The Rewards and Costs of Stronger Passwords in a University: Linking Password Lifetime to Strength

Ingolf Becker, Simon Parkin, and M. Angela Sasse, ***University College London***

https://www.usenix.org/conference/usenixsecurity18/presentation/becker

This paper is included in the Proceedings of the 27th USENIX Security Symposium.

August 15–17, 2018 • Baltimore, MD, USA

ISBN 978-1-939133-04-5

Open access to the Proceedings of the 27th USENIX Security Symposium is sponsored by USENIX.

The Rewards and Costs of Stronger Passwords in a University: Linking Password Lifetime to Strength

Ingolf Becker, Simon Parkin & M. Angela Sasse
University College London
{i.becker,s.parkin,a.sasse}@ucl.ac.uk

Abstract

We present an opportunistic study of the impact of a new password policy in a university with 100,000 staff and students. The goal of the IT staff who conceived the policy was to encourage stronger passwords by varying password lifetime according to password strength. Strength was measured through Shannon entropy (acknowledged to be a poor measure of password strength by the academic community, but still widely used in practice). When users change their password, a password meter informs them of the lifetime of their new password, which may vary from 100 days (50 bits of entropy) to 350 days (120 bits of entropy).

We analysed data of nearly 200,000 password changes and 115,000 resets of passwords that were forgotten/expired over a period of 14 months. The new policy took over 100 days to gain traction, but after that, average entropy rose steadily. After another 12 months, the average password lifetime increased from 146 days (63 bits) to 170 days (70 bits).

We also found that passwords with more than 300 days of lifetime are 4 times as likely to be reset as passwords of 100 days of lifetime. Users who reset their password more than once per year (27% of users) choose passwords with over 10 days fewer lifetime, and while they also respond to the policy, maintain this deficit.

We conclude that linking password lifetime to strength at the point of password creation is a viable strategy for encouraging users to choose stronger passwords (at least when measured by Shannon entropy).

1 Introduction

The expiration of passwords for machine accounts has had a long history. Tracing back to 1979, expiration was a tool to stop users sharing accounts on the first university computers [33]. This was not a need borne of security – it was a management mandate to allow for proper accounting of computation time. However the notion has been appropriated to serve security, spread by various international government guidelines that have since prescribed the expiration of passwords [9, 12]. Various justifications for password expiration have been found: the longer a password is 'alive', the higher the chance of compromise and the need to reset passwords (due to sustained attacks or inevitable leakage), or, that expiration limits the portability of a compromised password, as old passwords may be replicated on other services for convenience [8, 15, 28].

These myths have been thoroughly debunked. The security benefits of password expiration are marginal at best [16, 49]. Users regularly choose new passwords that are very similar to a previous password (through for instance incremental changes to a number in a sequence of passwords) [48, 49]. Further, passwords of sufficient strength can be combined with background protections to be strong enough in most scenarios: a password which can resist 10^6 guesses is all but uncrackable in an online attack scenario [25], if combined with sensible throttling [25, 45]. To defend against the offline attacks a password is required to withstand 10^{14} guesses.

This body of research has now informed practical advice, and a change of guidelines. Both the National Institute of Standards and Technology (NIST, US, [26]) and the National Cyber Security Centre (NCSC, UK, [34]) now prescribe that passwords should not expire unless there is evidence of compromise.

A holistic view of password policy management is required in practice. For example, a user's choice to re-use passwords across separate accounts is rational when there are simply too many passwords to remember [29]. Users may apply strategies to group accounts by perceived importance and assign a password to each group [24].

Against this background of prior knowledge on password expiration, we were invited to study the new password policy implemented at our home institution. The choice of password strength estimation and parameters

were not made by the authors. The new policy allows users to select any password of character length 8 or more with an estimated information entropy (Shannon entropy, a poor measure of cracking resistance, but still widely deployed) of at least 50 bits (see Section 3.3 for the policy specifics). The new system retains the expectation that users will harden their accounts with strong passwords, but in a twist provides a reward of longer password lifetime for selecting stronger passwords. A password with an estimated entropy of 50 bits has a lifetime of 100 days, and every additional bit of entropy increases the lifetime by approximately 3 days, up to 350 days for 120 bits of entropy.

We then use the term *password strength* here as the number of days a password lives for before being expired, as this is a measure of account strength that is visible to both the users and managers of the system.

The research questions examined in this paper are:

RQ1 What effect does the password policy of variable expiration have on a user's choice of password?

RQ2 Are there identifiable groups of users with analytically different responses to the new password rules and introduction of the new policy?

RQ3 What can be discerned about the impact of a policy intervention at a large institution from system logs?

We believe that this research constitutes the largest analysis of password data from a single institution with over 100,000 enrolled users in the system, who change their passwords nearly 200,000 times and reset (forgotten or expired) their passwords 115,000 times over a period of 14 months. Our approach is novel as we analyse routine change and intentional reset events together, to understand individual users' journeys through adoption and continued use of the new system. This approach leverages the working relationship with the system managers, who allowed continuing access to the anonymised log data and kept us informed on events outside of the system which could impact use and hence the logs themselves (such as university-wide events).

We begin the remainder of the paper with a review of the related literature in Section 2. After an introduction to our methodology in Section 3 we describe and compare the general statistics of our dataset to prior studies on large password analysis (Section 4). This is followed by an analysis of the password change data in particular, answering our research questions in Section 4.4. We draw on 93 interviews with staff and students for anecdotal user feedback in Section 4.7. We then discuss the impact of the results in Section 5 and close with conclusions and recommendations in Section 6.

2 Related Literature

The related literature is divided into the following sections: we start with a discussion of password strength estimation, then focus on the user's role in password management and password studies.

2.1 Password strength estimation

Traditionally, password strength has been measured as the entropy of a password through a calculation involving a password's length and the different number of character classes it uses [30] (Shannon entropy, which is also the estimation technique our institution uses, albeit with a few modifications as described in Section 3.3). These estimates are however not representative of the cracking effort, as passwords are not actually chosen randomly [13]. This has led to the creation of strength meters inspired by password-cracking, which estimate the number of attempts required for a password to be guessed. The current state of the art is zxcvbn [46], which algorithmically accurately estimates the strength of weak ($< 10^4$ guesses) passwords with only 234kB of data. For stronger passwords the strength estimation error of zxcvbn increases, but it is still a better estimator of cracking resistance than information entropy. To accurately estimate the strength of stronger passwords, significantly more storage and processing power is required, however this is infeasible for real-time feedback [43].

2.2 The role of users in password security

A primary question that is easily ignored when conducting password research is the attacker's modus operandi, and consequent interactions with the state of security defenses. The main attack vectors of interest are online and offline attack. An online attacker performs attacks over a wire, while the offline attacker has access to the physical system. While an online attack can be rate limited, blacklisted, and actively monitored [4, 37, 44], none of these defenses are possible against an offline attack. This implies that the defensive requirements on the password are very different [22, 23]. For passwords to be resistant to offline attacks they realistically need to be able to withstand 10^{14} guesses. In the context of an organisation it is not sufficient for the mean password strength to achieve this level: an attacker is often satisfied when compromising any one account with access to an asset of value, hence every password needs to withstand such an attack, which is infeasible [27]. When the entire system is under attack, the defense should be centered on the system too, rather than offloading it to all the users, for example through Ersatzpasswords [3].

As researchers have identified the need to raise the minimum strength of passwords, a large number of studies have focused on helping and educating the user in choosing stronger passwords. Users have been subjected to immediate feedback and suggestions before submitting their password choices [38, 39] with varying degrees of success. Research has attempted to improve users' ability to remember passwords, for example by allowing much longer composite passwords [40], memory aids [47], or training [14]. Perhaps unsurprisingly, positive attitudes towards security correlate with stronger passwords [17]. Such interventions are often measured over a relatively short timeframe; a wide-reaching intervention such as a password system overhaul may require time. We then leverage the opportunity to measure behaviour through password change events over time (where this would be impacted by users' capacity to remember passwords and use longer passwords in practice).

2.3 Studying passwords in the wild

A considerable amount of password research has been conducted in a lab setting. This allows for great internal validity through the ability to control the environment and measure specific properties of users choices and behaviours around passwords. However, Fahl et al. found that only about half of passwords gathered in a lab study are comparable to users' real-world passwords [20]. This problem is not specific to password studies, a large number of lab-based studies in security suffer from a lack of ecological validity. However, studying security perceptions in the real-world comes with its own issues [31]. Fortunately there are a number of password studies that are conducted in live environments.

The first scientific dissemination of password data was conducted on leaked password datasets [19, 45]. More recently Bonneau pushed the scientific principles of conducting password research by legitimately and rigorously analysing passwords of 70 million Yahoo! users [7]. The flurry of data breaches at large online services have fuelled research by providing extremely large datasets. Yet in all of these cases the user is often a customer of the organisation, with two consequences: service password policies tend to bow to the need for accessibility, as services that make access difficult don't have as many customers [21]. Users may not assign much value to these accounts, unless their personal data/money is stored there.

Apart from our research, the only other comparable study of password behaviour in a work environment with high value passwords to study is by Mazurek et al. [32]. Here the entire plaintext password database of over 25,000 accounts was available to the researchers (although considerable security precautions were taken to limit access to the plaintext passwords). The authors discover significant correlations between a number of demographic and behavioural factors and password strength, and we will be comparing our demographic findings to this research primarily.

Related to passwords, Parkin et al. studied a static password expiration policy of 100 days in a university, contrasting the analysis of helpdesk-related system events over a period of 30 months to findings from a small set of 20 interviews with system users [35]. Users appreciated the need for security and strong passwords, but their attempts to create strong passwords were frustrated by usability issues not directly apparent from system events (such as an inability to know in advance what the system would accept as a valid password).

Zhang et al. studied 31,075 passwords belonging to 7,936 university accounts in order to analyse the dependency between consecutive passwords [49]. We contrast their main results to our data in Section 4.

2.4 Password policy

A comprehensive overview of the last 30 years of password policy research is given by Zhang-Kennedy et al. [50]. Ever since "Users are not the enemy" there has been a sustained effort to design security policies for the user, taking into account their strengths and limitations. Strength aspects such as length and composition, as well as management aspects such as change-it-often, do-not-reuse, do-not-write-down and do-not-share-with-anyone have been either entirely revised or are at least strongly challenged [11, 12, 25, 26, 34].

User capability, user inclusion in their own and others' security, and a holistic approach to defensive security then together serve as indicators for identifying a *sustainable*, *workable*, and ultimately *secure* password system. With this in mind, we design the analysis of the password dataset in a way that considers the rewards (and costs) for (i) the user, and (ii) the organisation.

3 Methodology

Here we describe the methodology for analysing the logs of the password change system at UCL. We were not involved in the design of the policy or the choice of password strength estimator. We were approached by the IT services department who were eager to collaborate on exploring the scientific value of their policy design and its impact on the system's users. This lead to a productive working relationship for this project, which helped us to reason about the results and discuss possible causes for data patterns outside of the password system itself. This is especially important given the complexities not only of the data and the systems to which the data applies, but

also the institution, being that is has tens of thousands of account holders with varying levels and modes of interaction with the system.

The main contribution of this work is a scientific analysis of the effect of the policy. The analysis is informed by consideration of the cost of the policy to users.

3.1 The interface

The password change/reset interface is web-based. The new password has to be typed twice. Below the second password entry box are a password strength meter and a text field that displays the new password's lifetime in days. Both meter and days of password lifetime update on any change to the first new password form field. For passwords of < 50 bits of entropy the strength meter states *Too weak* and the password cannot be submitted. Passwords of lifetime 100 to 163 days are stated to be of *Medium* strength (yellow strength bar). Between 164 and 223 days a password is considered to be *Strong* (green bar), and beyond that the password is classed as *Very strong* (dark green bar).

3.2 The dataset

We received access to the password change and reset logs, which consisted of timestamps, anonymised user IDs, action performed (i.e., change/reset/etc), the integer password lifetime of the new password (100–350), as well as some coarse demographics information for the 100,000 users. We received IRB approval for our approach to log analysis, alongside in-person interviews with a subset of system users (see Section 4.7) (UCL Ethics ID 5336/007). Regarding the dataset, we had no individually-identifying information (an arrangement made with the system owners at point of data access), as well as only a single number for the user's password strength (i.e., not the password itself or any element of it). The password log data was stored on encrypted drives, and regular extensions to the dataset over time were transferred and stored securely.

The policy came into effect in October '16 and users began using the new system from that date when next requiring to change or reset their password. As the previous policy's expiration was set to 150 days, all active passwords will have been transferred to the new policy by April '17 (so that in effect it was a soft transition). Although we continue to have access to new data, we are confident that 14 months of complete data is sufficient, for the following reasons:

- The dataset includes at least one academic year's worth of data and regular events in an academic year, such as school closures and holidays;
- All currently active passwords were set on the new system;
- There are approximately six months of system events for the annual intake of new students (academic year starts in September to October, as seen for instance in Figure 2), who were never exposed to the previous policy.

3.3 Calculation of entropy

The minimum password requirements involve a complex combination of a number of fixed rules. Passwords are initially checked against static requirements. Passwords are required to: include at least one character from three of four possible character types (lowercase character, uppercase character, number, and symbol); be between 8 and 30 characters long, and; not contain the user's username or parts of their real name. The entropy of a password is then calculated by estimating the information entropy of the password by multiplying the size of the character class of each of the characters [2]. A number of factors decrease the entropy: repeated characters; lexicographically subsequent characters as well as the presence of a substring of the password in a dictionary of size $306,000$. Common character substitutions are also checked against the dictionary.

3.4 Uses of a password

Studying adoption and use of the system over time is important, where understanding new authentication systems in terms of how easy they are to learn is critical [8]. The password studied should be the only password staff and students require to access necessary services for work or study respectively. UCL uses one password for all of their services. This includes access to timetabling, e-learning resources, university e-mail, logging on to physical desktop machines, and WiFi. The frequency of use of this password is expected to vary naturally for different user types, who use different services, and access them from different machines (the most simple differentiation being a device they manage themselves or a fixed-place common-access machine). While users may resort to password managers to store their password for use in browsers, students (Undergraduate, Postgraduate and Medical) accessing the machines in university computer rooms will still have to type the password. Similarly, administrative staff work on a university computer and therefore have to regularly type the password to log in to and unlock their machines. Research staff and students however may have the flexibility to type their password very infrequently, especially if (a) they are using devices which they themselves manage and which no other user would have access to, and (b) they can complete their

work or study activities with minimal or ad-hoc access to services managed through the single-sign-on system. Ad-hoc access may be governed by the nature of the work done by distinct specialised groups, hence we are also interested in adoption and use differentiated by faculty/department.Users may then balance the convenience of accessing a system with the security of the mechanism that facilitates access to that system [6].

3.5 Perceived value of a password

Individuals in organisations will strive to protect their account if they perceive and understand a need to keep their organisation secure [1]. The UK's Universities and Colleges Information Systems Association (UCISA) distinguishes between the information security roles and competencies for distinct groups in universities [42]. Assuming that system users are aware of responsibilities like those described in the guide, they may have distinct attitudes towards the security of their accounts, and the associated passwords. Researchers may for instance have access to sensitive data, whereas administrators and teaching staff alike may manage staff and student records. Students may have access to their own information, but also the university's IT infrastructure; postgraduate students might have access to research data.

By considering factors which may influence the perceived value of a user's password, the scope of RQ2 is refined. Given both the frequency of use and the perceived value of accounts, we expect students to have weaker passwords than other groups, and researchers to have stronger passwords. We also expect administrative staff to value their account security while balancing any increases to password strength (delaying password change) with lower time cost per system authentication event. Regular enactment of security tasks over a working day may push users in an organisation to find ways to reduce the burden of security that relates to their primary productive work [6]. We test these hypotheses in Section 4.4.

3.6 User interviews

In addition to the password log analysis, 93 users of university systems were interviewed between February and March '17 (53 students and 40 staff). Users who had changed their password in the prior 2-3 months, or who had just received a reminder to change their password, were invited for interview. This framing allowed for the possibility that participants would not know that there was a new password policy.

The study was advertised via staff and student newsletters, and flyers positioned around the main university campus. Interviews were approximately 30 minutes in duration, and included discussion of: services accessed through university login; perceptions of passwords and security in relation to university-related tasks, and; participants views of the university's password system. A computer displaying the interface of the new system supported the interview (as described in Section 3.1). Participants were provided with a £15 voucher for completing the interview.

The average participant age of staff members and students were 34.6 and 22.8 respectively. Student participants had been at the organisation on average for approximately two years (including many who had joined the university just before the new system was deployed); staff participants had used the university systems for on average of approximately five years. Participants represented a range of schools and divisions (including administrative functions).

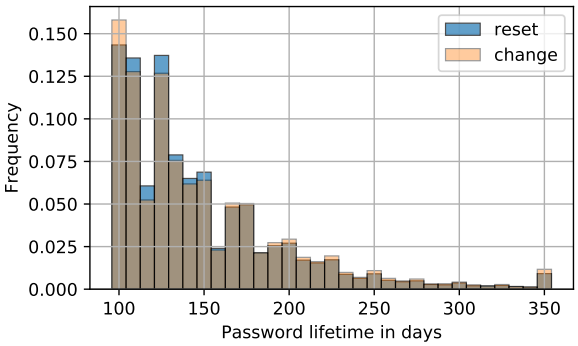

Figure 1: Normalised frequency of password lifetime. The mean frequency is 147.74 and 146.60 days for changes/resets respectively.

4 Results

In this section we describe the properties of user passwords found in the data, as well as characterise the adoption and usage behaviour for the new system across the user population and specific groups. We put our results in the context of existing research and highlight the impact of the policy on user behaviour.

Figure 1 describes the distribution of strength of all passwords observed in the university. The two distributions of password resets (when a password has been forgotten or it has expired) and changes (when the user still knows the previous password) are virtually identical. The histogram is strongly skewed to the left and decays rapidly, apart from approximately 1% of passwords that achieve the maximum strength of 350 days.

It is interesting to compare this distribution to the password strength distribution of Mazurek et al.'s study per-

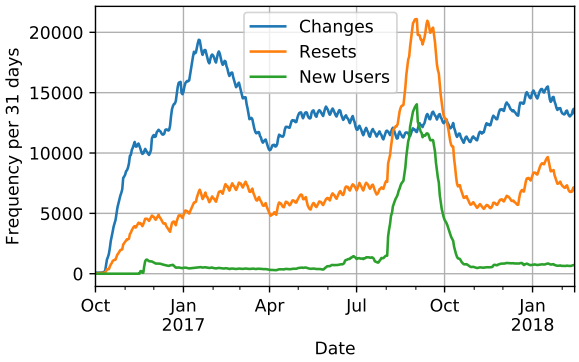

Figure 2: 31-day moving average of the number of password changes and resets, as well as the number of new users joining the university and using the system for the first time. The legend is in order of final values.

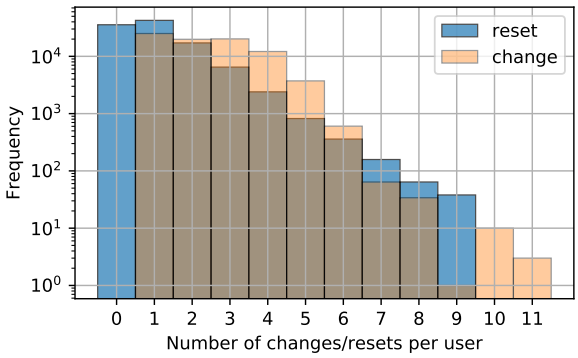

Figure 3: Distribution of the number of changes and resets the users in the dataset have made. Mean frequency is 2.41 and 1.08 for changes and resets respectively. 66% of users have reset their password at least once.

formed at Carnegie Mellon University (CMU) [32, Figure 7, page 11]. Their measured password strengths approximate a uniform distribution between 10^9 (100 days) and 10^{14} (225 days) guesses, and only 42% of passwords are guessed in 10^{14} guesses. Their estimated mean password entropy is 36.8 bits, compared to 69.64 bits here.

There are two systematic explanations for these stark differences. First, the mean password entropy reported by Mazurek et al. is calculated by state-of-the-art brute-forcing, compared to an information theoretic approach chosen by our IT department that only weakly correlates to actual password strength. Thus, our entropy estimates are likely large over-estimations [46, Fig. 8]. Secondly, the entropy estimate in our analysis is the same estimate used for providing feedback to the user in the form of the password meter (principally the fullness of the bar), and the weakest allowed password has an entropy of 50 bits. This explains the high concentration of passwords with 100 days lifetime, compared to the study performed at CMU; where policy and strength meter are not linked to the measured guessing strength.

The same explanations also apply to the differences between our analysis and Bonneau's analysis of cracking attempts of the Yahoo! password dataset [7, Figure 6 in particular]. Their identified cumulative distribution is aligned with our data, although Bonneau achieves a 50% success rate with 10^6 guesses.

4.1 Noteworthy events during the study

As with any study of an active real-world system, there are external events that have an effect on the system being studied. As we cannot control for these events, they should be acknowledged in the analysis. Further, external events can be leveraged to understand if there are particular kinds of events which can influence the adoption and use of an authentication system at a large organisation. Figure 2 highlights three families of events.

From the deployment of the new system in October '16 the userbase of the new system slowly grows as users change or reset their passwords (where this forces them to use the new system and hence appear in the dataset). Secondly, there is a peak of password resets in Jan-Feb 2017, which corresponds to the expiration of all passwords of users who joined the university in September '16 and had a fixed lifetime of 150 days. We expected that the rate of resets would decrease once users became familiar with the new system. This did not happen, indicating that familiarity with the system does not reduce the need to reset. The third event of note refers to the peak of new user being onboarded to the system in September '17 in time for the new academic year, where over 10,000 new students joined the university. This also causes the simultaneous peak in the number of changes, as setting an initial password is classified as a change.

4.2 Password change behaviour

The effect of the password policy on changes and resets is shown in Figures 3 and 4. In the full period studied, more users (66%) had to reset their password than not – on average, a user had to reset their password 1.08 times. Users may have to reset their passwords for two reasons: if they have forgotten their original password, or if their password has expired. The cost of a reset is significantly higher than a change, as it requires either physical presence at the institution's help desk or using a phone-based reset system. Over the period studied, the mean number of password changes and resets per user is 3.5. This is investigated further in Section 4.3.

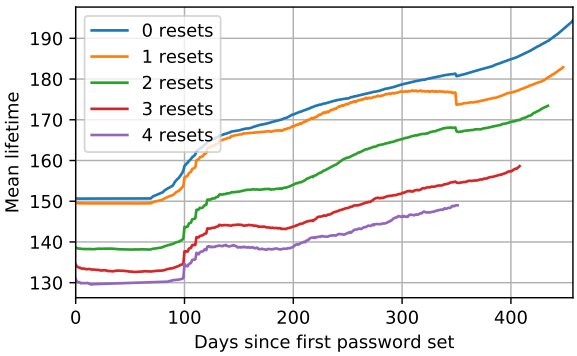

Figure 4: Average password lifetime of unexpired passwords by number of password resets. After 100 days the weakest passwords expire and users choose stronger passwords, which accounts for the steep rise. This pattern repeats after another 100 days. At 350 days users change their previously strongest passwords to one that is as strong or weaker password, causing a pronounced dip in the average password expiration.

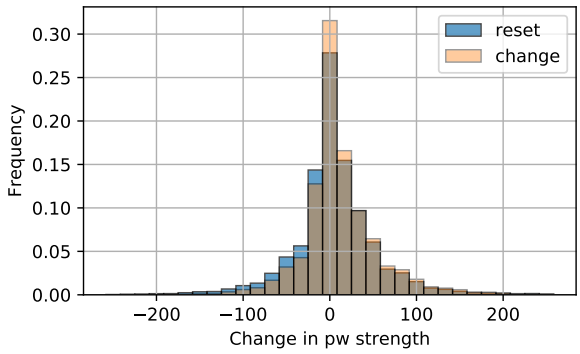

Figure 5: Distribution of the change in the password lifetime after the password change/reset. Mean change is 11.97 and 4.55 days for changes and resets respectively.

There is a strong positive correlation between each user's previous password strength and the likelihood of that same user resetting their password before expiration (i.e., forgetting the password, Spearman's $\rho = 0.95$, $p < 10^{-15}$). A user with a password lifetime of more than 300 days is four times as likely to forget their password than a user with a password with a 100 day lifetime. The minimum reset frequency per day of actual password lifetime is achieved with passwords which have a 100 day lifetime. Most resets however occur shortly after passwords have been set, and not after a user has been using a password for 100 days. Having a relatively strong password on the system then incurs the additional cost of potentially needing to reset that password. This may not only negate the advantages of having a strong password in the first place, but results like these can also inform predictive helpdesk/support provisioning [36], i.e., if users are encouraged to maintain stronger passwords, they may require more helpdesk support to reset passwords.

This is in contrast with Figure 4: The more password resets a user will have had, the weaker their password choice. While the average password lifetime of all groups is increasing as the users renew their password, the division between users with 0 or 1 reset and users with more resets remains pronounced, separated by at least 10 days of lifetime. This analysis suggests that one reset per year does not affect the system's performance, but two or more resets do (which applies to 27% of users). While system owners should obviously try to minimise the number of resets required, it appears one reset per year per user is an acceptable upper bound.

The answer to our first research question is alluded to in the mean password strength change of 12.73 days (as shown in Figure 5). This shows positive increases in password strength on consecutive password changes and resets on average. One common finding in password expiration research is that when forced to changed one's password, the new password will be similar to the old one. Figure 5 indicates that this effect may also be present in our dataset: 20% of changed passwords have identical expiration as their previous password, and 36% vary within 3 bits of entropy.

These figures vary slightly during the period of time analysed here, with a gradual increase to 28% in February (3 months after the change in policy) but returning to 20% in June and remaining constant from then on. Prior literature has examined this behaviour: Adams et al. found that 50% of their participants varied some element of their password when creating new passwords. Zhang et al. study behaviours at greater scale, by analysing 7,700 accounts and developing an efficient transformation algorithm to test for related passwords. The authors are then able to break 17% of their accounts within 5 guesses, and 41% within 3 sec of CPU time ($\approx 10^7$ guesses, our estimate) [49]. While we cannot determine the true dependence between current and prior passwords in our dataset, the strength proxy (through Figure 5) may suggest a similar proportion of related passwords.

4.3 Time dependence of subsequent changes/resets on prior lifetimes

Users are sent an email reminding them of their password's impending expiration 30, 20, 10, 4 and 1 day(s) in advance. The effect of the reminder is shown in Figure 6 with a bin size of 10 days. 10% of users act upon the

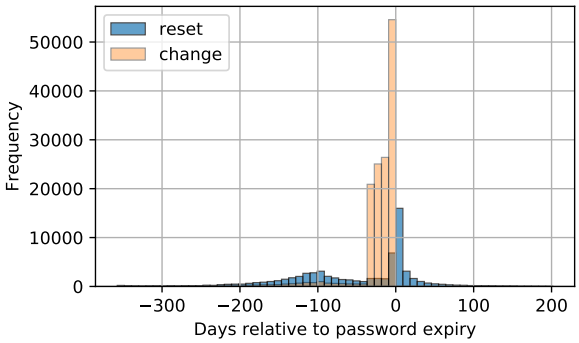

Figure 6: The frequency of password changes by the number of days relative to password expiration (day 0). The mean time for changes is −22.18 days and the mean time for resets is −52.09 days.

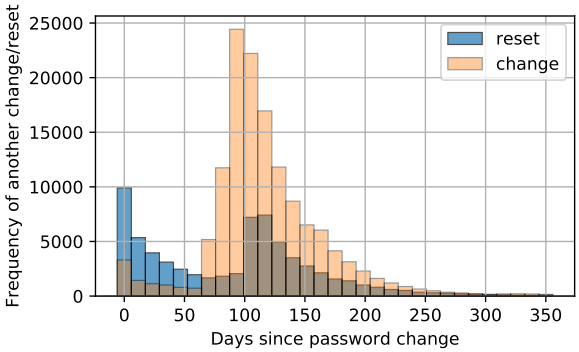

Figure 7: The distribution of the time between consecutive password changes. The mean time for changes is 117.16 days and the mean time for resets is 90.48 days.

reminder on average within 24 hours and subsequently change their password. Each following expiration warning causes an immediate increase in change rates, with the largest peak on the day of expiration, where another 13% of users change their password. This is followed by users resetting their passwords immediately after expiration, presumably after having been denied access to university resources. The general effect of these frequent reminders for the organisation is that the average user changes their password 22 days before expiration – essentially reducing the lifetime of their password voluntarily. This indicates that users in this institution change or reset passwords in response to reminders, and seldom voluntarily. This might be the case for users changing their password before even receiving the first 30-day advance warning of expiration, as can be seen in Figure 6.

Figure 7 is an analysis of the same time series as Figure 6, but anchored at the time of password creation rather than expiration. The main observation here is the strong concentration of password resets in the immediate proximity of password creation: users often forget their newly set password. The passwords created by reset within the first 48 hours after changing a password have a mean password strength of 6.9 days less than their previous password. This suggests that some users choose a weaker password due to forgetting the previous one (where in fact some users may be choosing weaker and weaker passwords in a cascade). The change rate initially decays before exhibiting the shape of a gamma distribution starting at 70 days – at the time of the first expiration warning email for passwords of 100-day strength, peaking at just before day 100, when a large number of user passwords expire.

These results imply that users reset their passwords primarily for two reasons: failure to recall the password, and the forced expiration of the password by the system. This is in line with personal password behaviours observed elsewhere [29]. These drivers are in contrast to instances where users would reset their password for primarily security reasons (such as believing that their password has been compromised).

4.4 Password change time series

In this section we study the password strength measure over time. The results answer two of our research questions: 'What effect does the password policy of variable expiration have on user's passwords – given the freedom, how will users choose?' (RQ1), and 'Are there contextual circumstances of groups of users which may influence their choice of password strength?' (RQ2). In Figures 2 and 8 to 10 we apply the same 31-day moving window to smooth out fluctuations due to weekly patterns (e.g., weekends, when most users are not actively using the system).

Figure 8 shows the evolution of the university's mean password strength over time. Initially we observe a small drop in strength between November '16 and February '17 (after the adoption of the policy), as users become accustomed to the new system. After this, the mean strength increases from 145.5 days to 170.1 days – an increase by 6.9 bits of entropy. This strongly suggests that users have adapted slowly to the new password policy, and eventually make use of their ability to increase password lifetime by strengthening their passwords.

The 'steady state solution' is an approximation of the attractor of the password change distribution. It is calculated by performing a linear regression on users' previous (x) and new password lifetimes (y). The solution of this linear regression for $y = x$ identifies the attractor. Users with previous passwords weaker than this attractor tend to reduce the lifetime of their new password, and vice versa.

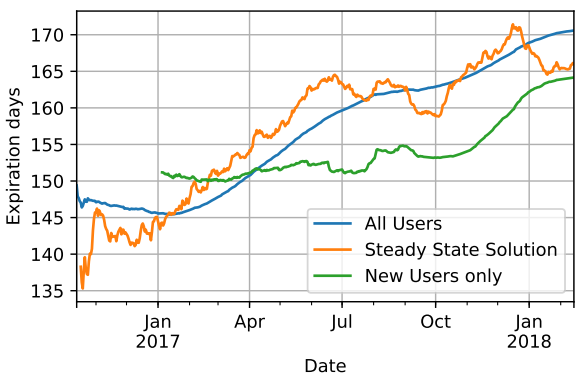

Figure 8: 31-day moving average of the mean password strength of all users and new users. The 'steady state solution' estimates the average strength of passwords in the system if users were to continue making their passwords stronger (or weaker) consistently with how they did so in the current measurement window. The legend is in order of final values.

The evolution of the mean password strength is underpinned by cyclical behaviours. A quarter of users have a password lifetime of less than 110 days (see Figure 1), and have to change their passwords on average every 80 days (see Figure 6), but every time they do, they increase their average password strength. This manifests twice in Figure 8: at the start of the deployment of the new system where there are no existing users (the increase in password strength is delayed until February '17); and again with the enrollment of over 10,000 new users who set their first password around September '17 (see Figure 2), in time for the start of the new academic year. As this large number of users have all set their initial passwords in a short time frame, their first regular password change occurs from November '17 onwards. Their change behaviour also causes the temporary plateau around September '17 and the subsequent increase of the mean password strength of all users, which is a statistically significant increase (paired t-test, $t(10892) = -47.19$, $p = 0$).

The 'steady state solution' gives us insights into the password changing trend over time: for example, if users had continued to choose new passwords in the same manner as they did in April '17, the mean password lifetime of the university would settle at 156 days. However, as the steady state solution continues to increase, it appears that the users are still responding to the policy. The artifacts of the cyclical changes are also evident in the trend.

The relatively small drop in the steady state solution after January '18 aligns with an increase in password resets at this time (see Figure 2). This could be due to users having forgotten their passwords after returning from the Christmas break. As new users have yet to catch up to the password strength of existing users, it is likely that the mean password strength in the university will increase further.

As we do not have data for the users' password strength before the adoption of the new password change system and policy, we are unable to do a rigorous before-after comparison of strength data that takes into account all factors that may have contributed to this change – for example the old system did not give any feedback on their password strength. This implies that interface design for the password creation/reset process may also have a part to play in users increasing their password strength (where a subset of users migrating between the old and new systems provided feedback in Section 4.7).

As the new users have not had experience of the previous system, and as there have been no other initiatives by the university to encourage stronger passwords, we consider the increase in users' average password expiration likely to be a consequence of the policy, answering RQ1. It appears to have taken around 150 days for the effect of the policy to start to achieve its aims.

4.5 Password change time series by school

We are fortunate to have some coarse demographic information for each user recorded in the data. Figure 9 compares the evolution of password strength for selected schools. The users of each school have together made at least 11,000 password changes; we calculated bootstrapped, bias-corrected and accelerated [18] confidence intervals for each of the schools. The 95% confidence intervals were within 1% of the mean for all schools in Figure 9 from January 2017 onwards. We have hence omitted the confidence intervals. For brevity, we omitted a number of smaller schools closely aligned with the university mean.

Throughout all schools there is a statistically significant positive increase in password strength (in-sample t-test, $p = 0$). The school of Education displays the lowest increase of 18 days, while Maths and Physics increased their password strength by 27 days. The differences between schools are also pronounced, with passwords in Engineering being 13.4 days (4 bits) stronger than in the school of Education. It is of note that the university's Education school has been part of the university for only a few years. A joint linear regression of the password strength changes of all faculties predicting the password strength was conducted. Each school contributed statistically significantly, explaining 82% of variance ($R^2 = 0.816$, $F(6,49201) = 36320$, $p < 10^{-10}$).

In previous research, only Mazurek et al. compare different university units for their respective password strength. Their password cracking algorithm managed

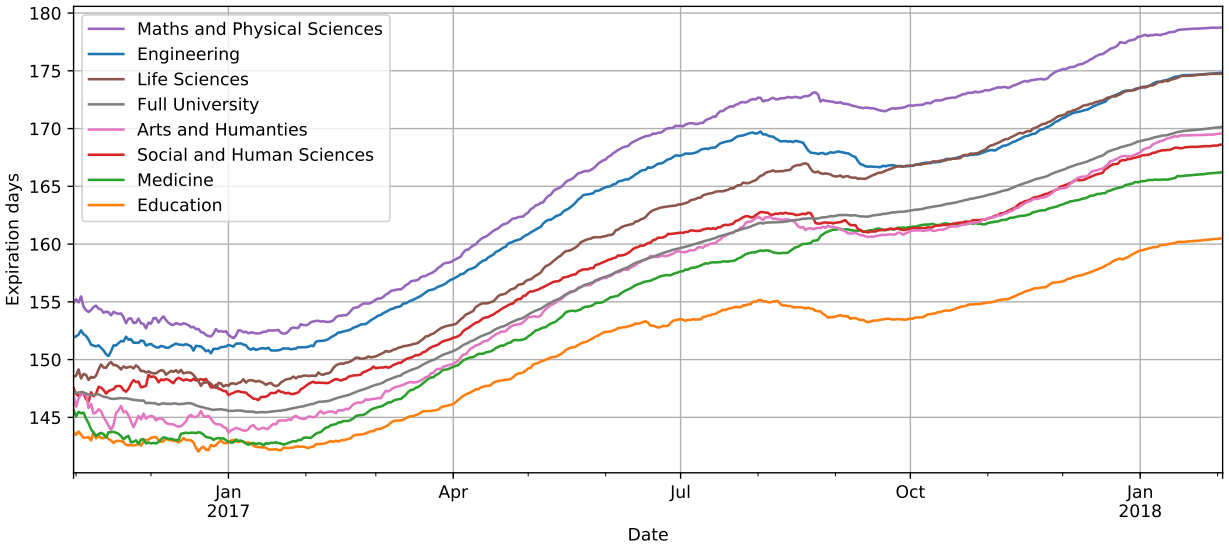

Figure 9: 31-day moving average password expiration for selected schools over time. The legend is in descending order of the final expiration values.

to predict in 3.8×10^{14} guesses the passwords of 38% of computer science accounts and 61% of business school accounts. They then performed a Cox regression on password survival times, reporting a 1.83 times chance of password compromise for business school passwords than for computer science.

In a naive model, 3.8×10^{14} guesses could be estimated as fully eliciting 48.43 bits. Given that the weakest allowed password in our university has an entropy of 50 bits, we expect 2.59% of Engineering accounts and 2.92% of School of Education accounts to be compromised after 3.8×10^{14} guesses. If we increase the attacker's brute force capacity to 60 bits (10^{18} guesses), the expected proportion of accounts which may be compromised increases to 36% and 44% respectively.In either case School of Education passwords are 1.13 and 1.22 times as likely as Engineering passwords to be guessed.

4.6 Password change time series by relationship

In addition to an analysis by school/faculty, we are also able to differentiate between the different roles of individuals within the university. The evolution of the respective user group's password strength can be found in Figure 10. Relationships with less than 5,000 / 2% of the total password changes/resets have been omitted. As for the previous graph, all user groups show an upward trend in their password strength over time. There are also significant variations between the groups, with Teaching/Research staff exhibiting password strengths 21 days stronger than Postgraduate students.A linear regression predicting the password strength depending on the relationship types was carried out. Each type of relationship contributed statistically significantly, explaining 89% of variance ($R^2 = 0.893$, $F(13, 12559) = 7957$, $p < 10^{-10}$).

The differences are in line with the hypotheses in Section 3: there appears to be both a positive correlation between password strength and likely value attached to the account (see Section 3.5), and a negative correlation between password strength and frequency of use. For example, Teaching/Research staff are likely to value their account security highly (using their accounts to access research and teaching data, which undergraduate students for instance would not.We observe that this group has the highest average password strength.

Administrative staff may value their account security highly too, but they also have a high frequency of use of the password, which may act to moderate their password strength. An interesting group to investigate in further research are the Alumni. These users are very different to the rest of the population: their account usage is low, so a long password expiration time will help minimise the frequency of password changes/resets; being potentially remote to the university, they may perceive the potential cost of a forgotten password as being much higher.

The results presented in this section answer our initial research questions: users have responded to the freedom of choosing their password lifetime slowly, but have in time increased their password lifetime considerably. The user population has needed time to adapt to the change in authentication protocols; 14 months after the interven-

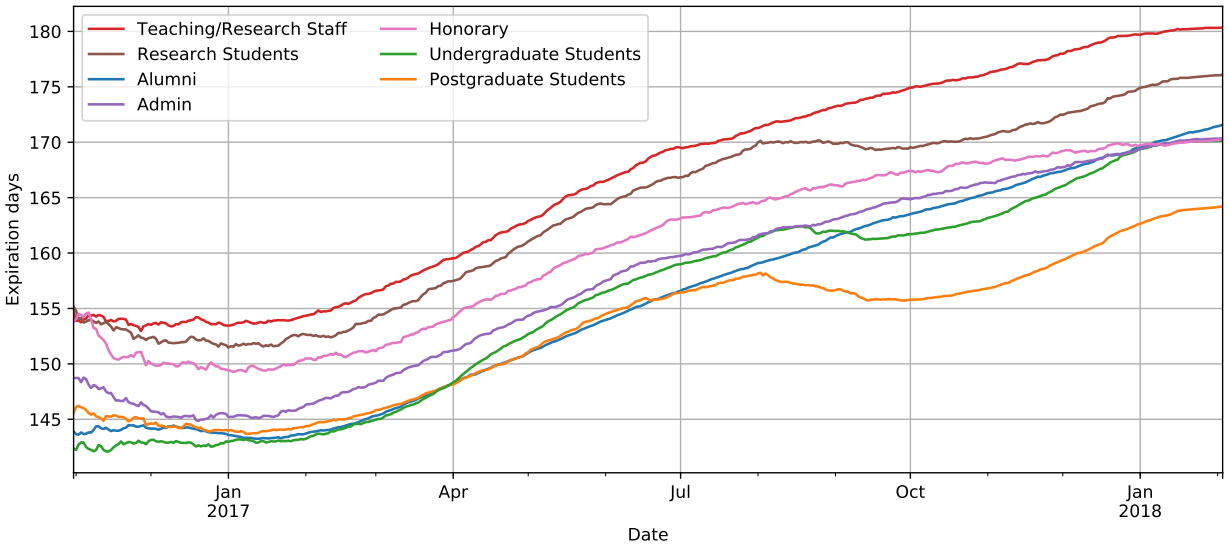

Figure 10: 31-day moving average password expiration for various relationships with the university over time. The legend is in descending order of the final expiration values.

tion, the password strength of all user groups has yet to plateau. We have identified differences in how users react to the policy change, by analysing the evolution of password strength between different subgroups (role and division). Other work has demonstrated that security preparedness and perceptions can differ between roles and divisions in a large organisation [5].

4.7 User feedback

Here we present a preliminary summary and discussion of field notes taken by interviewers (see Section 3.6). Feedback from the 93 interview participants informs the view of factors which may influence decisions around the construction and use of passwords on the studied system. We discuss general observations, with representative participant quotes. Participant identifiers signify E## (Employee/Staff) or S## (Student).

A few participants reported changing their password-related habits in response to the new system. This included beginning to store the password in a password manager, or as with E19, making a written note:

> *"Well, normally I just memorise it. This time around I did actually write it down when I changed it last week. Because it was so much longer than normal. Because previously they were eight characters. Now I think my password is like twelve characters. And it had to be that long to get the security up too. Because of now they rate it like low, medium, strong securities. So I had to keep adding characters to get it to say strong. So it is longer than I normally have."*

Many participants appreciated the flexibility of the new password policy. Some had however used the new system but not explored the differences between it and the old system; the differences between systems – and policies – were not immediately apparent to all participants. With the introduction of the new system, participants were split as to whether they believed passwords should be expired or remain valid indefinitely.

There was a general even split among interview participants as to whether they saw a link between password age and password strength. The data supports this, as a year after deployment users' average password strength has yet to settle (as notable in Figure 8). This could potentially be as much about discovering the features of the new system as it is about skillfully using it. For those who were aware of it, some did see it as an incentive to make a stronger password, such as E20:

> *"If they say if you make a stronger password you can keep it for longer, maybe it would help. [...] It wasn't clear that it was contingent on the strength of your password. I don't know if it is."*

Conversely, E25 found it difficult to create a valid password that was not labelled 'Weak':

> *"I probably tried about 6-7 passwords before I got to the one that it would accept ... It [the password meter] just kept not getting past the failed point ..."*

Others would consider password length alongside the need to type the password many times, and as a result would aim for a 'Medium'-strength password of around 8-12 characters. E30:

> *"Or trying to find a better password that would work. It does get harder because I had to change it so many times ... trying to think of a password. In a way it is not good that you are supposed to change a password. You run out of ideas of what to use. It's good that they are aware of your security but it does get a bit stressful."*

A number of participants commented that although they had created a longer password than before, they immediately reset their password as they found it too complicated to type, such as S17:

> *"even though I could remember it wasn't practically very helpful if you have to put in you know twenty characters. It's not great. So then I changed it to something that was shorter and last a little less time I just could remember that."*

This aligns with our findings in Figure 4 and Section 4.2, and also with Mazurek et al.'s engagement with system users [32]: those finding longer passwords unworkable will act to find a solution which is workable, abandoning the potential for longer lifetimes.

The summatory findings indicate that there may be a number of factors influencing password choice which are not represented in the dataset. The analysis in Section 4 was based on the available *data*, and the available *data fields*. Future collaboration will explore how the design of password system logs can be augmented to provide a more directly holistic view.

5 Discussion

There are hidden costs of the change in policy that should be considered. The intervention took time to gain traction, and it may have been that this time could have been shortened in some way. In some cases, users were voluntarily changing their passwords to a weaker combination of characters, taking time to learn how to *skillfully* choose stronger passwords (i.e., sustain stronger passwords over successive change events). The analysis informing Figure 4 uncovered that over 27% of users have had to reset their passwords more than once per year, and that these users have passwords with much shorter expiration. It could be that system usability hinders the adoption of the policy for a proportion of users.

As noted by Adams & Sasse, [1], most users in an organisation will want to behave securely, where insecure behaviour arises as they try to manage excessive demands in their workplace (where security would be just one of those demands). That the changes across different departments and user groups follow relatively similar patterns suggests that there was a collective change in password use, perhaps due to a collective culture towards security or influence from how peers are seen to behave.

From a security perspective, the implications of our results are clear. In the current format of the policy, the weakest possible password is strong enough to withstand an online attack (need to withstand 10^6 guesses); the increase in strength has not been pronounced enough to protect against offline attacks [23]. Rather than improving robustness to a wider range of attacks, the intervention has identified each user's individual threshold for trading off password complexity for password lifetime. It is a combination of the subjective cost optimisation of the individual's time (time spent both resetting and authenticating), acceptance of the perceived effort in managing a complex password, and their perceived value of their account. As different individuals interact differently with the university, this optimisation varies across user groups, as in Figures 9 and 10.

From a cost-benefit analysis, the policy has increased cost through increased individual effort cost and organisational support cost due to resets. The benefits for users rely on their perceptions: our user interviews found that the possibility of longer lifetimes was welcomed, and perceived this as an improvement considering their previous experiences of organisational password policies.

Here we have considered the different contexts in which users interact with the password policy. A further hidden cost arises from the interruption of the primary task from expiration of passwords, the reminder emails, and the planning of when to next change one's password (as one might be about to travel or go on leave, for instance). In studying the use of passwords and support of users in a large organisation, Brostoff [10] identified a range of 'costs' related to the expiry of passwords, such as designing new passwords, re-design of a candidate password if the system does not permit it, and amending any recall aids such as written notes. Brostoff's results also suggest that users may confuse prior and current passwords, where having had expired passwords then contributes to the daily cost of entering a current password correctly. The extra reward perceived for a stronger password must be greater than the cumulative additional time (i.e., perceived effort) required to correctly enter the password when it is needed. This is to say nothing of the frustration that may be caused in recalling and entering passwords, and the batching of tasks that may occur to reduce the regularity of password entry events [41]. A

similar approach to the work described in [41], of asking users to complete diaries – or otherwise report on their experience of using the system – may more clearly identify the workload caused by the authentication system.

5.1 Limitations

Our main limitation stems from studying passwords 'in the wild': our study did not have a control group. This means we are unable to observe if users would choose stronger passwords without the presence of the greater lifetime incentive. However, the existing literature [49, 50] suggests that users choose new passwords that are similar to previous ones, rather than continuously act themselves to improve the strength of their password.

We did not have log data for users prior to deployment of the new system. However, new users who were unaware of the old system behaved similarly to the existing population, suggesting that effects are due to the new policy rather than the change in systems.

6 Conclusion

Here we evaluated the impact of a new password policy upon 100,000 users at our university. In what is a novel policy designed by system managers, users were able to choose passwords with lifetime varying from 100 (50 bits of entropy) to 350 days (120 bits of entropy).

While the security community is moving away from prescribing password expiration, we have found that users 'play the game' and adapt their passwords in order to receive longer lifetimes. Results show that the intervention took over 100 days to gain traction, and that users took over 12 months to move from a lower-than-initial average 146-day (63 bits) to a higher 170-day (70 bits) password lifetime. The policy had both apparent and potential costs for individual users: 66% of users had to reset – as opposed to routinely change – their passwords, often multiple times. The average user had 3.5 passwords over the duration of the study. Users who are forced to reset their password more than once a year compensate by choosing significantly weaker passwords. Depending on the implementation of the reset procedure, both the actual and user-perceived cost may be high.

The analysis has revealed different levels of engagement with the policy. Had the system been monitored more directly for the impact upon users, the high reset rate and varied degrees of adoption amongst different user groups could have been seen as *early indicators* of the need for further support. It should also be noted that the policy intervention described in this paper gave users a choice in balancing delayed expiration and cost of managing a stronger password, rather than forcing the policy on them [29]. We continue to work with the system managers to analyse new log data, and to explore how user needs and challenges can be anticipated.

6.1 Policy interventions

One take-away here is that conclusions about the impact of an intervention should not be drawn based on immediate improvement or lack thereof. Other studies of the impact of behaviour change caused by security policies – in particular, lab studies – should measure interventions at meaningful intervals over a suitably long period of time, where arguably this would be a continuous activity.

When designing a new intervention, practitioners should consider how to measure the effectiveness of a change and the associated impact on users. After an intervention is deployed it may benefit from being monitored and *calibrated*, towards reducing problems and reward secure behaviour, where dynamic policy that reacts to users is far from being a common capability.

We have found that users will generally change their password in response to password expiry warnings and reminders; warning users too early effectively reduces the password lifetime. This potentially confuses the boundaries and meaning behind what password expiry is for, and what password expiry warnings are intended to achieve. Similarly, some of the cost of password resets can be avoided by allowing expired passwords to be changed, rather than going through a reset procedure.

Considering our findings regarding password resets and voluntary password changes, *a reward* of a longer password lifetime is not the same as an *optimal reward*; this opens up avenues of research to find optimally secure and workable defenses. In an ideal scenario we envision a deployment of a policy linking password expiration with password strength only if the weakest acceptable password is below the 10^6 guesses threshold identified by Florêncio et al. [23]. Passwords would then expire in line with the expected online guessing resistance of the password; if a password is stronger than the online guessing threshold it should not unconditionally expire.

Acknowledgements

We would like to thank our university, in particular Tom Crummey, Tim Purkiss and Noshir Homawala, for offering the opportunity to study this novel policy. Albesë Demjaha, Julienne Park, and Nissy Sombatruang contributed to the collection and initial analysis of user interviews. We would also like to thank Steven Murdoch and Sebastian Meiser for feedback on early versions of this paper. The authors wish to also extend thanks to the USENIX Security review committee and Lujo Bauer in particular.

References

[1] A. Adams and M. A. Sasse. Users are not the enemy. *Commun ACM*, 42(12):40–46, 1999. DOI: 10.1145/322796.322806.

[2] O. Alistratov. Data::Password::Entropy. Version 0.08. 2010.

[3] M. H. Almeshekah, C. N. Gutierrez, M. J. Atallah, and E. H. Spafford. ErsatzPasswords: ending password cracking and detecting password leakage. In *Proc. 31st annual computer security applications conference*. In ACSAC 2015. ACM, New York, NY, USA, 2015, pp. 311–320. ISBN: 978-1-4503-3682-6. DOI: 10.1145/2818000.2818015.

[4] M. Alsaleh, M. Mannan, and P. C. Van Oorschot. Revisiting defenses against large-scale online password guessing attacks. *IEEE transactions on dependable and secure computing*, 9(1):128–141, 2012. DOI: 10.1109/TDSC.2011.24.

[5] A. Beautement, I. Becker, S. Parkin, K. Krol, and M. A. Sasse. Productive Security: A Scalable Methodology for Analysing Employee Security Behaviours. In *Twelfth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Denver, CO, 2016.

[6] A. Beautement, M. A. Sasse, and M. Wonham. The compliance budget: managing security behaviour in organisations. In *Proc. workshop on New Security Paradigms (NSPW)*, 2008, pp. 47–58. DOI: 10.1145/1595676.1595684.

[7] J. Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *Proc. IEEE symposium on security and privacy (S&P)*. IEEE Computer Society, Washington, DC, USA, 2012, pp. 538–552. DOI: 10.1109/SP.2012.49.

[8] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano. The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. In *Proc. IEEE symposium on security and privacy (S&P)*. IEEE, 2012, pp. 553–567. DOI: 10.1109/SP.2012.44.

[9] S. Brand and J. Makey. Department of defense password management guideline. (CSC-STD-002-85). Department of Defense Computer Security Center, 1985.

[10] S. Brostoff. Improving password system effectiveness. Doctoral Thesis. University of London, 2005.

[11] S. Brostoff and M. A. Sasse. ''Ten strikes and you're out'': Increasing the number of login attempts can improve password usability. In *Proc. CHI Workshop on HCI and Security Systems*, 2003.

[12] W. E. Burr, D. F. Dodson, and W. T. Polk. Electronic authentication guideline. (NIST SP 800-63v1.0.1). DOI: 10.6028/NIST.SP.800-63v1.0.1. Gaithersburg, MD: National Institute of Standards and Technology, 2004.

[13] X. d. C. d. Carnavalet and M. Mannan. From very weak to very strong: analyzing password-strength meters. In *NDSS*. Vol. 14, 2014, pp. 23–26.

[14] D. Charoen, M. Raman, and L. Olfman. Improving end user behaviour in password utilization: an action research initiative. *Syst pract act res*, 21(1):55–72, 2008. ISSN: 1094-429X, 1573-9295. DOI: 10.1007/s11213-007-9082-4.

[15] W. Cheswick. Rethinking passwords. *Commun. ACM*, 56(2):40–44, 2013. ISSN: 0001-0782. DOI: 10.1145/2408776.2408790.

[16] S. Chiasson and P. C. v. Oorschot. Quantifying the security advantage of password expiration policies. *Des. codes cryptogr.*, 77(2):401–408, 2015. DOI: 10.1007/s10623-015-0071-9.

[17] Y.-Y. Choong and M. Theofanos. What 4,500+ people can tell you – employees' attitudes toward organizational password policy do matter. In *Human aspects of information security, privacy, and trust*. In LNCS. Springer, Cham, 2015, pp. 299–310. DOI: 10.1007/978-3-319-20376-8_27.

[18] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Vol. 1. Cambridge university press, 1997. ISBN: 978-0-511-80284-3.

[19] M. Dell'Amico, P. Michiardi, and Y. Roudier. Password strength: an empirical analysis. In *Proc. IEEE INFOCOM*, 2010. DOI: 10.1109/INFCOM.2010.5461951.

[20] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On the ecological validity of a password study. In *Proc. ninth symposium on usable privacy and security (SOUPS)*. ACM, New York, NY, USA, 2013. DOI: 10.1145/2501604.2501617.

[21] D. Florêncio and C. Herley. Where do security policies come from? In *Proc. sixth symposium on usable privacy and security (SOUPS)*. ACM, New York, NY, USA, 2010, 10:1–10:14. DOI: 10.1145/1837110.1837124.

[22] D. Florêncio, C. Herley, and B. Coskun. Do strong web passwords accomplish anything? In *Proc. 2nd USENIX workshop on hot topics in security*. In HOTSEC'07. USENIX Association, Berkeley, CA, USA, 2007, 10:1–10:6.

[23] D. Florêncio, C. Herley, and P. C. Van Oorschot. An administrator's guide to internet password research. In *Proc. USENIX LISA*, 2014.

[24] D. Florêncio, C. Herley, and P. C. Van Oorschot. Password portfolios and the finite-effort user: sustainably managing large numbers of accounts. In *Proc. USENIX security*. USENIX Association, San Diego, CA, 2014, pp. 575–590.

[25] D. Florêncio, C. Herley, and P. C. Van Oorschot. Pushing on string: the 'don't care' region of password strength. *Commun. ACM*, 59(11):66–74, 2016. DOI: 10.1145/2934663.

[26] P. A. Grassi, M. E. Garcia, and J. L. Fenton. Digital identity guidelines: revision 3. (NIST SP 800-63-3). DOI: 10.6028/NIST.SP.800-63-3. Gaithersburg, MD: National Institute of Standards and Technology, 2017.

[27] C. Herley. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proc. workshop on new security paradigms workshop (NSPW)*, 2009, pp. 133–144.

[28] C. Herley and P. V. Oorschot. A research agenda acknowledging the persistence of passwords. *IEEE security & privacy*, 10(1):28–36, 2012. DOI: 10.1109/MSP.2011.150.

[29] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proc. SIGCHI conference on human factors in computing systems (CHI)*. ACM, New York, NY, USA, 2010, pp. 383–392. DOI: 10.1145/1753326.1753384.

[30] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess again (and again and again): measuring password strength by simulating password-cracking algorithms. In *2012 IEEE symposium on security and privacy*, 2012, pp. 523–537. DOI: 10.1109/SP.2012.38.

[31] K. Krol, J. M. Spring, S. Parkin, and M. A. Sasse. Towards robust experimental design for user studies in security and privacy. In *Learning from authoritative security experiment results (LASER) workshop*, 2016.

[32] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, and B. Ur. Measuring password guessability for an entire university. In *Proc. CCS*. ACM, New York, NY, USA, 2013, pp. 173–186. DOI: 10.1145/2508859.2516726.

[33] R. Morris and K. Thompson. Password security: a case history. *Commun. ACM*, 22(11):594–597, 1979. ISSN: 0001-0782. DOI: 10.1145/359168.359172.

[34] NCSC. Password guidance: simplifying your approach. Guidance. UK National Cyber Security Centre, 2016.

[35] S. Parkin, S. Driss, K. Krol, and M. A. Sasse. Assessing the user experience of password reset policies in a university. In *Technology and practice of passwords*. In LNCS. Springer, Cham, 2015, pp. 21–38. DOI: 10.1007/978-3-319-29938-9_2.

[36] S. Parkin, A. v. Moorsel, P. Inglesant, and M. A. Sasse. A Stealth Approach to Usable Security: Helping IT Security Managers to Identify Workable Security Solutions. In *Proc. 2010 Workshop on New Security Paradigms (NSPW)*. ACM, New York, NY, USA, 2010, pp. 33–50. DOI: 10.1145/1900546.1900553.

[37] B. Pinkas and T. Sander. Securing passwords against dictionary attacks. In *Proc. 9th ACM conference on computer and communications security (CCS)*. ACM, New York, NY, USA, 2002, pp. 161–170. DOI: 10.1145/586110.586133.

[38] S. M. Segreti, W. Melicher, S. Komanduri, D. Melicher, R. Shay, B. Ur, L. Bauer, N. Christin, L. F. Cranor, and M. L. Mazurek. Diversify to survive: making passwords stronger with adaptive policies. In *Thirteenth symposium on usable privacy and security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 2017. ISBN: 978-1-931971-39-3.

[39] R. Shay, L. Bauer, N. Christin, L. F. Cranor, A. Forget, S. Komanduri, M. L. Mazurek, W. Melicher, S. M. Segreti, and B. Ur. A spoonful of sugar?: the impact of guidance and feedback on password-creation behavior. In *Proc. 33rd annual ACM conference on human factors in computing systems (CHI)*. ACM, New York, NY, USA, 2015, pp. 2903–2912. DOI: 10.1145/2702123.2702586.

[40] R. Shay, S. Komanduri, A. L. Durity, P. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Can long passwords be secure and usable? In *Proc. SIGCHI conference on human factors in computing systems (CHI)*. ACM, New York, NY, USA, 2014, pp. 2927–2936.

[41] M. Steves, D. Chisnell, M. A. Sasse, K. Krol, M. Theofanos, and H. Wald. Report: authentication diary study. (NIST IR 7983). National Institute of Standards and Technology, 2014.

[42] Universities and Colleges Information Systems Association (UCISA). *Chapter 8: roles and competencies*. Of *Information Security Management Toolkit*, 2015.

[43] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, D. Kurilova, M. L. Mazurek, W. Melicher, and R. Shay. Measuring real-world accuracies and biases in modeling password guessability. In *USENIX security*. USENIX Association, Washington, D.C., 2015, pp. 463–481.

[44] P. C. Van Oorschot and S. Stubblebine. On countering online dictionary attacks with login histories and humans-in-the-loop. *ACM trans. inf. syst. secur.*, 9(3):235–258, 2006. DOI: 10.1145/1178618.1178619.

[45] M. Weir, S. Aggarwal, M. Collins, and H. Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proc. 17th ACM conference on computer and communications security (CCS)*. ACM, New York, NY, USA, 2010, pp. 162–175. DOI: 10.1145/1866307.1866327.

[46] D. L. Wheeler. Zxcvbn: low-budget password strength estimation. In *25th USENIX security symposium (USENIX security 16)*. USENIX Association, Austin, TX, 2016, pp. 157–173.

[47] J. Yan, A. Blackwell, R. J. Anderson, and A. Grant. Password memorability and security: empirical results. *IEEE security & privacy*, 2(5):25–31, 2004. DOI: 10.1109/MSP.2004.81.

[48] E. v. Zezschwitz, A. D. Luca, and H. Hussmann. Survival of the shortest: a retrospective analysis of influencing factors on password composition. In *IFIP conference on human-computer interaction (INTERACT)*. In LNCS. Springer, Berlin, Heidelberg, 2013, pp. 460–467. DOI: 10.1007/978-3-642-40477-1_28.

[49] Y. Zhang, F. Monrose, and M. K. Reiter. The security of modern password expiration: an algorithmic framework and empirical analysis. In *Proc. 17th ACM conference on computer and communications security (CCS)*. ACM, New York, NY, USA, 2010, pp. 176–186. DOI: 10.1145/1866307.1866328.

[50] L. Zhang-Kennedy, S. Chiasson, and P. v. Oorschot. Revisiting password rules: facilitating human management of passwords. In *2016 APWG symposium on electronic crime research (eCrime)*, 2016. DOI: 10.1109/ECRIME.2016.7487945.