Let's go in for a closer look: Observing passwords in their natural habitat

Sarah Pearman* Jeremy Thomas* Pardis Emani Naeini*
Hana Habib* Lujo Bauer* Nicolas Christin*
Lorrie Faith Cranor* Serge Egelman† Alain Forget‡

*Carnegie Mellon University, †International Computer Science Institute, ‡Google, Inc.
{spearman,thomasjm,pardis,hana007,lbauer,nicolasc,lorrie}@cmu.edu
egelman@icsi.berkeley.edu
aforget@google.com

ABSTRACT

Text passwords—a frequent vector for account compromise, yet still ubiquitous—have been studied for decades by researchers attempting to determine how to coerce users to create passwords that are hard for attackers to guess but still easy for users to type and memorize. Most studies examine one password or a small number of passwords per user, and studies often rely on passwords created solely for the purpose of the study or on passwords protecting low-value accounts. These limitations severely constrain our understanding of password security in practice, including the extent and nature of password reuse, password behaviors specific to categories of accounts (e.g., financial websites), and the effect of password managers and other privacy tools.

In this paper we report on an *in situ* study of 154 participants over an average of 147 days each. Participants' computers were instrumented—with careful attention to privacy—to record detailed information about password characteristics and usage, as well as many other computing behaviors such as use of security and privacy web browser extensions. This data allows a more accurate analysis of password characteristics and behaviors across the full range of participants' web-based accounts. Examples of our findings are that the use of symbols and digits in passwords predicts increased likelihood of reuse, while increased password strength predicts decreased likelihood of reuse; that password reuse is more prevalent than previously believed, especially when partial reuse is taken into account; and that password managers may have no impact on password reuse or strength. We also observe that users can be grouped into a handful of behavioral clusters, representative of various password management strategies. Our findings suggest that once a user needs to manage a larger number of passwords, they cope by partially and exactly reusing passwords across most of their accounts.

CCS '17, Dallas, TX, USA
 978-1-4503-4946-8/17/10.
DOI: 10.1145/3133956.3133973

1 INTRODUCTION

Text passwords are ubiquitous and have been a topic of interest for usability and security researchers for many years. Researchers have reported for decades that users, despite good-faith efforts and interest in the security of their information, struggle to comply with password creation and management guidelines [1, 22, 31] and fail to create secure passwords [20, 36, 46]. As the number of accounts per user and the amount of data protected by passwords increases, so does the motivation for password-cracking attacks. To protect against these attacks, modern password guidelines suggest that passwords should be at least eight characters in length, should not contain common and easily-guessed words [19], should contain multiple character types, and, ideally, should be randomly-chosen [17]. Furthermore, users should create distinct passwords—all meeting these complexity requirements—for all of their accounts [48]. Given the complexity required by modern password advice, combined with the number of accounts that a frequent Internet user possesses, password management places unrealistic demands on human memory [4, 43].

Researchers have sought to understand users' current password management strategies and limitations to inform the design of secure systems and interfaces that account for the human in the loop [25]. However, many of these studies have depended on self-reports in surveys or on other indirect measurements [21, 42, 46], or have solely focused on one password per user, such as the passwords revealed in password leaks from a particular website [33] or passwords specifically created for a study [30, 52, 54]. The ability to examine all of the passwords individual users use in their daily online activities, as well as the broader context in which they are used, is critical to understanding important security properties, such as the extent to which users reuse their passwords, how much of a password users reuse, whether users understand the concept of higher- and lower-value accounts, and to what extent password managers improve the strength or usability of passwords.

To date, there have been few quantitative *in situ* studies that encompass all or most of a user's passwords. Most notable is Wash et al.'s field study that investigated password behaviors captured over six weeks from the daily online activities of 134 participants [50]. Unlike previous work, this study was able to investigate how people reuse passwords across different websites. They found that their participants reused each of their passwords for 1.7–3.4 websites. They also found that passwords that were entered frequently or that

were more complex were reused more often. While providing valuable insights, the study did not examine the degree to which people *partially* reuse passwords, e.g., by creating a "new" password by appending a digit to an old one. Furthermore, their use of entropy as a proxy for password strength leads to the question of whether their findings hold for other measures of password strength.

In this paper we describe a comprehensive, longitudinal, *in situ* study of passwords and password behaviors. In particular, leveraging an ongoing longitudinal study of multiple user behaviors called the Security Behavior Observatory (SBO) [14, 15], we examine—with careful attention to privacy and approval from our ethics review board—the password use information and hashed passwords of 154 participants over an average of 147 days (21 weeks). We also collect a wide range of computer usage information, ranging from the presence of password managers and web browser privacy extensions to the presence of malware and update status of software on each participant's computer.

More specifically, to examine how passwords and parts of passwords are reused across accounts, participants' web browsers were instrumented to compute and store the hash values of passwords and of three-character-or-longer substrings of passwords that participants entered into any web site. Before hashing the passwords, the browser recorded the length, entropy, domain on which it was used, and counts of each character type for each password observed. To measure passwords' resistance to guessing attacks, an open-source client-side neural-network password guesser [35] was deployed to participants' computers, which computed and recorded the strength of each password.

Relying on this data-collection approach, we believe we are able to get a longer, broader, and more accurate look than has previously been possible at the characteristics of passwords and at password behaviors across the full range of participants' web-based accounts. In particular, our sample is more diverse than the Wash et al. [50] sample, which was made up entirely of university students, and we also examine partial reuse. Furthermore, our analyses use a more accurate strength measurement than Shannon entropy used by Wash et al. Additionally, the SBO dataset allows us to correlate password behavior to other security behavior.

We used this data to further explore the occurrence of password reuse in our participants' everyday online activities. Specifically, we attempt to address questions such as: To what extent do users partially reuse passwords across websites? Are there certain attributes of a password that correlate with more or less reuse? Are stronger passwords used differently than weaker ones? Are there clusters of users that are similar in their password reuse habits? Are there passwords that are similar to each other in how they are used and constructed? Do participants' engagement in other password security behaviors serve as possible predictors of password strength and password reuse?

Our analysis revealed several insights into how people use passwords in their daily activities. We found that passwords that included digits and special characters were more likely to be reused on multiple sites. Passwords used on job- or work-related sites or on shopping sites were also more likely to be reused. Contrary to Wash et al.'s findings [50], we found that stronger passwords were less likely to be reused and that the frequency with which a password was entered did not affect its likelihood of reuse.

Additionally, among exactly- or partially-reused passwords, passwords used on educational, financial, government, or "portal" websites tended to be reused on fewer total domains. Longer passwords, if reused, also tended to be reused on fewer domains. Reused passwords containing digits, however, tended to be reused on more total domains.

Furthermore, our analysis revealed that the use of password managers or autofill functions did not have discernible effects on password reuse or password strength and that a number of other security behaviors that we observed are not reliable indicators of reuse. We also observed that users can be grouped into a handful of behavioral clusters, representative of various password strength selection strategies, suggesting that once a user needs to start managing a large number of passwords, they cope by reusing—both partially and exactly—across multiple websites.

The remainder of this paper proceeds as follows. In Section 2, we provide an overview of prior work studying password strength, password guidance, and password behaviors. In Section 3, we then describe the details of our longitudinal study and data collection. We present our results in Section 4, describing our analysis of user behavior patterns, methods for clustering groups of users and passwords, and models for predicting reuse. In Section 5, we discuss the implications of our findings.

2 RELATED WORK

Text passwords have been the *de facto* standard for authentication between end-users and computer systems since the 1960s [34]. In the intervening 50 years, numerous researchers have shown how user errors stemming from poor usability (e.g., use of low entropy passwords, poor storage habits, memory lapses, etc.) result in poor security [1, 26, 29, 36]. Despite their known shortcomings, passwords are still in widespread use and are unlikely to go away at any point in the immediate future [24]. Thus, the focus of this paper is on the use of text-based passwords; we consider prior work on alternatives to text-based password authentication (e.g., graphical passwords, biometrics, etc. [4]) to be out of scope.

Given the acknowledged permanence of textual passwords [23], our goal is to better understand how people use and make decisions about them *in situ* so that authentication systems can be better designed around human limitations. Towards accomplishing that goal, we use this section to outline prior research that has been performed to better understand what constitutes a strong password in light of modern attacks, how to guide users towards selecting stronger passwords, and how users currently use passwords.

2.1 Password Strength

Ever since Morris and Thompson's seminal work on password cracking [36], there has been an arms race to increase password strength beyond an attacker's ever-increasing ability to crack passwords. These efforts must rely on quantifiable definitions of password strength, which continues to be an elusive term to define. For instance, one naïve approach is based on Shannon entropy [40]:

$$H = \log_2(c^l)$$

Here, the entropy of a password is defined in bits as a function of the character classes used (c = 26 for only lower case letters, 52

for upper and lower case letters, 62 for alphanumeric passwords, etc.) and its length (l). NIST guidelines from 2006 use this formula as a starting point for estimating the password strength of user-generated passwords of different lengths. The guidelines use heuristics to assign entropy values to each character in a password string. However, the guidelines warn that the formula should only be taken as "a rough relative estimate of the likely entropy of user chosen passwords" [5].

Of course, we know that in practice a password such as "password1," which can be easily guessed by a cracker with knowledge of the distribution of users' password choices, is much weaker than "s723ja0xp" (the naïve entropy approach estimates both as being 46.5 bits). In the past decade, several high profile data breaches have provided researchers with the data to refute this approach. For instance, Weir et al. showed this by examining several million passwords that were leaked from various major websites [51]. One key insight from this research is that passwords are not chosen randomly, and therefore strength should be calculated based on how similar one password is to another. Schechter et al. proposed using "popularity" as a metric to prohibit the use of various passwords on a system, once they become commonly used, so as to maintain a wide distribution of passwords between all of the system's accounts [39].

Kelley et al. showed that the number of guesses it takes to crack a password can be used as an effective strength method [27]: they employed this method to compare the relative strengths of passwords created under differing composition policies. Ur et al. showed that while "guessability" is likely the most effective metric for password strength, care must be taken since various cracking algorithms will yield varying results based on configuration options and method employed [47]. Thus, unless multiple approaches are examined, researchers' estimates of password strength may not accurately generalize to the real world.

Dell'Amico et al. further demonstrated the effectiveness of certain password guessing techniques in cracking even strong passwords. They also suggest that there may be a point of "diminishing returns" such that the probability of successfully guessing a password no longer justifies the cost of continuing an attack [9]. Bonneau examined millions of passwords from a major webmail provider and observed that the most common passwords do not even appear to vary much by language or other demographic factors [3]. He also observed that while having a payment card associated with the account increases password strength (likely due to users rationally putting more effort into mitigating the increased risk of account compromise), these passwords were still highly susceptible to offline guessing attacks.

2.2 Password Guidance

Assuming that well-defined strength metrics can be agreed upon, how should these be used to promote stronger passwords? Bishop and Klein suggested that password strength be measured at account creation, so that proposed passwords can be proactively checked [2]. Proctor et al. studied this recommendation by enforcing additional composition requirements beyond a minimum length (e.g., adding symbols and numbers), and observed that contrary to expectations, they had a negligible effect on memorability [37].

Yan et al. performed a study to examine how different types of password advice would result in stronger or weaker passwords, and how this advice would impact memorability [52]. Predictably, they observed that when instructed to either construct a password using a mnemonic phrase or random character assignment, the resulting passwords were significantly stronger than those created in the control condition, wherein no advice was given. When it came to measuring memorability, participants who constructed random passwords reported having a much tougher time remembering them, often resorting to writing them down. These results were consistent with results earlier work by Zviran and Haga [54].

Komanduri et al. examined the effects of different composition policies on a large scale [30]: they recruited over 5,000 participants to construct and then attempt to remember passwords constructed under varying minimum requirements. Ultimately, they observed that longer passwords that do not require special characters (e.g., symbols and/or numbers) generally are both stronger and more memorable than shorter passwords with more stringent composition policies. While textual passwords based on word phrases may seem like they embody greater entropy, Kuo et al. showed that with a properly constructed dictionary, they are still susceptible to guessing attacks [32]. Furthermore, Shay et al. demonstrated several usability challenges related to system generated passphrases in comparison to equivalent strength random passwords [41].

Forget et al. showed that "persuasive technology" could be used to guide users through the process of strengthening their passwords [13]. Their system suggested that users include additional symbols at arbitrary positions. A user study showed significant results, though for usability reasons, the authors recommended that systems suggest users add no more than three additional characters.

In addition to composition policies, graphical meters are another way of offering users password guidance. Ur et al. showed that meters do result in stronger passwords, but that the particular design of the meter is inconsequential [45]. Egelman et al. corroborated these results, but showed that they hold only when users are forced to change a password for an account they perceive as worth protecting [10]. Otherwise, users are likely to simply reuse a password that they use elsewhere.

One shortcoming of meters is that they tend to rely on simple heuristics, rather than actual measures of strength. This leads to inconsistency in the feedback provided on different websites [6]. To provide a data-driven approach to password feedback, Melicher et al. constructed a neural network that outputs the number of guesses likely needed to guess a given password [35]. Since it does not have the computational overhead of other guessing approaches, it can be implemented client-side to provide realtime feedback in the form of a strength meter.

Another problem with meters is that when they indicate that a password is "weak," they often do not explain what can be done to make it "strong." Ur et al. used iterative design to create a meter that conveys better guidance to users and showed it to be effective [44].

Up until very recently, experts frequently told users to periodically change their passwords. However, research by Zhang et al. called this practice into question, because "new" passwords are frequently predictable modifications of "old" passwords [53]. Chiasson and van Oorschot also found in a quantitative analysis that the security benefit of password expiry was "minor at best" and

"questionable in light of overall costs" [8]. Florêncio and Herley made the case that *most* advice offered to users about passwords is misguided, because it does little to address current attack vectors and because better system administration practices are often a much more effective solution [12].

In 2017, the National Institute of Standards and Technology published new authentication guidelines that differed from previous guidelines in their emphasis on usability. The new guidelines discourage complex password policies and arbitrary password expiration periods [19].

2.3 Password Behaviors

Confronted with large numbers of passwords, users develop coping strategies. Hayashi and Hong performed a diary study of daily password use in 2011 [21]. From their 20 participants, they observed that participants entered passwords an average of 75 times during a two-week period, which the authors estimated corresponded to over 11 online accounts. Their participants reported using various aids to remember their passwords for around 40% of their accounts. The number of passwords that users are expected to manage has likely increased over the years. For instance, Gaw and Felten performed a study five years earlier and found that most users had up to three passwords [16].

In 2007, Florencio and Herley performed arguably the first large-scale study of password behaviors [11]: they instrumented participants' web browsers to record password reuse across websites. They concluded, based on three months of data collection from half a million users, that the average web user has under seven unique passwords that are each reused across four websites. Since then, others have used large data sets of leaked passwords to examine users' password choices (e.g., [33]).

Another stream of research lies in gathering qualitative data to better understand users' methods and attitudes toward password creation [46]. Shay et al. examined 470 university computer users' attitudes about their experiences with more stringent password requirements recently introduced by the university [42]. They observed that while users were predictably annoyed by the new policies, they ultimately believed that they served the greater good of increasing security. Inglesant and Sasse observed that similar tensions between usability and promoting security occur in the workplace [25].

Most relevant to our work is a study performed by Wash et al., who instrumented 134 students' web browsers to analyze password usage on the web over a six-week period [50]. Their study provided unique insights into password construction, use, and reuse. We use a similar methodology to answer additional questions about *in situ* password use among a more generalizable sample. Additionally, our software instrumentation allows for a deeper analysis into the password reuse habits of users, with an examination into the reuse of password substrings. Furthermore, we analyze the use of passwords in a broader security context by observing correlations of password reuse and strength with other security behaviors, such as the use of privacy-enhancing browser extensions and password managers.

3 METHODOLOGY

For our analysis, we use data collected by the Security Behavior Observatory (SBO), a longitudinal study of the security behaviors of Windows computer users [14, 15].

Study participants use their own home computers, which are instrumented with data collection software. The software suite is composed of system-level processes that collect a variety of security-related metadata, including information regarding system configuration, system events, operating system updates, network packets, and installed software, as well as browser extensions that collect data including browsing history, browser settings, and presence of browser extensions.

The data collection software is designed to run passively without interfering with users' normal activities. In order to maximize the ecological validity of the study, participants in the SBO are not prompted or instructed to change their behavior in any way beyond what is necessary to install and run the data collection software.

The SBO has been recruiting continuously since 2014. Participants may leave the study at any time, so the data collection period varies for each user. The SBO has collected data from approximately 512 machines in the period between fall 2014 and summer 2017 and is currently collecting data from approximately 200 machines.

The SBO protocol is approved by our institution's ethics boards. Participants receive $30 for enrolling as well as $10 per month for participation.

A researcher conducts an enrollment phone call with each participant during which the researcher explains the consent form and study protocol and provides the participant with the opportunity to ask questions. The consent form explains the types of data that may be collected from the computer, including network traffic, input from devices connected to the computer, and interaction with websites. After the consent process is completed, a researcher assists participants with the installation of the data collection software and browser extensions. If there are other users of the computer being enrolled in the study, those users must also complete the consent process before data collection can begin. Each participant is also required to complete a short demographic survey at the time of enrollment.

Data is encrypted in transmission and stored securely on a hardened server accessible only to maintainers and collaborating researchers. Participants may leave the study at any time.

3.1 Password Data Collection

In January 2017, we updated the SBO browser extensions for the Google Chrome, Mozilla Firefox, and Vivaldi browsers. We added functionality to securely collect metrics regarding the use and composition of passwords entered within the browser. To collect all passwords entered by participants, we identified every HTML input field on every browser event (such as clicks, key presses, page loads) and filtered using heuristics to extract unique password submission events. For each password, we collected a salted (one-way) hash of the password text and composition metadata such as character length and number of characters in each character class (uppercase, lowercase, special characters, and digits). Using hashes of the password text allows us to analyze password reuse patterns and password attributes without collecting plaintext passwords. In

addition to the hashes that permitted analysis of password reuse, for each password we also collected the length, the number of characters from each character class, hashes of each password's substrings of three or more characters, and a calculated measure of strength.

During the password data collection period that began in January 2017 and ended in July 2017, the SBO has received data from 294 distinct browsers associated with 224 SBO participants. (Other SBO participants either used unsupported browsers such as Internet Explorer or Microsoft Edge or had technical issues preventing browser data transmission.) Of those 224 participants, we excluded 28 who reported during the consent process that there were other users of their computers. Since we are primarily concerned with understanding individuals' management of their own password portfolios, and particularly individuals' decisions regarding password reuse, we did not want to confound the analysis by receiving passwords from multiple users of the same machine. From the remaining 196 participants, we also excluded 42 who had not had the updated browser extension running for at least 28 days and/or had not sent any password inputs. Ultimately, we analyzed password data from 154 users.

We considered all of the passwords typed by our 154 participants into a web browser running our extension during the study period, with a few exceptions. Since passwords are hashed and salted on a per-browser basis, we were unable to assess reuse of passwords across different browsers. If a user sent data from two or more browsers during the observation period, we only considered the user's primary browser, which we identified as the one that had sent data on more days. Some cases of duplicate records occur due to unrelated technical issues, since a new browser record may be created if the SBO software or the browser is uninstalled and reinstalled, and others occur when users use more than one browser regularly (e.g., using Chrome for certain tasks and Firefox for others). 54 secondary browser records were discarded from 43 participants with multiple browser records. On average, the excluded browsers were used on approximately 26 days during the observation period, whereas the main browsers included in this analysis were used on approximately 84 days.[1]

To avoid analyzing mistyped or incorrect passwords, we applied filtering logic to limit each participant to at most one password for each website. (Even if a user changed a password during our data collection period, this logic counted only one password per website.) For each website, we observed the set of all passwords submitted and selected the password with the highest number of submissions throughout the term of the study. In the case of an equal number of submissions, we then selected the password submitted on the highest number of days. And in the case both methods failed to distinguish a password, we then selected the password with most submissions across all websites. In the rare case, less than 3% of all passwords, where multiple participant passwords still remained for a website, we selected the most recent of the submissions. This approach is similar to the methodology used by Wash et al. [50].

[1]The browser extensions for the software are installed for all browsers installed on the users' computer at the time of enrollment, even browsers that the users report that they do not use.

3.2 Password Reuse

We divided password reuse into *exact* and *partial* reuse. We identified exact reuse by examining hashes of the full contents of password fields in HTML form elements. If the same hash value appeared across two or more domains for a given user, we considered the corresponding password to be exactly reused. We identified partial reuse by computing hashes for all substrings of length four or more in each password. We considered a password to be partially reused if it includes a four-character (or more) substring that also appears in a different password belonging to the same user on a different domain.

We chose to identify passwords as "partially reused" for this analysis only if they shared a substring of four or more characters in part because we were concerned about identifying coincidental reuse of trigrams that might not constitute meaningful reuse of a significant portion of a given password. By limiting partial reuse to only four-character substrings, we may miss some passwords that contain common substrings with unique characters inserted in the middle. However, we checked for instances of passwords that share two 3-character substrings that we did not label as partial reuse and found only two instances across our entire data set.

Passwords may be both exactly and partially reused. For example, if *password1* was identified as a user's correct password on Website 1 and on Website 2, *password1* would be considered to be exactly reused. If *pass1234* then appeared on Website 3, that would mean that *password1* was also partially reused, since a four-character substring of that password ("pass") appears on a different domain. We refer to passwords that are both exactly and partially reused as having *exact-and-partial* reuse. To distinguish passwords that are either exactly or partially reused but not both, we refer to *only-exact* and *only-partial* reuse.

We also introduce the notion of *unique passwords*, which correspond to the set of passwords belonging to a user, excluding those that are exactly reused. In the above example, the set of unique passwords is {*password1, pass123*}, while the set of passwords is {*password1, password1, pass123*} since *password1* is used both on Websites 1 and 2.

3.3 Password Strength Measurement

To analyze password strength, we used an open-source implementation of a client-side model of password guessing based on neural networks [35]. The guesser was trained on publicly available password datasets to provide an estimated number of guesses (a *guess number*) to crack passwords of 8 to 32 characters in length.

3.4 Detection of Password Autofill

As all browsers included in this study provide password storage or "remember" functionality, we instrumented our collection software to detect the number of key presses within each password field encountered in the browser. Combining this keystroke data with our password length measurements, we distinguished the entries participants manually typed from those autofilled by either the browser or third-party password management software. For example, if we observed an eight-character password submitted with zero keystrokes, then we assume the browser or some other software provided some form of autofill for the submission. We added

this instrumentation several months weeks after we began data collection, limiting us to approximately seven weeks of keystroke data. We collected keystroke data for 546 passwords (329 unique passwords) entered on 305 distinct domains by 90 users. Of those 546 passwords, we observed 311 being autofilled at least once, and 240 were autofilled on 100% of all observations within the period during which keystroke data was collected.

3.5 Website Categories

To determine the categories of websites on which we observed password entries, one of our researchers manually coded 1,030 domains and created a codebook of 15 website categories. We crowdsourced the coding of 1,450 additional website categories on Mechanical Turk, where three workers selected a category for each domain using our category list and coding instructions. In the 308 cases where there was not agreement between at least two of the three Mechanical Turk workers, a researcher re-coded the website category. We then used a script to combine domain names that were actually just variations of the same domain so that we could accurately assess password use within and across domains. Ultimately, we observed and analyzed password entries on 2,077 distinct domains across 154 users.

3.6 Third-Party Data

We employed third-party blacklists and databases as necessary to identify malicious or risky files, downloads, page visits, and events in the SBO dataset. In particular, for this analysis, we compared browsing data to blacklists gathered from the Google Safe Browsing API to detect downloads of dangerous programs [18]. Additionally, in order to detect malware and potentially unwanted programs present on users' computers, we compared file hashes from users' filesystems to results in VirusTotal's database, which compiles virus scan results from multiple commercial and open-source antivirus products [49]. File hashes were classified as malware or potentially unwanted if they were flagged by 25% or more of the scanners whose results were aggregated by VirusTotal.

4 RESULTS

We begin by describing the demographics of our participants. Next, we provide an overview of our participants' password use and reuse behaviors, as well as the characteristics of their passwords. We describe five password reuse behavior patterns that we observed. Then we discuss password strength distributions and characteristics of reused passwords. Finally, we explore correlations with security behaviors and intentions.

4.1 Demographics

Our participants' ages ranged from 19 to 79, with a median of 26 and a mean of 31.5. Thus, our sample skews markedly younger than the general population. However, given the bias towards younger users (usually convenience samples of university students) that is commonly seen in behavioral research, including usable security and password research, the fact that this sample does include at least some older users is of value. Additionally, our sample is biased towards female users, with 60.4% of users self-identifying as female. Our participants had varying education levels, with 16.9% having completed a graduate (Master's or Doctoral) degree, 40.3% a Bachelor's degree, and 5.2% an Associate's degree as their highest levels of education. A minority of our participants (27.9% of 154) completed some college, and 8.4% completed high school or a GED.

4.2 Passwords in the Wild

We observed a total of 4,057 passwords that our 154 participants submitted to 2,077 different web domains. When we count each of a participant's passwords only once, we find 1,522 unique passwords. Table 1 provides a summary of the participants and passwords observed in this study. Table 2 shows the distribution of website domain categories to which passwords were submitted.

4.2.1 Password Characteristics. On average, participants submitted a password 1.40 times per day. Including all passwords, participants had an average password length of 9.92 characters with their average password composed from 2.77 character classes including 2.70 digits, 5.91 lowercase letters, 0.84 uppercase letters, and 0.46 special characters. The average strength of all passwords in our dataset was on the order of 10^{12} guesses. The distribution of average password strength per user is shown in Figure 1.

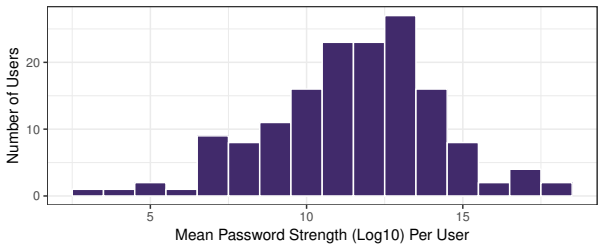

Figure 1: Histogram showing average password strength per user. Password strengths were calculated using the neural network guesser and then averaged for each user.

Additionally, 14% of participants used one password that had a guess number that was at least two standard deviations or more above the mean strength of all their passwords. Half of these outlying passwords were neither exactly or partially reused.

As shown in Table 2, website categories for which these passwords were used varied; half were created for either a shopping or educational website, while others were used on social network, government, research, portal, or tech/software/filesharing websites. Figure 5 shows how password reuse varies by category.

4.2.2 Reuse Characteristics. The average participant used 9.88 unique passwords and submitted passwords to 26.34 different web domains, resulting in a median domain-to-password ratio of 2.39. From our set of unique passwords, 1,578 in total, 511 or 32% were exactly reused. We observed partial reuse in 833 or 53% of all passwords. Combined, we observed any form of reuse in 951 or approximately 60% of all unique passwords.

We found that most participants reused the majority of their passwords on multiple accounts. As seen in Figure 2, a quarter of our participants maintained a set of passwords in which over 90% of the passwords exhibited either partial or exact reuse. We observed exact reuse in 67% and partial reuse in 63% of the average participant's

Table 1: Summary statistics computed for each participant across all 154 participants in our data set. We first computed means for each participant. Then we computed the mean, median, standard deviation, minimum and maximum values of these participant means. "All Passwords" statistics include all instances of reused passwords. "Unique Passwords" statistics include each of a participant's passwords only once, regardless of how many times they were reused. Active days refer to days in which participants were observed using their main web browser. Domains per password, exact reuse password, and reused substring refer to the number of domains on which each password, reused password, or reused substring was observed.

	All Passwords					Unique Passwords				
Statistic	**Mean**	**Median**	**SD**	**Min**	**Max**	**Mean**	**Median**	**SD**	**Min**	**Max**
Passwords	26.30	24.00	18.00	1.00	103.00	9.88	8.50	6.55	1.00	36.00
Password entries per day	1.40	1.17	1.45	0.01	9.81	-	-	-	-	-
Password entries per active day	2.11	1.85	1.66	0.09	13.30	-	-	-	-	-
Page visits per day	81.30	66.20	76.30	0.26	427.00	-	-	-	-	-
Page visits per active day	123.00	97.80	89.90	7.20	490.00	-	-	-	-	-
Days in study	147.00	168.00	54.70	31.00	217.00	-	-	-	-	-
Days active	84.10	80.00	49.30	4.00	207.00	-	-	-	-	-
Domains per password	2.71	2.39	1.37	1.00	9.20	-	-	-	-	-
Domains per exact reuse password	5.99	5.24	3.28	0.00	21.50	-	-	-	-	-
Domains per reused substring	3.73	3.10	2.40	0.00	14.00	-	-	-	-	-
Percentage non-reused passwords	21.20%	15.00%	21.90%	0.00%	100.00%	40.20%	40.00%	21.00%	0.00%	100.00%
Percentage only-exact-reused passwords	15.90%	1.89%	24.70%	0.00%	100.00%	10.10%	1.85%	17.30%	0.00%	100.00%
Percentage only-partial-reused passwords	11.80%	8.96%	14.00%	0.00%	66.70%	25.40%	25.00%	20.10%	0.00%	100.00%
Percentage exact-and-partial reused passwords	51.10%	60.00%	30.30%	0.00%	94.70%	24.40%	25.00%	16.40%	0.00%	71.40%
Password length	9.92	9.53	1.54	7.33	15.70	10.20	9.82	1.81	7.25	16.80
Password character classes	2.77	2.72	0.49	1.83	4.00	2.68	2.67	0.40	2.00	4.00
Password digits	2.70	2.25	1.51	0.50	8.90	2.69	2.33	1.54	0.40	11.30
Password lowercase letters	5.91	6.00	1.91	1.24	13.00	6.06	6.00	1.90	1.83	13.00
Password uppercase letters	0.84	0.75	0.60	0.00	3.75	0.94	0.78	0.62	0.00	3.75
Password special characters	0.46	0.37	0.35	0.00	1.83	0.53	0.46	0.44	0.00	3.08
Password guesses (log10)	11.50	11.90	2.70	3.11	18.20	11.90	12.00	2.47	4.39	17.50

Table 2: Number of distinct domains in each category.

Category	**Count**	**%**
Shopping	360	17.3
Educational	297	14.3
Financial	197	9.5
Jobs/Work	195	9.4
Tech/Software/Filesharing	156	7.5
Research	144	6.9
Hobby/Interest/Game	121	5.8
Health/Fitness	79	3.8
Social Network	62	3.0
News/Media/Entertainment	47	2.3
Government	43	2.1
Portals (sites like google.com with diverse sets of uses and functionalities, often including email)	25	1.2
Adult	6	0.3
Other	179	8.6
Unknown	166	8.0

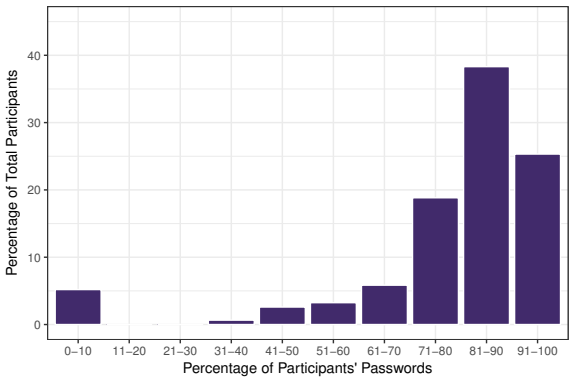

Figure 2: Percentage distribution of participants' passwords that are reused (partially or exactly). Nearly 40% of participants reuse 81-90% of their passwords.

passwords. Furthermore, the average participant reused, partially or exactly, 79% of their passwords.

From the partially reused passwords we extracted 603 unique shared substring values with lengths ranging from 4 to 20 characters. These 603 substrings appeared 1,704 times within 833 or 53% of all passwords. Therefore, among partially reused passwords, the average password is related to 3.66 other passwords, by an average of 2.05 different substrings. As shown in Figure 3, most partially reused passwords include a shared string of 4 to 8 characters, although some include longer shared strings. In Figure 4, we see the distribution by length of the non-shared portion (the remaining characters that are not a part of the shared substring) of partially reused substrings. This highlights the large amount of partial reuse

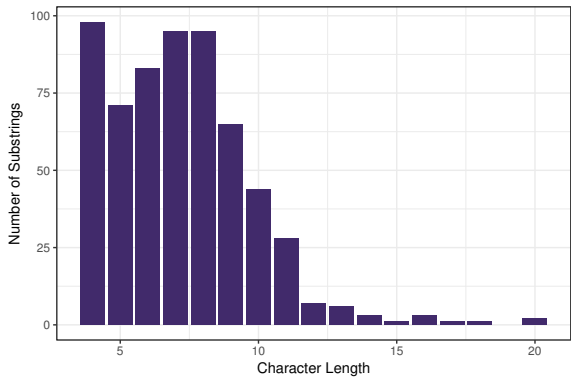

Figure 3: Character length distribution of shared portion of partially reused passwords. Length 4 substrings are most common, followed closely by length 7 and 8 substrings. The mean length of reused substrings per participant ranged from 4.0 to 13.5, with an overall mean across participants of 7.18 characters (median = 7, SD = 1.72).

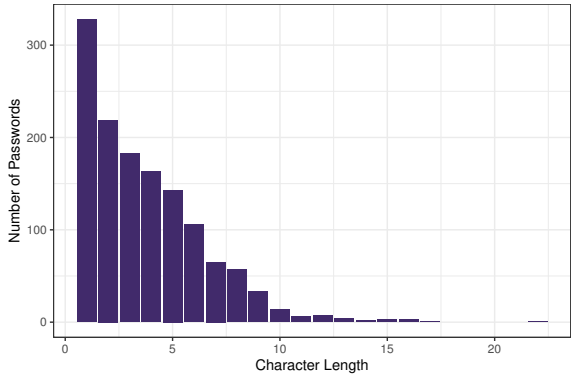

Figure 4: Character length distribution of non-shared portion of partially reused passwords. The most common non-shared portion length is 1 character. The mean length of non-reused substrings per participant ranged from 0.5 to 9.0, with an overall mean across participants of 3.25 characters (median = 3.08, SD = 1.81).

involving small changes of only a handful of characters with one character differences most common (22% of partially reused passwords). We identified 682 instances of the substring used as a prefix (within 61% of partially reused passwords) and 530 instances of the substring used as a suffix (within 52% of partially reused passwords). These overlap, as many partially reused passwords share substrings with more than one other password.

We also investigated whether participants tend to reuse passwords differently within the same category of website, rather than across categories. We found, overall, that password reuse was rarely limited to a specific category of website: only 2.64% of reused (exactly or partially) passwords were reused within the same category. Focusing specifically on financial websites, where passwords likely protect high-value accounts, most financial passwords were reused

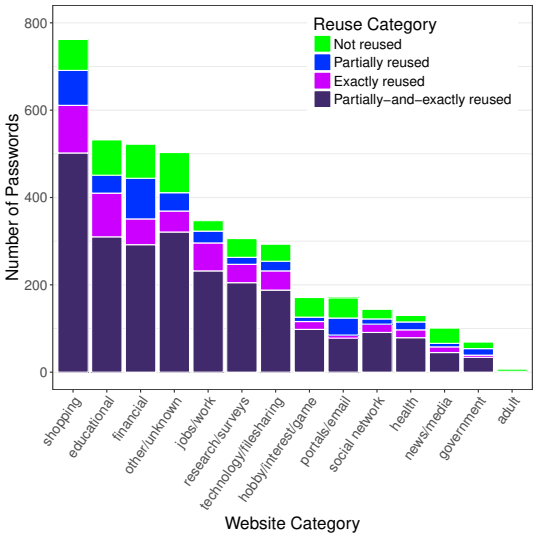

Figure 5: Observed passwords by category. Our participants entered passwords on more shopping websites than on any other category of site. Each category is further broken down by the extent of password reuse for that category. Partial and exact reuse is most common across all categories.

and most of that reuse included reuse in different categories. Specifically, 85.06% of passwords used on financial sites were reused, and 95.50% of those passwords were reused for other types of websites.

4.3 Groups of Password Reuse

To investigate the strategies people use to create passwords, we analyzed the reuse behaviors of our participants in further detail. Taking into consideration all of a user's passwords, we identified the proportions of their passwords that were only-exactly reused, only-partially reused, exactly-and-partially reused, and neither exactly or partially reused. Participants were then grouped by their dominant strategy, which we determined to be the type of reuse observed for at least 50% of their passwords. We distinguished five password reuse strategies among our participants, including a mixed strategy.

Unique Password Creators: A group of 10 (6.5%) of participants followed the strategy of creating unique passwords (neither exactly nor partially reused) for at least 50% of the passwords they created. Almost all of these participants had few online accounts and used their passwords infrequently. With the exception of one participant who had 39 online accounts, all unique password creators entered passwords on eight or fewer domains. Additionally, these participants had online activity for only an average of 17% of the days they were enrolled in the study.

Partial Password Re-users: Five people (3.2%) followed the strategy of only partially reusing passwords for at least 50% of the passwords they created. They had, on average, four total passwords, and were active in the study for 29% of the days that they participated. Interestingly, this group generally did not exactly reuse passwords, as passwords that were only partially reused or unique accounted for 80% of this group's passwords.

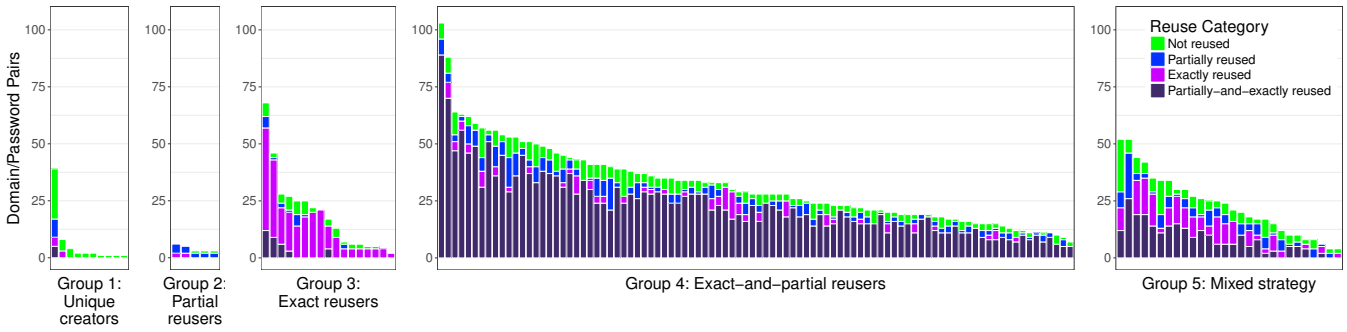

Figure 6: Users grouped by password reuse strategies: Unique password creators (Group 1), Partial password reusers (Group 2), Exact password reusers (Group 3), Exact-and-partial password reusers (Group 4), and Mixed-strategy users (Group 5).

Exact Password Re-users: Seventeen (11%) of our participants followed the strategy of only exactly reusing a password for at least 50% of the passwords they created. These users varied greatly in their percentage of days active online. On average, 72% of their passwords were exactly reused, and did not share substrings with their other passwords. Participants in this group maintained an average of 1.6 distinct reused passwords that were each used across an average of 9.9 different domains.

Exact-and-partial Password Re-users: Ninety-four (61%) of our participants had a set of passwords such that at least 50% were both exactly-and-partially reused across websites. Participants in this group were generally active online, having 32 different online accounts on average. Most (72%) of the passwords these participants used were both exactly-and-partially reused. On average, participants in this group had a set of 3.3 passwords which they exactly or partially reused on 7.4 different domains.

Mixed Strategy Users: Twenty-eight participants (18%) seemed to have mixed strategies for creating passwords, choosing to exactly reuse a password on some domains, partially reuse a password on others, exactly reuse a partially reused password, or occasionally making a unique password for the domain. These participants were also active online, maintaining an average of 22 different online accounts. Passwords that exhibited any form of reuse comprised 75% of these participants' passwords, on average.

In our participant sample, we found that people who have a large number of accounts cope by reusing passwords, either consistently as "exact-and-partial password re-users" or somewhat less consistently as "mixed strategy users." Only 20% of users have a pure strategy when creating passwords (to either exactly reuse, partially reuse, or have no reuse), and those users generally have a small number of passwords.

4.4 Password Strength Distributions

Each participant is using, on average, 9.88 unique passwords. To better understand how users select these passwords, we analyzed the distribution of password strength (expressed by its guess number) over each participant's passwords. We used the Kolmogorov-Smirnov test to calculate distances between distributions corresponding to different pairs of users. We then used hierarchical clustering based on these distances to group users into clusters with similar distributions. Hierarchical clustering converged to four clusters in which participants had similar password strength distributions. These four clusters (1–4) contain the majority (121/154, or about 78.5%) of our users.

Table 3 provides, for each cluster, summary statistics of the strengths of the passwords used. For each participant, we compute the average password strength, and then describe the distribution of these average password strengths across the entire cluster. At first glance, Clusters 1 and 2 appear to have relatively similar, medium-to-high strength passwords; Cluster 3 users seem to pick slightly weaker passwords, and Cluster 4 users appear to use much weaker passwords.

These summary statistics, however, only tell part of the story. Widely different distributions may have, indeed very similar summary statistics. In Figure 7, we plot, for each user, within each cluster, the distribution of their passwords strength. The x-axis corresponds to the logarithm in base 10 of the guess number of a given password; in other words, "10" means that the corresponding password has a guess number of 10^{10}.

Cluster 1, as seen in Figure 7, contains distributions covering nearly the complete range of strength values 0 to 10^{30}. More interestingly, most of the per-user distributions appear to have a relatively narrow couple of peaks—meaning that their passwords all fall within a couple of strength tiers, the largest of which, are, for the most part, between 10^{10} and 10^{15} guesses.

This "multi-modal" behavior is consistent with the partial reuse we frequently observed. Namely, these multiple modes could be the result of people picking a couple of "base passwords," and deriving their other passwords from these base passwords, resulting with similar guess numbers for all derived passwords.

Cluster 2 users, on average, pick stronger passwords than Cluster 1 users. However, we see that the peaks are much "flatter," and, overall, the distributions of passwords chosen by each user are far more spread out. Cluster 2 users may reuse less, or the modifications they make to their base passwords could have a more drastic spread-out effect.

Users in Cluster 3 behave very similarly to those in Cluster 1, with passwords whose strength distribution have a couple of modes; however they pick comparatively weaker passwords than users in Cluster 1.

Table 3: Strength of cluster participants' average passwords.

Cluster	n	Mean	Min	25%	Median	75%	Max	SD
1	38	12.00	10.10	11.38	12.00	12.58	13.50	0.76
2	45	14.00	12.00	13.05	13.80	14.71	17.50	1.33
3	26	10.40	8.59	10.02	10.60	10.84	13.60	1.02
4	12	8.83	8.06	8.50	8.73	9.26	9.53	0.47

Finally, Cluster 4 shows users whose password strengths also follow bi-modal (or tri-modal) distributions. However, overall, the passwords chosen appear to be very weak.

To summarize, Clusters 1, 3 and 4 users pick passwords that, for the most part, are centered on a couple (1–3) of strength levels. The difference between these different clusters is the average strength in their respective user passwords: Cluster 1 users pick generally stronger passwords than Cluster 3 users, who in turn pick stronger passwords than Cluster 4 users. On the other hand, Cluster 2 users pick a broad range of passwords of varying strength, generally leaning toward stronger passwords.

4.5 Characteristics of Reused Passwords

We next examine the passwords we collected to determine whether there are similarities in how they are reused (Section 4.5.1); and whether reuse is affected by passwords' syntactic properties, strength, or the categories of sites where the passwords are used (Section 4.5.2).

4.5.1 Clustering passwords by reuse characteristics. We first study whether passwords—independently of who created them—have any notable similarities based on how they are reused. In particular, we cluster passwords according to their reuse characteristics (e.g., how often they were exactly or partially reused; whether they were reused mostly within or across categories of accounts) and then examine the clusters for patterns.

To perform this analysis, we apply k-means clustering to all the passwords we collected. (Here, if the same password is used on two accounts, it counts as two passwords.) In this clustering, each password is described according to the following seven dimensions.

- Exact reuse: Fraction of a user's accounts on which this password was exactly reused.
- Partial reuse: Fraction of a user's accounts on which this password was partially reused.
- Entries per day: Average number of times this password was used.
- Within-category reuse: Fraction of passwords in the same category of website for which this password is used that constitute exact or partial reuse of this password.
- Other-category reuse: Average fraction of passwords used for other categories of websites that constitute exact or partial reuse of this password.
- Span of category reuse: Fraction of categories in which this password is exactly or partially reused.
- Days site visited: Fraction of days (relative to days within study) on which the user visited the user visited pages within this site (possibly without logging in).

As is standard, we attempted to cluster for increasing values of k starting with $k = 2$, observing the change in the within-clusters sum of squares errors as k increased [28].

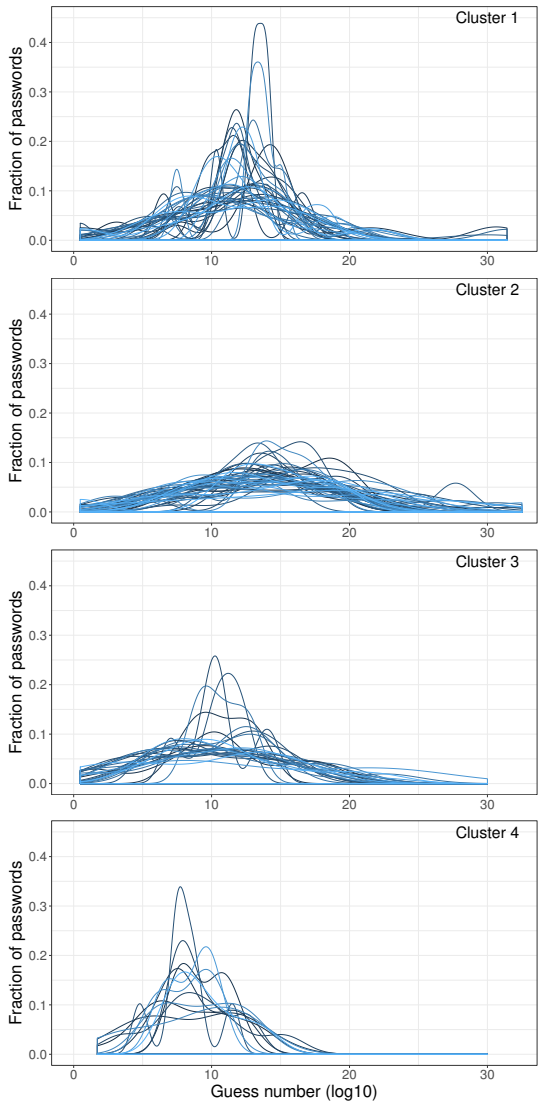

Figure 7: Password strength distribution, by user cluster. Each line represents the distribution of a given user, and each figure represents a given cluster.

We used the elbow method to determine the optimal value of k [28], which led us to cluster passwords into seven clusters, shown in Table 4. While the clustering does not appear to reveal any intuitively significant high-level trends, a few clusters stand out.

Clusters 6 and 7 have the highest strengths among all the clusters—the average number of guesses needed to crack passwords in these clusters is more than a magnitude larger than it is for our overall set of passwords. Both clusters represent passwords for sites that participants visit often (on more than 60% of days), although Cluster 6 passwords are typed in an order of magnitude times more often (1.48 entries per day compared to 0.18 times per day).

Cluster 2 is the next in order of strength, and still substantially stronger than average for all passwords. Passwords in this cluster

Table 4: Descriptions of clusters of passwords resulting from k-means clustering with $k = 7$. "N" is the number of passwords in the cluster; "Part." the number of participants whose passwords are represented in the cluster; "Guess" the log of the average guess number of passwords in the cluster. Remaining columns show the length; number of characters in each character class; and average number of character classes.

			Clustered Attributes							Non-clustered Attributes					
Cluster	**N**	**Part.**	**Exact**	**Partial**	**Entries /Day**	**Within**	**Other**	**Span**	**Days Visited**	**Guess**	**Len.**	**Digits**	**Spec.**	**Upper**	**Classes**
All	4057	154	0.15	0.17	0.05	0.16	0.68	0.38	0.10	11.4	9.87	2.73	0.46	0.82	2.72
1	988	113	0.09	0.09	0.03	0.11	0.89	0.28	0.05	11.17	9.72	2.82	0.51	0.91	2.88
2	297	65	0.09	0.05	0.05	0.74	0.26	0.11	0.15	12.81	10.13	3.08	0.84	0.82	2.72
3	586	144	0.00	0.00	0.03	0.00	0.00	0.00	0.07	11.61	10.37	2.58	0.47	1.12	2.27
4	934	59	0.37	0.13	0.03	0.16	0.84	0.57	0.06	10.65	9.85	2.59	0.25	0.64	2.72
5	998	86	0.11	0.44	0.03	0.15	0.85	0.59	0.06	11.56	9.48	2.68	0.41	0.65	2.76
6	36	27	0.12	0.17	1.48	0.17	0.64	0.27	0.72	13.10	10.67	2.92	0.78	1.28	3.19
7	218	81	0.11	0.18	0.18	0.13	0.79	0.35	0.63	13.01	10.56	3.04	0.67	1.03	3.02

are reused less than Cluster 6 and 7 passwords, but most reuse (> 74%) is within the same category of website. The sites on which these passwords are used are visited less frequently than Clusters 6 and 7.

The weakest cluster—4—has the highest average values for percentage of websites with exact reuse. The sites on which they are used are not frequently visited—approximately one day in 20.

Interestingly, Cluster 3—passwords that are never reused (exactly or partially)—nearly matches the average strength of the entire population.

Overall, in this case the clustering reveals few startling or definitive insights, although it does suggest a few trends (or notable absences thereof): First, with small exceptions, there seems to be little link between reuse type and strength. The exception is that a cluster of passwords (Cluster 4) that are exactly reused much more often than most is also by far the weakest. Second, passwords that are entered most frequently seem approximately average in terms of reuse characteristics but are much stronger than average.

4.5.2 Modeling password reuse. To better understand what factors might contribute to users' decisions to reuse or not reuse passwords, we constructed regression models that attempt to predict (1) whether passwords will be reused wholly, partially, both wholly and partially, or not at all; and (2) for passwords that are reused, how much (i.e., on how many domains) they would be reused.

Based on previous findings regarding reuse [50], we expected that the number of times a password was entered and password complexity and strength would be important predictors of reuse. Other factors we included in the models were the length of the password, the presence of each character class, the strength of the password, and the category of the site for which the password was used.

Explaining whether a password will be reused. Our first model, a multi-level logistic model with intercepts permitted to vary on the user and domain levels, attempted to explain whether a password would be reused (partially or exactly) or not. The model is shown in Table 5.

Properties of passwords that were most strongly correlated with reuse were the presence of digits and special characters. The model suggests that the presence of at least one digit makes a password much more likely to be reused, multiplying the odds of reuse by more than 12 (odds ratio 12.30).[2] The presence of special characters also increases the odds of reuse (odds ratio 2.69). These findings are consistent with previous work that found that complexity was correlated with reuse [50].

Several categories of web sites also had large but different effects. The odds of reuse were tripled for passwords used on shopping sites (odds ratio 3.16). Perhaps surprisingly, the odds of reuse were also approximately tripled for passwords used on job- or work-related sites (odds ratio 3.06).

Some of the categories of websites that one might expect would have an effect on password reuse—e.g., financial sites, which one would expect would have more unique passwords—were not shown as significant in our model.

Finally, password strength was revealed as a statistically significant factor in predicting reuse. For each order-of-magnitude increase in the number of guesses needed to guess a password, the odds of reuse decrease by approximately ~9% (odds ratio 0.91). In other words, the odds of reuse for a password that is one order of magnitude stronger than average would be 91% of the odds of reuse of a password of average strength, and the odds of reuse for a password that is three orders of magnitude stronger than average would be 75.4% of the odds for a password of average strength. This is unlike previous work, where a different measure of password strength was found to be positively correlated with reuse [50].

Explaining extent of reuse. Additionally, we developed a linear multi-level model to attempt to predict, for reused passwords only, the number of domains on which the password would be reused. This model had intercepts varying on domain and user and, besides variables included in the logistic model, also included multiple user-level variables, including whether a user was a student, educational level, gender, programming language knowledge, and age. This model is described in Table 6.

Among the subset of passwords that are partially or exactly reused (the dataset on which this regression was run), passwords with digits are likely to be reused on more domains. Among reused

[2]The odds ratio is computed as e^x, where x is the coefficient shown in the "Estimate" column in Table 5.

Table 5: Logistic multi-level model predicting password reuse. In this and other regression tables, rows ending with * describe statistically significant factors.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.37	0.89	1.54	0.12
log10(length)	-0.50	0.98	-0.51	0.61
uppercase	-0.21	0.15	-1.36	0.17
digit	2.51	0.22	11.59	<0.01*
special	0.99	0.17	5.73	<0.01*
log10(entries)	0.07	0.15	0.45	0.65
log10(nng)	-0.09	0.02	-4.30	<0.01*
cat:adult	-1.69	1.70	-0.99	0.32
cat:educational	0.44	0.28	1.59	0.11
cat:financial	0.13	0.29	0.46	0.65
cat:gov	-0.85	0.48	-1.76	0.08
cat:health	0.27	0.44	0.62	0.53
cat:hobby/int./game	-0.11	0.37	-0.31	0.75
cat:job/work	1.12	0.36	3.06	<0.01*
cat:news/media	-0.27	0.50	-0.53	0.60
cat:portals	-0.74	0.52	-1.43	0.15
cat:research	0.59	0.36	1.62	0.11
cat:shopping	1.15	0.29	3.93	<0.01*
cat:socialnetwork	0.56	0.47	1.18	0.24
cat:tech/filesharing	0.71	0.34	2.08	0.04

passwords, passwords without digits are exactly reused on an average of 8.22 domains and partially reused on an average of 18.17 domains, whereas passwords with digits are exactly reused on an average of 14.52 domains and partially reused on an average of 12.77 domains. When averaging across all passwords, including those with no reuse, passwords without digits are exactly reused on an average of 4.66 domains and partially reused on an average of 10.33 domains, while passwords with digits are exactly reused on an average of 12.44 domains and partially reused on an average of 10.94 domains.

A number of other factors each made passwords likely to be reused on fewer domains. Longer passwords tended to be reused on fewer domains. Stronger passwords, more-frequently-entered passwords, and passwords containing uppercase letters also tended to be reused on fewer domains, but the effects of these variables were quite small. The category of website on which the password was used also sometimes had a small effect: passwords used on portals or on educational, financial, or government sites were all likely to be reused on fewer domains overall.

4.6 Correlations with Other Security Behaviors

The SBO reports data on general security behavior, such as the presence of security extensions (e.g., password managers, anti-viruses) or suspected malware. Table 7 shows the security behavior attributes we analyzed for our study participants. We used this data in conjunction with password-related outcomes—i.e., whether a password was unique or reused, the number of domains a password was reused on, and password strength—to determine if we can predict user behavior with respect to passwords from other security behaviors.

Table 6: Linear multi-level model predicting amount of password reuse. Note: The dependent variable here is transformed to the power of 0.3 based upon a Box-Cox normality plot.

	Estimate	SE	df	t	Pr(>|t|)
(Intercept)	2.64	0.28	248.53	9.33	<0.01*
log10(length)	-0.88	0.17	3153.72	-5.07	<0.01*
uppercase	-0.07	0.02	3065.75	-3.13	<0.01*
digit	0.35	0.04	3085.45	8.99	<0.01*
special	-0.01	0.02	2925.19	-0.26	0.80
log10(entries)	-0.06	0.02	535.42	-3.18	<0.01*
log10(nng)	-0.01	0.00	3142.90	-3.65	<0.01*
cat:adult	0.18	0.31	3025.60	0.58	0.56
cat:educational	-0.10	0.03	1327.15	-3.01	<0.01*
cat:financial	-0.12	0.03	392.84	-3.82	<0.01*
cat:gov	-0.18	0.06	1972.52	-2.83	<0.01*
cat:health	-0.07	0.05	1738.54	-1.45	0.15
cat:hobby/int./game	-0.05	0.05	2778.78	-1.03	0.30
cat:jobs	0.01	0.03	1208.64	0.18	0.86
cat:news/media	-0.04	0.06	1443.44	-0.64	0.52
cat:portals	-0.17	0.05	50.54	-3.30	<0.01*
cat:research	0.07	0.04	661.79	1.95	0.05
cat:shopping	-0.02	0.03	811.62	-0.74	0.46
cat:socialnetwork	-0.01	0.05	1022.10	-0.23	0.82
cat:tech/filesharing	-0.03	0.04	1560.88	-0.71	0.48
student	0.07	0.14	127.60	0.53	0.60
ed:Some college	0.17	0.17	128.07	0.95	0.34
ed:Associate's	-0.07	0.27	128.45	-0.26	0.80
ed:Bachelor's	0.30	0.17	127.59	1.73	0.09
ed:Master's	0.01	0.20	128.72	0.04	0.97
ed:Doctoral	0.76	0.42	123.74	1.79	0.08
gender:Male	-0.01	0.10	130.14	-0.07	0.94
proglangs:Yes	-0.06	0.11	129.46	-0.50	0.62
age	0.00	0.00	128.07	0.72	0.47

Table 7: Security behaviors overview.

	No	Yes
Has password manager	135	19
Has security/privacy extensions	53	101
Dangerous downloads	136	18
Malware detected	128	26

4.6.1 Predicting Reuse with Other Security Behaviors. We first ran a multi-level logistic regression to attempt to predict simply whether a password would be reused (exactly or partially) or not. To that effect, the dependent variable in the logistic model in Table 8 is a binary variable ("reuse") which is coded to 1 if a password is partially or exactly reused, and 0 otherwise.

Since we did not have autofill detection data for all password inputs, this regression model and the others below examine a smaller subset of our password data, comprised of 546 passwords (329 unique passwords) entered on 305 distinct domains by 90 users.

We included five different security behavior variables: whether the user had a password manager, whether they had security- or privacy-related browser extensions, whether they had assented

to downloading programs flagged as dangerous (according to the Google Safe Browsing API), whether malware or adware was detected on the machine (using data from VirusTotal) at some point during the observation period, and the percentage of total entries of the password that were performed with autofill (rather than by manually typing the password). As shown in Table 8, no factors in this model are statistically significant predictors of whether a password will be reused.

Table 8: Logistic multi-level model: Security behaviors as predictors of password reuse.

	Est.	Std. Error	z value	Pr(>|z|)
(Intercept)	2.00	0.45	4.43	<0.01*
has password mgr	0.23	0.52	0.45	0.66
has sec./priv. exts	0.21	0.35	0.58	0.56
dangerous downloads	0.72	0.66	1.10	0.27
malware detected	0.34	0.43	0.78	0.43
percent autofilled	-0.00	0.00	-0.35	0.72

Table 9: Logistic multi-level model: Security behaviors and average amount of daily web browsing as predictors of password reuse.

	Est.	SE	z	Pr(>|z|)
(Intercept)	-0.12	0.80	-0.15	0.88
has password mgr	0.14	0.51	0.27	0.79
has sec./priv. exts	0.05	0.35	0.16	0.88
dangerous downloads	0.72	0.66	1.10	0.27
malware detected	0.20	0.42	0.47	0.64
percent autofilled	-0.00	0.00	-0.24	0.81
log10(avg navs / day)	1.27	0.45	2.83	<0.01*

Table 10: Linear multi-level model: Security behaviors and average amount of daily web browsing as predictors of amount of password reuse (among reused passwords only). Note: The dependent variable here is transformed to the power of 0.26 based upon a Box-Cox normality plot.

	Est.	SE	df	t	Pr(>|t|)
(Intercept)	1.25	0.22	97.68	5.61	<0.01*
has password mgr	0.22	0.14	74.71	1.56	0.12
has sec./priv. exts	0.06	0.10	78.51	0.55	0.59
dangerous downloads	0.26	0.15	82.06	1.74	0.08
malware detected	-0.09	0.12	75.42	-0.76	0.45
percent autofilled	-0.00	0.00	445.25	-1.73	0.08
log10(avg navs / day)	0.41	0.12	89.04	3.27	<0.01*

We wondered whether the user's amount of web browsing could be an omitted variable that might explain other factors including presence of malware (due to increased exposure) and might also affect password reuse and other aspects of password behavior. Thus, we constructed a model that also included average navigations

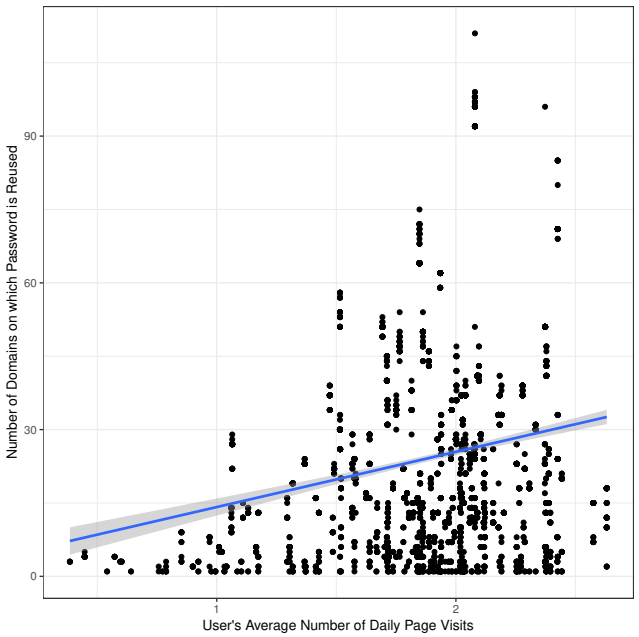

Figure 8: Relationship between user's average daily page visits and number of domains on which a password is reused (for all partially- or exactly-reused passwords).

observed per day (which was calculated as total page visits observed divided by days of observation) as a regressor.

The results, shown in Table 9, indicate that the average number of page visits per day was a significant predictor of reuse. An increase in the log10 value of the user's average navigations per day (e.g., increasing average navigations per day from 100 to 1000) would more than triple the odds of password reuse (odds ratio 3.56).

We also constructed a linear model to attempt to predict, for reused passwords, the number of domains that they would be reused on.

In a model with the four security behaviors plus average daily navigations, summarized in Table 10, the number of average daily navigations is again the only variable indicated to be statistically significant with $p < 0.05$. An one-increment increase in the log10 value of the user's average number of daily page visits (e.g., increasing average navigations per day from 100 to 1000) would predict an increase of 0.41 in the number of domains on which a given password belonging to that user would be reused. The plot in Figure 8 depicts the relationship between the user's average daily page visits and the number of domains on which a password is reused.

4.6.2 Predicting Password Strength with Other Security Behaviors. In addition to measuring possible correlations between security behaviors and password reuse, we also constructed a linear multi-level model that attempts to predict password strength based on security behaviors. Results from this model are shown below in Table 11.

The presence of a password manager did not have a statistically significant predictive effect on password strength in either

Table 11: Linear multi-level model: Security behaviors and average amount of daily web browsing as predictors of password strength.

	Est.	SE	df	t value	Pr(>|t|)
(Intercept)	10.77	0.72	210.79	15.03	<0.01*
has password mgr	-0.51	0.97	77.88	-0.52	0.60
has sec./priv. exts	0.83	0.70	73.62	1.18	0.24
dangerous downloads	0.59	1.03	81.35	0.58	0.57
malware detected	-1.03	0.83	66.26	-1.23	0.22
percent autofilled	0.00	0.00	471.87	1.01	0.31
log10(avg navs / day)	0.85	0.84	87.94	1.02	0.31

model, nor did the percentage of entries of the password that were performed with autofill.

The detection of dangerous downloads, the presence of security- or privacy-related browser extensions, the detection of malware on a user's machine, and the average number of navigations per day also did not have significant predictive effects on password strength.

5 DISCUSSION

Our results show that password reuse—in both exact and partial form—is extremely rampant. Participants in our study have passwords for 26.3 web domains on average, and they appear to deal with the problem of creating and recalling these passwords by partially or exactly reusing approximately 80% of their passwords across domains. While previous work had found high rates of exact password reuse [11, 50], our study suggests that the problem may be even worse that previously thought when partial reuse is taken into account. We observe that on average 16% of a participant's passwords are exactly reused, 12% are partially reused, and an additional 51% are both exactly and partially reused. Thus many participants have clusters of both partially and exactly reused passwords that share common substrings.

Most participants (122 users, 79.2%) adopted hybrid strategies incorporating both exact and partial reuse in order to manage their passwords. Some participants did display simpler strategies of password reuse: 6.5% of participants mostly used unique passwords, 3.2% mostly partially reused passwords, and 11% mostly exactly reused passwords. However, those participants tended to have lower levels of activity and fewer accounts. Participants with larger numbers of accounts tended to either exactly-and-partially reuse their passwords or to employ mixed strategies, presumably in order to cope with the memory demands of their larger password portfolios.

Password managers are increasingly recommended to help users generate random and unique passwords for a large number of accounts [7, 38]. However, similar to Wash et al., we found no statistically significant effect of the presence of a password manager or the use of autofill functionality on the frequency of password reuse. We also found no statistically significant effect of the presence of a password manager or the use of autofill on password strength.

However, we observed only 19 participants who had installed password managers, and although we were able to observe whether some passwords were autofilled, we were not able to determine whether those were autofilled by third-party password managers or by native browser functionality. We are also unable to account for any users that may be accessing password managers on their mobile devices or for whether users were utilizing the password generator functions of their password managers.

If participants are using password managers of any kind to randomly generate and store most of their passwords, we would expect those participants to have consistently strong passwords and very little password reuse. The only participants in our data set with little password reuse had a small number of passwords. Thus we suspect that participants are either not using password managers, or using them only to store the passwords they create themselves rather than to generate and store random passwords. Further investigation is needed to determine whether password managers are able to effectively serve users' needs and relieve the memory demands of modern password portfolios while also encouraging higher security. Changes to password managers may be needed to better facilitate their use as random password generators for non-expert users. Our research and Wash et al.'s findings both suggest that password managers may not be panaceas in their current forms.

Based on previous findings regarding password reuse [50], we expected that frequency of password entry would be an important predictor of reuse. This was not confirmed by our models. Furthermore, the model shown in Table 6 indicates that more frequently entered passwords were actually reused with slightly less frequency, although this effect size is small. In addition, while previous work found that password strength as measured by entropy is positively correlated with reuse, we found that password strength as measured by guessability does not positively correlate with reuse; in fact, we find a weak negative correlation.

The properties we found most strongly positively correlated with password reuse were the presence of digits and special characters. However, we also found that stronger passwords were less likely to be reused. We speculate that passwords that contain digits and special characters lend themselves to reuse because they are likely to satisfy password policies on more domains than passwords without digits or special characters. However, the mere presence of digits and special characters does not necessarily ensure that a password is strong, especially when those digits or special characters are placed in predictable locations. Stronger passwords are generally longer and contain digits, capital letters, and special characters in unpredictable places, which may make them harder to remember or type. It is also possible that users create stronger passwords for accounts they value more, and thus they choose not to reuse them as often.

We also found some effects of website category on password reuse. Passwords used on government websites tended to be reused on fewer domains, which may be because users consider government websites more important in terms of security or may also be related to relatively stringent password composition or expiration requirements on government websites. More surprisingly, we found that passwords used on shopping and job search websites are more likely to be reused and are reused on larger numbers of domains. This is somewhat surprising considering that shopping website passwords may protect sensitive credit card data and that job- and work-related sites may contain other information that

users might want to keep secure, such as payroll and employment information.

Past work has shown that users cope with the unreasonable memory demands imposed by advice to create unique, strong passwords in part by reusing passwords.

Here, by observing a relatively diverse sample of users on their own home computers, in their natural environments, we are able to observe unprecedented detail regarding users' reuse strategies.

6 ACKNOWLEDGMENTS

This work was partially funded by the NSA Science of Security Lablet at Carnegie Mellon University (contract #H9823014C0140); the National Science Foundation, grant CNS-1012763 (Nudging Users Towards Privacy); and the Hewlett Foundation, through the Center for Long-Term Cybersecurity (CLTC) at the University of California, Berkeley. This work was also partially supported by NATO through Carnegie Mellon CyLab, and by a hardware donation from NVIDIA. We thank Rick Wash and Emilee Rader for providing us with detailed information about their field study, Alex Davis for his guidance on the regression analyses, and the reviewers for their assistance in improving the paper.

REFERENCES

[1] Anne Adams and M. Angela Sasse. 1999. Users are not the Enemy. *Commun. ACM* 42, 12 (December 1999), 41–46.

[2] Matt Bishop and Daniel V. Klein. 1995. Improving System Security via Proactive Password Checking. *Computers & Security* 14, 3 (1995), 233–249.

[3] Joseph Bonneau. 2012. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *2012 IEEE Symposium on Security and Privacy*. 538–552. DOI: http://dx.doi.org/10.1109/SP.2012.49

[4] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. 2012. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *2012 IEEE Symposium on Security and Privacy*. 553–567. DOI: http://dx.doi.org/10.1109/SP.2012.44

[5] William E. Burr, Donna F. Dodson, and W. Timothy Polk. 2006. *NIST Special Publication 800-63: Electronic Authentication Guideline.* Technical Report. NIST.

[6] Xavier de Carné de Carnavalet and Mohammad Mannan. 2015. A Large-Scale Evaluation of High-Impact Password Strength Meters. *ACM Trans. Inf. Syst. Secur.* 18, 1, Article 1 (May 2015), 32 pages. DOI: http://dx.doi.org/10.1145/2739044

[7] Andrew Chaikivsky. 2017. Everything You Need to Know About Password Managers. *Consumer Reports* (2017).

[8] Sonia Chiasson and Paul C. van Oorschot. 2015. Quantifying the Security Advantage of Password Expiration Policies. In *Designs, Codes, and Cryptography.* http://chorus.scs.carleton.ca/wp/wp-content/papercite-data/pdf/chiasson2015desi-expiration.pdf

[9] Matteo Dell'Amico, Pietro Michiardi, and Yves Roudier. 2010. Password Strength: An Empirical Analysis. In *Proceedings of IEEE International Conference on Computer Communications (INFOCOM).*

[10] Serge Egelman, Andreas Sotirakopoulos, Ildar Muslukhov, Konstantin Beznosov, and Cormac Herley. 2013. Does My Password Go up to Eleven? The Impact of Password Meters on Password Selection. In *Proceedings of the ACM Computer-Human Interaction Conference.*

[11] Dinei Florêncio and Cormac Herley. 2007. A Large-Scale Study of Password Habits. In *Proceedings of the International World Wide Web Conference (WWW).* Banff, Alberta, Canada, 657–665.

[12] Dinei Florêncio, Cormac Herley, and Baris Coskun. 2007. Do Strong Web Passwords Accomplish Anything?. In *Proceedings of the 2nd USENIX workshop on Hot topics in security.* USENIX Association, Berkeley, CA, USA, Article 10, 6 pages. http://portal.acm.org/citation.cfm?id=1361419.1361429

[13] Alain Forget, Sonia Chiasson, Paul C. van Oorschot, and Robert Biddle. 2008. Improving Text Passwords through Persuasion. In *Proceedings of the 4th symposium on Usable privacy and security (SOUPS '08).* ACM, New York, NY, USA, 1–12. DOI: http://dx.doi.org/10.1145/1408664.1408666

[14] Alain Forget, Saranga Komanduri, Alessandro Acquisti, Nicolas Christin, Lorrie F. Cranor, and Rahul Telang. 2014. *Security Behavior Observatory: Infrastructure for Long-Term Monitoring of Client Machines.* Technical Report 14-009. Carnegie Mellon University CyLab. https://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab14009.pdf

[15] Alain Forget, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, Marian Harbach, and Rahul Telang. 2016. Do or Do Not, There Is No Try: User Engagement May Not Improve Security Outcomes. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016).* Denver, CO, 97–111.

[16] Shirley Gaw and Edward W. Felten. 2006. Password Management Strategies for Online Accounts. In *Proceedings of the second symposium on Usable privacy and security (SOUPS '06).* ACM, New York, NY, USA, 44–55. DOI: http://dx.doi.org/10.1145/1143120.1143127

[17] Google. 2017. Creating a Strong Password. (2017). https://support.google.com/accounts/answer/32040?hl=en

[18] Google. 2017. Google Safe Browsing. (2017). https://developers.google.com/safe-browsing/

[19] Paul A. Grassi, James L. Fenton, Elaine M. Newton, Ray A. Perlner, Andrew R. Regenscheid, William E. Burr, Justin P. Richer, Naomi B. Lefkovitz, Jamie M. Danker, Yee-Yin Choong, Kristen K. Greene, and Mary F. Theofanos. 2017. NIST Special Publication 800-63B: Digital Authentication Guideline. (2017). https://pages.nist.gov/800-63-3/sp800-63b.html

[20] Yael Grauer. 2017. 2016's Worst Passwords Are Just As Bad As 2015's (So Please Tell Me Yours Is Not on the List). *Forbes* (Jan. 2017).

[21] Eiji Hayashi and Jason Hong. 2011. A Diary Study of Password Usage in Daily Life. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11).* ACM, New York, NY, USA, 2627–2630. DOI: http://dx.doi.org/10.1145/1978942.1979326

[22] Cormac Herley. 2009. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *Proceedings of the 2009 workshop on New security paradigms workshop.* ACM, 133–144.

[23] Cormac Herley and Paul C. van Oorschot. 2012. A Research Agenda Acknowledging the Persistence of Passwords. *IEEE Security Privacy* 10, 1 (Jan 2012), 28–36. DOI: http://dx.doi.org/10.1109/MSP.2011.150

[24] Cormac Herley, Paul C. van Oorschot, and Andrew S. Patrick. 2009. *Passwords: If We're So Smart, Why Are We Still Using Them?* Springer-Verlag, Berlin, Heidelberg, 230–237. DOI: http://dx.doi.org/10.1007/978-3-642-03549-4_14

[25] Philip Inglesant and M. Angela Sasse. 2010. The True Cost of Unusable Password Policies: Password Use in the Wild. In *Proc. ACM CHI'10.* 383–392.

[26] David L. Jobusch and Arthur E. Oldehoeft. 1989. A Survey of Password Mechanisms and Potential Improvements. Part 2. *Computers & Security* 8, 7 (1989), 675–689.

[27] Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio López. 2012. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy.* 523–537.

[28] David J. Ketchen and Christopher L. Shook. 1996. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strategic Management Journal* 17, 6 (1996), 441–458. DOI: http://dx.doi.org/10.1002/(SICI)1097-0266(199606)17:6⟨441::AID-SMJ819⟩3.0.CO;2-G

[29] Daniel V. Klein. 1990. Foiling the Cracker: A Survey of, and Improvements to, Password Security. In *The 2nd USENIX Security Workshop.* USENIX Association, Berkeley, CA, 5–14.

[30] 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 Conference on Human Factors in Computing Systems (CHI).* ACM, Vancouver, BC, Canada, 2595–2604.

[31] 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 (SOUPS '06).* ACM, New York, NY, USA, 67–78. DOI: http://dx.doi.org/10.1145/1143120.1143129

[32] Cynthia Kuo, Sasha Romanosky, and Lorrie Faith Cranor. 2006. Human Selection of Mnemonic Phrase-Based Passwords. In *Symposium on Usable Privacy and Security.* ACM, New York, NY, USA, 67–78. DOI: http://dx.doi.org/10.1145/1143120.1143129

[33] David Malone and Kevin Maher. 2012. Investigating the Distribution of Password Choices. In *Proceedings of the 21st International conference on the World Wide Web (WWW '12).* ACM, New York, NY, USA, 301–310. DOI: http://dx.doi.org/10.1145/2187836.2187878

[34] Robert McMillan. 2012. The World's First Computer Password? It Was Useless Too. *Wired* (January 2012). https://www.wired.com/2012/01/computer-password/

[35] William Melicher, Blase Ur, Sean M. Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Fast, Lean, and Accurate: Modeling Password Guessability Using Neural Networks. In *Proceedings of the 25th USENIX Security Symposium.* Austin, TX, 175–191. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/melicher

[36] Robert Morris and Ken Thompson. 1979. Password Security: A Case History. *Commun. ACM* 22, 11 (November 1979), 594–597. https://spqr.eecs.umich.edu/courses/cs660sp11/papers/10.1.1.128.1635.pdf

[37] Robert W. Proctor, Mei-Ching Lien, Kim-Phuong L. Vu, E. Eugene Schultz, and Gavriel Salvendy. 2002. Improving Computer Security for Authentication of Users: Influence of Proactive Password Restrictions. *Behavior Res. Methods, Instruments, & Computers* 34, 2 (2002), 163–169.

[38] Neil J. Rubenking. 2017. The Best Password Managers of 2017. *PC Magazine* (2017).

[39] Stuart Schechter, Cormac Herley, and Michael Mitzenmacher. 2010. Popularity is Everything: A New Approach to Protecting Passwords from Statistical-Guessing Attacks. In *Proc. HotSec'10.*

[40] Claude E. Shannon. 1951. Prediction and Entropy of Printed English. *Bell Systems Technical Journal* 30, 1 (1951), 50–64.

[41] 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 '12)*. ACM, New York, NY, USA, Article 7, 20 pages. DOI: http://dx.doi.org/10.1145/2335356.2335366

[42] Richard Shay, Saranga Komanduri, Patrick Gage Kelley, Pedro G. Leon, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2010. Encountering Stronger Password Requirements: User Attitudes and Behaviors. In *Proc. SOUPS'10.*

[43] Elizabeth Stobert and Robert Biddle. 2014. The Password Life Cycle: User Behaviour in Managing Passwords. In *Symposium on Usable Privacy and Security (SOUPS)*. Menlo Park, CA.

[44] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. 2017. Design and Evaluation of a Data-Driven Password Meter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3775–3786. DOI: http://dx.doi.org/10.1145/3025453.3026050

[45] 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 Conference on Security Symposium (Security'12)*. USENIX Association, Berkeley, CA, USA, 5–5. http://dl.acm.org/citation.cfm?id=2362793.2362798

[46] 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 *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Ottawa, Canada, 123–140. https://www.usenix.org/system/files/conference/soups2015/soups15-paper-ur.pdf

[47] Blase Ur, Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Saranga Komanduri, Darya Kurilova, Michelle L. Mazurek, William Melicher, and Richard Shay. 2015. Measuring Real-World Accuracies and Biases in Modeling Password Guessability. In *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, Washington, D.C., 463–481. https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/ur

[48] US-CERT. 2016. Choosing and Protecting Passwords. (2016). https://www.us-cert.gov/ncas/tips/ST04-002

[49] VirusTotal. 2017. VirusTotal API. (2017). https://developers.virustotal.com/v2.0/reference

[50] Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. 2016. Understanding Password Choices: How Frequently Entered Passwords are Re-used Across Websites. In *Symposium on Usable Privacy and Security (SOUPS)*. 175–188. https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-wash.pdf

[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 '10)*. ACM, New York, NY, USA, 162–175. DOI: http://dx.doi.org/10.1145/1866307.1866327

[52] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. 2004. Password Memorability and Security: Empirical Results. *IEEE Security and Privacy* 2 (September 2004), 25–31. Issue 5. DOI: http://dx.doi.org/10.1109/MSP.2004.81

[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 '10)*. ACM, New York, NY, USA, 176–186. DOI: http://dx.doi.org/10.1145/1866307.1866328

[54] Moshe Zviran and William J. Haga. 1993. A Comparison of Password Techniques for Multilevel Authentication Mechanisms. *Comput. J.* 36, 3 (1993), 227–237.