Just Take It! Exploring 2FA Usability Challenges Through Authentication Log Analysis

Dylen Greenenwald
University of Illinois Urbana-Champaign
Champaign, IL, USA
dgree21@illinois.edu

Sameer Patil
University of Utah
Salt Lake City, UT, USA
sameer.patil@utah.edu

ABSTRACT

Two factor authentication is vital for organizational security. Without it, uninformed users' account security and privacy is naked and primed for abuse by malicious actors. Despite its theortetical merit, most large-scale implementations of two factor authentication have been known to impose substantial burdens on usability. Without a usable system to ensure user ordinary security and privacy, account risk remains as high as ever due to lowered guard as a result of user frustration.

In this study, we provide a secondary analysis of a large-scale dataset. Here, we analyze approximatly 96 million logs to further identify the usability challenges encountered by typical users at large public universities. We find that ordinary users frequently encounter issues with the two factor authentication process, creating long failure sequences as a result of troubles accessing university systems. We learn that simple mitigations may have a large impact on user frustration and organizational security at large.

ACM Reference Format:
Dylen Greenenwald and Sameer Patil. 2024. Just Take It! Exploring 2FA Usability Challenges Through Authentication Log Analysis. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Large organizations with tens of thousands of users face the important challenge of providing solid security for users of all types of backgrounds. Unfortunately, there doesn't exist a simple solution that spans such a wide range of demographics. Numerous approaches to organization-wide two factor authentication have been proposed in order to address this challenge, but they all face their own issues. Security keys impose the burden of keeping track of an often extraneous device while push notifications, authenticator apps, SMS and voice calls all require a user to have their phone equipped and handy at all times. While there may not currently exist a simple solution to ease these immediate burdens, various measures may be taken to lighten the authentication load on the average user. As such, we must throughly understand exactly what challenges are imposed by existing mechanisms.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Two factor authentication (2FA) is a long-standing mechanism against common attacks against user security and privacy. The general idea behind it is to combine something you know (e.g. credentials) with something you have (e.g. phone, security key, etc). This strategy is meant to prevent stolen credentials from being used by attackers who would have otherwise compromised a victim's account. While almost any online service runs at least *some* non-negligible risk under the threat of takeover, large-scale organizations (e.g. public universities) have an elevated level of risk. In addition to needing to protect basic personally identifiable information (PII), there are often subservices as part of these organizations that store highly sensitive user information (e.g. financial data, social security number). Sensitive data aside, many of these organizations have strong reputations to uphold; even the smallest breach or data leakage could have a drastic impact on revenue, enrollment, or other important metrics for organizational success. There are a variety of known attack vectors for malicious actors to steal a user's credentials, the most common of which is through *phishing*. Despite 2FA's existence to address this problem and its high success rate against known attackers (over 90% in [4]), data breaches happen all too often. This is because the implementation of these extra layer(s) of authentication can impose a burden on users, causing them to rush the authentication process and thereby lowering their guard to potential threats seeking to obtain their second factor.

Since account takeover has been such a long-standing threat, there have been many mechanisms proposed to solve this problem. Push notifications provide users a way to verify their login from their smart phones by selecting an option from a smart phone notification that indicates they are indeed the party attempting to authenticate with their account. Authenticator apps (e.g. LastPass, Google Authenticator) achieve the same end by providing a 6 digit numerical code that changes every 30-60 seconds, which a user must supply upon successful login. Security keys contain cryptographic signatures that enable users to verify their authentication through an external hardware root of trust, often touching a dedicated disk that generates a One-Time Passcode (OTP) derived from a Public Key Infrastructure (PKI) private key. Other, less secure, mechanisms include SMS ("Secure" Messaging System), and phone calls to provide codes that verify a user's login. These methods have been shown to be vulnerable to a variety of different attacks [5, 6].

In this work, we intend bridge the gap within existing two factor authentication implementations by applying a careful eye to the underyling causes for failures within existing systems. We detect and analyze a large number of failure sequences in a dataset of approximately 96 million authentication logs. Here, rather than

focusing on improving technical systems to improve account security, we look at usability challenges that may weaken the extra defense provided by 2FA. We find that many users face substantial usability challenges during various phases of 2FA deployment, with some suffering as many as 20 or more failed consecutive attempts at authenticating through their second factor.

2 EVALUATION

Here we present the dataset, our methodology, and several concrete findings resulting from our analysis.

2.1 Dataset

In order to contextualize our method and findings, we briefly provide pertinent information about our dataset here.

We reuse the data from Abbott and Patil [1]. This dataset consists of approximately 96 million log entries from Indiana University Bloomington, each of which contains numerous fields describing a unique attempt to verify correctly supplied university credentials through the provision of a second factor of authentication. The relevant fields include timestamp, user ID, factor, IP address, result, reason for result, device, country, state, and city. Note that there are a few extra fields included in the data, but seeing it was not utilized for this work, its discussion is unnecessary.

It is also important to note that this data was collected during multiple stages of 2FA deployment over the course of roughly two years. These stages have differing characteristics (e.g., optional 2FA, mandatory for select systems, and mandatory for all systems). Due to time constraints, we were not able to take these into account for our analysis. Still, these three distinct periods are likely to have differing usability challenges, especially when user familiarity with the process is taken into account.

2.2 Methodology

Given the relatively short timeline of this project, our analysis focused on the detection, processing, and analysis of *failure sequences*. We define a **failure sequence** as a series of "rapid" login attempts for a single user account, all of which result in failing the second factor of authentication with the potential exception of the final attempt in the series.

We utilize the Pandas [7] library toward the end of effectively managing such a large dataset. We iteratively process the logs, one (variably sized) `.csv` file at a time. Unfortunately, we were unable to consistently process the data due to the exploratory and short-term nature of this project. The remainder of the methodology was developed using a "testing" data processing strategy wherein we use this inconsistent lazy loading technique.

Since the data is organized by timestamp, we must group the login entries in order to perform a user-specific failure sequence analysis. Thus, after loading a log file into a data frame, we utilize the 'groupby' API, detecting failure sequences one user at a time. Note that for our purposes, a "rapid" login attempt is defined as one happening within 30 minutes of the previous attempt by that user; we refer to this window as the failure buffer throughout the rest oft he paper. We record data as we go, tracking whether or not the failed login attempts are consistent (i.e., they utilize the same second factor, and/or originate from the same device, location, IP address). We extract all of this information so we can obtain a fuller picture of specific usability challenges not captured by previous work (e.g. [1, 2, 8]). After processing all failure sequences, we then analyze all of the failure sequence results as an aggregate whole, encoding temporal patterns across the entire dataset.

2.3 Results

As previously noted, the most major limitation of this study is the confined timeframe to conduct the research. To reiterate, the caveat that qualifies these results (or the lack thereof) is the fact that the project idea was not settled until the tail end of the REU experience (approximately week 6 of the 10 week program). With that said, we report here what we do find. Additionally, we discuss potential implications and mitigations here rather than a dedicated discussion section.

Perhaps the most surprising finding is the largest number of failures in a single sequence - 23. This statistic, by itself, paints a dire picture of usability challenges faced by users during early adoption of 2FA at IUB. We posit that similar patterns are likely to be present in other large public universities, and potentially other large organizations due to the heterogenous nature (primarily with regard to technical background) of demographics of these institutions.

Additionally, we find that consistency across attempts within a single failure sequence is quite low. Approximately 27% of attempts within a single failure sequence utilize the same factor (i.e., the mechanism being used to verify the login attempt). Furthermore, only 38% originate from the same device. These figures suggest that users attempt to remedy usability challenges by switching the means by which they authenticate. This indicates that there was something inherently difficult about the process itself, independent of the choice of factor or the user's chosen machine.

By contrast, we also find that over 97% of login attempts within a single failure sequence originate from the same IP address and location. This is to be expected as it is unlikely that a user is going to travel when they need to access a particular university resource.

3 RELATED WORK

Multifactor authentication (MFA) has been the defacto framework for ensuring organizational security for a long time. Related work can be broadly divided into its security and usability.

3.1 Security of Multifactor Authentication

There have been a myriad of different studies that evaluate purely the security of multifactor authentication systems. Doerfler et al. studied Google's implementation of risk-based authentication [4], showing that additional authentication challenges have over 90% success against known attackers compared to single factor authentication systems. Lee et al. examined vectors for authentication circumvention when contacting cellular carriers [5], finding that insecure procedures were commonplace and easily exploitable. Mulliner et al. looked at the evolution of SMS-based one-time passwords (OTPs) and found that, due to systemic architectural changes, they are no longer secure and needed improvement [6]. As a result, authentication systems that use SMS-based OTPs as a second factor have undermined security.

3.2 Usability of Multifactor Authentication

As various systems for multifactor authentication are adopted by large organizations, whether it be universities or corporate institutions, the amount of research on its usability has been increasingly studied in recent years. Das et al. studied the usability of Yubico's Yubikey, a universal second factor (U2F) security key [3]. They found that despite hypothetical merit, users struggled with simple authentication ceremonies involving the Yubikey. Weidman and Grossklags gauged the usability of a large-scale transition from hardware tokens to personal devices as a secondary authentication factor [9], finding no major differences in usability.

3.2.1 Log Analysis Studies. There is an enormous body of existing research on two factor authentication, and a similarly large subset of this focuses on usability. Despite this, few studies on usability interpret objective data through programmatic analysis to draw conclusions. Reynolds et al. conducted a multifaceted study on Duo 2FA usability by combining qualitative survey-based analysis with a concrete analysis of log data across two large public US universities [8]. Abbott and Patil performed a similar analysis, differing primarily in the period over which the logs were recorded and the fact that it focused solely on a single large public US university. Our work differs from other existing studies on log analysis by capturing temporal patterns from the logs rather than aggregate, row-by-row statistics, providing a more comprehensive picture of usability challenges prevalent in university 2FA systems through log data analysis.

4 CONCLUSIONS

Usability in security is of paramount importance, as users are often the weakest link in any given security system. As such, we must strengthen this link by providing ease of use. Multifactor authentication is no exception to this rule.

To the best of our knowledge, we present the first temporally-based analysis of authentication log data on a large-scale university over an extended period of time. We find that users experience excessive failures when attempting to access various university services, in spite of various attempts made to overcome device- and factor-specific obstacles. While our findings are limited, we are confident that continued work on this dataset, and in this space in general, will result in a renewed understanding of simple mitigations for common usability challenges related to multifactor authentication.

REFERENCES

[1] Jacob Abbott and Sameer Patil. 2020. How Mandatory Second Factor Affects the Authentication User Experience. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376457

[2] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. 2018. It's not actually that horrible: Exploring Adoption of Two-Factor Authentication at a University. 1–11. https://doi.org/10.1145/3173574.3174030

[3] Sanchari Das, Andrew Dingman, and L. Camp. 2018. *Why Johnny Doesn't Use Two Factor A Two-Phase Usability Study of the FIDO U2F Security Key.* 160–179. https://doi.org/10.1007/978-3-662-58387-6_9

[4] Periwinkle Doerfler, Kurt Thomas, Maija Marincenko, Juri Ranieri, Yu Jiang, Angelika Moscicki, and Damon McCoy. 2019. Evaluating Login Challenges as a Defense Against Account Takeover. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 372–382. https://doi.org/10.1145/3308558.3313481

[5] Kevin Lee, Benjamin Kaiser, Jonathan Mayer, and Arvind Narayanan. 2020. An Empirical Study of Wireless Carrier Authentication for SIM Swaps. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association, 61–79. https://www.usenix.org/conference/soups2020/presentation/lee

[6] Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean-Pierre Seifert. 2013. SMS-Based One-Time Passwords: Attacks and Defense - (Short Paper). In *International Conference on Detection of intrusions and malware, and vulnerability assessment.* https://api.semanticscholar.org/CorpusID:11866425

[7] Pandas. 2024. https://pandas.pydata.org/

[8] Joshua Reynolds, Nikita Samarin, Joseph Barnes, Taylor Judd, Joshua Mason, Michael Bailey, and Serge Egelman. 2020. Empirical Measurement of Systemic 2FA Usability. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 127–143. https://www.usenix.org/conference/usenixsecurity20/presentation/reynolds

[9] Jake Weidman and Jens Grossklags. 2017. I Like It, but I Hate It: Employee Perceptions Towards an Institutional Transition to BYOD Second-Factor Authentication. In *Proceedings of the 33rd Annual Computer Security Applications Conference* (Orlando, FL, USA) *(ACSAC '17)*. Association for Computing Machinery, New York, NY, USA, 212–224. https://doi.org/10.1145/3134600.3134629