Quantifying costs of enhanced security in multifactor authentication

Seth Hastings[1], Tyler Moore[*1], Neil Gandal[1,2], and Noa Barnir[2]

[1]*College of Engineering and Computer Science, The University of Tulsa, 800 S. Tucker Dr., Tulsa, OK, 74104, USA*

[2]*Berglas School of Economics, Tel Aviv University, Chaim Levanon St 55, Tel Aviv-Yafo, 6997801, Israel*

Abstract

Multifactor authentication (MFA) is one of the most important security controls, topping most lists of cyber hygiene activities advocated by experts. While the security benefits may be substantial, less attention has been paid to the impact on users by the added friction introduced by the more stringent precautions. In this paper, we construct and analyze a dataset of authentication logs from a University population spanning two years. We focus on opportunity costs experienced by users: (1) log-in failures and (2) the time spent away from IT applications following a failed authentication before attempting to re-authenticate. The second measure captures how user frustration can manifest by avoiding or delaying future engagement after experiencing failures. Following an exogenous change in MFA policy from a deny/approve mobile notification to a more cumbersome two-digit code mobile notification confirmation, we show that there are significant increases in the number of log-in failures and in time spent away following failures when using mobile MFA. We also briefly examine which types of users had the greatest difficulty adjusting to the more secure mobile MFA procedure.

We gratefully acknowledge support from the US National Science Foundation (NSF) Award No. 2147505 and Award No. 2452738 and the US Israel Binational Science Foundation (BSF) Award No. 2021711. We thank the editor and two referees whose comments and suggestions greatly improved the paper. We also thank seminar participants at Boston University, Hebrew University, and the Politecnico in Milano for helpful comments and suggestions.

[*]Forthcoming, Information Systems Frontiers 2025. Corresponding author: tyler-moore@utulsa.edu

1 Introduction

In response to growing threats and increased regulatory pressures, organizations have sought to strengthen their cybersecurity posture. They are allocating more resources towards cybersecurity initiatives, and adopting new security controls to mitigate elevated risks. Such investments have undoubtedly brought benefits in terms of reduced exposure to attacks. However, increased security can also introduce opportunity costs. Some legitimate tasks may now be blocked, from emails mistakenly caught in a spam filter to accounts being locked out following mandatory password changes. Additionally, even when working properly, security controls introduce friction that can slow task performance and frustrate users. Such opportunity costs are often overlooked, but they are critically important because they may add up to substantial losses and can even alter behavior to be less productive or secure.

As organizations seek to strengthen their cybersecurity posture, changes often come first to how authentication works. The Cybersecurity and Infrastructure Security Agency (CISA) recommends four critical steps individuals and organizations can take to strengthen security [6]. The first item of the four is to "turn on multifactor authentication" (MFA). For individuals, the process can be as simple as tweaking a configuration setting. For firms, the process can be a bit more involved, as it requires changes to how enterprise IT infrastructure is configured and operated. Nonetheless, organizations are increasingly supporting MFA. Most often, they are actually mandating its use throughout the enterprise [1].

MFA provides an excellent opportunity to study the opportunity costs of cybersecurity controls. That is because authentication affects everyone and is highly visible to users. Moreover, MFA significantly alters the steps users must take to use an enterprise IT system. When MFA works well, it can be seamless. Enrolled users provide a second factor (often a mobile device) and carry on with their tasks as before. However, when users fail to authenticate, they cannot complete their intended task. This can happen because they forgot their second factor, got a new phone, or for a variety of other reasons. Correcting the problem can be time consuming and costly, often requiring manual assistance from IT staff.

While we fully expect that the benefits of MFA to outweigh the costs, the burden imposed is often not explicitly accounted for. In this paper, we empirically analyze the opportunity costs of MFA in a deployed setting. Once opportunity costs are identified, it becomes possible to take steps that minimize them. As we will show, choices in how the technology is deployed can greatly impact how users respond and the resulting magnitude of the costs imposed.

Increasingly, the technologies deployed by enterprises generate large amounts of "data exhaust" that could be mined for insights into user behavior [9]. We leverage a very large dataset of Microsoft Azure Active Directory sign-in logs (now known as Microsoft Entra ID) from a University between 2021–2023. Using these data, we examine the opportunity costs associated with the adoption of a more onerous multifactor authentication process. Critically, at the end of the 2021–2022 academic year, the University changed the MFA procedure for mobile use with the authenticator app from a deny/approve "push" notification to a more cumbersome two-digit code which needs to be entered into the authenticator app when prompted on the login screen. This was especially cumbersome for users using Mobile MFA who attempted to login from a mobile device. This is because both the authentication app and the login window had to be open at the same time and users had to switch between them. Figure 1 provides screenshots. This exogenous change allows us to examine the added costs associated with a more secure mobile MFA method. In the

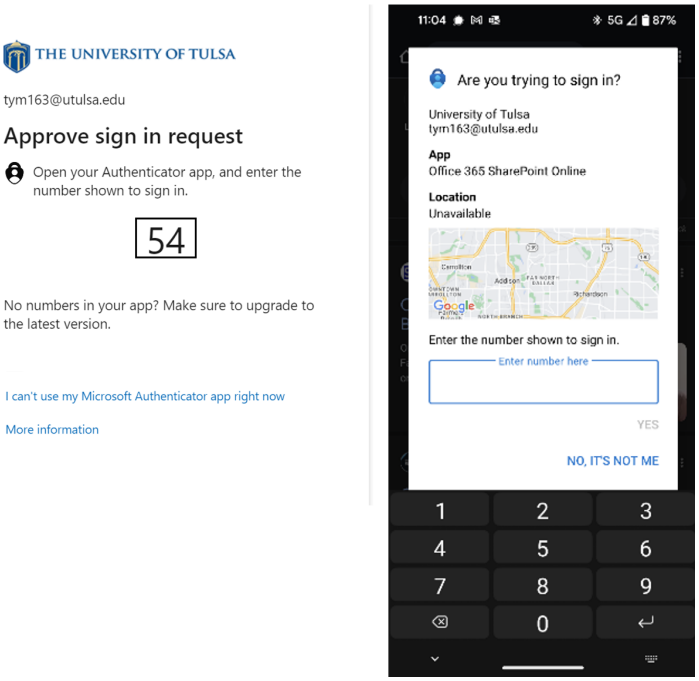

Figure 1: Changed MFA procedure. Interstitial prompt with two-digit number (left) and phone-based application for entering the code (right).

case of text messages, no change was made.

We focus on two measures that serve as proxies for increased opportunity costs associated with the change in MFA policy. (1) The first measure is the number of login failures users experience. (2) The university employs a single-sign-on system and tracks all authentication attempts to any university service. Hence, the second measure we employ is how long a user remains without access to IT resources following a failure. In particular, we measure the *time away* following a failed login until the user attempts to login again. If users become frustrated, distracted or disengaged after failing to authenticate, then they may take longer to reengage. Hence, both failed logins and time away are promising measures of the opportunity cost from onerous security measures.

We first report descriptive statistics. This section clearly shows that there were significant increases in the number of log-in failures and in time spent away following failures when using mobile MFA following the exogenous change. We then employ "fixed effects" econometric models to analyze how these costs changed over time.[1] The econometric results confirm the descriptive data "results" and provide us with estimates of the effect of changes in the MFA procedure on the number of failures and time away. Although we have very limited data on user characteristics, we do know the time of day for each attempted login. We find that users who were primarily active from 8:00 am to 5:00 pm during the week had the greatest difficulty adjusting to the new mobile MFA procedure. These users likely contain more staff members than faculty or students.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes our derived event logs, defining the most relevant attributes. Section 4 discusses our data and provides descriptive statistics. In section 5, we conduct the econometric analysis and provide our results.

[1]These models explicitly take into account that there are repeated observations on users. This enables us to examine how user costs increased from enhanced security changes to how MFA was deployed.

Section 6 briefly concludes.

2 Related Work

Existing work investigating multifactor authentication use has studied adoption rates, usability, and user attitudes towards the technology. This work is discussed in Tables 4 and 5 in the Appendix.

Our work is very different than that of the existing literature because we are interested in how an exogenous change in the MFA authentication process affected the number of login failures and the time away following a failure. Further our analysis is different, since we measure these costs using "fixed effects" econometric models. The previous work in the literature has not exploited the panel nature of the data (i.e., repeated observations of the same users). We are also interested in examining which type of users had trouble adjusting to the MFA change and examining changes in user behavior following the change in MFA policy.

3 Methodology for Constructing Authentication Events

Whereas most prior work studying authentication usage has surveyed users about their experience, we seek to go straight to the source: authentication logs. Through a partnership with the IT department at the authors' university we obtained access to anonymized Entra ID authentication logs for analysis, approved by the Institutional Review Board (IRB) under protocol 24-02.

Interpretable user experience is buried in raw security logs, with combinations of values for different attributes indicating meaningful states. A sign-in log entry contains around 36 attributes representing a single system interaction. Hence, a small period of user interaction can generate many log entries, sometimes dozens per minute. Critically, many of these entries represent back-end processes that users do not directly experience. By inspecting these logs carefully, we constructed a set of 38 *row codes* that capture critical information about an authentication attempt, and are used to characterize an attempted individual login. This allows us to discard irrelevant entries and consolidate significant interactions as we construct events. The construction methodology itself is described in detail in a technical paper ([10]). Hence, we do not discuss it here. We define an *event* as follows:

> The occurrences reflected in log data that are directly experienced by a user, beginning when an authentication to a particular application is initiated, and terminated upon the eventual success or failure of the authentication attempt.[2]

Each event captures the number of errors encountered before eventual success or failure, as well as the type of errors involved, the type of authentication used, as well as whether the attempted login was from a desktop/laptop or a mobile device

During a login attempt, a user can experience one or more errors, from misconfigurations to failed passwords or MFA prompts, before ultimately succeeding in the authentication. Errors are assigned to three primary categories: User and Configuration Errors which are split by attribution, and Interrupts. User Errors are those error codes generated by invalid or missing user input, such

[2]If there is a lapse of activity great than 90 seconds, we also define this as a failure. The results are robust to changing the length of the lapse in activity.

as failure to answer an MFA prompt or incorrect password entry. Configuration Errors encompass errors that are not due to user error, such as developer errors or issues with the user's account status. Interrupts occur when the system needs to take further action during an authentication flow, such as when the token presented has expired, and the user must be redirected to use their second factor. These "Interrupt Errors" do not indicate adverse events or impediment to normal usage flows, and instead serve as flags for various operations. We also track the number of times the user input their password during the authentication event. Time Away measures the gap in time between a failed authentication to a service and the next attempted login[3].

4 Data and Descriptive Statistics

4.1 Time Periods for Analysis

Our data is from November 15 2021 to May 31 2023. We divided the data into four periods:

- Academic year 2021–22: from (November 15 2021 — May 31 2022)
- Summer 2022 (June 1 2022 — August 15 2022)
- Early Academic year 2022–23: (August 16 2022 — November 14 2022)
- Academic year 2022–23: from (November 15 2022 — May 31 2023)

In the analysis, we employ data from the two "partial" academic years covering the November 15 to May 31 period to keep the dates consistent across samples. Our results are qualitatively unchanged if we include the data from August 16 2022 — November 14 2023 in the 2022-2023 academic year.

We examine what happened to the number of login failures and "Time Away" (TA) following the 2021–2022 academic year. This made authentication more secure, but with a "cost" in that authentication became more complicated. From our standpoint, this yields a natural experiment and enables us to compare the before and after periods and the effect of an (exogenous) increase in mobile MFA authentication procedures on Time Away and the number of failures.

We are particularly interested in how this change affected time away and the failure rate. The explanatory variables (factors) we employ in the analysis are discussed when we present our models.

4.2 Descriptive Statistics

The first step when analyzing a large data set is to cut the data in many ways and look for patterns. When we examined the data by academic year at the event level, we were struck by the significant increase in TA and the number of failures during the 2022-23 academic year for attempted logins using mobile MFA relative to the 2021-2022 academic year.

Following a discussion with the University IT department, we learned that following the 2021-22 academic year, the mobile MFA authentication process was changed. It was changed from a (1) push notification, where users simply had to approve or deny that they were trying to login to a (2) two-digit approval system requiring the user to enter a number shown in the login process into a mobile authenticator. The effect of this change is well illustrated by the descriptive statistics for

[3]Time away is similar to "recovery time" reported by [15], except our measure does not discriminate between successful and failed follow ups; it simply captures the gap between interactions after a failed login.

	21-22 academic year		22-23 academic year	
	Mean		Mean	
Time Away (minutes) - Mobile MFA	34.9	0	81.3	170.0
Time Away (minutes) - Text MFA	10.3	0	24.3	0
Failure Rate (Mobile MFA)	10.2%		17.9%	
Failure Rate (Text MFA)	2.6%		4.8%	

Table 1: Descriptive statistics: Event level Data

failures and time way at the event level. Below we show comparisons on these measures when(i) mobile MFA was employed and (ii)) when Text MFA was employed.

In Table 1, we report the descriptive data at the event level when mobile or text MFA is used.

Descriptive statistics at the event level for the mobile MFA login procedure show a very significant absolute and percentage increase in mean TA from approximately 35 minutes per event in the 2021-22 academic year to approximately 81 minutes per event in the 2022-23 academic year. More importantly, the table shows that the 90^{th} percentile of the distribution of TA increased dramatically from 0 minutes in the 2021-22 academic year to approximately 170 minutes per event in the 2022-23 academic year. Thus, a non-trivial percent of users have struggled with the enhanced MFA procedure for Mobile MFA.[4]

Table 1 shows that the failure rate (the percent of times an authentication attempt was not successful) increased significantly in the second academic year when the mobile MFA procedure changed: The failure rate with mobile MFA rose from 10.2 percent in the 2021-22 academic year to 17.9 percent in the 2022-23 academic year. This is a very large absolute increase.

Table 1 also shows that MFA using text messages is much less problematic for users and there was a much smaller change from the 2021-22 Academic year to the 2022-23 Academic year. The mean Time Away was approximately 10 minutes when using Text MFA login procedure in the 2021-22 academic year and approximately 24 minutes per event in the 2022-23 academic year.

Importantly, the 90^{th} percentile of the distribution of 'time away" for text messages was zero in both the 2021-22 academic year and the 2022-23 academic year. Additionally, the differences between these two methods in mean time away was 25 minutes (35-10) in the first academic year and 57 minutes in the second academic year.

While the failure rate was higher for Text MFA in the second period (4.8% in the second period vs. 2.6% in the first period), it was much lower than when Mobile MFA was used. Further, the differences between these two methods in the failure rate (by academic year) was 7.6% (10.2-2.6) in the first academic year and 13.1% (17.9.-4.8) in the second academic year. Hence, the difference nearly doubled in the second year.[5]

[4]The increased mean authentication delay (denoted elapsed), on the other hand, is virtually unchanged: From approximately 3.5 seconds per event in the 2021-22 academic year to approximately 4.2 seconds per event in the 2022-23 academic year. Hence, we do not focus on this variable as we noted in the introduction.

[5]Once we control (in the regressions) for whether the login attempt was from a mobile or desktop/laptop device, there is virtually no change in the failure rate between the periods when using Text MFA.

5 Econometric Analysis

We now turn to the formal analysis, in which we use (i) Time Away and (ii) log-in failures as the dependent (or response) variables. To ensure that our results are not due to new users, as discussed, we only include users that were active in both academic years. Since there is little change in faculty and staff users from year to year and since most undergraduate students are at University for four years, most of the users (around 90%) are repeat users.

5.1 Fixed Effect Models

We have panel data, that is, repeated observations on each individual. Having a panel rather than cross-sectional data (one data point on each individual) is advantageous, since a cross-section cannot control for time-invariant individual characteristics, like user attitudes towards risk. Such unobservable factors are included in the error term in cross-sectional analysis. If these unobserved effects are correlated with the right-hand-side variables of the estimation equation, the estimates from the cross-sectional analysis will be biased. However, we eliminate this problem by using fixed effect models. We now describe the fixed effect model.

The equation we start with is the following:

$$Y_{it} = \alpha_i + X_{it}\beta + \delta_t + \epsilon_{it}. \tag{1}$$

The dependent variable Y_{it} is (say) the sum of TA for user i at time t, where time is at the aggregated weekly level.[6]

The explanatory variables in X_{it} are observable time-varying factors that likely affect Time Away and β are coefficients to be estimated. The vector $\alpha_i = \alpha + A_i\eta$ is such that α is a constant and A_i is the vector of unobserved time-invariant user characteristics. An example is user attitudes towards risk. The key is that the user characteristics in the vector A_i is do not change over time. As we show below, we do not need to know the value of these characteristics in order to estimate the model. δ_t is the week effect. Finally, ϵ_{it} is an error term.

The following equation expresses the mean values at the level of the user, where the mean is computed over time from equation (1).

$$\bar{Y_i} = \alpha_i + \bar{X_i}\beta + \bar{\delta} + \bar{\epsilon_i} \tag{2}$$

Subtracting (2) from (1) yields:

$$Y_{it} - \bar{Y_i} = (X_{it} - \bar{X_i})\beta + (\delta_t - \bar{\delta}) + (\epsilon_{it} - \bar{\epsilon_i}) \tag{3}$$

Since the vector $\alpha_i = \alpha + A_i\eta$ does not depend on time, it drops out in equation (3) which are the deviations from the mean. Equation (3) is the fixed effects model we will estimate.[7]

We employ a variable (denoted "Post") in X_{it} that takes on the value zero if the data are in the first academic year and one if the data are in the second academic year. We interact "Post" with all of the other explanatory variables. In this way, we analyze both years together, which is preferred to estimating both years separately, since we can easily see the differences between the first and second year. Our results are robust to running separate regressions for each year.

[6] In fixed-effect analysis, the data must be in time periods (say a day or week).

[7] See Angrist (2009) for more a detailed discussion of fixed effects models [3].

5.2 Variables in the Analysis

The variables we employ (and their definitions) in the analysis (at the weekly level) are as follows:

- Dependent Variables:
  - Time Away (TA): The sum of the Time Away in minutes for that user during the period, which is a week in our analysis.[8]
  - The sum of the number of failures for that user during the week.

  Independent Variables:
- IEs: The number of Interrupt Errors during the period.
- CEs: The number of Configuration Errors during the period.[9]
- Text-MFA: The Number of Logins when a text message MFA procedure was used during the period.
- Mobile-MFA: The Number of Logins when a mobile app MFA procedure used during the period.
- Pw-uses: The number of Password Entries (whether correct or incorrect) during the period.
- Mobile entries is the number of attempted logins from a Mobile device.[10]
- Period: The week number
- Post is a binary variable that takes on the value zero if the data are in the first academic year and one if the data are in the second academic year. We interact Post with all of the independent variables.

We are mainly interested in how (i) the different MFA uses (Text message, Mobile app) and (ii) whether the user attempted to login from a mobile or desktop device affected (I) the number of failures and (II) "Time Away", the time in between a failed authentication attempt and the next attempt to authenticate. The other variables are primarily controls.

Overall in both academic years, 13 percent of attempted logins used text message MFA procedures, while 17 percent of attempted logins used mobile MFA procedures. The remainder of the attempted logins were primarily from a Remembered device. In many cases, when using a remembered device,the user did not have to use MFA.[11] The breakdown among these categories did not change from year to year.

The formal analysis is at the weekly level. The dependent variables are failures and TA, which is defined as the time in minutes between a failed login and the subsequent attempt to login for that user for each event in the week. We add the TA and number of failures as well as all independent variables for each event to get the totals at the weekly level for each user.

We employ a log/log functional form (which employs the natural logarithm (ln) of each variable). This functional form typically gives better results in terms of the explanatory power of the model when the variables employed have skewed distributions. This is true in our case as well, since the raw data is quite skewed.

[8]When calculating the mean TA for descriptive statistics, we limited TA to 1000 minutes. We do this so not to "distort" the means, as several values reach 14,000 minutes. In the regressions, we do not restrict TA. Because we have so many observations, and because we are running a log/log model, nothing in the results changes if we restrict Time Away to 1000 minutes in the econometric analysis.

[9]Nothing changes in the analysis if we combine the interrupt and configuration errors into one variable of "non-user" errors.

[10]This is regardless of whether text MFA or Mobile MFA was employed.

[11]In these case there are virtually no login failures. Less than two percent of total attempted logins used either Phone Call or OATH MFA procedures.)

The overall R-squared, which measures the explanatory power of the model (and ranges from 0 to 1) is 0.504 for the log/log model with Time Away as the dependent variable and 0.509 when using the number of failures as the dependent variable. [12]

In (natural) logarithm form, the variables are as follows:

- ln-TA is the natural logarithm of the sum of "Time Away" (ln(TA + .001)[13]
- In-Failures = ln(Failures + .001)
- ln-Ies = ln(Ies + .001)
- ln-Ces = ln(Ces + .001)
- ln-Rem-Device = ln(Rem-Device +.001)
- ln-Text-MFA = ln(Text MFA + .001)
- ln-Mobile-MFA = ln(Mobile MFA + .001)
- ln-Pwuses = ln(Pw-uses + .001)
- ln-Mobile-entries = ln(Mobile-entries + .001)

5.3 Regression Results: Time Away as the Dependent Variable

The results in the first column of Table 2 (in the appendix) show that in the case of Time Away, other things being equal, the estimated coefficient associated with the number of mobile MFA uses per week is positive (0.16). The Table shows that this result is statistically significant at the 99 percent level of confidence for the first academic year. That is, more mobile MFA login attempts per week, other things being equal, leads to significantly more Time Away in that week.

Strikingly, in the case of the 2022-23 academic year, the estimated coefficient associated with the number of mobile MFA uses is (0.25=0.16+0.09). The difference in the coefficient estimates between the two years (0.09) is statistically and economically significant as shown in the Table.

Since we are estimating a log-log model, this means that a 100 percent increase in the number of Mobile MFA uses leads to a 25 percent increase in Time Away in the 2022-23 Academic year vs. 16 percent in the 2021-2022 academic year. This means that, conditional on the same number of mobile MFA uses, there is significantly more Time Away in 2022-23 than in 2021-2022. Thus, controlling for other factors, the change in mobile MFA policy (which made it more secure) greatly increases the weekly Time Away when Mobile MFA is used, relative to the effect in the academic year 2021-22.

Importantly, the results show that, other things being equal, the estimated coefficient associated with the number of text messages MFA uses is much smaller in both academic years (0.027 in year one and 0.042 in year two). Other things being equal, there is a very small change in Time Away in the second period (relatively to the first period) when text MFA is employed.

In the case of attempting to login in from a mobile device (whether it is using text MFA or Mobile MFA), the number of attempted logins from a mobile device had virtually no effect on Time Away in the first period. (The estimated coefficient is 0.005.) However, this coefficient is much larger (0.094=0.005+0.089) in the second period reflecting the fact that logins from a mobile device became more cumbersome in the second period.

[12] Unsurprisingly, the overall R-squared is much lower when estimating linear/linear models. This was what we expected given the skewed distribution of the data.

[13] Since these variables can take on the value zero, we add a very small number (.001) in order to create the logarithms.Nothing changes if we add a slightly larger value than 0.001.

5.4 Regression Results: Number of Failures as the Dependent Variable

In the case of the number of failures as the dependent variable, the results are qualitatively the same. The results in the second column of Table 2 show that in the case when the dependent variable is the number of failures, other things being equal, the estimated coefficient associated with the number of mobile MFA uses per week is positive (0.10) and is statistically significant for the first academic year. That is, more mobile MFA login attempts per week, other things being equal, leads to significantly more Time Away in that week. In the case of the 2022-23 academic year, the estimated coefficient associated with the number of mobile MFA uses (0.14=0.10+0.04) is 40 percent larger than the coefficient associated for the 2021-2022 academic year. Again, the difference in the coefficient estimates between the two years is both statistically and economically significant. Thus, controlling for other factors, the change in mobile MFA policy (which made it more secure) greatly increases the number of failures when Mobile MFA is used, relative to the effect in the academic year 2021-22.

Similarly to the case when time away is the dependent variable, the results show that, other things being equal, the estimated coefficient associated with the number of text message MFA uses is much smaller in both academic years (0.022 in year one and 0.027 in year two) and that there is virtually no change in the second year.

In the case of attempting to login in from a mobile device (whether it is using text MFA or Mobile MFA), the number of attempted logins from a mobile device had virtually no effect on the number of failures in year one. (The estimated coefficient is 0.006.) The estimated coefficient associated with the number of attempted logins from a mobile device is much larger (0.051=0.006+0.045) in the second period. This again reflects the fact that logins from a mobile device became more cumbersome in the second period.

5.5 Different Types of Users

In this section we examine how different types of users were affected by the change in Mobile MFA policy. We do not know the identity of the users and do not know if they are faculty, staff or students. However, we can proxy for these groups. It is probably likely that many of the University staff primarily use the online system during work hours, which we defined to be 8:00am - 5:00m pm. Hence, we divided the users as follows:

- Group 1 - less than 1/3 of their logins in 2021-22 academic year occurred during "work hours".
- Group 2 - Between 1/3 and 2/3 of their logins in 2021-22 academic year occurred during "work hours".
- Group 3 - More than 2/3 of their logins in 2021-22 academic year occurred during "work hours".

It is likely that Group 3 consists includes much of the University staff, while Group 1 has a greater percentage of students and faculty members.

In the case of Time Away, other things being equal, the estimated coefficient associated with the number of mobile MFA uses per week is 0.14 for Group 1 and 0.19 for Group 3 in the first academic year. In the case of the 2022-23 academic year, the estimated coefficient associated with the number of mobile MFA uses is 0.22. (0.22=0.14+0.08) In the case of Group 3, the estimated coefficient associated with the number of mobile MFA uses in the second academic year is 0.30

(0.30=0.19+0.11). Thus the difference between the groups essentially nearly doubles in the second academic year from 0.05 (0.19-0.14) to 0.08 (0.30-0.22). Group 3 users had much greater difficulty adjusting to the new mobile MFA policy.

In the case of the Number of failures, other things being equal, the estimated coefficient associated with the number of mobile MFA uses per week is 0.09 for Group 1 and 0.12 for Group 3 in the first academic year. In the case of the 2022-23 academic year, the estimated coefficient associated with the number of mobile MFA uses is 0.12. (0.12=0.09+0.03) In the case of Group 3, however, the estimated coefficient associated with the number of mobile MFA uses in the second academic year is 0.17 (0.17=0.12 +0.05). Thus the difference (0.03 vs. 0.05) nearly doubles in the second academic year. See Table 3. Again, this shows that group 3 users had much greater difficulty adjusting to the new mobile MFA policy.

6 Concluding remarks

Multifactor authentication is widely touted as one of the most important security controls organizations can deploy to improve cybersecurity. While the benefits of MFA are well understood, the burdens they impose are not. This paper sets out to fix this discrepancy.

Using a large dataset gathered from a University with mandatory multifactor authentication requirements, we studied login failures the time users spend away from IT systems and services following a failed authentication attempt. In particular, we investigated the impact of a change in policy to a more secure and onerous configuration requiring users to input codes to app-based (Mobile) MFA. We find that the number of login failures and time away increases substantially for Mobile MFA following this policy change. This suggests that the opportunity costs imposed by the more secure configuration are high for a non-trivial number of users. This is especially the case when the attempted login is from a mobile device.

The most important take-away in our opinion is that the increased opportunity costs from the more secure MFA security measure was very large. The significant increase in both the number of failures and time away in the second period when using mobile MFA is something that decision makers must take into account when improving security. We are not arguing against improved security, but rather that users who have difficulty should be helped so that the adjustment does not result in a large increase in opportunity costs. In this paper, we have shown a way that such users can be identified. Unfortunately, many University IT departments are concerned only with enhancing security and not making the adjustment process easier for those who struggle. This has implications far beyond University users. A cyber security expert at a major bank told us that bank agents spend a disproportionate amount of time helping a small group of users who have trouble with MFA to access and make changes to their account. The time spent on this is a large opportunity cost, both to users and the bank agents. Hence, we suggest that all institutions identify and help the users who have trouble adjust to increased MFA security.

Declarations

Ethics approval and consent to participate Through a partnership with the IT department at the authors' university we obtained access to anonymized Entra ID authentication logs for analysis,

approved by the Institutional Review Board (IRB) under protocol 24-02.

Consent for publication Granted.

Availability of data and material The raw, anonymized log data has been shared with one of the authors, but has not been approved for further dissemination. Upon publication, we will request for approval to share derivative datasets used for analysis.

Competing interests Nothing to declare.

Funding We gratefully acknowledge support from the US National Science Foundation Award No. 2147505 and award No. 2452738 and the United States-Israel Binational Science Foundation Award No. 2021711.

Authors' contributions All authors whose names appear on the submission 1) made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work; 2) drafted the work or revised it critically for important intellectual content; 3) approved the version to be published; and 4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

[1] Jacob Abbott and Sameer Patil. How Mandatory Second Factor Affects the Authentication User Experience. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, April 2020. Conference Name: CHI '20: CHI Conference on Human Factors in Computing Systems ISBN: 9781450367080 Place: Honolulu HI USA Publisher: ACM.

[2] Bushra A. Alahmadi, Louise Axon, and Ivan Martinovic. 99% False Positives: A Qualitative Study of {SOC} Analysts' Perspectives on Security Alarms. In *USENIX Security Symposium*, pages 2783–2800, 2022.

[3] Joshua Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

[4] Haibo Bian, Tim Bai, Mohammad A. Salahuddin, Noura Limam, Abbas Abou Daya, and Raouf Boutaba. Uncovering Lateral Movement Using Authentication Logs. *IEEE Transactions on Network and Service Management*, 18(1):1049–1063, March 2021. Conference Name: IEEE Transactions on Network and Service Management.

[5] 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 Computing Systems*, pages 1–11, 04 2018.

[6] Cybersecurity and Infrastructure Security Agency. 4 things you can do to keep yourself cyber safe, 2022. https://www.cisa.gov/news-events/news/4-things-you-can-do-keep-yourself-cyber-safe.

[7] David Freeman, Sakshi Jain, Markus Duermuth, Battista Biggio, and Giorgio Giacinto. Who Are You? A Statistical Approach to Measuring User Authenticity. In *Proceedings 2016 Network and Distributed System Security Symposium*, San Diego, CA, 2016. Internet Society.

[8] Mathieu Garchery and Michael Granitzer. Identifying and Clustering Users for Unsupervised Intrusion Detection in Corporate Audit Sessions. In *2019 IEEE International Conference on Cognitive Computing (ICCC)*, pages 19–27, Milan, Italy, July 2019. IEEE.

[9] Gerard George, Martine R Haas, and Alex Pentland. Big data and management, 2014.

[10] Seth Hastings, Corey Bolger, Philip Shumway, and Tyler Moore. Transforming raw authentication logs into interpretable events. *Workshop on SOC Operations and Construction (WOSOC 2024)*, 2024.

[11] Brian Lindauer. Insider Threat Test Dataset, 9 2020.

[12] Che-Wei Liu, Peng Huang, and Henry Lucas. Centralized it decision making and cybersecurity breaches: Evidence from u.s. higher education institutions. *Journal of Management Information Systems*, 37:758–787, 07 2020.

[13] Ken Reese. Evaluating the usability of two-factor authentication. Technical Report 2018-06-01, Brigham Young University, 2018. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=7869&context=etd.

[14] Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, and Kent Seamons. A usability study of five two-factor authentication methods. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*, SOUPS'19, page 357–370, USA, 2019. USENIX Association.

[15] Joshua Reynolds, Nikita Samarin, Joseph D. Barnes, Taylor Judd, Joshua Mason, Michael Bailey, and Serge Egelman. Empirical measurement of systemic 2FA usability. In *USENIX Security Symposium*, 2020.

[16] J. J. Sonneveld. Profiling users by access behaviour using data available to a security operations center. Master's thesis, University of Twente, January 2023. https://essay.utwente.nl/94221/.

[17] Nengwen Zhao, Honglin Wang, Zeyan Li, Xiao Peng, Gang Wang, Zhu Pan, Yong Wu, Zhen Feng, Xidao Wen, Wenchi Zhang, Kaixin Sui, and Dan Pei. An empirical investigation of practical log anomaly detection for online service systems. *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 1404–1415, August 2021. Conference Name: ESEC/FSE '21: 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering ISBN: 9781450385626 Place: Athens Greece Publisher: ACM.

A Regression Tables

Table 2: User Fixed Effects Regressions

Dependent Variables	**In Time Away**	**In Failures**
Explanatory Variables		
ln Mobile app MFAs	0.161***	0.103***
	(0.006)	(0.003)
ln Mobile app MFAs*Post	0.092***	0.040***
	(0.008)	(0.005)
ln Text message MFAs	0.027***	0.022***
	(0.005)	(0.003)
ln Text message MFAs*Post	0.015**	0.005
	(0.007)	(0.004)
ln Interrupt Errors	1.102***	0.636***
	(0.011)	(0.006)
ln Interrupt Errors*Post	0.121***	0.041***
	(0.015)	(0.008)
ln Configuration Errors	1.231***	0.682***
	(0.027)	(0.013)
ln Configuration Errors*Post	-0.268***	-0.124***
	(0.036)	(0.018)
ln Password Entries	0.116***	0.084***
	(0.010)	(0.006)
ln Password Entries*Post	0.221***	0.185***
	(0.026)	(0.015)
ln Mobile Entries	0.005	0.006***
	(0.004)	(0.002)
ln Mobile Entries*Post	0.089***	0.045***
	(0.007)	(0.004)
Adjusted R^2	0.504	0.522
Observations	210,167	210,167

Notes: Robust Standard errors are clustered at user level, (ln=natural log). ** (**) significant at 99% (95%) level.

Table 3: Fixed Effects Regressions for different types of users

Dependent Variables	**ln Time Away**	**ln Failures**
Explanatory Variables		
ln Mobile app MFAs*Group 1	0.138***	0.089***
	(0.009)	(0.005)
ln Mobile app MFAs*Post*Group 1	0.081***	0.033***
	(0.010)	(0.006)
ln Mobile app MFAs*Group 2	0.151***	0.098***
	(0.010)	(0.006)
ln Mobile app MFAs*Post*Group 2	0.087***	0.038***
	(0.011)	(0.006)
ln Mobile app MFAs*Group 3	0.189***	0.121***
	(0.010)	(0.006)
ln Mobile app MFAs*Post*Group 3	0.107***	0.049***
	(0.011)	(0.006)
ln Text message MFAs*Group 1	0.018***	0.014***
	(0.007)	(0.004)
ln Text message MFAs*Post*Group 1	0.028***	0.014***
	(0.09)	(0.005)
ln Text message MFAs*Group 2	0.010	0.013***
	(0.07)	(0.004)
ln Text message MFAs*Post*Group 2	0.027**	0.010*
	(0.010)	(0.006)
ln Text message MFAs*Group 3	0.054***	0.039***
	(0.008)	(0.005)
ln Text message MFAs*Post*Group 3	-0.013	-0.009
	(0.011)	(0.006)
Adjusted R^2	0.504	0.509
Observations	210,167	210,167

Notes: The regressions also include controls for the natural logarithms of Interrupt Errors, Configuration Errors, Password Entries, Mobile Entries, and their interactions with the post variable. Robust Standard errors are clustered at user level, (ln=natural log). *** (**) significant at 99% (95%) level.

Topic	Focus	Findings
Usability of MFA in a University setting [13]	Examined the usability of MFA in the context of a University and proposed a four-phase model of user behavior that describes the adoption and use cycle.	They specifically examined the usability of Yubikeys, surveying the participants after they had setup the devices, and following up after four weeks for a semi-structured interview. They found that while most users recognized the potential security benefit, some did not find the additional trouble worthwhile when used for non-critical accounts.
Usability of MFA in a University setting [5]	In a study at a study from Carnegie Mellon University, they examined the deployment of an MFA system utilizing Duo. They explored user behaviors and opinions around mandatory adoption and analyzed usage data including over one million authentication attempts and many help-desk tickets.	Combined with two surveys, they found 40% of participants had prior 2FA experience, and more than half were using CMU Duo on a weekly or greater basis. They found that while making adoption mandatory increased negative perception over those who adopted voluntarily, attitudes towards adoption improved across all six constructs between pre- and post-activation of Duo.
Usability of MFA in a University setting [14]	Conducted over a 2-week period, they examined the usability and qualitative reactions to various forms of multi-factor authentication across the set of test platforms.	They found that system usability scores (SUS) for each method did not match the ranking of the tools' mean time to authenticate. They partially attribute this to differences in the types of errors that occur with each second factor method.
Usability of MFA in a University setting [15]	They published a study tracking users through the first 90 days of MFA use in a university setting, and reported on most commonly failed modes of MFA, as well as average times to authenticate and the time between a failed attempt and subsequent successful login, dubbed "recovery time".	Their results showed a large divide in several measures of success, such as recovery time. While most errors led to a recovery time of under a minute, 20% of users failed to authenticate after such an error until the next day. They also report on 2nd factor usage rates, which showed a large preference for push notification. Help desk tickets were also collected and examined qualitatively.

Table 4: Overview of Related Work.

Topic	Focus	Findings
Identify anomalous behavior [16]	They investigated deviation from baseline cluster location as an indicator of insider threats.	Using the "Insider Threat" data set from [11], they detected 80% of insider threats within the ITA administration group. When applying the methodology to real-world data, cluster consistency dropped by 50%, which they partially attribute to the differences in granularity in the most relevant features from each data set. For more work on clustering behavior, see [7, 8].
Identify anomalous behavior [12]	They used a private dataset of 4 million logs to demonstrate a behavior-based authentication compromise detection model. They used only two features to model users: consecutive failures and login time. Although the topic is different, the paper is particularly relevant given the use of failed logins.	The gap between failed attempts has no time cap, so a single event may span a large time frame, which is not true to the user experience. The probabilistic model employed demonstrated a good true positive false positive trade off with high prediction accuracy at a low computational cost. See [4] for similar work focusing on identifying lateral movement.
Security Operations Center (SOC) diagnosis [2]	They surveyed SOC practitioners.	They and found that there are an excessive number of security alerts across organizations, and this high alert load combined with low interpretability results in analyst fatigue, human error, and burnout.
Security Operations Center (SOC) diagnosis [17]	They investigated log anomaly detection systems	They found that log data was used in over 30% of incident diagnoses, with indicators that this portion would be larger if the logs had greater interpretability. They emphasize poor interpretability as a limiting factor in both the accuracy and actionability of generated alerts, and advocate for systems that combine a level of domain knowledge with the raw data to produce logs and alerts than are more easily interpreted.

Table 5: Overview of Related Work (continued).