Empirical Measurement of Systemic 2FA Usability

Joshua Reynolds, *University of Illinois at Urbana-Champaign and University of California, Berkeley and International Computer Science Institute;* Nikita Samarin, *University of California, Berkeley and International Computer Science Institute;* Joseph Barnes, Taylor Judd, Joshua Mason, and Michael Bailey, *University of Illinois at Urbana-Champaign;* Serge Egelman, *University of California, Berkeley and International Computer Science Institute*

https://www.usenix.org/conference/usenixsecurity20/presentation/reynolds

This paper is included in the Proceedings of the 29th USENIX Security Symposium.

August 12–14, 2020

978-1-939133-17-5

Open access to the Proceedings of the 29th USENIX Security Symposium is sponsored by USENIX.

Empirical Measurement of Systemic 2FA Usability

Joshua Reynolds[†‡] *Nikita Samarin*[‡] *Joseph Barnes*[†] *Taylor Judd*[†]
Joshua Mason[†] *Michael Bailey*[†] *Serge Egelman*[‡]

[†]*University of Illinois at Urbana-Champaign* [‡]*University of California, Berkeley and International Computer Science Institute*
{joshuar3, joshm, mdbailey}@illinois.edu {nsamarin, egelman}@berkeley.edu

Abstract

Two-Factor Authentication (2FA) hardens an organization against user account compromise, but adds an extra step to organizations' mission-critical tasks. We investigate to what extent quantitative analysis of operational logs of 2FA systems both supports and challenges recent results from user studies and surveys identifying usability challenges in 2FA systems. Using tens of millions of logs and records kept at two public universities, we quantify the at-scale impact on organizations and their employees during a mandatory 2FA implementation. We show the multiplicative effects of device remembrance, fragmented login services, and authentication timeouts on user burden. We find that user burden does not deviate far from other compliance and risk management time requirements already common to large organizations. We investigate the cause of more than one in twenty 2FA ceremonies being aborted or failing, and the variance in user experience across users. We hope our analysis will empower more organizations to protect themselves with 2FA.

1 Introduction

Two-Factor Authentication (2FA) is being widely implemented in an attempt to combat the billions of dollars lost yearly to cybercrime and fraud worldwide [21]. A 2019 worldwide survey of over 1,000 executives found that eight in ten organizations are using two-factor authentication, and 96% of executives expect their company to expand their 2FA use [1]. As these organizations integrate a new authentication mechanism into the everyday routine of mission-critical systems, they need to understand and prepare for its impact on their personnel.

Prior research has shown that the rollout and daily use of 2FA have unique and inherent usability challenges, and organizations need to understand them to plan effectively when adopting 2FA. For example, Strouble et al. estimated in 2009 that the U.S. Air Force was losing about 14 work-years per year to missing 2FA cards [33]. Prior lab studies and user surveys [3–5, 7–14, 16, 17, 19, 25, 28, 29, 31, 33, 34] have identified issues and pain points in both the setup and daily use of 2FA systems. However, prior work has focused mainly on individual devices, user interface choices, and specific user populations rather than overall organizational impacts.

There are two important questions when organizations are estimating 2FA system integration costs. First, what systemic usability effects are evident across a 2FA system at scale? Second, what factors plausibly explain variance observed in the systemic usability of 2FA across organizations?

Three studies took qualitative approaches to answering these questions. Colnago et al. [7], Abbott and Patil [2], and Dutson et al. [13] examined 2FA deployments at large universities. Because these studies focused primarily on survey methods to measure these challenges, we investigated to what extent quantitative evidence corroborated these findings. Partnering with two large public universities' security teams we quantify at scale the impacts of these issues using anonymized records, including over 35 million 2FA login attempts, thousands of support tickets, telephony charge records, enrollment dates, and account credential compromise records. These universities were the University of Illinois at Urbana Champaign (UIUC) and the University of California, Berkeley (UCB) and they both use Cisco's "Duo" two-factor authentication service. Our contributions include confirming some prior findings, contradicting others, and providing new insights across organizations and their implementation choices.

Our results support Dutson et al.'s observation that 2FA's optional reliance on the phone system was repeatedly cited as an annoyance [13]. Comparing the error rates among second factors, we find that telephony-2FA is the most error-prone. From support ticket text, we learned that telephony issues were the second factor which most often drove users to seek technical support. All three studies document errors stemming from desyncing, misreading, and mistyping hardware token codes. Looking at technical support tickets, our findings corroborate Colnago et al. and Abbot and Patil, in that new user enrollment generated the largest support burden [2, 7].

Tracking new users through the first 90 days of 2FA use, we largely corroborate the finding that 2FA has a quick learning curve. Our qualitative data from support tickets match the Colnago et al. finding that asking users to utilize their own personal device for 2FA bothers a small number of people due either to ideological limitations on what their employer can demand, or annoyance about needing yet another app.

At the same time, while Colnago et al. [7] found that one of the most commonly reported 2FA inconveniences is the extra time it requires, we observed that most users are probably only spending about ten minutes per year on 2FA. This estimate comes from an analysis of the frequency and type of 2FA ceremonies at each university combined with Reese et al.'s and Lang et al.'s measured time to complete 2FA ceremonies [19, 29]. We believe that the discrepancy is due to user perceptions of the process, rather than the actual time lost to 2FA ceremonies. This low time cost also might help explain Colnago et al.'s finding that the burden of 2FA was perceived by regular users to be lighter than they feared, as measured by a pre-adoption survey. Dutson et al. and Colnago et al. also suggested longer device remembrance timeouts as a method of reducing user burden [7, 13]. We simulated both longer and shorter device remembrance windows to learn the theoretical impact on user burden and observed diminishing returns from increasing timeouts. We also found that the 2FA login frequency at UIUC, which allowed no device remembrance, was very similar to that of UCB because of other factors. We show that the practical impact of different client devices, fragmented authentication services, and web session timeouts can have just as large an impact as device remembrance. For example, adding device remembrance to a system with short session timeouts will have a larger effect than adding remembrance to a system that already has long web session timeouts.

While Dutson et al., Abbot and Patil, and Colnago et al. both identify a plethora of errors occurring in 2FA [2, 7, 13], to the best of our knowledge, we are the first to break down the frequency and variance of these errors, and compare them across second factor types and user populations. Both Colnago et al. and Dutson et al. found that 2FA users find 2FA to be easy to use, but annoying. We found that this annoyance could be driven by the failure of more than 1 in 20 2FA ceremonies. Furthermore, by observing the time between user errors and their next successful login, we learn how much time 2FA errors waste. Most 2FA errors take about a minute to resolve. However, for 20% of users, a successful login is not usually observed again until hours or days later.

We hope this information and our recommendations will enable more organizations to make an informed choice whether to adopt 2FA to strengthen their authentication systems.

2 Background and Related Work

Two-factor authentication (2FA) combines any two of: something you know (e.g., a password), something you have (e.g., a smartphone), or something you are (e.g., your fingerprint). Current 2FA systems typically use "something you know," like a password or a public key, as the first proof of identity in an authentication ceremony. Common second factors include SMS/phone calls, physical tokens, biometrics, standalone one-time password (OTP) generators, OTP applications, and push notifications. These are a heterogenous mix of secondary identity proofs, which are included in O'Gorman's and Bonneau et al.'s classifications of authentication mechanisms [5, 24].

The goal of adopting a 2FA system is to make stolen account credentials useless for attackers who do not possess the second factor of authentication. Stolen credentials would otherwise grant this attacker access to critical systems.

2.1 Strengths and Weaknesses of 2FA

2FA has the potential to drastically reduce account compromise for an organization. Doerfler et al.'s study of Google's authentication system [12], which intelligently adds extra authentication challenges, including 2FA, saw a success rate over 90% against known attackers. Their system optionally employs 2FA among other signals, such as CAPTCHAs, browser fingerprinting, and geo-fencing to detect abnormal logins. It can take advantage of the stronger guarantees of 2FA, but reduces the user burden by considering factors that usually require no user interaction. A machine-learning algorithm presents authentication challenges of increasing difficulty when a login attempt is classified as abnormal.

Each 2FA systems has a different attack surface. For example, an attacker can act as a "reverse proxy," relaying credentials from a phishing page in real time to the legitimate login site. The attacker can then man-in-the-middle any second factors that rely on SMS, OTPs, phone calls, or push notifications. In addition to being vulnerable to reverse proxy phishing, telephony-based 2FA can lead to permanent account compromise when paired with phone network infrastructure attacks [20, 22, 23, 32]. However, methods like U2F and WebAuthn's incorporation of browser-validated domain information mitigate reverse-proxy threats and do not rely on the phone system. Biometric second factors have the unique challenge of irrevocability [26].

2.2 Known Usability Issues with 2FA

Two-Factor authentication systems combine the usability characteristics [30] of multiple authentication schemes. Some challenges are specific to new users, others to specific populations [8, 25, 31, 33]. These past works' findings identify

individuals' difficulties and suggest better design decisions using qualitative data, whereas the goal of our work is to quantify these effects of 2FA adoption across organizations.

Lang et al. expressed their support for security keys as a 2FA method [19] and measured the time taken to authenticate with security keys vs. time taken to authenticate with other one-time-password (OTP) options. Reese et al. also reported results of measurements of the time taken for various other second factor devices [28, 29]. Reese et al.'s measurements form the basis for our estimations of total user time spent authenticating across our datasets. Lang et al. also report overall counts of support tickets submitted by Google employees over time during Google's internal adoption of security keys. Strouble et al. found in 2009 that the U.S. Air Force lost a combined total of 14 person-years per year to lost 2FA cards [33]. Das et al. compared various MFA Solutions (Duo, Microsoft, Google, Okta, and Authy) by the ratings and sentiments of user reviews of their respective apps [9]. They found general user discontent with the leading MFA solutions and suggested improvements to account recovery, second factor migrations, user training, and risk communication.

2.3 Studies of 2FA Impact on Organizations

Prior work studying the organizational impacts of 2FA has primarily relied on survey methodologies to identify prevalent issues, gauge user perceptions, and suggest system design improvements. We present a complementary view of these impacts from a log analysis perspective and directly compare our findings. We are aware of two prior large-scale studies of 2FA deployments at private universities performed by Colnago et al. and Dutson et al. [7, 13]. Abbott and Patil [2] also performed a concurrent study at a public university. Our quantitative-first approach gives us an overlapping, but distinct vantage point on systematic 2FA usability. Our work complements Colnago et al. and Dutson et al. by drawing conclusions primarily from two separate universities using an order of magnitude more logs, which tell a subtly different story than self-reported data. Further, adding analysis from two other universities and comparing with Abbott and Patil shows which findings appear to be most generalizable.

Dutson et al. surveyed 4,275 of approximately 38,500 students, faculty, and staff at Brigham Young University (BYU) one year after 2FA was mandated. Colnago et al. surveyed 1,251 of approximately 20,000 students and staff members before a mandatory 2FA adoption at Carnegie Mellon University (CMU). After adoption had taken place, they surveyed 796 2FA users for comparison. Colnago et al. also reported some analysis of "over 1 million 2FA authentication logs from over 13,000 users" as well as aggregate data about 2FA-related support tickets. Abbott and Patil performed three surveys at various stages of 2FA deployment (n=83, 195, 287) at Indiana University Bloomington. They also analyzed 1,600 support call transcripts and 90 million 2FA logs. Our study offers two new points of comparison which sometimes support, challenge, or expand past findings. For example, both Dutson et al. and Colnago et al. found overall that 2FA users find 2FA easy to use, but annoying. Our analysis supports these conclusions, and we present timing analysis to estimate how much user time 2FA takes a user every year, as well as how long it takes users to recover when they encounter an error in their 2FA process.

Dutson et al. and Colnago et al. identified issues appearing to cause the most errors. Both Abbott and Patil and we use our log analysis to break down these errors by the 2nd factor choice they affected and their relative frequencies. We expand these findings with the addition of data from two more institution plus aggregate information on the campus demographics most impacted by errors. Whereas Colnago et al. were somewhat limited in their analysis of technical support tickets by relying on others' classifications, we sampled and categorized the most common issues from the actual text. We compare our categorization with that of Abbott and Patil's analysis. We corroborate the idea that setup and new 2nd factor setup causes the most tickets, and we add analysis of which second factors caused the most support calls.

Colnago et al. further showed that the burden of 2FA turned out to be lighter than respondents feared in their pre-adoption surveys. Tracking new users through the first 90 days of 2FA use, we largely corroborate the finding that 2FA has a quick learning curve.

Both studies identified gaps in user understanding of the system which they felt could be corrected with improvements to new-user orientation or the user interface. To minimize annoyance, Colnago et al. and Dutson et al. suggested the idea of using 2FA only as needed to protect critical systems and using remembrance of previously authenticated devices to reduce user burden. Augmenting Colnago et al's reporting of the overall usage and effect of device remembrance on user burden, we examine the distribution of user benefit from this option as well as simulate the effects of various remembrance timeouts on user burden. We find that device remembrance is only part of the story, with multiplicative effects also stemming from session timeouts and the lack of universal single-sign-on systems.

3 Methodology

To measure the costs and benefits of large 2FA deployments, we partnered with the account security teams at UIUC and UCB to examine their records and logs kept during their 2018 implementations of 2FA using Cisco's Duo service. In this section, we describe the data records that were kept at each university, as well as our procedure for data cleaning before beginning our analysis. Both universities are large and diverse organizations servicing tens of thousands of students. Both

are using Cisco's Duo system for 2FA. Each has thousands of full-time and part-time employees engaged in professions as diverse as teaching, research, management, maintenance, accounting, fire safety, emergency response, groundskeeping, healthcare, IT, etc. However, the overall population skews towards highly educated students and educators. We therefore analyze differences among subpopulations in Section 5.

UCB provided 32,366,721 anonymized 2FA log events from June 27, 2018 to June 26, 2019, which showed the results of 2FA ceremonies initiated after a user successfully entered their username and password into various services. UIUC shared 1,985,601 anonymized telephony charge records, a log of 6,467,262 2FA events from June 13, 2018 to March 31, 2019, 17,085 2FA-related support tickets, student/employee status data for 38,536 of their 77,931 users, a small survey of early adopters, and engineering time tracking data for the 2FA project. Both universities also shared promotional and informational posters and emails they used to communicate with users during their mandatory 2FA adoption which proceeded in phases through the Summer and Fall of 2018. To protect the privacy interests of 2FA users, UCB, and UIUC we established a joint IRB protocol across UCB and UIUC and we are unable to make this data publicly available.

3.1 Data Cleaning

To ensure data quality, we performed several data cleaning procedures on the 2FA logs and support tickets. We removed duplicate records from 2FA logs (238,338 from UIUC, 156,728 from UCB) as well as malformed logs (2,108 from UCB). We also removed the records of a single user at UCB identified by their security team as a runaway testing script which was responsible for 913,180 failed login events. This left a total of 37,523,629 usable log events with 6,228,924 from UIUC and 31,294,705 from UCB. A sample of the log format can be seen in Appendix A.

Support tickets can be generated by alert scripts, user emails, and user phone calls. Automated 2FA signup alerts accounted for 9,724 of 17,085 support tickets in our dataset. Further, 640 tickets were automatically created when email vacation responders replied to 2FA mass announcements. This left 6,721 user-caused tickets for our analysis, of which 6,169 were handled by the general help desk and 552 were escalated to or raised with the security team, specifically.

Because some of these tickets could be sensitive or embarrassing to the creator, we created an anonymization plan to discard personally identifying information as part of our IRB protocol. We used pattern matching and name lists to redact names, addresses, titles, numbers, etc. This method had false positives and false negatives. Whenever we encountered a ticket with persisting personally identifying information (PII), we stopped our analysis, removed it, and resumed. Whenever we report a quote where our system removed PII, we include the mark "[PII]."

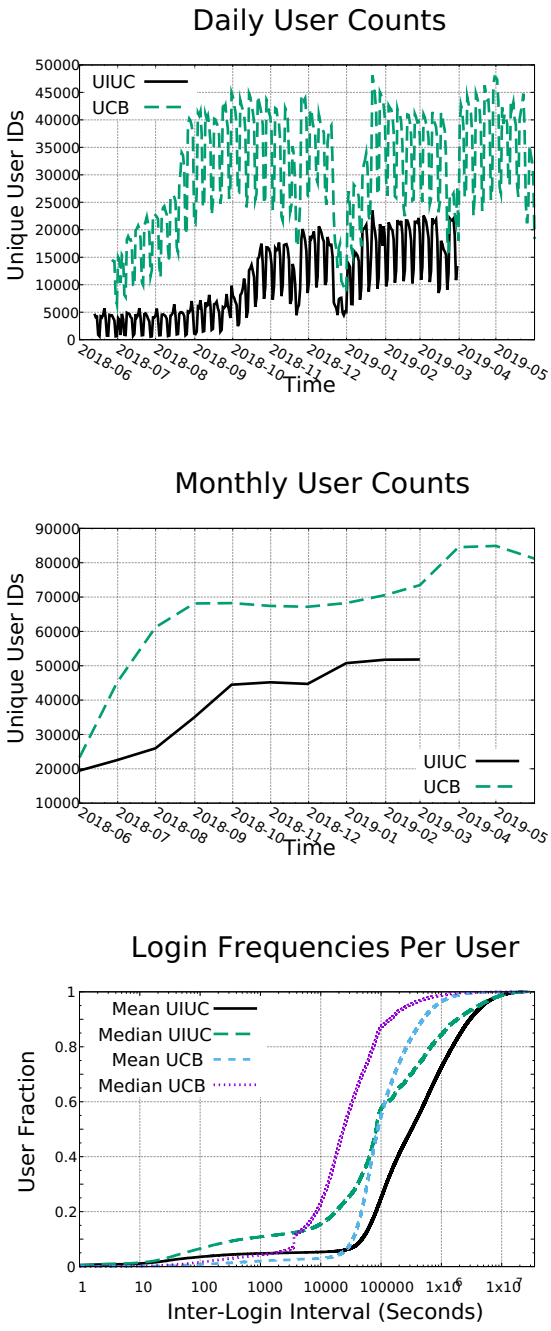


Figure 1: ***Timeline of Unique User IDs Aggregated Daily and Monthly and the Distribution of Login Frequency Per User****— The timeline of daily unique user IDs shows high usage during work weeks and periodic dips on weekends and university holidays. UCB provided a full year of logs, and UIUC provided 9 months of logs. A log was generated every time a user succeeded or failed a 2FA ceremony, and was necessarily preceded by a successful password authentication. The user base increases as more personnel are required to use 2FA and new people join the organization. Steeper upward trends consistent with the gradual 2FA rollout at both institutions are visible in late 2018.*

2nd Factor Type	Time (s)	Count UIUC	Count UCB	Hrs/Yr UIUC	Hrs/Yr UCB	Hrs/User-Year UIUC	Hrs/User-Year UCB
App Push	11.8	2,884,875	5,967,112	11,820.0 hrs	19,721.0 hrs	9.1 min	5.5 min
Phone Call	20.8	865,559	1,272,396	6,251.3 hrs	7,412.5 hrs	4.8 min	4.2 min
SMS/Code	18.4	1,688,161	1,970,448	10,785.5 hrs	10,154.6 hrs	8.3 min	5.8 min
U2F/Yubikey	9.7	204,489	46,427	688.7 hrs	126.1 hrs	0.5 min	0.1 min
Total	-	-	-	29,525.5 hrs	37,414.1 hrs	22.7 min	15.6 min

*Table 1: **Estimated User Time Spent Authenticating**—Using measurements of the time to authenticate using various 2FA methods by Reese et al. and Lang et al. [19, 28, 29], we estimate the total time spent on 2FA per person and overall at these organizations. Their measurements did not use exactly the same interface and systems, so we applied the time measured from the most similar devices in the record. A key difference in UIUC and UCB is that the former has no device remembrance policy, whereas the latter can remember a device for 30 days.*

3.2 Baseline Authentication Behavior

Observed 2FA patterns are, by necessity, strongly tied to existing traditional authentication patterns. We consider the generalizability of our organizations by comparing with studies of traditional authentication patterns. The number of unique users per day and per month for both organizations is given in Figure 1. An average workday sees about 20K of 78K users logging in at UIUC and 40K of 105K users at UCB. The monthly aggregation displays the forced adoption curve of 2FA at both universities as the number of active users rises. Users at UCB re-authenticate several times per day, while users at UIUC usually authenticate every few days. There are also about 35% of users at UIUC and about 20% of users at UCB who log in less than monthly.

4 Systemic Usability of a 2FA Deployment

Understanding the baseline authentication behavior at each university, we begin measuring the user burden evident in these 2FA deployments. We begin by asking how much time a user should expect to have to spend on 2FA. Based on Colnago et al. and Dutson et al.'s suggestions to reduce this overall time using device remembrance [7, 13], we investigate the theoretical and observed benefits of device remembrance.

Next, we ask how much of a burden 2FA errors are causing to these organizations. How often are users resorting to account recovery options? How often do 2FA ceremonies end in failure, and why? How much time does it take a user whose 2FA login fails to solve their problem and log in successfully? We also investigate which problems most commonly force users to seek technical support assistance.

4.1 Time Taken By Authentications

How much extra user time is spent when 2FA is added to their authentication routine? While an individual 2FA ceremony may be fast, the total time over a year may be burdensome. Colnago et al. found that their survey respondents were most annoyed about the time taken by 2FA [7]. However, based on our analysis, we estimate that the average user only spends tens of minutes per year or less on these 2FA systems. There is also a subset of users who end up authenticating far more than their peers. A breakdown of measured user burden by subpopulation will be presented in Section 5.

We estimated this by counting the total number of 2FA authentications divided by the type of second factor used per person. We then leveraged Reese et al.'s published empirical timing estimates for four of five tested types of 2FA [28]. We averaged their findings with the findings for employees and customers of Google as reported by Lang et al. [19] Redmiles et al. [27] also measured SMS 2FA timing, but do not report timing information directly. These estimates show the time users take between successfully entering their username/password and completing the 2FA ceremony. We totaled the user time required at each university to authenticate millions of times per year overall. Because their data includes users learning to use the system, we chose to make our estimates based on the median times reported by Reese et al. This is necessarily a rough estimate because Reese et al.'s users only had two weeks to learn the system, were a smaller sample size, and were using a different custom 2FA system. Lang et al. had a large sample size, but still has a different UI to that our users were given. Further, this estimate is limited by an imperfect mapping of the measured 2FA methods to the 16 distinct second factors labeled in our dataset.

The results aggregated across the organization and normalized to time-per-year are displayed in Table 1. We also report a cumulative distribution function (CDF) of the time required of each user in Figure 2. Organizations should expect users to spend between 10 minutes and an hour per year on 2FA—even if nothing ever goes wrong. At organizations as large as our universities, this aggregate time could be valued at hundreds of thousands of dollars per year (based on an hourly wage). In practice, organizations make these kinds of investments for many kinds of mandatory trainings and programs aimed at reducing overall liability.

At the per-user level, we know the number of logins are not evenly distributed, so we also calculated the distribution of time taken per user. Overall, only about 10% of people at both universities spend more than an hour per year on 2FA.

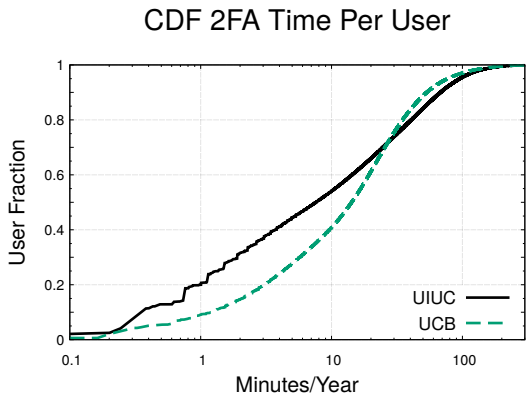


Figure 2: ***CDF of Time Spent Per User****— The total estimated annual time spent on 2FA per person at each university. Based on our authentication frequency data combined with Reese et al.'s and Lang et al.'s past measurements of median 2FA ceremony duration [19, 29]. 90% of people are likely to be spending an hour or less on 2FA per year on average.*

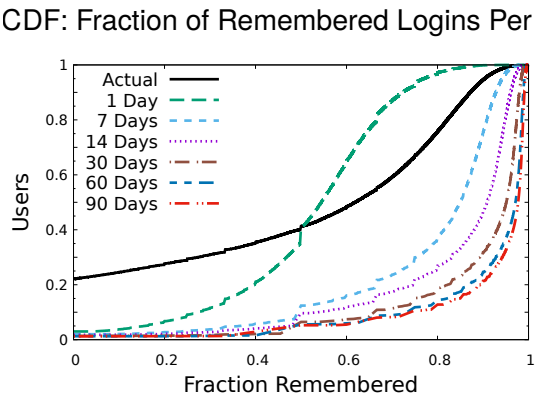


Figure 3: ***Ideal vs. Actual Remember Me at UCB****— We simulated the ideal effects of a "Remember Me" feature for 2FA of different lengths on the 6 months of 2FA login data from UCB. In the idealized simulation, the organization has true single sign-on and the user uses exclusively one device. In reality, fragmented authentication systems and users on various devices lessen the benefits of a "Remember Me" feature. This figure demonstrates that while increasing the device remembrance timeout does decrease user load, the benefit scales inversely with the timeout period.*

4.2 Device Remembrance

One possible mitigation to reduce user burden in authentication is to remember trusted devices on which a successful 2FA has recently taken place. This saves users time on their personal devices and reduces the overall impact of 2FA. Fortunately, our two universities have very different device remembrance policies, which allows us to compare their effects. We report the usability effects from UIUC which had no device remembrance in comparison with UCB which chose an optional 30-day remembrance policy. [1]

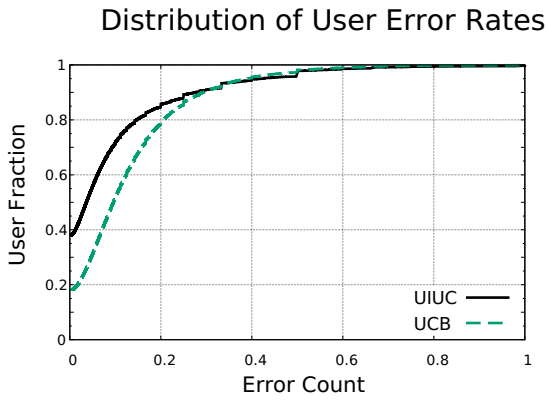
Because 70% (21.1M of 30.0M) of logins at UCB were remembered, tens of thousands of hours of users' time was saved. By the previous timing estimation, a 30-day "remember me" policy has saved approximately 80K person-hours per year for UCB by eliminating 70% of 2FA events. Paid at $20 per hour, an organization with 100K users would experience a yearly indirect cost at between approximately $400K–600K.

The usability benefit was not, however, uniform across users. A CDF of the device remembrance rates per user is given in Figure 3. Colnago et al. reported an overall remembrance rate of 49% with only 55% of users taking advantage of the feature at Carnegie Mellon University [7]. Abbott and Patil reported about 20% remembrance at Indiana University Bloomington, and describe some UI issues that make this feature harder to find [2]. At UCB, by contrast, 80% of users are benefitted by remember me and the overall remembrance rate is 70% with 60% of users able to skip 2FA for at least 50% of their logins. Colnago et al.'s qualitative data revealed that 20% of users were unaware of the remembrance feature, 10% reported being unable to use it, and 12% chose to avoid it.

There are differences between the expected and measured impact of device remembrance. We knew to expect that a 30-day remembrance period does not reduce user burden by 30x due to fragmented login systems, browser cookie deletion rules, user ignorance, and multiple devices [7, 13]. But, we still might expect the average user time per year to be 70% lower at UCB than at UIUC. However, this was not the case. Table 2 shows that UCB users spend about 32% less time on 2FA per year on average. The hidden factor is web service timeouts: UIUC's web services time out after 8-12 hours, whereas users must re-authenticate at UCB after 15-30 minutes of inactivity.

We also ran a simulation to compare the expected impact of a device remembrance policy to the measured impact. We used a six-month period from UCB's data (after the adoption window was over) in which to run a simulation of various remembrance window sizes. For each user, we counted the number of times they would have had to authenticate if it were only required every N days where N = [1,7,14,30,60,90]. This was based on the timing of these users' actual recorded login events (see Figure 1). The results are presented as a CDF in Figure 3. The difference between the predicted impact and the measured impact is due to users blocking 3rd party cookies, changing machines, changing browsers, and not choosing to (or not knowing how to) be remembered. From our data, we could not reliably differentiate a session timeout from other causes of session renewal. The simulation

[1] Each organization arrived at this policy based on differing threat models, and this work will not evaluate which policy provides better protection.



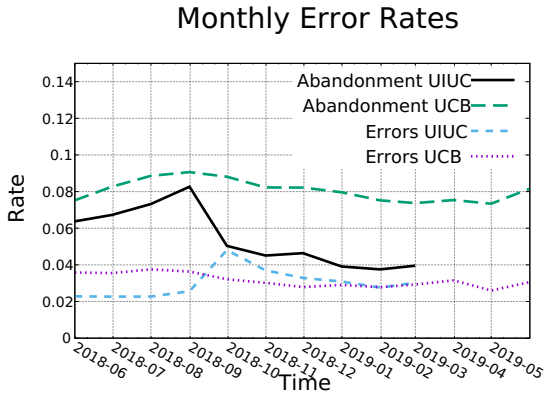


*Figure 4: **Error Rates: Over Time and Per-User Distribution**— Early adopters at UIUC largely matched the error rate at UCB, but as 2FA was forced onto the rest of the users (throughout late 2018), abandonment became far less common and error rates rose. Fewer than 20% of users saw errors more than 20% of the time.*

demonstrates the expected diminishing returns of increasing device remembrance timeouts. The number of required re-authentications scales inversely with the remembrance time.

4.3 Errors in 2FA Ceremonies

We observed that more than one in twenty 2FA ceremonies did not end successfully. This observation was concerning because logs were created only after a user successfully entered their username and password. The first graph in Figure 4 shows the errors over time in the system broken down by user cancellations/abandonment and other errors.

We examined errors by aggregating unsuccessful login attempts from UCB's logs by their reason for failure. Table 2 presents the reasons for failure as well as the second factor device classes they affected. The highest error rate was caused by users canceling or abandoning their interaction, followed by users entering invalid passcodes. This aligns with the

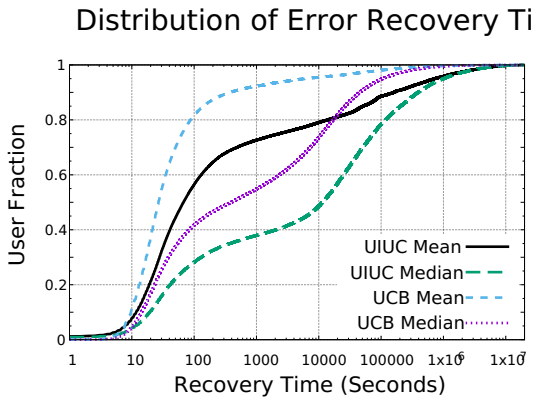


*Figure 5: **CDF of User Error Recovery Times**— This graph shows the mean and median error recovery time per fraction of users. This mean and median are the mean and median of individual users' recovery times. "Recovery time" is the time difference between a failed 2FA login and the next successful login.*

findings of Abbott and Patil [2].

To see whether these errors were common to all users, we also present the distribution of error rates per user in Figure 4. Sixty percent of users experienced 1 to 100 errors and 40% saw no errors at UIUC. Seventy-five percent of users experienced between 1 and 100 errors and 20% of users did not experience errors at UCB. Forty-five percent of users at UIUC and 60% of users at UCB saw error rates under 20%, while more than one in seven users at both universities experienced errors more than 20% of the time. The overall lesser error counts at UIUC may be due to the lack of device remembrance—leading to more frequent logins.

We investigated whether our samples' proximity to mandatory 2FA adoption periods at each institution led to elevated error rates. However, Figure 4 shows that error and abandonment rates at UCB were relatively stable. The early adopters at UIUC shared a similar error rate to the overall steady state of their UCB counterparts. The one difference observed from this perspective was that at the time when UIUC faculty and graduate students were forced to enroll, session abandonment fell. Simultaneously, errors temporarily peaked.

4.4 Recovery Time from Failure

To better understand how much time users spend locked out when experiencing errors, we measured the time between an authentication failure and the next successful attempt. We call the difference between the timestamp of a failed 2FA attempt and the next subsequent successful login the "recovery time." Where there were repeated failures, only the time between the first failure and the next success were counted. Note, that

Failure Cause	Affected 2nd Factors	Count UIUC	Fraction UIUC	Count UCB	Fraction UCB
User Canceled	n/a	87,676	19.22%	558,562	48.19%
No Response	Phone, Duo Push	199,327	43.71%	278,202	24.00%
Invalid Passcode	SMS, Tokens, Passcode, Bypass	153,850	33.73%	187,777	16.20%
Anomalous Push	Duo Push	0	0.00%	77,176	6.66%
Deny Unenrolled User	n/a	0	0.00%	14,546	1.25%
Error	U2F, Phone, Duo Push	18,689	4.10%	21,173	1.83%
No Keys Pressed	Phone	24,293	5.33%	15,300	1.32%
User Mistake	Duo Push	1,671	0.37%	3,357	0.29%
Locked Out	n/a	1,394	0.31%	753	0.06%
Call Timed Out	n/a	0	0.00%	1,797	0.16%
User Marked Fraud	Duo Push	52	0.01%	165	0.01%
Misc Invalid Request	Phone, Duo Push, or n/a	715	0.16%	271	0.02%
Total	Any	487,676	100%	1,159,079	100%

*Table 2: **Causes of Aborted and Failed 2FA Ceremonies at UCB**—The fraction shown is the fraction of total errors at that university which were of each specific type. The leading causes of 2FA failures were timeouts (No Response) and users cancelling their authentication ceremony (User Cancelled). The next leading cause were incorrect passcodes, which includes users who mistype passcodes from SMS, the help desk, a hardware token, or a backup passcode. "No Keys Pressed" indicates a user or their voicemail answered the phone, but did not send a keypress to authorize access. "Deny unenrolled user" is an error triggered when someone is forced to start using 2FA, but has not yet set up any second factors. If users dismiss a Duo Push notification, they can choose to mark the event as a "User Mistake" or fraud. Only UCB enabled a feature to block multiple push notifications from being sent at once. Blocked duplicate requests failed with the code "Anomalous Push." "Error" is a miscellaneous category.*

this metric will capture actual user struggle as well as effects like user distraction. Another source of error could be from users beginning multiple simultaneous 2FA ceremonies and succeeding with one before another timed out.

Average recovery times at both organizations were 10–100 seconds. The full distribution is shown in Figure 5. However, the median recovery time is split between the 10-100 and the 10,000–100,000 seconds range (≈3-28 hours). This means that individual users' recovery times are left-skewed. Hours pass before 40% of users next successfully log in (by their individual median response times). The worst 20% of user's median recovery times indicate that their failed or aborted logins were not successfully retried until at least the next day.

These recovery delays may indicate a productivity cost if important tasks are postponed or forgotten. As one user wrote in a support ticket:

> "Today around 2:20pm I attempted to log into the wiki. I selected Duo Push. Nothing appeared on my phone and after about a minute of sitting and waiting, I got this response: Login timed out…I pushed Send Me a Push again and got this message: Shibboleth has encountered an error…After that, I started over and everything worked that time. (But I have forgotten why I was going to the wiki.)" (HELPDESK-2003)

4.5 Problems Causing Support Tickets

Some problems arising from 2FA were concerning enough that users created tickets with the engineering help desk either online, by email, or by phone. Because previous work has already established the existence of usability problems using rigorous qualitative methods, our goal was mostly to learn which problems were severe enough to be escalated to the level of needing technical support. We conducted a qualitative analysis on support tickets supplied by UIUC using the grounded theory approach. We iteratively performed open coding on a random subset of 6,721 tickets to design a codebook containing 13 codes. The subset was of size 200 and a different subset was drawn at each iteration to avoid sampling bias. Using the resulting codebook, we applied the codes to another random drawn subset of 500 tickets. The results of this process are shown in Table 3. Two researchers independently coded the dataset before resolving any conflicts, yielding a Kupper-Hafner agreement score of .79 ("substantial agreement") [18]. We chose the Kupper-Hafner statistic over Cohen's kappa because our codes were not mutually exclusive, a fundamental assumption for Cohen's kappa [6]. The disagreements were resolved by consensus among the coders before final reporting.

4.5.1 Enrollment and setup issues

The highest proportion of all support tickets are related to 2FA enrollment and setup issues (34.40%±5.48%). These tickets indicate that many users were confused about the nature of 2FA and unable to identify it as a source of error when performing their ordinary tasks. Therefore, a lot of support effort was aimed at explaining what 2FA is and how to initially set it up. As an example, one user said:

> "I am having trouble getting into my school email. I keep getting this message: "Access Denied. The

Code	Notes	# Count	Prevalence	99% Confidence Interval
Setup/Enrollment	Someone requesting help to enroll and setup for themselves or others	172	34.40%	±5.48%
Un-Enrollment	Someone requesting to stop using or disable 2FA for their account	10	2.00%	±1.62%
Update	Someone needing to register a new device or phone number	69	13.80%	±3.98%
Availability	2FA device is lost, dead, without service, broken, etc.	42	8.40%	±3.20%
Recovery Issues	Couldn't get recovery email, prove identity, or refused to share PII	41	8.20%	±3.17%
Phone/SMS	Problem centered on using telephony for 2FA	51	10.20%	±3.49%
App	Problem centered on the Duo Mobile app	26	5.20%	±2.56%
Smartphone	Unclear if user was using app or telephony or clearly both	23	4.60%	±2.42%
Token	Problem centered on using a hardware token	18	3.60%	±2.15%
Feedback	Feature requests, policy complaints, and negative opinions	16	3.20%	±2.03%
Positive Opinion	User expressed support or gratitude for the 2FA system	0	0.00%	±0.00%
New Factor	User tried a new 2nd factor type	16	3.20%	±2.03%
Misc Issue	Unspecified issues, blank tickets, misc. issues	171	34.20%	±5.47%

*Table 3: **Codebook for 2FA Support Tickets from UIUC Part 1**—500 support tickets were coded by two independent researchers. We present the estimated prevalence of these issues across all 6,721 support tickets alongside a 99% confidence interval for proportions. Codes were not all mutually exclusive. We report an extrapolation to the presence of these themes in the full population of tickets with a proportional confidence interval calculated at a .99 confidence level. Overall our agreement was significant to strong with Kupper-Hafner's interrater agreement for non-mutually-exclusive coding (0.79—indicating substantial agreement).*

> username you have entered cannot authenticate with Duo Security. Please contact your system administrator." I wondered if you could help." (HELPDESK-5216)

Certain categories of users were particularly disadvantaged, as they had to be physically on campus in order to enroll in 2FA. The support staff provided an enrollment link to users off-campus, however several users struggled to find the email:

> "I'm an off campus student, and the email that I received a few weeks ago [PII] that I would be receiving a [PII] for [PII] 2FA registration. I never got that link. Can this be sent to me?" (SEC-356)

Others did not know there would be such an email and were concerned that they might have to physically travel to campus to enroll in 2FA:

> "I keep getting [PII] asking me to update my password to 2FA. When I attempt to do so, I get a message that I must be connected to the [BLINDED] network to process it. Today I got an [PII] saying that if I don't update by [BLINDED], my account will be shut off. What am I supposed to do? I have to travel to [BLINDED] to change my password?" (HELPDESK-1998)

Some users also required additional assistance setting up their second factors, including the Duo Mobile app, phones, and hardware tokens. Although most of these issues were resolved easily, other problems were more involved:

> "Helped client enroll in 2FA with a non-smartphone. I first set it up as a landline/basic phone but that option does not allow texts, so I ended up adding my own phone number (which I removed later), removing his number, then adding his number again as a smartphone so that he could use both the call and text options." (HELPDESK-2698)

These tickets demonstrate that a lot of assistance from the support staff is required during the pre-enrollment and initial enrollment stages. Furthermore, providing adequate online resources to users that facilitate the process of 2FA enrollment would likely lead to a reduction in the number of issues experienced by the users.

4.5.2 Updates and recovery issues

Another major source of issues arose when existing users of 2FA had to register a new device or update their phone number (13.80%±3.98%). In many cases, users were locked out of their accounts as they did not have access to their previous device to use the Duo Mobile app or to their old phone numbers to receive a text or a call. In these instances, users had to reach to support staff to obtain a bypass code, which allowed them to access their 2FA settings. For instance, a member of the support team described one such problem:

> "This person has a new phone number to authenticate with for 2FA. I had TL [PII] send them a bypass token and gave them instructions for updating their account's phone number." (HELPDESK-2207)

An additional problem occurred when users had no secondary non-university email registered with their account to receive the bypass codes (8.20%±3.17%). In this case, the support staff also had to verify the claimed identity of the users, which was not always possible. Moreover, it resulted in an additional burden on the staff, as users had to follow up at a later time in order to obtain the bypass code:

> "User called and said he had a new phone number so he needed a bypass to change it in [BLINDED]. I tried to generate a code for him but the bank account did not match so he said he would find it and call back." (HELPDESK-2459)

Sometimes, the users themselves were reluctant to share the information required to verify their identity. For example:

> "I informed the customer of the things we needed in order to send a verification code, but she was not comfortable sharing the last four of her bank account number so she hung up." (HELPDESK-4911)

These support tickets indicate that users should be prompted to provide a secondary communication channel (e.g., non-organizational email) during enrollment to facilitate assistance when they are locked out of their accounts. It is also important that organizations have a mechanism for identity verification that users deem acceptable and non-intrusive.

4.5.3 Second factors and availability issues

When it comes to second factors chosen by users, 51 support tickets are focused on issues with telephony, i.e. calls and SMS (10.20%±3.49%), 26 are related to the Duo Mobile app (5.20%±2.56%), and 18 are centered on the hardware token (3.60%±2.15%). In 23 support tickets users mentioned problems with their device (4.60%±2.42%), and it was not clear from the context whether the Duo Mobile app or telephony was impacted, such as in this user support request:

> "I need to log on ASAP but I don't have access to my phone. There appears to be no option to bypass or send a temporary [PII] to my email address." (HELPDESK-5710)

The context of these tickets varies, depending on whether the issue occurs during the enrollment or the usage of 2FA. Although some tickets do not list any specific problem, device availability is a major theme that emerged from the support tickets (8.40%±3.20%). While all second factors could be affected by a lack of availability, most users experienced problems when they left their device at home or could not receive a call or a text (due to lack of cellular service, international travel, etc.). These problems were aggravated when users were unable to prove their identity to the support staff in order to obtain a bypass code, such as in this case:

> "Client wanted to login to 2FA, but the "call me" was registered with his home phone, and he was not at home. I told him we could send a bypass if he provided the last 4 digits of his back[*sic*] account set up with university direct deposit, but he did not know it. After a minute of searching, he hung up, seemingly upset." (HELPDESK-2220)

Sometimes users were confused about the requirement of an Internet connection to use 2FA. Although the Duo Mobile App requires an Internet connection to receive a push authentication request, it can also be used to obtain a time-based one-time password (TOTP), which does not require an Internet connection or mobile service. Nevertheless, some users were possibly unaware of this functionality:

> "[PII] a PhD studying and is traveling abroad. [PII] having trouble login to the system since my phone number is not available." (HELPDESK-752)

4.5.4 Miscellaneous issues

Other support tickets cover a wide range of topics including assistance setting up a new type of second factor (3.20%±2.03%), un-enrollment requests (2.00%±1.62%), and feedback (3.20%±2.03%). We applied the code 'feedback' to tickets that include feature requests, policy complaints, and negative opinions, as all three aspects came together most of the time:

> "2FA is important for critical/sensitive systems and when accessing systems from off campus, but to implement it across the board for all systems is too much. It's too invasive and starts interfering with productivity. Sometimes the cure in fact IS worse than the disease." (SEC-41)

> "Why can't it be more similar to banking authentication? [PII] have to do the [PII] new device/browser combination and after that it never requires a second factor authentication. [PII] eliminates the nuisance and frustration of having to go through a many stepped process just to download homework assignments and watch lectures." (SEC-52)

Moreover, although tickets that expressed positive opinion appeared during our open coding, none of the 500 tickets that we randomly selected for subsequent coding conveyed support for the implemented 2FA system. An example of a support ticket we encountered during open coding that expresses positive opinion is:

> "I appreciate the fact that 2FA is mandatory. [PII] is a very important tech and I use it wherever possible." (HELPDESK-5665)

4.5.5 Comparison to Related Work

Colnago et al. reported help desk ticket classification statistics provided by CMU's technical support staff [7]. While 2FA help desk tickets were normally less than 5% of their help desk's workload, they swelled to 25% during the mandatory 2FA adoption period. They did not have access to the ticket text, and thus limited their analysis to the categorization done

2nd Factor Choice	UIUC	UCB
Duo Push	6.27%	5.31%
Phone Call	7.36%	6.28%
Duo App Passcode	6.27%	2.55%
SMS Passcode	6.87%	12.21%
Hardware Token Passcode	1.27%	0.13%
Help Desk Bypass	9.65%	31.26%
U2F Token	0.74%	0.73%
Yubikey Passcode	1.97%	1.57%
WebAuthn	-	0.39%
Total	6.11%	5.40%
Remembered Devices	0.00%	0.00%
Unknown Passcode	100.00%	100.00%
2nd Factor n/a	15.56%	75.00%
Overall	7.85%	8.94%

*Table 4: **Comparison of Error Rates With Each Second Factor**—A comparison of the frequency of errors for each second factor type at UIUC and UCB. The **Total** line excludes remembered devices and errors occurring before a second factor was selected.*

by the support staff, a quarter of which were categorized as "Incidents," "Fraud," "Locked Out," and "Broken or Replacement Token." Confusion with the Duo app caused 18% of tickets, and another 18% were attributed to hardware tokens. It is unclear whether the latter were problems with hardware tokens, or merely people seeking to obtain the free hardware token CMU offered. Another 39% were miscellaneous 2FA problems labeled "Request" and "User Questions or Consultation" and "Add to All-SP service."

Similar to our findings, Abbott and Patil found that the largest concerns for users were related to setting up 2FA and finding configuration information (81% of analyzed transcripts), including information about registering an additional device (28%), accessing accounts when the device used for 2FA is inaccessible (16.58%), obtaining a physical token (15%), and interacting with the Virtual Private Network (VPN) for off-campus access (6%). We complement their findings by including information about the specific 2FA factor that contributed to the problem. Dutson et al.'s survey found that the most common issue (52%) reported was losing the phone registered for telephony 2FA or with the Duo app on it [13]. This agrees with our finding that telephony support tickets were the most common device-attributable cause.

4.6 Account Recovery

Many of the miscellaneous 2FA support tickets we analyzed ended with the support staff issuing a temporary account recovery token to bypass 2FA. When users encounter a 2FA problem they cannot resolve easily, they can opt to use this account recovery workflow. UIUC allows up to 24 bypass tokens to be generated by technical support staff or sent to a personal recovery email address. UCB users can only receive a bypass from technical support staff. Using this bypass indicates an issue that prevented a user from using the regular 2FA workflow, *e.g.*, forgetting a phone at home. How often do users resort to this bypass? How common is it among users to have been driven to this workflow at some point?

We examined 2FA bypass tokens at each university per unique user ID. Once obtained, a token is valid for multiple authentications over three days. Only about 5% of users resorted to this bypass, with only about 2% using one twice or more. The group of users at UIUC with the highest per-person recovery rate was the College of Media with an average of 2.7 bypasses per person. The maximum 2FA bypasses by a single user during the time period was 198.

Our analysis revealed that a small number of users tried to use this bypass as their primary authentication method until they ran into yearly maximum limits:

> "Client called because they had run out of bypasses/had requested the maximum amount of temporary passcodes." (HELPDESK-6119)

> "Client called in, does not have her phone with her right now, and has run out of bypass codes." (HELPDESK-5536)

5 Variance in Usability

Our second research question asks what factors beyond the previously discussed system design choices plausibly explain observed variances in usability. Because our methodology was observational rather than experimental, we cannot establish a causality. However, looking at users choices of second factors as well as user demographics suggest plausible explanations for some of this variance. Some demographics of users use the 2FA system in distinct ways, and some second factors are more problematic than others. We expect that the variance observed in each of these dimensions may combine to explain why the error rates at UIUC and UCB differ.

5.1 2FA Preferences and Second Factors

We begin by describing the choice of second factor devices by users at each university. At UIUC, the relative usage of each factor changed over time (Figure 6). Early adopters of 2FA used TOTP/HOTP/Recovery codes about as often as push notifications for 2FA. At the time when all faculty, staff, and graduate students were forced to use 2FA, push notifications and SMS messages became the most common 2FA choices.

At UCB, the distribution of 2nd factor choices was stable. Push notifications are by far the most common 2FA choice at UCB and represent a consistent 60% of logins. Phone calls and app codes are tied for about 15% of logins.

We observed that users did not tend to switch between the second factors that they used. The median user at UIUC tried only two second factor options. The median user at UCB tried

Demographic	#users	#auths	%F-Mdn	%F-Mn	#2nds	%phone	%SMS	%appcode	%app	%Hard	%Yubi	Recovery
Technical Departments	12543	119	5.8	8.7	2.71	13.9	23.0	3.8	46.2	6.4	2.3	1.0
Non-Technical Departments	4360	123.5	6.5	10.0	2.78	17.8	27.4	3.4	38.5	5.6	2.0	1.3
Administration	2287	141	5.3	8.4	2.78	14.8	19.4	2.3	37.3	11.9	5.2	0.6
Sensitive Payroll/HR/Legal	838	111.0	6.4	10.2	2.8	17.3	23.7	2.5	28.9	14.0	3.5	1.0
Misc Offices	33450	109.0	5.8	9.0	2.63	14.4	25.1	3.8	41.4	8.9	1.7	1.1
Facilities	3804	39.5	5.6	10.4	2.23	18.9	27.9	2.6	25.7	20.1	0.4	0.7
Student	18718	131.0	5.9	8.4	2.67	12.0	27.8	4.8	49.4	3.1	0.2	1.3
Faculty	10607	82	5.2	8.6	2.56	15.7	20.1	1.9	29.3	20.7	4.4	0.5
Staff	3317	134	6.5	9.5	2.91	19.2	23.7	3.6	42.0	4.2	2.2	1.1
IT	1178	177.0	5.0	7.8	2.85	11.0	14.6	2.1	55.4	6.5	4.8	0.7
Overall	33450	109.0	5.8	9.0	2.63	14.4	25.1	3.8	41.4	8.9	1.7	1.1

*Table 5: **2FA Usage by Organizational Role**—A breakdown of 2nd factor usage rates, failure rates, and most common 2nd factor choices among various organizational roles. These roles are not mutually exclusive. The table shows the count of users in the group, median count of authentications per user, median and mean failure rates, and the breakdown of 2nd factor choices.*

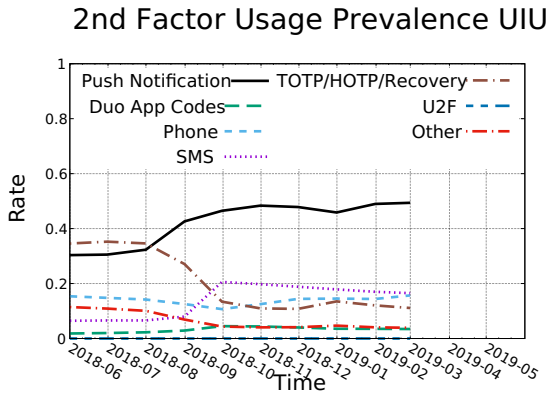


*Figure 6: **CDF of 2FA Method Among UIUC Users**— At forced adoption time, the prevalence of Duo App code logins fell, replaced by SMS and Push Notifications. Apart from that infusion of new users, the relative usage of the different factors were fairly stable. UCB's distribution did not change over time is not shown.*

three. Duo App push notifications were tried at least once by 75% of UCB users , but only by 47% of UIUC users. SMS was tried by 60% of UIUC users, but only by 30% at UCB. Half of UCB's users tried the offline 2FA code generation feature of the Duo App, but only 10% tried it at UIUC. Three percent of support tickets at UIUC were resolved by trying a new form of 2FA. UCB managed to get at least 50% of their users to try a system that would not fail when traveling or without cell service.

This variety of user 2FA choices stands in contrast with Colnago et al.'s findings because CMU does not allow telephony-based authentications. They reported users using 89% push notifications, 5% app codes, and 5% various tokens with users using an average of 1.3 types of 2FA.

We also investigated this error rate in the context of the second factor choices users made at each university. We report the error rates broken down by second factor choice in Table 4. Telephony (phone and SMS) factors had the greatest error rate. U2F token users had the lowest error rate. Unfortunately, failures in entering codes were not always attributed in the logs to a particular second factor.

5.2 Demographics

Another plausible explanation for variance in 2FA usability is users' expertise or the sensitivity of their tasks. To understand what roles within an organization behave differently than others, we analyzed several of our previous indicators in conjunction with generalizable categories of organization members. The results are reported in Table 5.

Because a technical background might have an effect on 2FA effectiveness, we first compared members of academic departments of science, mathematics, engineering, medicine, etc. from departments of law, psychology, sociology, history, etc. We found little difference among the two

populations, showing no evidence of an affect correlating to technical experience. The largest difference was a 10% higher preference for the Duo app over phone and SMS authentications among the more technical departments.

IT staff and civil service staff showed the lowest failure rates when using 2FA, but the effect size is small with less than a 3% overall difference in failure rates among the mean and median members of these groups.

The median number of authentications during our measurement was around 120 for groups except Faculty, IT, and Facilities workers. The median IT worker authenticated about 50% more often than others. The median faculty member authenticated about 25% less often than the overall mean. Facilities and public safety workers authenticated less than half as often as other employees and students.

It appears that hardware tokens are used more often among populations where personal funds are not required to purchase the device. Facilities employees, faculty, and staff working with sensitive data are the most likely to authenticate with a hardware token or Yubikey. Faculty had access to hardware tokens paid for by their departments. Students and overall staff are the least likely to use hardware tokens. For the students, purchasing tokens at UIUC is an extra expense ($10-$40 depending on token type).

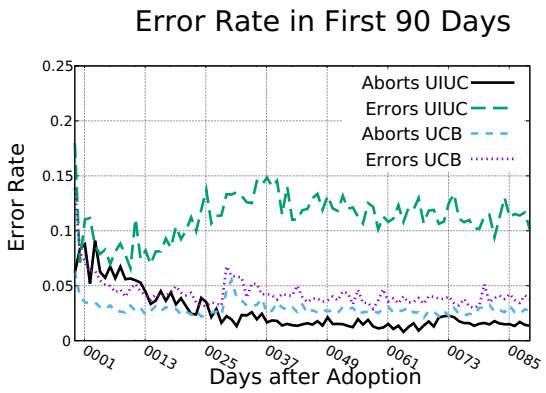


*Figure 7: **Error Rates in First 90 Days Since Adoption**— Daily error rate since first recorded enrollment event per user. UIUC users have an elevated error rate throughout the first 90 days. UCB users' error rates conform to normal error rates after about a month*

5.3 Learning Curve for New Users

We might expect errors to be much higher for users recently joining the system. Figure 7 shows the error rate of users for whom we had 2FA registration information on for each of their first 90 days using the system. There were prominent learning effects at UCB only on the first day, where there was an elevated percentage of incorrectly typed codes. However, at UIUC, There is a clear trend throughout the first month where abandonment rates slowly fall, to be replaced by errors.

6 Discussion

We discuss the findings an organization should consider when planning to implement or improve their 2FA system based on what we learn from these case studies. Our results indicate the day-to-day cost of 2FA to be similar to other compliance and risk-management programs common to large organizations. We caution that 2FA can exacerbate user frustrations with fragmented authentication systems, low or no device remembrance, and short session timeouts. Fragmented authentication systems can also lead to integration challenges as 2FA is turned on across various populations and services. We conclude with an acknowledgement of the limitations of our work and identify open questions for future inquiry.

6.1 Low Compliance Cost of 2FA

The total compliance cost of 2FA in terms of organization time is similar to that of other common risk-mitigation and compliance initiatives, such as trainings in ethics, legal compliance, first aid, etc. In this way, 2FA is not an unusual burden in terms of total time taken per user. However, 2FA differs from these other compliance initiatives in that it becomes an extra task along the critical path to many primary tasks. It therefore appears that user annoyance with 2FA evident in prior work is unlikely to be due to the overall time 2FA takes. However, users may still be experiencing the impression of a long time spent due to their 2FA frequency or due to the recovery cost of errors. As Hauer et al. recently explored [15], users' percieved level of availability differs from actual availability. Furthermore, as more services support or require 2FA, the combined burden across all their services may scale beyond users' patience.

6.2 Multiplicative Effects on the User Burden

Despite making very different choices about session management, UIUC and UCB users end up spending a similar amount of time on 2FA. UIUC allowed no device remembrance during our data collection period, while UCB had a 30-day window available. However, UCB has much stricter session timeout rules than UIUC.

Neither organizations has a single sign-on service that spans every service they operate. This means that users authenticate extra times for each. This was captured in one of UIUC's security tickets where one admin explained to another:

> "First we agree that there are too many prompts... But the real problem... is actually with how we do ...SSO on campus. Currently we have 3 major (and more [PII]) web authentication

> systems on campus: SiteMinder, Shibboleth, and ADFS. None of them share session information with one another... Introducing 2FA has shined a bright spotlight on this problem." (SEC-10803)

A conversion to full single-sign-on (SSO) would reduce the users' burden. One early adopter reported that other organizations have seen this benefit:

> "Colleagues at other institutions report that their 2FA implementation was not nearly so difficult, and that they're only prompted on each device once per 2 wks or less often. Having to do it for each service at least once a day is incredibly cumbersome."(EARLY-73)

UIUC has longer application session timeouts than UCB, which times users out after 15 or 30 minutes of inactivity, depending on the application's sensitivity. This means that UIUC users have a lower authentication burden in general. But this is offset by UCB's choice to allow a 30-day device remembrance window. UIUC users were not allowed to use device remembrance, but individual sessions lasted longer.

It should be feasible to tune the parameters of these two timeouts to reduce user burdens of authentication to the minimum required by the sensitivity of an individual application. Instead of a blanket 2FA remembrance, less sensitive application access requests could be allowed with a remembered device, while especially sensitive apps require 2FA sooner. Session management changes have similar potential to cut down on the user burden as improvements to 2FA ceremony workflows, themselves.

6.3 Limitations

Our study has limitations. We intend our analysis and comparisons to supplement prior findings in many aspects of the user burden added by 2FA, and our findings do not represent a complete measurement of user inconvenience using 2FA. Both of our partner organizations are universities, whose members have either student or employee relationships with the organization. This does not allow us to study 2FA in the context of customers or users of a free service. Both of our organizations contract with the same vendor for 2FA. While their vendor, Cisco's Duo Security, is a leading 2FA vendor, costs and impacts with another vendor's 2FA solution may vary. In general, integration with a specific organization's workflow, practices, vendor software, etc., may be expected to effect 2FA usability.

6.4 Future Work

Open questions in this area include the effectiveness of 2FA at protecting organizations from abuse, measuring the distractive impact of 2FA ceremonies in users' workflows, and encouraging the adoption of better second factors.

Measuring the extent to which 2FA has blocked an attacker from using stolen credentials was something we were unable to do from our vantage point. We observed from account compromise records at UIUC that the rate at which user credentials were stolen did not differ before and after 2FA, as would be expected. What remains to be measured is which of these compromises led to an attacker gaining control of that account. For now, records of successful 2FA logins on compromised accounts indicate either a benign login, or a successful 2FA phish. By the same logic, unsuccessful 2FA logins on compromised accounts likewise indicate either a frustrated attacker or a benign user mistake. Doefler et al. were able to measure this specifically at Google by leveraging a blacklist of known attackers [12]. Developing a method that does not rely on prior knowledge of attackers would allow other organizations to also measure their 2FA's effectiveness.

Work users experience these interruptions on a daily basis, and it would be informative to quantify their productivity impact. 2FA created extra daily distractions for tens of thousands of people at each university. Past work indicates that 2FA distractions incur non-monetary costs on employees' well-being. Zijlstra et al. [35] found that people compensate for time lost to distraction. However, this compensation process incurs an emotional and well-being cost. This idea would support Colnago et al.'s finding that users' initial negative perception of 2FA fades into the background within months of 2FA adoption [7]. But, these interruptions are ongoing and incur an emotional and well-being cost that may explain the annoyance reported by users.

Future work could also try to encourage users to move away from less desirable 2nd factors. After considering the findings of Dutson et al., Colnago et al., Abbott and Patil, and this work, it is surprising that each institution studied has a distinctly different split of 2nd factor choices by their users [2, 7, 13]. The existence of these differences indicates that either environmental factors or design choices by 2FA implementers have the potential to greatly impact 2nd factor selection and drive users to the most desirable 2nd factor options first. UCB's identity team specifically tried to educate users to use the Duo app for push notifications and code generation, and ended up with much higher usage of the app than UIUC.

Telephony 2FA is reliant on the security of the phone network, the slowest method, the most error-prone, incurs recurring charges, and causes the greatest support burden. It generates extra telephone charges equal to about a dollar per user. Problems with telephony-based 2FA were twice as common as any other 2nd factor choice ($10.20\% \pm 3.49\%$). This burden is not proportional to its popularity, which is far exceeded by the use of push notifications. Finally, telephony 2FA has long been known to be vulnerable to direct attacks on phone networks or social engineering attacks on service providers [20, 22, 23, 32]. The choice to incorporate user-owned and controlled devices into the authentication

system also requires extra support effort. While for many users the system is plug-and-play with their devices, some users now need extra technical support when transitioning to a new smartphone or phone number.

7 Acknowledgements

We would like to thank the anonymous reviewers and our shepherd, Elissa Redmiles, for their guidance in improving this paper. We would also like to acknowledge the support of UIUC Technology Services, UCB's CalNet Identity and Access Management team, Nathan Malkin, Ester Cha, Greg Snow, Jeremy Rosenberg, Kaylia Reynolds, Rakib Hasan, Julia Bernd, Alisa Frik, Paul Murley, Simon Kim, Zane Ma, and Deepak Kumar. This work was partially funded under NSF Grants 1528070 and 1817249. The first author was also partially supported by the State Farm Doctoral Fellowship program. The second author was also partially supported by Center for Long-Term Cybersecurity at U.C. Berkeley.

References

[1] 2019 Thales access management index. *Thales eSecurity*, 2019.

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

[3] Claudia Ziegler Acemyan, Philip Kortum, Jeffrey Xiong, and Dan S. Wallach. 2FA might be secure, but it's not usable: A summative usability assessment of Google's two-factor authentication (2FA) methods. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 1141–1145. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[4] B.S. Archana, Ashika Chandrashekar, Anusha Govind Bangi, B.M. Sanjana, and Syed Akram. Survey on usable and secure two-factor authentication. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 842–846. IEEE, 2017.

[5] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy*, pages 553–567. IEEE, 2012.

[6] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[7] 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*, page 456. ACM, 2018.

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

[9] Sanchari Das, Bingxing Wang, and L. Jean Camp. MFA is a waste of time! Understanding negative connotation towards MFA applications via user generated content. In *Proceedings of the Thriteenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2019)*, 2019.

[10] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A comparative usability study of two-factor authentication. *arXiv preprint arXiv:1309.5344*, 2013.

[11] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Gregory Norcie. Two-factor or not two-factor? A comparative usability study of two-factor authentication. *Computing Research Repository*, 2013.

[12] Periwinkle Doerfler, Kurt Thomas, Maija Marincenko, Juri Ranieri, Yu Jiang, Angelika Moscicki, and Damon McCoy. Evaluating login challenges as a defense against account takeover. In *The World Wide Web Conference*. ACM, 2019.

[13] Jonathan Dutson, Danny Allen, Dennis Eggett, and Kent Seamons. "Don't punish all of us": Measuring user attitudes about two-factor authentication. In *4th European Workshop on Usable Security (EuroUSEC)*. IEEE, 2019.

[14] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, 30(4):208–220, 2011.

[15] Tamás Hauer, Philipp Hoffmann, John Lunney, Dan Ardelean, and Amer Diwan. Meaningful availability. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*, pages 545–557, 2020.

[16] Mike Just and David Aspinall. On the security and usability of dual credential authentication in UK online banking. In *2012 International Conference for Internet Technology and Secured Transactions*, pages 259–264. IEEE, 2012.

[17] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M. Angela Sasse. "They brought in the horrible key ring thing!" Analysing the usability of two-factor authentication in UK online banking. *arXiv preprint arXiv:1501.04434*, 2015.

[18] Kupper Lawrence L. and Hafner Kerry B. On assessing interrater agreement for multiple attribute responses. In *Biometrics*. International Biometric Society, 1989.

[19] Juan Lang, Alexei Czeskis, Dirk Balfanz, Marius Schilder, and Sampath Srinivas. Security keys: Practical cryptographic second factors for the modern web. In *International Conference on Financial Cryptography and Data Security*, pages 422–440. Springer, 2016.

[20] Kevin Lee, Ben Kaiser, Jonathan Mayer, and Arvind Narayanan. An empirical study of wireless carrier authentication for SIM swaps.

[21] McAfee. Economic impact of cybercrime. https://www.mcafee.com/enterprise/en-us/solutions/lp/economics-cybercrime.html, 2018.

[22] Collin Mulliner, Ravishankar Borgaonkar, Patrick Stewin, and Jean-Pierre Seifert. SMS-based one-time passwords: Attacks and defense. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 150–159. Springer, 2013.

[23] Collin Mulliner, Nico Golde, and Jean-Pierre Seifert. SMS of death: From analyzing to attacking mobile phones on a large scale. In *USENIX Security Symposium*, volume 168, 2011.

[24] Lawrence O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

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

[26] Salil Prabhakar, Sharath Pankanti, and Anil K. Jain. Biometric recognition: Security and privacy concerns. *IEEE security & privacy*, 1(2):33–42, 2003.

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

[28] Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armnkecht, Jacob Cameron, and Kent Seamons. A usability study of five two-factor authentication methods. In *Fifteenth Symposium on Usable Privacy and Security*, 2019.

[29] Kendall Ray Reese. Evaluating the usability of two-factor authentication. *BYU Masters' Thesis*, 2018.

[30] Karen Renaud. Quantifying the quality of web authentication mechanisms: A usability perspective. *Journal of Web Engineering*, 3(2):95–123, 2004.

[31] Joshua Reynolds, Trevor Smith, Ken Reese, Luke Dickinson, Scott Ruoti, and Kent Seamons. A tale of two studies: The best and worst of YubiKey usability. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 872–888. IEEE, 2018.

[32] David Rupprecht, Katharina Kohls, Thorsten Holz, and Christina Pöpper. Breaking LTE on layer two. In *IEEE Symposium on Security & Privacy (SP)*, 2019.

[33] Dennis D. Strouble, Gregory Schechtman, and Alan S. Alsop. Productivity and usability effects of using a two-factor security system. In *Annual Conference of the Southern Association for Information Systems*, 2009.

[34] Ding Wang and Ping Wang. On the usability of two-factor authentication. In *International Conference on Security and Privacy in Communication Networks*, pages 141–150. Springer, 2014.

[35] Fred R.H. Zijlstra, Robert A. Roe, Anna B. Leonora, and Irene Krediet. Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, 72(2):163–185, 1999.

A Appendix - Duo 2FA Log Sample Format

ID	Time	UserID	Integration	Result	Reason	2nd Factor	Type
000001	12:04:54 10/10/18	ID:24424	CalNet 2-Step Verification	SUCCESS	Valid Passcode	Duo Mobile Passcode	Authentication
000002	12:08:13 10/10/18	ID:10353	CalNet Account Manager	SUCCESS	User Approved	Phone Call	Authentication
000003	12:18:07 10/10/18	ID:73278	CalNet 2-Step Verification	FAILURE	Invalid Passcode	-	Authentication
000004	23:18:57 10/12/18	ID:73278	-	SUCCESS	User Approved	Duo Push	Enrollment
000004	23:18:57 10/12/18	ID:73278	sts.illinois.edu	FRAUD	User Marked Fraud	Duo Push	Authentication

*Table 6: **Sample 2FA Log Data**—For clarity, we provide a mock-up of the data available across the logs shared by UIUC and UCB. IP addresses, names, and device names were anonymized and the university identity teams retained the key. Columns not directly reported on (such as integrated Splunk server IDs, and anonymized client IP addresses, and anonymized device names) have been omitted to be concise.*