An Empirical Analysis of Enterprise-Wide Mandatory Password Updates

Ariana Mirian
amirian@ucsd.edu
UC San Diego
La Jolla, USA

Grant Ho
grho@ucsd.edu
UC San Diego & Univ. of Chicago
La Jolla & Chicago, USA

Stefan Savage
savage@cs.ucsd.edu
UC San Diego
La Jolla, USA

Geoffrey M. Voelker
voelker@cs.ucsd.edu
UC San Diego
La Jolla, USA

ABSTRACT

Enterprise-scale mandatory password changes are disruptive, complex endeavors that require the entire workforce to prioritize a goal that is often secondary to most users. While ample literature exists around user perceptions and struggles, there are few "best practices" from the perspective of the enterprise—either to achieve the end goal or to minimize IT costs. In this paper, we provide an empirical analysis of an enterprise-scale mandatory password change, covering almost 10,000 faculty and staff at an academic institution. Using a combination of user notifications logs, password update records, and help desk ticket information, we construct an empirical model of user response over time. In particular, we characterize the elements of the campaign that relate to ideal and non-ideal outcomes, including unnecessary user actions and IT help desk overhead. We aim to provide insight into successes and challenges that can generalize to other organizations implementing similar initiatives.

ACM Reference Format:
Ariana Mirian, Grant Ho, Stefan Savage, and Geoffrey M. Voelker. 2023. An Empirical Analysis of Enterprise-Wide Mandatory Password Updates. In *Annual Computer Security Applications Conference (ACSAC '23), December 4–8, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3627106.3627198

1 INTRODUCTION

Enterprise-wide mandatory password updates are inevitably fraught affairs. Typically driven by either a change in circumstances (*e.g.*, evidence of a data breach) or security policy (*e.g.*, requirements for longer or more complex passwords), such mandates require that all members of an organization update their Single-Sign On (SSO) passwords within a set time period. These dual requirements of completeness and timeliness are particularly challenging given the limited resources of IT service departments. Scale requires that instructions be delivered via mass communication (*e.g.*, email), yet they must contend with a broad spectrum of understanding, capability, and incentives in the user population. Unsurprisingly, there are few established best practices for how to achieve these goals, and limited empirical data about how to most effectively enact this change at enterprise scale.

This paper seeks to address this deficit through the empirical analysis of a mandatory password update event at our institution, one which required almost 10,000 faculty and staff to take independent action. Using data from this experience, we explore how the *operational requirements* of coordinated enterprise-scale password changes — timeliness, completeness, and staff overhead — interact with the behavioral and organizational aspects of the problem that have the potential to create friction.[1] We are guided by concretely motivated questions that, were the answer understood, would directly inform operational practice, such as: How long does it take to effect institution-wide password updates? What impact do notifications have on user compliance? What factors predict efficient password updating behavior and how significant is the staff overhead in managing user problems during the process?

Our work combines detailed records of user notification events, password update logs, and IT help desk reports, to empirically deconstruct the synchronized password update process across our campus population. In doing so our work makes three primary analysis contributions:

- *Communication Effectiveness.* We demonstrate the effectiveness of repeated email requests in driving timely password update behavior—characterizing how much of the population is responsive to serial pleas over time and what subset is not reached and/or motivated by such efforts. We also analyze the effects of Web-based interstitial login reminders in galvanizing this unresponsive remainder into action.
- *Completeness hazards.* It is common during such updates to track the fraction of user accounts that have complied with the password update edict. After correcting for inactive accounts (*e.g.*, for users who have left the institution), we identify the small subset of users who are ultimately unable to meet the password update burden. We show that this set is over-represented in business units whose job function does not require regular computer use.

ACSAC '23, December 4–8, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0886-2/23/12.
https://doi.org/10.1145/3627106.3627198

[1] We specifically *do not* focus on issues such as how password policies interact with password strength, which has an extensive literature [7, 13, 18, 28, 33, 38, 49, 51–53].

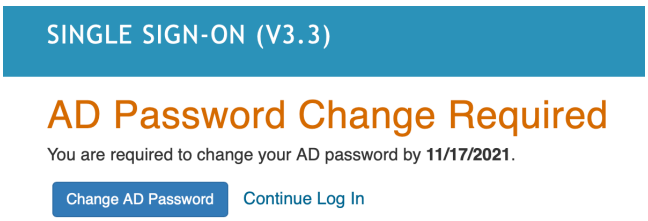


Figure 1: Example of the browser intercept message that campus's SSO portal displayed to users who had not updated their passwords by mid October 2021.

- *Quantified IT overhead.* Finally, we explore the costs to IT organizations in supporting universal mandatory password updates, using the number of help desk tickets as a proxy for the IT staff time that must be spent to help shepherd users through the process.

From these results we provide guidelines for reasoning about mandatory password update costs in terms of effectiveness and IT staff effort. We believe this is a pragmatic example of a more general and analytical approach to managing enterprise IT security processes.

2 BACKGROUND

This study describes a natural experiment driven by a security policy directive that required all users at our university to update their campus Active Directory passwords, used for Single-Sign On (SSO) across a range of university IT services (*i.e.*, organizational e-mail, calendaring, financial services, etc.). For a variety of reasons, campus faculty and staff were prioritized in this effort and thus our work focuses on the experience of this population.

In the summer of 2021, our campus Information Technology Services (ITS) team enacted a campaign to reach out to affected employees, inform them of this policy, and direct them to online resources for updating their passwords.[2] These resources included two self-service Web portals: one for updating passwords after a valid login and one for (re)setting a password without a valid login (requiring employee specific identification). As well, employees using "managed" Windows or Mac devices were able to update their SSO password locally with a valid login.

Employees *new* passwords were required to be different from their previous password, to be at least 12 characters in length, to not include their username as a substring, and to utilize three of four character classes (uppercase, lowercase, numbers, symbol).[3] Our work does not concern the quality of the resulting passwords, but we document these requirements to the extent that the additional burden may have caused some users to delay or fail to change their password as directed.

The password update campaign consisted of three kinds of actions performed by ITS staff: asynchronous reminder emails, synchronous login intercepts, and actively resetting non-compliant users' passwords to random strings ("scrambling"). Initially, a campus-wide email was sent to all employees on August 10th notifying them of the upcoming password update requirement. There were two stages of correspondence after this initial email. The first consisted of a set of four email messages (we refer to them as communications) that were sent to disjoint "waves" of users that were staggered in time. Waves were segregated based on the first letter of a user's last name: A–B, C–G, H–N, O–Z. Each subsequent wave increased in size as the ITS team became increasingly confident about their ability to manage technical or user issues that arose.

Wave	Comm #1	Comm #2	Comm #3	Comm #4
Wave 1	2021/08/18	2021/08/25	2021/09/01	2021/09/08
Wave 2	2021/08/25	2021/09/01	2021/09/08	2021/09/15
Wave 3	2021/09/01	2021/09/08	2021/09/15	2021/09/22
Wave 4	2021/09/08	2021/09/15	2021/09/22	2021/09/29

Table 1: Dates for the email communications sent during each of the four waves.

Each wave received the same set of email messages that were staggered by one week, as shown in Table 1. If a user updated their password, they did not receive subsequent communications.[4]

The first three email communications were very similar to each other, and the fourth differed slightly. The first email served as the initial notification, informing users that they needed to update their password, and that their deadline was four weeks from the initial email. The second and third email reiterated the deadline and requirement to update the password. The fourth email ("last wave communication") did NOT mention any deadline, but instead informed users that this was their *final* notification, and that they should "Avoid account access complications and change your AD password now".

The second stage of the campaign started roughly one month after the last communication of Wave 4. During this second stage, users who had not updated their password received an active notification (an "SSO intercept") each time they logged into a campus service. These intercept messages were initially rolled out to a small subset of users and gradually deployed to all users who had not updated their password. As seen in Figure 1, the login intercept told users that they were required to update their AD password by a certain date (and provided an inline button that, if clicked, brought them to the password update portal). After the deadline passed, this intercept became modal and would not allow a login without a password update.

Finally, two more email notifications were sent to users, which we refer to as the "Final" notifications and "Scramble" notifications. These notifications were sent in conjunction with the later stages of the login intercept to further convince users to update their password. The "Final" email communication told users that "Unless changed, your AD password will expire on <Deadline>", while the "Scramble" notifications informed users that "Your AD Account password will be removed on <Deadline> and you will lose access to all AD-accessed university systems...".

[2]We were not involved in the design or implementation of this password change campaign and are simply studying its effects retrospectively.

[3]This requirement, as well as a further filter against using "known compromised" passwords provided by a third-party service, were enforced mechanically by rejecting new passwords that did not comply with these requirements.

[4]Because communication lists were constructed by querying the Active Directory (AD) system for password update information, updates were not strictly atomic. Thus, a user who updated their password after the second communication list was constructed, but before it was sent, would still receive the reminder even though they had already updated their password.

Any users who had not updated their password after receiving these final messages and SSO intercepts had their account password "scrambled" (*i.e.*, set to a random value) by an ITS administrator. Such users would thus be unable to login to the vast array of campus IT services and would need to trigger the password reset mechanism themselves or with the help of the ITS help desk. We were given the list of users whose passwords had been scrambled as well as the date and time of the scrambling.

Our university has a number of closely-affiliated but semi-independent organizations, such as separately endowed research institutes and a medical center, which have their own IT infrastructure. A small subset of accounts in our data set reflect "secondary accounts" of users who have a primary appointment at one of these sister organizations, but who happen to have an account in main campus's IT systems as a result of joint initiatives. As we discuss later (§ 5), these users might not frequently check or use these secondary campus accounts, since their day-to-day online activities could revolve around an account at their home organization.

3 ETHICS

Our analysis does not expose any vulnerabilities, nor does it indirectly create harms by virtue of its results. The benefits of our research include better understanding the dynamics around mandatory password policy changes, how to do so more efficiently and, by generalization, improving mass compliance with other changes in security policy. Our analysis is based on secondary use of data already routinely logged by our institution's IT services group and this data is de-identified for our analysis. Further, we only pursue analyses of population aggregates and do not present results about individual users (even de-identified). Our project has been reviewed by our institutional review board (IRB) and considered exempt. Additionally, our work takes place with the full knowledge of our institution's CISO and with the associated IT staff (our work is driven, in part, by helping this organization understand how to better manage their security communications).

4 METHODOLOGY

In this section, we discuss our university authentication process, the data sources we used, and the set of accounts we focus on.

4.1 Authentication into Campus Services

Our institution uses Active Directory (AD) for basic authentication and Duo for two-factor authentication for all major systems. Thus, users accessing campus services ranging from email to payroll first need to login using an Active Directory username and password, and then authenticate via Duo (typically a phone-based app) to access their service. Our Duo deployment supports a *remembrance window* of seven days which, if configured, reduces the two-factor authentication requirement to once per week *per device*.

4.2 Data Sources

We conduct our analysis using four data sources from August 2021 to March 2022: Splunk logs, email correspondence logs, Active Directory metadata, and Help Desk tickets. We explain each of these data sources in further detail.

Splunk Logs. Our institution collects various logs of user activity and stores them in Splunk, a third-party service for capturing, indexing, and querying system log information. For this study we use Splunk-managed event logs from our campus' Active Directory and Duo deployments.

The Active Directory logs contain password update information—notably Windows Event IDs 4724 (account password reset attempt) or 4723 (account password change attempt) paired with 4738 (account changed)—as well as metadata about the password update itself (*i.e.*, who initiated the change). The event codes and metadata allow us to differentiate password updates into four different semantic categories: a password change by a user via a campus self-service online password change portal, a password change by a user via the user's campus-administered machine (*e.g.*, via the Windows Sign-in/Password dialog), a password reset by a user via the password change portal, and an administrative reset (*e.g.*, help desk, departmental IT support).

The Duo logs contain every Duo authentication success and failure for users on campus. For password updates initiated by the online password update portal, users must already be authenticated via both Active Directory and Duo. For password resets initiated via the portal, no authentication is necessary (although failed authentications appear in the logs if the user attempted to authenticate but forgot their password).

Email Correspondence and Scrambled Accounts. The campus security team notified users about the new password update requirements via a series of email messages (§ 2). These messages used Emma [14], an email marketing service which incorporates a tracking pixel into messages to identify when each email is delivered, opened, or bounced. This team provided the Emma logs to us, as well as which accounts were ultimately scrambled and when.

Active Directory Metadata. Each user profile in Active Directory has additional metadata, including their Organizational Unit (OU). This metadata indicates user roles and departmental affiliations. We use this profile information to correlate behavior with user demographics in our analysis.

Help Desk Tickets. Finally, we used aggregate statistics collected from logs of campus Help Desk tickets to help understand the IT staff burdens created by the password update campaign. As discussed in more detail in Section 6.1, our data consists of coded tickets (*i.e.*, tagged as related to password updates) from the period in question that are de-identified and tagged with associated OU membership. This process produced 919 password update related tickets submitted by 762 distinct users.

4.3 User Population

For the purposes of our study, we focus specifically on active users who successfully received the password update correspondences. Concretely, we consider users that satisfy the following criteria:

1) Users successfully contacted. We only consider users who were successfully contacted by the email notification campaign. We consider users "successfully contacted" if the email tracking service indicates that they received (although not necessarily opened) all messages in the notification campaign until they updated their password. This avoids confounding effects caused by non-human

Category	Number of Users
Single Change Users	7925 (81.3%)
Multiple Change Users	1291 (12.2%)
Nonresponsive Users	528 (5.42%)
Total	9744 (100%)

Table 2: Distribution of the different kinds of users.

accounts or the minority of users who, for one reason or another, have no working email point of contact.

2) Users are active. Like any large organization, ours has user accounts that are accessible but largely inactive (*e.g.*, "email-for-life" accounts). Since we are interested in the behavior of active users—those for whom password expiration will have a direct impact on their activity—we restrict the account population to accounts that have had at least one successful login authentication (both Active Directory and Duo two-factor) during the campaign.

Table 2 summarizes the user population we consider in this study. Among 9,744 users, 7,925 (81.3%) of them updated their password exactly once during the password campaign ("single change" users), 1,291 (13.2%) updated their password more than once ("multiple change" users), and 528 (5.42%) did not update their password by the communicated deadline ("nonresponsive" users, whose passwords were scrambled by the IT staff due to their failure to act in a timely fashion). We note that most users are "single change" users—changing their password once during this campaign—with a smaller percentage deviating from this behavior and incurring additional costs (either on individual users or the IT organization).

5 USER RESPONSIVENESS

In this section we analyze how the users in our study responded to the password update campaign. In particular, we explore the following research questions:

- RQ1: Were repetitive emails effective in prompting user change?
- RQ2: Were login intercepts effective in prompting user change?
- RQ3: In what ways did multiple change users react differently than single change users?
- RQ4: Which users utilized password reset more than a password change?
- RQ5: Which organizational units were slower in updating their passwords?

We focus on this set of questions when analyzing user responsiveness to understand which actions are most effective for the organization (RQ1, RQ2, RQ3, RQ5), which update mechanisms are utilized the most and thus should be embraced (RQ4), and why organizations should use different mechanisms for certain employee groups to more effectively promote password updates (RQ5).

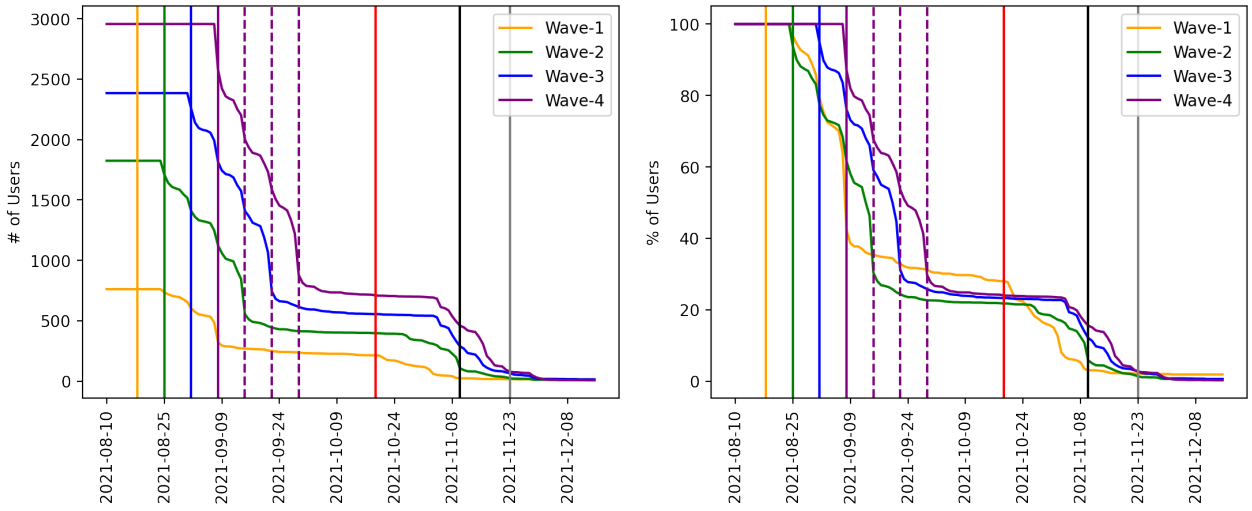
5.1 Single Change Users

We begin by examining the behavior of single change users. Figure 2 shows the password change behavior of these users over time. The left graph (a) shows four curves, each corresponding to one of the communication waves. Each curve shows, on a daily granularity, the remaining number of users in that wave who still need to change their password. The right graph (b) shows the same results, but with each wave normalized to the number of users in that wave: the curves show the percentage of users in each wave who have yet to change their password. The solid vertical lines correspond to the start of various actions taken by campus during the campaign, including when campus sent initial email communications to each wave (first four solid vertical lines), intercepted logins (solid vertical line at October 19, 2021), and final/scrambled notifications (last two solid vertical lines). For a subset of users who had not yet updated their password, the IT staff began scrambling their passwords on November 16, 2021, prior to sending email notifications. Additionally, we note that each wave received four communications, but the communications were staggered by a week and thus are overlapped. We denote the trailing last communications in the final wave with dashed lines. For reference, Table 1 shows the dates of each communication in the different waves.

From the timeline in Figure 2, we define periods of user activity based on user response to the various notifications. Each wave begins with a "responsive" period that engages with responsive users until seven days after the final communication for a given wave. Each wave then has an "idle" period between the email notifications and the first use of the login portal intercept. Finally, each wave ends with an "intervention" period engaging with unresponsive users and spanning the login intercepts, expiration warning email communications, and account scrambling. The intervention period is the same for each wave (October 19th until December 15th), but the responsive period and idle period are shifted by each wave start date. For example, the responsive period for Wave 1 is August 18th to September 15th and the idle period is September 15th to October 19th, while for Wave 2 the responsive and idle periods are from August 25th to September 22nd and September 22nd to October 19th, respectively. Combining the waves, 71.0% of these users changed their password during the responsive period, 5.28% changed during the idle period, and 23.7% changed during the intervention period.

RQ1: Repetitive emails are effective in prompting a majority of user updates but have diminishing returns. As seen in Figure 2 by the stairstep shape of the curves from August 25 to September 29, multiple email communications were effective for the majority of users. An immediate question for an organization planning to use email notifications is how many iterations to perform. We measure effectiveness of each iteration by quantifying the number of users who initiated a password update within a week of a given communication. For our campus, multiple communications was clearly impactful. The first three communications resulted in a roughly uniform response from users proportional to the size of the wave, roughly 15%.[5] An interesting question is whether a fifth communication would have induced a similar response as the previous four. Given the much smaller response of the fourth communication (around 5% across each wave) and subsequent email notifications, we speculate that a fifth email would only have further diminishing returns.

[5] An exception is the first communication of the first wave, which does not appear to have prompted any password updates. Upon investigating, this apparent lack of response was due to a data collection error in that timeframe.



(a) Number of users in each wave **(b) Percentage of users in each wave**

Figure 2: (a) The number of single change users *without* a password update in the different waves over time, and (b) the same results showing the percentage of users in each wave. The first four solid lines denote the beginning of the communication series for each wave. The line at October 19, 2021 denotes the start of login intercept, while the line after November 8, 2021 denotes the start of final notifications, and the line at November 23, 2021 for scramble notifications.

Our results suggest that our organization clearly needed to be proactive throughout the campaign. The initial email communications were effective for roughly 70% of users. Subsequently, very few remaining users changed their passwords during the idle period, and these remaining users only started reacting again once campus activated the login portal intercept.

RQ2: Login intercepts are an effective tool for user updates. While our organization used login intercepts for well over a month (from October 19, 2021 to November 11, 2021), they staggered their use for the different waves. Moreover, towards the end of the campaign they continued displaying login intercepts in addition to sending a final round of email warnings (note that the IT staff sent the first batch of final email warnings on November 9, 2021). To more clearly assess the impact of login intercepts, we examine their impact on just users in the first two waves, users who had the longest exposure and response to just the login intercepts (before the final email warnings were sent). For this time period preceding November 9, 88% of the remaining non-updated users in Wave 1 responded to the portal intercept and successfully updated their password. For Wave 2, 51% of the remaining users updated their password during the login intercept period (note that Wave 2 users had one fewer week in which to respond compared to Wave 1 users). Email notifications are clearly effective for the majority of our population, but require action out of context. The portal intercept, in contrast, happens when the user is in the process of logging in, and was successful in leading users to update their password.

5.2 Multiple Change and Nonresponsive Users

We next examine multiple change and nonresponsive users. We are interested in understanding the differences in experiences, to better understand how to make this process more effective.

	Responsive	Idle	Intervention
% First Change	57.25%	8.33%	34.33%
% Second Change	22.23%	12.79%	60.69%
% Third Change	18.30%	10.21%	64.68%

Table 3: Breakdown of the first, second, and third password changes for multiple change users across the different time periods of the campaign.

To begin, when compared to single change users, multiple change users have more than one password update, suggesting these users experienced more friction with the password update process.

Of the 1,291 multiple change users, 72.03% have two password changes, 18.20% have three, and 9.76% have more than three. For simplicity, we focus on the 90.23% of multiple change users that have two or three password changes since they capture the bulk of this population. Table 3 shows which period during the campaign the user made the update for each password update.

RQ3: Multiple change users are less responsive to email communications than single change users. However, multiple change users have similar password update attempts as nonresponsive users. Compared with the single change users, the multiple change users are less responsive to the email communications: 71% of single change users update their password during the responsive period, but only 57% of the multiple change users make their initial password update during the period. Correspondingly, more multiple change users (34%) wait until the intervention period than single change users (23%) before making an update.

The majority of the second and third password updates for the multiple change users happen later in the intervention period (60%

and 64%, respectively), rather than closely associated with the first password update in the responsive period. We originally suspected that most users who had multiple password updates had issues involving multiple devices. For instance, they might first change their password on their laptop, but then soon after attempt to login via their phone (e.g., which had the older password cached). At that point the most expedient action would be to reset their password via their phone so that they could continue to login. This scenario would lead to multiple password updates in quick succession, but the long duration between the first and subsequent password updates for the multiple change users indicates this explanation does not hold for most of them.

Three other situations could explain the behavior of multiple change users and their delayed subsequent password updates. The first is that the users were confused because they have multiple accounts on our campus (*e.g.*, a faculty account on main campus and another account on the health campus). If a user has two accounts, changed the password on their first account, and then received an intercept for their second account, they may not have paid attention to the account targeted in the notification and instead re-initiated a password change on their first account.

We explored this hypothesis by comparing the anonymized legal name attributed to each user account and counting how many single change, multiple change, and nonresponsive users have user accounts with the same legal name. Overall, there are less than 50 instances where two different user accounts have the same legal name, indicating 1) the legal name attribute is not correct, 2) most users do not have multiple accounts, or 3) their additional accounts are hosted on separate IT infrastructure that we do not have access to (see note about various infrastructures in § 2).

The second hypothesis is that users became confused about messaging: they forgot whether they changed their password, were reminded about the password change out of band, and re-initiated a change. Given the granularity of our data, we unfortunately could not explore this hypothesis further.

The third hypothesis is that these users were attempting to change their passwords with a different modality. For example, single change users might be users who primarily use a desktop/laptop, whereas multiple change users might have been predominantly mobile device users. Unfortunately, our dataset does not include sufficient information to definitively test or validate this hypothesis.

We finally compare multiple change and nonresponsive user reactions. Among the nonresponsive users 68.62% had two changes, 29.20% had three changes, and the remaining 7.56% had over three changes, a distribution similar to the multiple change users. If we use the number of changes as a proxy for how many issues a user faced (with a higher number of changes approximating more issues), then the nonresponsive users experience no more issues than multiple change users and simply encounter them in a later time period.

5.3 Password Update Mechanisms

We next investigate the different mechanisms that users selected to update their passwords, providing insight into time and energy spent on these updates. Recall that users can change or reset their password via a self-service Web portal, via their campus-managed work computer, or by invoking the help of campus administrative staff. To minimize procedural costs, organizations want to maximize the use of the first two methods and minimize the third.

Category	% Change	% Reset	% Both	% Admin Reset
Single Change	92.72%	7.28%	—	0.52%
Multiple Change	29.20%	13.90%	56.85%	23.86%
Nonresponsive	2.36%	77.12%	19.66%	27.22%

Table 4: Password change mechanisms across the three different user populations. Note that there are two ways to perform a reset, and thus Admin Reset is a subset of the Reset and Both columns.

RQ4: Multiple change and nonresponsive users utilize password resets more than single change users. Table 4 summarizes the actions taken by the three responsiveness categories of users in the study. Note that there are two ways for a user to execute a reset, and thus "Admin Reset" is a subset of the "Reset" and "Both" columns. Single change users, as desired, overwhelmingly perform their password change on their own: only 0.52% of these users require administrator assistance with updating their password. In contrast, multiple change and nonresponsive users require significant administrative help. Roughly a quarter of each user category (23.86% of multiple change users, 27.22% of nonresponsive users) initiate a password reset with the assistance of an administrator. To further reduce procedural costs, organizations can focus on reducing circumstances that lead to users making multiple changes. Nonresponsive users represent a difficult case since they generally have minimal interaction with campus already (§ 5.4).

As a final observation, in addition to the self-service Web portal and IT help desk service, our campus also allows users to change their password via the operating system of their work machine. More than 22% of the single change users updated their password using their work machine, and all of these updates were successful (the users were already logged in). Since this method is both effective and low cost, organizations should continue to support it and encourage its use.

5.4 User Role

Next, we explore how the password update behavior of users correlates with their role on campus. Recall that the account profiles for the users on our campus specify the Organizational Units (OUs) that the user is associated with.[6]

For the 50 largest OUs on campus by population, we calculate the percentage of single change, multiple change, and nonresponsive users in each OU who update their password. Using these values, we construct three distributions (one for each user responsiveness category) and calculate the Z-Score of each OU, which characterizes how far the value deviates from the mean. For each user responsiveness group (single change, multiple change, and nonresponsive), we identify OUs that are either above or below 1.96 standard deviations from the mean as outliers.[7] Among these outlier OUs, users

[6]Note that users can have multiple OU labels and we count the users in all OUs that they are associated with.

[7]Examining data that is above or below 1.96 standard deviations is considered common practice for finding outliers when using Z-Scores.

in the "Extension", "Instructors", and "Extension Business" OUs are over-represented in nonresponsive users, and under-represented in single change users. These OUs are interesting because they correspond to users who can perform their daily jobs without needing to interact with campus accounts or systems as often as other roles.

Focusing on only single change users, we again examine the 50 largest OUs, but this time across the three time periods (Responsive, Idle, Intervention). Specifically, we investigate if single change users in the intervention period differ from those in other periods. Using Z-Scores to identify outliers, we see that users in the "Building Services", "Recreation", and "Dining Services" OUs are over-represented in the intervention time period. Once again, a common thread among many of these OUs is that they correspond to users more on the periphery of the campus: users who may not need to interact with main campus systems regularly.

RQ5: Users in peripheral organizations take longer than their counterparts. Both of these findings reinforce the point that users who take longer or have difficulty updating their password are correlated with roles that may have less online interaction with main campus systems. In this light, it is not surprising that email notifications are less effective or that such users have more difficulty performing the update (*e.g.*, resulting in a disproportionate number of password scrambles). While this finding may seem obvious, it is still important to understand from an organizational standpoint, as this may change how the organization approaches these departments in future campaigns. In particular, organizations may want to target these users differently: *e.g.*, targeting such users earlier, or forgoing email reminders and using login intercepts from the start, or even using a different notification mechanism such as text messages.

6 HELP TICKET WORKLOAD

Although password update initiatives can improve the security of an organization, these efforts generate extra work for users and the IT staff, particularly when issues arise during the password update process. To better understand these associated costs, we analyzed changes in the volume of help desk tickets regarding password and account changes during the password update time period. In particular, we examine the following research questions surrounding the costs of enterprise password update campaigns:

- RQ6: Did the password campaign increase the number of help desk tickets?
- RQ7: What were the costs of different enterprise actions in terms of help desk ticket load?
- RQ8: Do users in different departments produce heavier help desk ticket loads?

6.1 Help Desk Ticket Data

Our university uses ServiceNow, a centralized ticketing service, to manage all help tickets and requests generated by users. Users can submit help tickets via a standard web portal or by emailing specific help aliases; additionally, users can call specific campus phone numbers to speak with support staff, who then manually create a ticket on behalf of the user during the assistance process. To identify help desk tickets related to the password update process, we created aggregate statistics from the ticket database related to the password update roll-out. Concretely, tickets that met the following criteria are involved in the analysis:

	Password Update Campaign		Prior Year	
All Waves	7.82%	(762 / 9,744)	2.21%	(215 / 9,744)
Wave 1	7.94%	(78 / 983)	2.24%	(22 / 983)
Wave 2	7.66%	(174 / 2,272)	2.60%	(59 / 2,272)
Wave 3	8.04%	(237 / 2,948)	2.37%	(70 / 2,948)
Wave 4	7.71%	(273 / 3,541)	1.81%	(64 / 3,541)

Table 5: Percentage of users with password help tickets one year apart.

(1) The ticket was assigned to the "Service Desk" team (which handles all password and account related issues).
(2) The ticket's customer was a user from the 9,744 users in the population we investigate (§ 4).
(3) The ticket was created between August 9, 2021 and February 1, 2022 (*i.e.*, between the start of the password reset notifications and approximately one month after the final password reset notification).
(4) The ticket satisfied the following keyword requirements: the ticket contained at least one word from each of two lists — ["password", "account"] and ["lock","reset", "change", "update", "sign in"] — and it also did *not* contain any "false positive" words identified based on manual sampling (*e.g.*, "compromise", "new", etc.).

In total, this search yielded 919 help desk tickets filed by 762 distinct users. For the remainder of this section, we refer to these 762 users as "ticket-filing users" and any password-update related ticket they file simply as a "help ticket". Over 85% of these users (653) filed only one help ticket during the update time frame, and 12% of users (93) submitted exactly 2 tickets. Among the remaining 3% of users (16), the maximum number of tickets filed by any single user was 12 tickets, and upon manual inspection it appeared that this user is an IT staff member who created help desk tickets on behalf of users who called the support hotline.

6.2 Changes to Help Ticket Volume

Using the volume and timing of tickets, we investigated how much additional work our institution's IT staff encounters as a result of initiating an enterprise-wide password update.

RQ6: Password updates increase the overall ticket volume by a factor of 3–4×. Table 5 shows the percentage of ticket-filing users during the password update time period (second column) and the percentage of ticket-filing users from this same exact population during the same time frame one year prior to the password change campaign (third column). We observe a 3–4× increase in the proportion of ticket-filing users during the password update time period (7.5–8%) when compared to the same set of users during the same time period in the prior year (1.8–2.6%). The proportion of users who submit tickets, and the relative increase over the preceding year, remains consistent across all wave groups.

RQ7: Actions lead to different ticket volumes. As described in Section 2, over the course of the password update roll-out, campus IT staff employed multiple types of actions to encourage users to update their password.

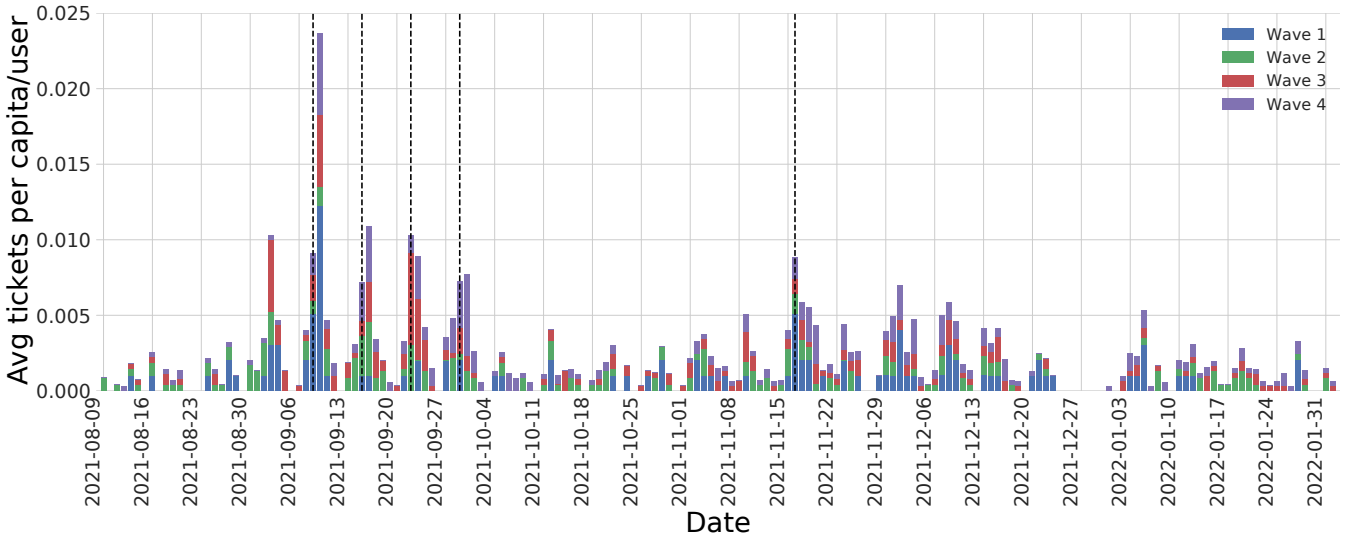


Figure 3: Ticket volume per day, normalized (divided) by the total number of users in each wave notification group.

Figure 3 displays the total volume of tickets that users from different wave groups submitted during each day, where the daily volume is normalized (divided) by the total number of users in each respective wave group. Figure 4 shows the cumulative fraction of tickets submitted by all users over time. As marked by vertical dashed lines in both figures, two types of actions led to noticeable increases in the volume of tickets.

First, we see large spikes in the proportion of users who submit tickets after each of the first four dashed lines; these dates correspond to when the IT staff sent their fourth ("last") communication email to users in each of the waves. These notifications stated that users must immediately change their passwords to "avoid account access complications". As we observed in Figure 2, this set of email messages galvanized a significant fraction of users into updating their password, which likely accounts for the increase in help ticket volumes immediately following these email notifications.

The last dashed line in Figure 3 corresponds to the date (Nov 16) when the IT staff began to automatically scramble the passwords of any user who had not yet updated their password. Unsurprisingly, this intervention led to a significant increase in the proportion of users who filed help desk tickets. Among the 528 nonresponsive users, 77 users filed password help tickets (14.6%); in contrast, only 7.6% (700) of the 9,216 single change and multiple change users without a password scramble submitted a help desk ticket. Furthermore, of the 77 nonresponsive users, only 8 users submitted a ticket prior to having their password reset by the IT team, which suggests that the vast majority of these users filed tickets as a result of the IT team's actions.

In contrast to these two actions, from October 19, 2021 to November 15, 2021, the IT staff configured the university's SSO login portal to display a browser interstitial message after every successful login to users who had not updated their password; from November 9 to November 16, the IT staff also sent out an additional email notification and the SSO portal continued to produce browser interstitial pop-ups. As seen in Figures 3 and 4, this institutional action generated noticeably fewer tickets than both the earlier email message notifications and the password scrambling: only 8% of tickets were submitted between October 19 and November 9 (the period where only active action was login intercept).

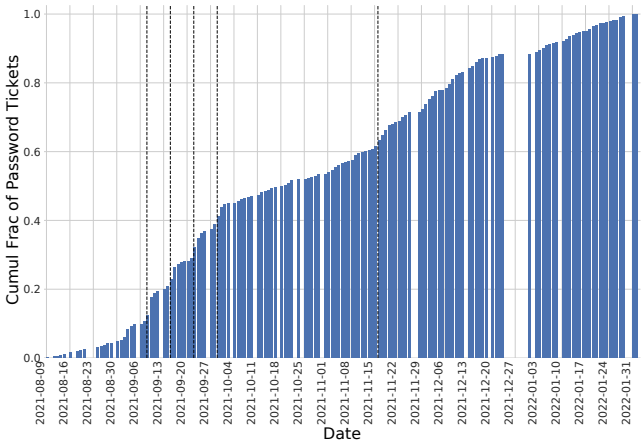


Figure 4: Cumulative fraction of password-update help tickets over time.

User Responsiveness	% w/ 1 Ticket	% w/ 2+ Tickets
Single Change Users	5.6%	0.6%
Multiple Change Users	18.6%	3.8%
Nonresponsive Users	14.6%	2.3%

Table 6: Proportion of single change, multiple change, and nonresponsive users (§ 5) who file exactly one ticket and multiple tickets.

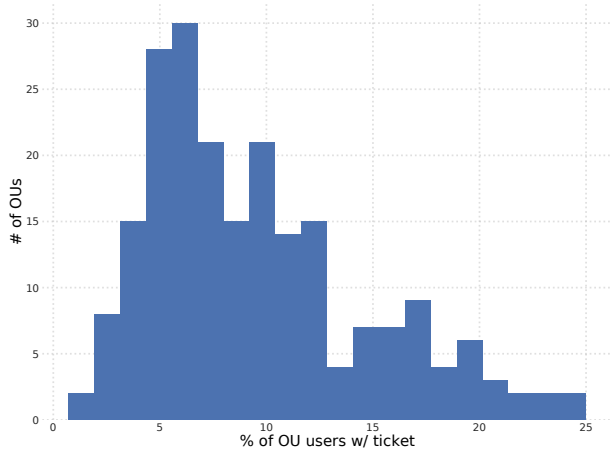


Figure 5: Percentage of users in each OU (proxy for department/role) who filed a password-related help desk ticket during the password update period.

We hypothesize that the SSO intercepts created a lower ticket volume because they presented a more concise message and direct, in-situ path to updating a user's password. Namely, whereas the email notifications contained a detailed description of the update roll-out and list of instructions for users to complete, the SSO intercept message displayed a short message with a link for the user to immediately update their password (as shown in Figure 1). Furthermore, users are more likely to successfully update their password independently because they only saw the SSO intercept message after successfully authenticating with their old password (which they then can use to change their password).

6.3 Help Ticket User Demographics

Update Responsiveness and Help Ticket Volume. We next explore whether the single change users' apparent efficiency at successfully resetting their password correlated with needing less help from IT staff members. As seen in Table 6, single change users in fact submit 3–6× fewer tickets than users in other categories: only 5.6% (445 / 7,925) single change users submit one help ticket, compared to 18.6% (240 / 1,291) multiple change users and 14.6% (77 / 528) nonresponsive users.

RQ8: Help ticket volume is non-uniform by OU. We also investigated whether a user's specific department (a proxy for job role and technical familiarity) correlated with the likelihood of them requesting help. As discussed earlier in Section 4, our institution uses Active Directory to manage information about users and their accounts, and each user has an associated set of Organizational Unit (OU) affiliations (*e.g.*, Computer Science department, Staff Tech Support, etc.). Users can and often do have multiple OU affiliations as a result of being affiliated with multiple departments or groups on campus. For our analysis, a member of campus's IT staff computed the set of OUs that each user in our dataset belonged to.

The 762 users who filed a password-reset related ticket span a total of 314 distinct OUs. Figure 5 shows the proportion of each OU's users that filed a password-related help ticket: for each OU, this proportion equals the number of ticket-filing users affiliated with the OU divided by the total number of users in our data set who had an affiliation with the OU (*i.e.*, if any of a user's OU affiliations match, then we count them as part of the OU). As seen by the right-skewed distribution, users in several OUs submit help tickets at over twice the rate as the median OU (8.19%).

OU	% of Notified OU Users
Teaching & Learning Commons	25.0%
Sponsored	24.4%
Academic Affairs	23.5%
Counseling	23.1%
IT Services	22.5%
Emeritus	21.7%
Provost (Div 1)	21.2%
Emeriti	20.6%
Provost (Div 2)	20.6%
Employment Community Outreach	20.0%

Table 7: Top 10 OUs with the highest proportion of active users (undergoing a password reset) who filed a password-reset related ticket (§ 6.3).

Table 7 shows the OUs with the highest proportion of users who submitted a help ticket (again, with some modifications to the OU names to blind our organization). Among these OUs, we note that the tickets submitted by users in IT Services correspond to members of the IT staff submitting tickets on behalf of users who contacted help / support out-of-band (*e.g.*, via a phone call to the help desk). A total of 301 OUs (not shown in Figure 5) had 0 affiliated users who submitted a password-reset related ticket; of these, only 31 OUs (10.3%) have more than 10 users and span a variety of different parts of campus with no clear thematic grouping (*e.g.*, they cover a variety of different academic departments and groups, such as postdocs, the campus registrar's OU, and OUs for technical institutes co-located and affiliated with campus).

Similarly, the OUs with the highest proportion of users who submit tickets lack easily discernible patterns: on one hand, this set of OUs contains both users with looser present-day affiliations to campus (*e.g.*, emeritus) as well as groups involved in day-to-day campus interactions (*e.g.*, Academic Affairs, the Copy Center, and staff in various Provosts' offices). Based on this heterogeneous mix across both high and low ticket-filing OUs, it appears that other underlying factors (beyond a user's department affiliation or working ground) may be more predictive in determining whether users will need help during the update process (*e.g.*, technical aptitude and familiarity).

7 RELATED WORK

Password research is a large field with influential work dating back to the 1970s [34]. Prior literature explores a diverse set of questions related to password security, ranging from password guessability and cracking, to optimal policies, to user mental models around passwords. Here we focus on work that is most closely related to our analyses and discuss the contributions of our work.

User Perceptions and Password Change Policies: Understanding users' mental models about password policies and their impact on security has a rich history [24, 27, 43–45, 48, 50]. Studies

have found that users report rarely changing passwords unless asked [26]. Other studies have further shown that users are generally proactive in changing their passwords when a deadline is provided, while others find that they postpone as long as possible [4, 36, 46]. Notably, Choong et al. found that positive behaviors lead to more secure behaviors, while negative attitudes lead to poorer security behaviors [9]. These findings were backed up by Becker et al. in a much larger follow-up study [4]. Understanding user mental models has shed light on the burden that these password policy changes incur on the user.

Due to concerns about user burden and also security of passwords, NIST changed their password change policy guidance in 2017 to no longer recommend periodic password changes [35]. Recent work from Gerlitz et al. in 2021 quantified how German companies created their password policies, and in 2023 found that a minority of companies surveyed still require periodic password changes [20, 21], contrary to current recommended practices. Furthermore, Lee et al. measured the top English website password policies and found many of them do not follow recommended guidelines, and Sahin et al. subsequently interviewed website administrators to determine the factors that go into password policy creation [29, 42].

Empirical analyses: More closely related to our work, a number of papers have conducted measurement studies to understand various aspects of the password lifecycle. For example, studies have examined password update metrics at scale and found that strength meters effectively prompted users to produce higher entropy passwords [49]. A number of studies have also examined and quantified password guessability and password reset policies of university populations at scale [33, 36]. Most recently, researchers have proposed a new system to securely study login attempts at a university in real-time to advance password research [6].

Organizational policy adoption: Outside the password literature, two other lines of research relate to our study. The first direction studies security communication — how the content and modality of security information plays a role in how it is acted upon [11]. This work includes studies on the efficacy of both user interface elements such as phishing toolbars [12] and browser TLS security indicators [2, 15–17, 40] as well as email-based vulnerability notifications [8, 23, 30–32, 47].

More closely related to our work is Amador et al., who explored prospect theory in the context of password changes [3]. While they employed participants via Mechanical Turk and Prolific and thus did not have the benefit of a large-scale natural experiment, they found that a negative framing prompted more secure passwords than a neutral or positive framing, indicating that different communications can have different effects on password creation. Outside of this work, we are not award of other literature that examines communications regarding enterprise password updates or resets.

The other key line of research focuses on the user overhead and adoption issues around new security technologies, such as two-factor authentication (2FA) [1, 10, 22, 41]. Notably, Colagno et al. found that ticket volume increased 5× during their mandatory 2FA adoption phase. While Abbot et al. [1] and Reynolds et al. [41] do not report the increase in tickets due to the policy change, they do characterize the support costs for incorporating two-factor authentication into their respective universities by analyzing related help desk tickets.

Contributions of Our Work: Our work lies in the intersection of multiple sub-areas of password research. While users at our institution were required to update their password, we do not measure any attributes of the password itself, but rather when and how the user complied with the enterprise-wide policy change. Unlike many studies that examine user actions via lab settings, our study provides a large-scale analysis of a real-world enterprise password update process, which provides an empirical basis for validating prior findings and exploring new findings about how users behave during a password update process from an enterprise perspective.

On one hand, one contribution of our work is to validate previous findings based on a large-scale, real-world dataset. For example, we empirically confirm that users are reactive, rather than proactive, in complying with a password policy change [5, 26]. We also validate that email reminders have no discernible effect for a subset of users, and that a majority of users changed their password before the final deadline [36]. Moreover, we find a 3–4× increase in ticket volume for a mandatory password change, similar to Colagno et al. who found that ticket volume increased 5× during their mandatory password change [10]. Finally, our study showed the need for more stringent, proactive reminders to ensure full user compliance, similar to Parkin et al. who found that some users specifically wait until the expiry period to reset their password [36].

Beyond providing validation of prior results, our study also sheds new light on previous findings. For example, we show that email reminders are not always as efficacious as previously claimed. One survey study from 2010 found that 98% of their participants changed their password after receiving multiple email messages, while in our much larger analysis only about 70% of users did so [46]. This difference may be due to the fact that these studies were conducted over a decade apart, and thus people interact with email very differently now than they did prior. Finally, while prior work has discussed the importance of a central management system (e.g., Active Directory) in easing the user experience of policy changes [19, 36], our work shows that a range of user difficulties still persist, indicating that centralized management systems ease, but do not erase, user difficulties, a useful finding for organizations in general.

8 GENERALIZABILITY, LIMITATIONS, AND DISCUSSION

Generalizability: Our work, like much prior work [4, 6, 33, 36], focuses on data from a single institution. In an ideal world, we would be able to more directly compare and contrast this data with similar data from other organizations, to further the generalizability of the claims presented [41]. Due to the inherent challenges of acquiring such data from one, let alone multiple organizations, both our work and many related works do not do so. However, we argue that our work provides a useful additional data point in the corpus of password research for two reasons. First, this large-scale, in-situ analysis validates previous claims that: 1) users are reactive to password change campaigns, not proactive as some admins may hope [5, 26, 36], 2) email remains a useful notification medium, but does not prompt all users to take action [36], and 3) while the number of Help Desk tickets does increase 3–4× due to the password change campaign, the overall volume is moderate and therefore is unlikely to overwhelm Help Desk staff [10]. In

addition to these validations, this study also provides insight into the efficacy of active intercept, prompting many users who were not reactive to email into changing their password. This finding is useful since active intercept can be deployed and evaluated at other organizations.

Moreover, we believe that many of these findings can generalize to industry and government, given similarities in login/password interfaces and enterprise IT infrastructure. For example, we believe the active intercept was effective because it grasped user attention at the time when the user was trying to complete an action [25, 37, 39]. Implemented in a different organization, we expect active intercepts will still have higher impact than email. Our work supports prior research that shows many people are reactive when it comes to changing security policies, another finding that we believe can generalize to other organizations [5, 26, 36].

Furthermore, our work is the most recent study to capture a security policy change at a large organization, representing almost 10,000 user actions. While situated at a university, these 10,000 users nonetheless span a wide range of job functions, backgrounds, and other demographic factors. While it is difficult to directly compare this work to others due to its empirical, post-deployment nature, we would be excited to see studies similar to ours in the future for further validation and comparison.

Limitations: As with any real-world study, there are limitations that need to be considered. One such limitation is that this analysis is restrospective and observational, which scopes the questions we can(not) answer, such as whether one specific notification directly causes a specific percentage change in password update rates. As an example, due to the observational nature of this study, we cannot precisely determine whether the lower user responsiveness to the fourth email in the initial email notifications was due to notification fatigue, or because the wording in the email changed. An additional challenge we faced was data granularity — since this study was retrospective, some data was not retained with enough granularity for us to answer further questions. For example, we know when the active intercept was introduced for a given wave, but individuals within the wave were introduced to the intercept on a rolling basis. It would have been helpful for us to understand what specific day an individual was introduced to active intercept, as well as how often they viewed it, to further quantify the effectiveness of the notification. Lacking that data we could only stick to high-level aggregates. Further, we were unable to explore hypotheses for multiple change users because of a lack of information about their device/browser modality. Finally, we also needed to account for confounding factors in the data. For example, it is difficult to ascertain the full effect of the SSO messaging for Waves 3 and 4, as it was deployed at a similar time as the final email messages.

Discussion: In this work, we empirically deconstructed a password update campaign at our university aimed at employees and staff. Though our work contained challenges, we still provide useful insight into the dynamics of a recent, large-scale, in-situ organizational policy change. Specifically, our key takeaways are:

(1) Email communications are effective but have diminishing returns, potentially due to user fatigue or lack of engagement with email as a communication mechanism. We observed that the first three reminder emails each prompted about 15% of users to change their passwords, but the fourth only elicited a response from an additional 5%. Alternatives (e.g., SMS messaging) or advance scheduling may be appropriate for users whose roles require less email engagement.
(2) Interceptive forms of communication can be incredibly effective, even for non-responsive users. They incur little cost in terms of IT support and they locate a password change request in the midst of an authentication action — a context in which the user is already prepared to enter their password — and in so doing removes the cognitive load of reading and understanding documentation and deciding how and when to schedule a future password change. While the intercept capability must be built and implemented, the cost afterwards in Help Desk tickets appears to be quite low, with large returns in user efficacy and IT staff efficiency. This finding is particularly useful because our organization had not implemented or measured active intercept before this campaign, though there is prior knowledge that interceptive communication can be more effective in appropriate settings [25, 37, 39].
(3) A proactive stance is needed. This campaign had an idle period of about a month between the initial and final notifications which IT admins had hoped would give users time to comply with this policy change. When examining the data, the opposite is true — we observe very little user action during this idle period, suggesting that reminders create a short attention window for this task and users are not "waiting" to change their password later.

Finally, in this study, our organization valued cost in tandem with efficacy, and thus implemented a longer password update campaign that was cognizant of the unknown burden that might be placed on IT staff. However, every organization has different constraints and incentives that define their operational logistics. For some organizations and situations, expediency trumps all. This study provides insight in uncovering the factors of a large-scale security update at our organization, and other organizations can use these results to design campaigns suited to their needs and circumstances.

ACKNOWLEDGMENTS

We thank our anonymous reviewers and shepherd for their time and invaluable feedback. We would also like to thank our collaborators within the UCSD IT organization for their time and help. In particular, we would like to thank James Dotson, Elaine Fleming, and Mark Hersberger for their help in providing access to data, as well as help in understanding specifics about the campaign. This work would not have been possible without their time and effort! Further, we would like to thank Edward Wade in his help in clarifying various aspects of institutional infrastructure. We would also like to thank Phillip Lopo and Mike Corn for their support for the project and ongoing collaborations.

Funding for this work was provided in part by National Science Foundation grants CNS-1705050 and CNS-2152644, the UCSD CSE Postdoctoral Fellows program, the Irwin Mark and Joan Klein Jacobs Chair in Information and Computer Science, and operational support from the UCSD Center for Networked Systems.

REFERENCES

[1] Jacob Abbott and Sameer Patil. 2020. How Mandatory Second Factor Affects the Authentication User Experience. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu, HI, USA, 1–13.

[2] Devdatta Akhawe and Adrienne Porter Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *Proceedings of the 22nd USENIX Security Symposium*. USENIX Association, Washington, D.C., 257–272.

[3] Joram Amador, Yiran Ma, Summer Hasama, Eshaan Lumba, Gloria Lee, and Eleanor Birrell. 2023. Prospects for Improving Password Selection. In *Proceedings of the Nineteenth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Anaheim, CA, 263–282.

[4] Ingolf Becker, Simon Parkin, and M. Angela Sasse. 2018. The Rewards and Costs of Stronger Passwords in a University: Linking Password Lifetime to Strength. In *Proceedings of the 27th USENIX Security Symposium*. USENIX Association, Baltimore, MD, USA, 239–253.

[5] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. 2020. (How) Do People Change Their Passwords After a Breach?. In *Proceedings of the 4th Workshop on Technology and Consumer Protection (ConPro)*. IEEE, 8 pages.

[6] Marina Sanusi Bohuk, Mazharul Islam, Suleman Ahmad, Michael Swift, Thomas Ristenpart, and Rahul Chatterjee. 2022. Gossamer: Securely Measuring Password-based Logins. In *Proceedings of the 31st USENIX Security Symposium*. USENIX Association, Boston, MA, 1867–1884.

[7] Joseph Bonneau. 2012. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *Proceedings of the 33rd IEEE Symposium on Security and Privacy*. IEEE, San Francisco, CA, USA, 538–552.

[8] Orçun Çetin, Carlos Gañán, Lisette Altena, Samaneh Tajalizadehkhoob, and Michel van Eeten. 2019. Tell Me You Fixed It: Evaluating Vulnerability Notifications via Quarantine Networks. In *Proceedings of the 4th IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, Stockholm, Sweden, 326–339.

[9] Yee-Yin Choong, Mary Theofanos, and Hung-Kung Liu. 2014. United States Federal Employees' Password Management Behaviors — a Department of Commerce Case Study. NISTIR 7991.

[10] 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. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Montréal, Canada, 1–11.

[11] Lorrie Faith Cranor. 2008. A Framework for Reasoning about the Human in the Loop. In *Proceedings of the 1st USENIX Workshop on Usability, Psychology, and Security (UPSEC)*. USENIX Association, San Francisco, CA, USA, 15 pages.

[12] Lorrie Faith Cranor, Serge Egelman, Jason I. Hong, and Yue Zhang. 2007. Phinding Phish: An Evaluation of Anti-Phishing Toolbars. In *Proceedings of the 14th Network and Distributed Systems Security Symposium (NDSS)*. The Internet Society, San Diego, CA, USA, 16 pages.

[13] Matteo Dell'Amico, Pietro Michiardi, and Yves Roudier. 2010. Password Strength: An Empirical Analysis. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*. IEEE, San Diego, CA, USA, 983–991.

[14] Emma. 2022. Email analytics. https://myemma.com/email-marketing-features/email-analytics/. [Accessed June 7 2022].

[15] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings: Comprehension and Adherence. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Seoul, Republic of Korea, 2893–2902.

[16] Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Emre Acer, Elisabeth Morant, and Sunny Consolvo. 2016. Rethinking Connection Security Indicators. In *Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Denver, CO, 1–14.

[17] Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhimedi, and Sunny Consolvo. 2014. Experimenting At Scale With Google Chrome's SSL Warning. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Toronto, Canada, 2667–2670.

[18] Dinei Florencio and Cormac Herley. 2007. A Large-Scale Study of Web Password Habits. In *Proceedings of the 16th International World Wide Web Conference*. Association for Computing Machinery, Banff, Alberta, Canada, 657–666.

[19] Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. 2014. An Administrator's Guide to Internet Password Research. In *Proceedings of the 28th USENIX Conference on Large Installation System Administration (LISA)*. USENIX Association, Seattle, WA, USA, 35–52.

[20] Eva Gerlitz, Maximilian Häring, and Matthew Smith. 2021. Please do not use !?_ or your License Plate Number: Analyzing Password Policies in German Companies. In *Proceedings of the Seventeenth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, 17–36.

[21] Eva Gerlitz, Maximilian Häring, Matthew Smith, and Christian Tiefenau. 2023. Evolution of Password Expiry in Companies: Measuring the Adoption of Recommendations by the German Federal Office for Information Security. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Anaheim, CA, 191–210.

[22] 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 *Proceedings of the 30th USENIX Security Symposium*. USENIX Association, 109–126.

[23] Carlos H. Ganan and Michel Eeten. 2017. Make Notifications Great Again: Learning How to Notify in the Age of Large-Scale Vulnerability Scanning. In *Proceedings of the 16th Workshop on the Economics of Information Security (WEIS)*. La Jolla, CA, USA, 1–15.

[24] Delbert Hart. 2008. Attitudes and Practices of Students towards Password Security. *Computing Sciences in Colleges* 23, 5 (May 2008), 169–174.

[25] Helen M. Hodgetts and Dylan M. Jones. 2007. Reminders, Alerts and Pop-ups: The Cost of Computer-Initiated Interruptions. In *Human-Computer Interaction. Interaction Design and Usability*, Julie A. Jacko (Ed.). Springer Berlin Heidelberg, Beijing, China, 818–826.

[26] Philip G. Inglesant and M. Angela Sasse. 2010. The True Cost of Unusable Password Policies: Password Use in the Wild. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Atlanta, GA, USA, 383–392.

[27] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. 2011. Of Passwords and People: Measuring the Effect of Password-Composition Policies. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Vancouver, BC, Canada, 2595–2604.

[28] Cynthia Kuo, Sasha Romanosky, and Lorrie Faith Cranor. 2006. Human Selection of Mnemonic Phrase-Based Passwords. In *Proceedings of the Second Symposium on Usable Privacy and Security*. Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, 67–78.

[29] Kevin Lee, Sten Sjöberg, and Arvind Narayanan. 2022. Password policies of most top websites fail to follow best practices. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, USA, 561–580.

[30] Frank Li, Zakir Durumeric, Jakub Czyz, Mohammad Karami, Michael Bailey, Damon McCoy, Stefan Savage, and Vern Paxson. 2016. You've Got Vulnerability: Exploring Effective Vulnerability Notifications. In *Proceedings of the 25th USENIX Security Symposium*. USENIX Association, Austin, TX, USA, 1033–1050.

[31] Frank Li, Grant Ho, Eric Kuan, Yuan Niu, Lucas Ballard, Kurt Thomas, Elie Bursztein, and Vern Paxson. 2016. Remedying Web Hijacking: Notification Effectiveness and Webmaster Comprehension. In *Proceedings of the 25th International World Wide Web Conference (WWW)*. Association for Computing Machinery, Québec, Canada, 1009–1019.

[32] Max Maaß, Henning Pridöhl, Dominik Herrmann, and Matthias Hollick. 2021. Best Practices for Notification Studies for Security and Privacy Issues on the Internet. In *Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES)*. Association for Computing Machinery, Vienna, Austria, 10 pages.

[33] Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. 2013. Measuring Password Guessability for an Entire University. In *Proceedings of the ACM Conference on Computer & Communications Security (CCS)*. Association for Computing Machinery, Berlin, Germany, 173–186.

[34] Robert Morris and Ken Thompson. 1979. Password Security: A Case History. *Commun. ACM* 22, 11 (Nov. 1979), 594–597.

[35] NIST. 2022. Digital Identity Guidelines Frequently Asked Questions. https://pages.nist.gov/800-63-FAQ/. [Accessed September 13 2023].

[36] Simon Parkin, Samy Driss, Kat Krol, and M. Angela Sasse. 2015. Assessing the User Experience of Password Reset Policies in a University. In *Proceedings of the 9th International Conference on the Technology and Practice of Passwords (PASSWORDS)*. Springer, Cambridge, UK, 21–38.

[37] Muhammad Aditya Pratama, Yayu Hizza Anisa, Nur Amilah, Arry Avorizano, and Rizki Edmi Edison. 2021. The Influence of Pop Up Notification on Visual Attention and Learning. *Education Quarterly Reviews* 4, 4 (2021), 9 pages.

[38] Robert Proctor, Mei-Ching Lien, Kim-Phuong Vu, E Schultz, and Gavriel Salvendy. 2002. Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, & Computers* 34, 2 (2002), 163–169.

[39] Pablo-Alejandro Quinones, Jigar Vora, Aaron Steinfeld, Asim Smailagic, Jeffery Hansen, Dan P. Siewiorek, Pete Phadhana-Anake, and Abhishek Shah. 2008. The Effects of Highlighting and Pop-up Interruptions on Task Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52, 3 (2008), 177–181.

[40] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *Proceedings of the ACM CHI*

Conference on Human Factors in Computing Systems. Association for Computing Machinery, Montréal, Canada, 1–13.

[41] Joshua Reynolds, Nikita Samarin, Joseph Barnes, Taylor Judd, Joshua Mason, Michael Bailey, and Serge Egelman. 2020. Empirical Measurement of Systemic 2FA Usability. In *Proceedings of the 29th USENIX Security Symposium*. USENIX Association, 127–143.

[42] Sena Sahin, Suood Al Roomi, Tara Poteat, and Frank Li. 2023. Investigating the Password Policy Practices of Website Administrators. In *Proceedings of the 44th IEEE Symposium on Security and Privacy*. IEEE, San Francisco, CA, USA, 552–569.

[43] Richard Shay, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Alain Forget, Saranga Komanduri, Michelle L. Mazurek, William Melicher, Sean M. Segreti, and Blase Ur. 2015. A Spoonful of Sugar? The Impact of Guidance and Feedback on Password-Creation Behavior. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Seoul, Republic of Korea, 2903–2912.

[44] Richard Shay, Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. Correct Horse Battery Staple: Exploring the Usability of System-Assigned Passphrases. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS)*. Association for Computing Machinery, Washington, D.C., 20 pages.

[45] Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2014. Can Long Passwords Be Secure and Usable?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Toronto, Ontario, Canada, 2927–2936.

[46] Richard Shay, Saranga Komanduri, Patrick Gage Kelley, Pedro Giovanni Leon, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2010. Encountering Stronger Password Requirements: User Attitudes and Behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS)*. Association for Computing Machinery, Redmond, Washington, USA, 20 pages.

[47] Ben Stock, Giancarlo Pellegrino, Frank Li, Michael Backes, and Christian Rossow. 2018. Didn't You Hear Me? — Towards More Successful Web Vulnerability Notifications. In *Proceedings of the Network and Distributed Systems Security Symposium (NDSS)*. The Internet Society, San Diego, CA, USA, 16 pages.

[48] Blase Ur, Jonathan Bees, Sean M. Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Do Users' Perceptions of Password Security Match Reality?. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, San Jose, California, USA, 3748–3760.

[49] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. How Does Your Password Measure up? The Effect of Strength Meters on Password Creation. In *Proceedings of the 21st USENIX Security Symposium*. USENIX Association, Bellevue, WA, USA, 16 pages.

[50] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In *Eleventh Symposium On Usable Privacy and Security (SOUPS)*. USENIX Association, Ottawa, Canada, 123–140.

[51] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. 2010. Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. In *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS)*. Association for Computing Machinery, Chicago, IL, USA, 162–175.

[52] Matt Weir, Sudhir Aggarwal, Breno de Medeiros, and Bill Glodek. 2009. Password Cracking Using Probabilistic Context-Free Grammars. In *Proceedings of the 30th IEEE Symposium on Security and Privacy*. IEEE, Oakland, CA, USA, 391–405.

[53] Yinqian Zhang, Fabian Monrose, and Michael K. Reiter. 2010. The Security of Modern Password Expiration: An Algorithmic Framework and Empirical Analysis. In *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS)*. Association for Computing Machinery, Chicago, IL, USA, 176–186.