. From this, we determined that using data dependent randomization was more effective in our context. Thus, we employed this strategy in generating the pool of models from which one was selected uniformly at random for each time epoch during the experiment. Using an estimation error, we compared performance of fixed model approach to randomized modeling approach. Based on the observed results, we confirmed that the randomized modeling approach was more robust against ML model stealing compared to a fixed modeling approach. We thus concluded that an adversary in possession of a mobile device protected by an active authentication implementation using randomized modeling would be more robust against being breached compared to the current state of the art where typically a fixed model is used. For future work, the authors will be testing this approach against a variety of different Active Authentication implementations and datasets. One key to improving robustness against ML

model theft is understanding the tradeoff dynamics between prediction accuracy and gained ML model robustness when different adversary query strategies are employed. To that end, as part of future work, the authors plan to study this dynamic in detail using a variety of datasets, Active Authentication implementations and querying strategies.

Chapter 5:

Comparison to prior work

5.1 Voluntary Multiple Authentication Factors Adoption

There have been several studies done in the past on the 2FA adoption [3, 7, 14, 15] and on usability [1, 2, 6, 7, 12, 13]. However, these studies focused more on the general users' demographics that tended to be younger users e.g., college setting or employees in a company setting, or younger active users on social media platforms.

Our study focused on the older adults thus offering a new and different perspective on this topic of voluntary 2FA adoption. Previous research investigated understanding the motivations as well as processes others employ when helping older users, especially relatives, with mobile technology and security issues [27]. The authors in [28] studied the privacy and security concerns that older users had with emerging technologies and recommended educational approaches and technical protections that incorporated these users' needs and preferences.

5.2 Resilient User Behavioral Modelling

There is a significant body of work on user behavior-based machine learning implementation of Active Authentication and on the use of randomization principles in designing secure computer systems. Our work combined these two domains by using the later

(randomization principles) to add robustness against model stealing attacks in the former (Active Authentication). Many researchers have proposed mitigation strategies against ML model stealing. However, most of this work has targeted generic machine learning as a service (MLaaS) implementations [75,79,95,102,103,106]. Our work on the other hand targeted and focused on Active Authentication domain. Use of randomization as a defense against adversarial attacks on machine learning has been covered in literature [68,70,75,79,87,89] but most studies thus far have focused on applying this technique at model training time or using synthetic datasets. Our work applied randomization principles during both model training time (by varying the input features) as well as during model application time (by randomly selecting a given model from a pool for each time epoch). Moreover, our experiments were run using a real-world dataset.

Our work made a direct link between a trained user ML model theft and evasion attack leading to an account takeover. We did this by drawing a parallel between a traditional credential such as username/password combination, and a trained user behavior ML model in Active Authentication context. Using a stolen traditional credential, an adversary can access a victim's resource that is protected by that credential whether it is on the same device or on a different device. Similarly, in this context, a stolen user trained ML model can be used to access that user's device protected by active authentication. This differentiated our work from other related research that focused on ML model theft attacks in general classification problems such spam filtering, malware detection and image classification [68,75,105,107,111].

Conclusion

Today, digital authentication has become a routine part of our daily lives. This means that Account Recovery process is critical for enabling that continued access to our accounts. But, Account Recovery, if not secured properly, can become an attack vector for adversaries looking to take over a legitimate user's account.

This dissertation presented a seamless approach to the protection of Account Recovery process leveraging Machine Learning based user behavioral analysis. First, we presented a background on Account Recovery and the methods currently used for achieving this process. We then studied the proactive mechanisms in use today and noted that the human element is the one commonality factor among all the proactive mechanisms in use. Solving the human behavior problem was just as important as solving the technical Account Recovery problem. To that end, we performed a study on the user voluntary adoption of multiple authentication factors that can be used as fallback mechanisms during Account Recovery process. Next, we proposed the addition of user behavioral modeling based active authentication as a critical part of the Account Recovery process. This accounted for the reactive protection mechanism that was aimed at flagging out an imposter should they circumvent the proactive mechanisms in place. Finally, to ensure resiliency against evasion attacks resulting from model theft, we studied use of randomized modeling techniques to determine which randomization strategy was most effective for providing this resiliency and privacy protection in the context of active authentication for Account Recovery.

References

[1] J. Voris, Y. Song, M. Salem, S. Hershkop and S. Stolfo, Active authentication using file system decoys and user behavior modeling: results of a large scale study, Computers \& Security, vol. 87, p. 101412, 2019. Available: 10.1016/j.cose.2018.07.021.

[2] Steven M. Bellovin. Who are you? IEEE Security & Privacy, 15(6), November/December 2017.

[3] Maximilian Golla, Grant Ho, Marika Lohmus, Monica Pulluri, and Elissa M. Redmiles 2021. Driving 2FA Adoption at Scale: Optimizing Two-Factor Authentication Notification Design Patterns. In 30th USENIX Security Symposium (USENIX Security 21) 109–126.

[4] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In IEEE Symposium on Security and Privacy, SP '12, pages 553–567, San Jose, California, USA, May 2012. IEEE.

[5] Jacob Abbott and Sameer Patil. How mandatory second factor affects the authentication user experience. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20. Association for Computing Machinery, 2020.

[6] Sanchari Das, Andrew Dingman, and L. Jean Camp. Why Johnny doesn't use two factor a two-phase usability study of the FIDO U2F security key. In International Conference on Financial Cryptography and Data Security (FC), 2018.

[7] Thanasis Petsas, Giorgos Tsirantonakis, Elias Athana-sopoulos, and Sotiris Ioannidis. Two-factor authentication: Is the world ready? Quantifying 2FA adoption. In Proceedings of the Eighth European Workshop on System Security, page 4. ACM, 2015.

[8] J. Reynolds, T. Smith, K. Reese, L. Dickinson, S. Ruoti, and K. Seamons, "A tale of two studies: The best and worst of YubiKey usability," in 2018 IEEE Symposium on Security and Privacy (SP). IEEE, 2018.

[9] Elissa M. Redmiles, Michelle L. Mazurek, and John P. Dickerson. Dancing pigs or externalities? Measuring the rationality of security decisions. In Proceedings of the 2018 ACM Conference on Economics and Computation, pages 215–232, 2018.

[10] MS 2019. One simple action you can take to prevent 99.9% of attacks on your accounts. Retrieved May 20, 2022 from https://www.microsoft.com/security/blog/2019/08/20/one-simple-action-you-can-take-to-prevent-99-9-percent-of-account-attacks/

[11] The White House 2021. Executive Order on Improving the Nation's Cybersecurity. Retrieved April 27, 2022 from https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/

[12] Christina Braz and Jean-Marc Robert. 2006. Security and Usability: The Case of the User Authentication Methods. In Proceedings of the 18th Conference on l'Interaction Homme-Machine (IHM '06). Association for Computing Machinery, New York, NY, USA, 199–203.

[13] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. 2013. A Comparative Usability Study of Two-Factor Authentication. https://arxiv.org/abs/1309.5344.

[14] P. Ackerman, “Impediments to adoption of two-factor authentication by home end-users,” SANS Institute InfoSec Reading Room, Sep 2014.

[15] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M. Angela Sasse. “They brought in the horrible key ring thing!” Analyzing the Usability of Two-Factor Authentication in UK Online Banking. In Workshop on Usable Security, USEC ’15, San Diego, California, USA, February 2015. ISOC.

[16] Twitter Transparency Center 2022. Account Security. Retrieved July 29, 2022 from https://transparency.twitter.com/en/reports/account-security.html#2021-jan-jun

[17] The Wallstreet Journal 2021. Tech Companies Push Users to Adopt Two-Factor Authentication. Retrieved May 12, 2022 from https://www.wsj.com/articles/tech-companies-push-users-to-adopt-two-factor-authentication-11635807088

[18] International Organization for Standardization (ISO): ISO 9241-11. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. Retrieved April 5, 2022 from https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

[19] Pew Research Center. Technology use among seniors. Retrieved June 2, 2022 from https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/

[20] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. It’s not actually that horrible: Exploring adoption of two-factor authentication at a university. In Proceedings of the 2018 CHI Conference on Human Factors in Compu-ting Systems, page 456. ACM, 2018.

[21] Verizon Data Breach 2021. Investigations Report. Retrieved April 12, 2022 from https://www.verizon.com/business/en-gb/resources/reports/dbir/

[22] Google 2022. Making you safer with 2SV. Retrieved March 4, 2022 from https://blog.google/technology/safety-security/reducing-account-hijacking/

[23] Florencio, D., and Herley, C. A large-scale study of web password habits. In International Conference on World Wide Web (WWW) 2007.

[24] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa Redmiles, and Blase Ur. “What was that site doing with my Facebook pass-word?” Designing Password-Reuse Notifications. In ACM Conference on Computer and Communications Security, CCS ’18, pages 1549–1566, Toronto, Ontario, Canada, October 2018. ACM.

[25] Anupam Das, Joseph Bonneau, Matthew Caesar, Ni-kita Borisov, and XiaoFeng Wang. The Tangled Web of Password Reuse. In Symposium on Network and Distributed System Security, NDSS ’14, San Diego, California, USA, February 2014. ISOC.

[26] Florian M. Farke, Lennart Lorenz, Theodor Schnitzler, Philipp Markert, and Markus Dürmuth. 2020. “You still use the password after all” – Exploring FIDO2 Security Keys in a Small Company. In

Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020). USENIX Association, 19–35.

[27] Tamir Mendel and Eran Toch. My Mom Was Getting This Popup: Understanding Motivations and Processes in Helping Older Relatives with Mobile Security and Privacy. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(4):147:1–147:20, December 2019.

[28] Alisa Frik, Leysan Nurgalieva, Julia Bernd, Joyce S. Lee, Florian Schaub, and Serge Egelman. Privacy and Security Threat Models and Mitigation Strategies of Older Adults. In Symposium on Usable Privacy and Security, SOUPS '19, pages 21–40, Santa Clara, California, USA, August 2019. USENIX.

[29] Stéphane Ciolino, Simon Parkin, and Paul Dunphy. Of Two Minds about Two-Factor: Understanding Everyday FIDO U2F Usability through Device Comparison and Experience Sampling. In Symposium on Usable Privacy 30 Sixteenth Symposium on Usable Privacy and Security USE-NIX Association and Security, SOUPS '19, pages 339–356, Santa Clara, California, USA, August 2019. USENIX.

[30] E. Hargittai and K. Dobransky. Old dogs, new clicks: Digital inequality in skills and uses among older adults. Canadian Journal of Communication, 42(2), 2017.

[31] S. L. Willis, K. W. Schaie, and M. Martin. Cognitive plasticity. In Handbook of Theories of Aging, pages 295–322. Springer, 2009.

[32] AARP 2019. Older Adults keep pace on Tech usage. Retrieved Jan 23, 2022 from https://www.aarp.org/research/topics/technology/info-2019/2020-technology-trends-older-americans.html

[33] K. Dobransky and E. Hargittai. Unrealized potential: Exploring the digital disability divide. Poetics, 58:18–28, 2016.

[34] M. Haight, A. Quan-Haase, and B. A. Corbett. Revisiting the digital divide in Canada: The impact of demographic factors on access to the Internet, level of online activity, and social networking site usage. Information, Communication & Society, 17(4):503–519, 2014.

[35] K. Zickuhr and M. Madden. Older adults and Internet use. Technical report, Pew Internet & American Life Project, June 2012.

[36] G. A. Grimes, M. G. Hough, E. Mazur, and M. L. Signorella. Older adults' knowledge of internet hazards. Educational Gerontology, 36(3):173–192, 2010.

[37] Lesa Lorenzen-Huber, Mary Boutain, L. Jean Camp, Kalpana Shankar, and Kay H. Connelly. Privacy, Technology, and Aging: A Proposed Framework. Ageing Interna-tional, 36(2):232–252, December 2010.

[38] Helena M. Mentis, Galina Madjaroff, Aaron Massey, and Zoya Trendafilova. The Illusion of Choice in Discuss-ing Cybersecurity Safeguards Between Older Adults with Mild Cognitive Impairment and Their Caregivers. In ACM Conference on Computer-Supported Cooperative Work and Social Computing, CSCW '20, pages 164:1–164:19, Virtual Conference, October 2020. ACM.

[39] Pew Research 2021. Mobile Fact Sheet. Retrieved February 3, 2022 from https://www.pewresearch.org/internet/fact-sheet/mobile/

[40] M. Kowtko, "Using assistive technologies to improve lives of older adults and people with disabilities," 2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT). pp. 1-6, doi: 10.1109/LISAT.2012.6223205.

[41] Gallup 2015. The New Era of Communication Among Americans. Retrieved December 2, 2021 from https://news.gallup.com/poll/179288/new-era-communication-americans.aspx

[42] Greater Senior Living 2022. Tablets and Computers for Seniors: The Only Guide You Need. Retrieved April 27, 2022 from https://www.greatseniorliving.com/articles/tablets-and-computers

[43] Pearson K. 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philosophical Transactions A 373:253–318

[44] de Winter, J. C. F., Gosling, S. D., & Potter, J. 2016. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. Psychological Methods, 21(3), 273–290. https://doi.org/10.1037/met0000079

[45] SMS Bandits Phishing Service. Retrieved August 12, 2022 from https://krebsonsecurity.com/2021/02/u-k-arrest-in-sms-bandits-phishing-service/

[46] Nicholas Micallef and Nalin Asanka Gamagedara Arachchilage. 2021. Understanding users' perceptions to improve fallback authentication. Personal Ubiquitous Comput. 25, 5 (Oct 2021), 893–910. https://doi.org/10.1007/s00779-021-01571-y

[47] Hang A, De Luca A, Hussmann H (2015) I know what you did last week! do you? dynamic security questions for fallback authentication on smartphones. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp 1383–1392

[48] CWE-640 Weak Password Recovery Mechanism for Forgotten Password. Retrieved July 27, 2022 from https://cwe.mitre.org/data/definitions/640.html

[49] Albayram Y, Khan MMH (2016) Evaluating smartphone-based dynamic security questions for fallback authentication: a field study. Hum-Centric Comput Inf Sci 6(1):16

[50] How the Coronavirus Outbreak Has and Hasn't - changed the way Americans Work. Pew Research. Retrieved August 27, 2022 from https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirus-outbreak-has-and-hasnt-changed-the-way-americans-work/

[51] Anvari A, Pan L, Zheng X (2020) Generating security questions for better protection of user privacy. Int J Comput Appl 42(4): 329–350

[52] Han JK, Bi X, Kim H,Woo SS (2020) Passtag: A graphical-textual hybrid fallback authentication system. In: Proceedings of the 15th ACM Asia Conference on Computer and Communications Security, pp 60–72

[53] The future of secure work for people and organizations, Dashlane Report, Retrieved September 3, 2022 from https://www.dashlane.com/resources/the-future-of-secure-work-for-people-and-organizations?utm_source=website&utm_medium=blog&utm_campaign=SMBTrendsReport

[54] Xu S, Chan A, Lorber MF, Chase JP (2020) Using security questions to link participants in longitudinal data collection. Prev Sci 21(2):194–202

[55] Joseph Bonneau, Elie Bursztein, Ilan Caron, Rob Jackson, and Mike Williamson. 2015. Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google. In Proceedings of the 24th International Conference on World Wide Web (WWW '15). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 141–150. https://doi.org/10.1145/2736277.2741691

[56] Rabkin, Ariel: Personal knowledge questions for fallback authentication: security questions in the era of Facebook. In: SOUPS '08. ACM, July 2008.

[57] M. A. Sasse, M. Steves, K. Krol, and D. Chisnell, "The Great Authentication Fatigue – And How to Overcome It," in International Conference on Cross-Cultural Design, ser. CCD '14. Heraklion, Crete, Greece: Springer, Jun. 2014, pp. 228–239.

[58] E. Stobert and R. Biddle, "The Password Life Cycle: User Behaviour in Managing Passwords," in Symposium on Usable Privacy and Security, ser. SOUPS '14. Menlo Park, California, USA: USENIX, Jul. 2014, pp. 243–255.

[59] Pal, B.; Daniel, T.; Chatterjee, R.; Ristenpart, T.: Beyond Credential Stuffing: Password Similarity Models Using Neural Networks. In: SP '19. IEEE, May 2019.

[60] Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, and Kent Seamons. A Usability Study of Five Two-Factor Authentication Methods. In Symposium on Usable Privacy and Security, SOUPS '19, pages 357–370, Santa Clara, California, USA, August 2019. USENIX.

[61] Account Take Over, Retrieved August 2, 2022 from https://www.transmitsecurity.com/blog/the-rise-of-account-takeovers-and-how-to-prevent-them

[62] Account Take Over, Retrieved August 3, 2022 from https://spycloud.com/solutions/targeted-attacks/

[63] FIDO Alliance White Paper: Multiple Authenticators for Reducing Account-Recovery Needs for FIDO-Enabled Consumer Accounts, Retrieved March 14, 2022 from https://fidoalliance.org/white-paper-multiple-authenticators-for-reducing-account-recovery-needs-for-fido-enabled-consumer-accounts/

[64] P. G. Inglesant and M. A. Sasse, The True Cost of Unusable Password Policies: Password Use in the Wild, in ACM Conference on Human Factors in Computing Systems, ser. CHI '10. Atlanta, Georgia, USA: ACM, Apr. 2010, pp. 383–392.

[65] S. Foo, S. C. Hui, P. C. Leong, and S. Liu, An Integrated Help Desk Support for Customer Services Over the World Wide Web – A Case Study, Computers in Industry, vol. 41, no. 2, pp. 129–145, Mar. 2000.

[66] The Coronavirus Outbreak Has Become the World's Largest Work-From-Home Experiment, TIME, Retrieved January 2, 2022 from https://time.com/5776660/coronavirus-work-from-home/

[67] Y. Song, M. Ben Salem, S. Hershkop, S. J. Stolfo, System level user behavior biometrics using fisher features and gaussian mixture models, in: Security and Privacy Workshops (SPW), 2013 IEEE, IEEE, 2013, pp. 52-59.

[68] B. Biggio, G. Fumera, and F. Roli. Adversarial pattern classiffication using multiple classiffers and randomization. In Lecture Notes in Computer Science, pages 500-509, 2008.

[69] J. G. Dutrisac and D. Skillicorn. Hiding clusters in adversarial settings. In IEEE Int'l Conf. on Intelligence and Security Informatics (ISI 2008), pages 185–187, 2008.

[70] K. Wang, J. J. Parekh, and S. J. Stolfo. 2006. Anagram: a content anomaly detector resistant to mimicry attack. In Proceedings of the 9th international conference on Recent Advances in Intrusion Detection (RAID'06). Springer-Verlag, Berlin, Heidelberg, 226–248.

[71] L. Chen, H. Hassani, A. Karbasi. Near-optimal Active Learning of Halfspaces via Query Synthesis in the Noisy Setting. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17). AAAI Press, 1798–1804.

[72] B. Biggio, G. Fumera, and F. Roli. Security evaluation of pattern classifiers under attack. IEEE Trans. on Knowledge and Data Eng., 99(PrePrints):1, 2013.

[73] U. von Luxburg. Clustering stability: An overview. Foundations and Trends in Machine Learning, 2(3):235–274, 2010.

[74] V. Patel, R. Chellappa, D. Chandra and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges", IEEE Signal Process. Mag., vol. 33, no. 4, pp. 49-61, Jul. 2016.

[75] H. Zheng, Q. Ye, H. Hu, C. Fang, and J. Shi. Bdpl: A boundary differentially private layer against machine learning model extraction attacks. In Proc. of ESORICS, volume 11735 of Lecture Notes in Computer Science, pages 66–83. Springer, 2019.

[76] K. Yoshida, T. Kubota, M. Shiozaki and T. Fujino, "Model-Extraction Attack Against FPGA-DNN Accelerator Utilizing Correlation Electromagnetic Analysis," 2019 IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), San Diego, CA, USA, 2019, pp. 318-318.

[77] L. Huang, A. D. Joseph, B. Nelson, B. Rubinstein, and J. D. Tygar. Adversarial machine learning. In 4th ACM Workshop on Artificial Intelligence and Security (AISec 2011), pages 43–57, Chicago, IL, USA, 2011.

[78] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar. Can machine learning be secure? In ASIACCS '06: Proc. 2006 ACM Symposium on Information, Computer and Communications Security, pages 16–25, NY, USA, 2006. ACM.

[79] I. M Alabdulmohsin, X. Gao, and X. Zhang. Adding robustness to support vector machines against adversarial reverse engineering. In CIKM, 2014. 16. R. Shokri, M. Stronati, and V. Shmatikov, "Membership inference attacks against machine learning models," CoRR, vol. abs\/1610.05820, 2016.

[80] R. Shokri, M. Stronati, and V. Shmatikov, “Membership inference attacks against machine learning models,” CoRR, vol. abs/1610.05820, 2016.

[81] G. Ateniese, G. Felici, L. V. Mancini, A. Spognardi, A. Villani, D. Vitali. Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers. IJSN 10, 3 (2015), 137–150.

[82] E. Bigdeli, B. Raahemi, M. Mohammadi and S. Matwin, "A fast noise resilient anomaly detection using GMM-based collective labelling," 2015 Science and Information Conference (SAI), London, 2015, pp. 337-344.

[83] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z Berkay Celik, and A. Swami. Practical black-box attacks against machine learning. In Asia CCS, 2017b

[84] S. Ben-David, U. von Luxburg, D. Pal. “ A sober Look at Clustering Stability” Learn. Theor., no. 4005, 5–19, 2006

[85] E. Levine et al. Resampling method for unsupervised estimation of cluster validity. Neural Computation, 13(11):2573 – 2593, 2001

[86] U. Moller et al. A cluster validity approach based on nearest neighbor resampling. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR), pages 892–895, Washington, DC, USA, 2006. IEEE Computer Society.

[87] Y. Vorobeychik, B. Li. 2014. Optimal randomized classification in adversarial settings. In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, 485–492.

[88] R.Schumann, I. Rehbein, Active Learning via Membership Query Synthesis for Semi-supervised Sentence Classification, in Association for Computational Linguistics (ACL), 2019, pp. 472-481

[89] M. Barreno, B. Nelson, A. D. Joseph, and J. Tygar. The security of machine learning. Machine Learning, 81(2):121 148, 2010

[90] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar, “Adversarial machine learning,” in Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, ser. AISec ’11. New York, NY, USA: ACM, 2011, pp. 43–58.

[91] L.Batina, S. Bhasin, D. Jap and S. Picek, “ CSI Neural Network: Using Side-Channels to recover Your Artificial Neural Netwrok Information”. arXiv:1810.09076v1, 2018.

[92] G. Menardi. 2011. Density-based Silhouette diagnostics for clustering methods. Statistics and Computing 21, 3 (July 2011), 295-308.

[93] P. Fogla, and W. Lee. Evading network anomaly detection systems:Formal reasoning and practical techniques. In ACM Conference on Computer and Communications Security(2006),pp.59–68.

[94] D. Lowd and C. Meek. Adversarial learning. In KDD, 2005.

[95] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart. 2016. Stealing Machine Learning Models via Prediction APIs. In USENIX Security.

[96] P. W. Koh, J. Steinhardt, and P. Liang. 2018. Stronger Data Poisoning Attacks Break Data Sanitization Defenses.

[97] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, \& V. C. Leung (2018). A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View. IEEE Access, 6, 12103-12117.

[98] R. N. Reith, T. Schneider, and O. Tkachenko. Efficiently stealing your machine learning models. In Proc. of WPES, pages 198–210. ACM, 2019.

[99] B. Wang and N. Z. Gong, "Stealing Hyperparameters in Machine Learning," in Proceedings of the 2018 IEEE Symposium on Security and Privacy (S\&P). IEEE, 2018.

[100] S. J. Oh, M. Augustin, B. Schiele, and M. Fritz, "Towards Reverse- Engineering Black-Box Neural Networks," in Proceedings of the 2018 International Conference on Learning Representations (ICLR), 2018.

[101] T. Zhang, S. S. M. Chow, Z. Zhou, and M. Li. 2016. Privacy-Preserving Wi-Fi Fingerprinting Indoor Localization. In International Workshop on Security., 2016

[102] M. Kesarwani, B. Mukhoty, V. Arya, and S. Mehta. 2018. Model Extraction Warning in MLaaS Paradigm. Computer Security Applications (2018).

[103] T. Orekondy, B. Schiele, M. Fritz. Prediction poisoning: Towards defenses against DNN Model Stealing Attacks. International Conference on Representation Learning (ICLR) (2020)

[104] J. E. Tapiador and J. A. Clark, "Information-Theoretic Detection of Masquerade Mimicry Attacks," 2010 Fourth International Conference on Network and System Security, Melbourne, VIC, 2010, pp. 183-190.

[105] T. Takemura, N. Yanai and T. Fujiwara, "Model Extraction Attacks against Recurrent Neural Networks," arXiv:2002.00123v1, 2020

[106] S. Szyller, B. G. Atli, S. Marchal, and N. Asokan. DAWN: Dynamic Adversarial Watermarking of Neural Networks. arXiv:1906.00830v3. 2019.

[107] B. G. Atli, S. Szyller, M. Juuti, S. Marchal, and N. Asokan. Extraction of complex dnn models: Real threat or boogeyman? arXiv:1910.05429v2. 2019.

[108] Bank My Cell. Retrieved on July 17, 2019, from https://www.bankmycell.com/blog/how-many-phones-are-in-the-world

[109] P. Laskov, M. Kloft.: A framework for quantitative security analysis of machine learning. In: AISec '09: Proc. of the 2nd ACM works. on Sec. and art. int. pp. 1–4. ACM, New York, NY, USA (2009)

[110] B. Nelson, B. I. Rubinstein, L. Huang, A. D. Joseph, S.-h. Lau, S. J. Lee, S. Rao, A. Tran, and J. D. Tygar, "Near-optimal evasion of convex inducing classifiers." in AISTATS, 2010.

[111] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. ˇ Srndi´c, P. Laskov, G. Giacinto, and F. Roli, "Evasion attacks against machine learning at test time," in ECML-PKDD. Springer, 2013.

[112] B Settles. Active learning literature survey. University of wisconsin-madison, madison, wi, 2009. Technical report, CS Tech

[113] L. Wang, X. Hu, B. Yuan, and J. Lu. Active learning via query synthesis and nearest neighbour search. Neurocomputing, 147:426 434, 2015.

[114] R. Murmuria, A. Stavrou, D. Barbará and D. Fleck, "Continuous authentication on mobile devices using power consumption touch gestures and physical movement of users", Proc. Int. Workshop Recent Adv. Intrusion Detection, pp. 405-424, 2015.

[115] N. Carlini, D. Wagner. Towards evaluating the robustness of neural networks. In Security and Privacy (SP), 2017 IEEE Symposium on, pages 39-57. IEEE, 2017.

[116] Z. Sitová et al., "HMOG: New behavioral biometric features for continuous authentication of smartphone users", IEEE Trans. Inf. Forensics Security, vol. 11, no. 5, pp. 877-892, May 2016.

[117] R. Feinman, R. R. Curtin, S. Shintre, and A. B. Gardner. 2017. Detecting Adversarial Samples from Artifacts. ArXiv e-prints (March 2017). arXiv:stat.ML\/1703.00410

[118] X. Wang, T. Yu, O. Mengshoel, and P. Tague. 2017. Towards continuous and passive authentication across mobile devices: an empirical study. In Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '17). ACM, New York, NY, USA, 35-45.

[119] A. Dmitrenko. 2018. DNN Model Extraction Attacks using Prediction Interfaces.

[120] Y. Li, H. Wang and K. Sun, "Email as a Master Key: Analyzing Account Recovery in the Wild," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 2018, pp. 1646-1654, doi: 10.1109/INFOCOM.2018.8486017.

[121] A. Senarath, N. Arachchilage and B. Gupta, “Security strength indicator in fallback authentication: Nudging users for better answers in secret question,” International Journal for Infonomics, vol. 9, no. 4, pp. 533-537, January 2017.

[122] N. Micallef and M. Just, “Using avatars for improved authentication with challenge questions.” in Proceedings of 5th International Conference on Emerging Security Information, Systems and Technologies (SECURWARE), pp. 121-124, France, August 2011.

[123] S. Schechter, S. Egelman, and R. W. Reeder, “It’s Not What You Know, But Who You Know: A Social Approach to Last-Resort Authentication,” in ACM Conference on Human Factors in Computing Systems, ser. CHI ’09. Boston, Massachusetts, USA: ACM, Apr. 2009, pp. 1983–1992.

[124] Alshaikh M. 2020. Developing cybersecurity culture to influence employee behavior: a practice perspective. Computers & Security 98:102003 DOI 10.1016/j.cose.2020.102003.

[125] Han JK, Bi X, Kim H,Woo SS (2020) Passtag: A graphical-textual hybrid fallback authentication system. In: Proceedings of the 15th ACM Asia Conference on Computer and Communications Security, pp 60–72

[126] Javed A, Bletgen D, Kohlar F, D¨urmuth M, Schwenk J (2014) Secure fallback authentication and the trusted friend attack. In: 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, pp 22–28

[127] Xu S, Chan A, Lorber MF, Chase JP (2020) Using security questions to link participants in longitudinal data collection. Prev Sci 21(2):194–202

[128] Denning T, Bowers K, Van Dijk M, Juels A (2011) Exploring implicit memory for painless password recovery. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp 2615–2618

[129] Schechter S, Reeder RW (2009) 1+ 1= you: measuring the comprehensibility of metaphors for configuring backup authentication. In: Proceedings of the 5th Symposium on Usable Privacy and Security, pp 1–31

[130] WEIR, M., AGGARWAL, S., DE MEDEIROS, B., AND GLODEK, B. Password cracking using probabilistic context-free grammars. In IEEE Security & Privacy (2009).

[131] Schechter S, Brush AJBernheim, Egelman S (2009) It's no secret. measuring the security and reliability of authentication via "secret" questions. In: 2009 30th IEEE Symposium on Security and Privacy. IEEE, pp 375–390

[132] Blaine Nelson, Benjamin Rubinstein, Ling Huang, Anthony Joseph, Shing-hon Lau, Steven Lee, Satish Rao, Anthony Tran, and Doug Tygar. Near-optimal evasion of convex-inducing classifiers. In AISTATS, 2010.

[133] Google Account Recovery process, Retrieved August 16, 2022 from https://support.google.com/accounts/answer/9412469?hl=en

[134] Ducktail Criminal Group, Retrieved September 3, 2022 from https://www.scmagazine.com/analysis/identity-and-access/ducktail-criminal-group-targets-facebook-business-with-malware-to-take-over-accounts

[135] Social Media Account Scammers, Retrieved September 3, 2022 from https://www.komando.com/social-media/social-media-recovery-scams/823369/

[136] Sketchy Account Recovery Services, Retrieved September 3, 2022 from https://www.vice.com/en/article/k7w39x/account-recovery-service-twitter-hacked-instagram-coinbase

[137] Facebook phishers threaten users with Page Recovery Help Support, Retrieved September 3, 2022 from https://www.malwarebytes.com/blog/news/2022/04/facebook-phishers-threaten-users-with-page-recovery-help-support

[138] Joshua Reynolds, Nikita Samarin, Joseph Barnes, Tay-lor Judd, Joshua Mason, Michael Bailey, and Serge Egelman. Empirical Measurement of Systemic 2FA Usability. In USENIX Security Symposium, SSYM '20, pages 127–143, Virtual Conference, August 2020. USENIX.

[139] IRS abandons facial recognition plan after firestorm of criticism, Retrieved September 2, 2022 from https://www.washingtonpost.com/technology/2022/02/07/irs-idme-face-scans/

[140] 5 Layers to Authenticating Users and Mitigating Fraud, Retrieved October 19, 2022 from https://www.okta.com/resources/infographic-5-layers-to-authenticating-users-and-mitigating-fraud/

[141] S.M. Bellovin. 2015. Thinking Security: Stopping Next Year's Hackers. Addison-Wesley.