Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition

Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann

University of Munich (LMU), Media Informatics Group, Munich, Germany
{emanuel.von.zezschwitz,alexander.de.luca,
heinrich.hussmann}@ifi.lmu.de

Abstract. In this paper, we investigate the evolutionary change of user-selected passwords. We conducted one-on-one interviews and analyzed the complexity and the diversity of users' passwords using different analysis tools. By comparing their first-ever created passwords to several of their currently used passwords (e.g. most secure, policy-based), we were able to trace password reuse, password changes and influencing factors on the evolutionary process. Our approach allowed for analyzing security aspects without actually knowing the clear-text passwords. The results reveal that currently used passwords are significantly longer than the participants' first passwords and that most participants are aware of how to compose strong passwords. However, most users are still using significantly weaker passwords for most services. These weak passwords, often with roots in the very first passwords the users have chosen, apparently survive very well, despite password policies and password meters.

Keywords: password, evolution, security, policy, survey, human factor.

1 Introduction

Secret alphanumeric strings, called passwords, have been used to restrict access to specific information or services since the early days of computing. However, while 20 years ago passwords were mainly used by professionals for specific use cases, the introduction of the World Wide Web in the middle of the 1990s led to an extensive spread of passwords in people's daily lives. In the 2000s, the popularity of new technologies like smartphones and tablets and the growing amount of web-based services reinforced the process and thus, users nowadays have to memorize a multitude of passwords compared to a decade ago.

By the end of the 1990s, researchers began to evaluate the influence of user behavior on alphanumeric passwords [1, 2]. Those early studies which were based on self-reported data found that alphanumeric passwords always comprise a trade-off between usability and security. User-chosen passwords are often optimized for memorability and therefore based on dictionary words, birthdays, and so on. This makes them easy to guess for unauthorized persons. Furthermore, most people reuse passwords for multiple accounts and hardly renew passwords they once generated [3].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 460–467, 2013.

More complex passwords are often written down or shared with other people and thus, do not necessarily lead to improved security [4].

To support users in the selection of secure passwords, many companies introduced guidelines, password policies, recommender systems and password meters. The effect of these mechanisms was evaluated in lab studies which found that users often choose the same numbers or symbols and insert these at the same positions to comply with such systems. Thus, password policies do not necessarily increase system security [5]. Furthermore, it was shown that increasing the length of a password has the biggest effect on security and that the recall of policy-based passwords takes significantly more time [6]. This challenges the benefit of extensive password policies.

In addition to self-reported data and lab studies, large databases of user-chosen real-world passwords became available in recent years (e.g. [7, 8]). The analysis of these password lists confirmed that user-chosen passwords are often very short and many passwords are based on names, dictionary words and other trivial strings [7].

The goal of our study was to answer the question if password selection and security awareness evolved since the users' first contact with alphanumeric authentication and which factors influenced this process. Our work contributes to the field by providing valuable insights into the individual-related password evolution. We interviewed 40 people with different demographic backgrounds to get insights into their personal history of password use. We compared early passwords with currently used ones and analyzed what influenced the evolution of these passwords. By using electronic password analysis tools, we were able to quantify the data and gather detailed information about the complexity and the distance of different passwords without actually requiring the password itself.

The results show that the quantity as well as the complexity of used passwords rose in recent years and that password policies did influence this development. However, people still reuse passwords for multiple services and adapt to new policies by simply inserting new characters to old passwords. We found that even if users know how to build secure passwords, they use simple ones more often, some of them being their first-ever created password.

2 Evaluation

The main goal of our study was to find out how passwords evolved over time in terms of length, complexity and quantity and which factors influence this process. Therefore, we conducted one-on-one interviews and collected quantitative data using password analysis tools.

2.1 Password Tools

We implemented two password tools to analyze the composition of a given word and the distance between two words. The tools were built using JavaScript and HTML. Our tools did not store or transmit any data, but displayed statistical results to our participants, who copied them into a questionnaire.

Distance Test. The graphical user interface consisted of two password fields and one text field. When two words were entered into the password fields, the Levenshtein distance of these two words was computed and displayed in the text field. The Levenshtein distance describes the minimum number of changes required to transform one character sequence into the other.

Composition Test. The graphical user interface of the composition test consisted of one password field, one 14 x 2 table and one text field. When a word was entered into the password field, the composition of this word was analyzed and displayed.

We analyzed the word length and counted lower-case letters, upper-case letters, numbers and symbols. In addition, we checked for middle numbers/symbols and for repeated, consecutive and sequential characters. The script was based on the script of www.passwordmeter.com[1].

2.2 Design and Procedure

The study was conducted in a public coffee shop. We used one-on-one interviews in combination with a questionnaire and the password tools. The decision to conduct the study outside of the lab was made to gather a wider demographic spectrum. Participants were recruited via flyers, which were distributed in the coffee shop. As an incentive, we paid each participant one drink from the menu and gave out a 5 Euro shopping voucher. The interview lasted for about 20 minutes on average.

Participants were seated at a table in front of a wall. The examiner sat at the same table and used a laptop to read out the questions and to enter the answers into the questionnaire. At the beginning, we explained the study goals, the technical background (e.g. the password tools) and that there was no way we could steal their passwords from them. In addition, we warned them not to disclose their passwords during the interview. After the introduction, the interview started.

After collecting demographical data, we investigated the general password experience. This involved questions like the year of the first password creation or the amount of actively used passwords. In the case that our participants were not sure about their very first password, they were allowed to use the first password that they could remember. Further questions analyzed the influence of password polices, password meters, etcetera. After the interview, the laptop was handed over to the participants and they were asked to analyze their passwords. For this task, we positioned a screen in front of the participant to prevent shoulder surfing while passwords were entered.

The participants analyzed the requested passwords (see Table 1) and copied the results from the text field of the password tool to the questionnaire. Instructions looked like "*please compare your most secure password to the password that you use most often*". Using this approach, we were able to analyze many aspects of the passwords without actually knowing the participants' passwords.

[1] The script is available for distribution under the GNU General Public License (GPL).

Table 1. Password categories analyzed during the interviews

Password category	Definition
First	The first password ever created
Most used	The password which is used most often
Most secure	The password which is rated most secure (by the participant)
Policy-based	A password which was created based on a given policy
Meter-based	A password which was influenced by a password meter

2.3 Participants

We interviewed 40 participants. The average age of 39 participants was 26 years (18-59). One participant did not reveal his age, but stated to be 40 to 50 years old. 16 participants were female, 24 were male. 38 participants had an academic background, 19 of them had a technical background (e.g. computer science). This might be influenced by the coffee shop being located in a university and business district.

3 Results and Discussion

The results are based on the qualitative answers of the 40 interviewees and the quantitative password analysis. The definition of the analyzed password categories is found in Table 1.

3.1 Experience

On average, the participants had their first contact with passwords in the year 2000 (SE: 1, min = 1994, max = 2008). Their average age at this time was 15 years (SE: 1, min = 5, max = 54). Asked for the reason to create a password, 27 participants stated they had signed up for an email account; seven protected a user account of an operating system and four participants created their first password to protect a mobile phone. Beside these services, gaming and online banking were mentioned. In the first year of password use, our participants had to deal with a mean of 1.5 passwords (SE = 0.1, min = 1, max = 3). Today, the average amount is 14.2 (SE = 3.8, min = 1, max = 150). However, only 5.1 (SE = 0.7, min = 1, max = 28) of these are used frequently.

3.2 Complexity

We conducted a one-way repeated-measures ANOVA to compare the complexity of the different passwords. Fig. 1 shows the results of the complexity analysis; the concrete values can be derived from Table 2.

The results reveal that there is a highly significant main effect of the password category on the length of the password, $F_{2.38,135.99} = 10.33, p < 0.01$. The within-subject contrasts show that the first passwords (M = 7.6) are significantly shorter than passwords of all other categories (all $p < 0.05$). Most secure passwords (M = 12.1)

have the most characters, but policy-based (M = 10.2) and meter-based (M = 10.7) passwords are not significantly shorter (all $p > 0.05$). In contrast, most-often used passwords (M = 8.7) comprise significantly less characters (all $p < 0.05$). This result indicates that, according to the amount of characters, passwords became more secure over time. However, even if users know how to create secure passwords, and this creation is supported by policies and password meters, most authentications are still performed using shorter and thus less secure passwords. According to our participants, secure passwords are only used for specific services whose data is rated sensitive (e.g. bank account).

Table 2. Mean values (and SE) of the different characters tested in the password analysis

Password category	Length	Lower-case	Upper-case	Numbers	Symbols
First	7.60 (0.35)	5.68 (0.50)	0.23 (0.10)	1.68 (0.34)	0.03 (0.03)
Most used	8.65 (0.27)	5.83 (0.42)	0.45 (0.14)	2.25 (0.29)	0.13 (0.05)
Most secure	12.13 (0.79)	7.65 (0.70)	0.95 (0.29)	2.85 (0.33)	0.68 (0.24)
Policy-based	10.18 (0.52)	6.44 (0.57)	0.72 (0.16)	2.67 (0.28)	0.36 (0.11)
Meter-based	10.65 (0.60)	6.52 (0.66)	1.03 (0.31)	2.68 (0.36)	0.42 (0.13)

The analysis of the password composition revealed significant main effects on the use of upper-case letters ($F_{2.52,73.09} = 3.80, p < 0.05$), numbers ($F_{2.89,83.68} = 4.64, p < 0.05$) and symbols ($F_{2.01,58.19} = 5.26, p < 0.05$). Interestingly, there is no significant difference on the use of lower-case letters ($p = 0.13$). This indicates that passwords were always based on lower-case letters, but recently created passwords additionally comprise numbers, upper-case letters and symbols. The post-hoc tests reveal that most secure passwords (M = 1.0) and policy-based (M = 0.7) or meter-based (M = 1.0) passwords include significantly more upper-case letters, than the firstly created ones (M = 0.2), all = $p < 0.05$. In contrast, most used passwords (M = 0.5) are not based on significantly more upper-case letters than the first passwords ($p > 0.05$). The analysis of the amount of numbers shows that compared to the first passwords (M = 1.7), passwords of all other categories use significantly more numeric characters (all = $p < 0.05$). However, most used passwords (M = 2.3), most secure passwords (M = 2.9), policy-based passwords (M = 2.7) and meter-based passwords (M = 2.7) do not differ significantly, $p > 0.05$. Looking at symbols reveals that both, the first (M = 0.0) and the most used passwords (M = 0.1) are composed of significantly less symbols than the rest of the passwords ($p < 0.05$). Most secure passwords contain the most symbols (M = 0.7).

An analysis of the simplicity of passwords according to the use of letters only reveals that 50% of the first passwords and 22.5% of the most used passwords consist of letters only. In contrast to that, only 6.3% meter-based, 7.7% policy-based and 7.5% of the most secure passwords use letters only. In addition, 12.5% of the first passwords were based on numbers only. One participant (2.5%) still uses a numeric password as most secure and most often password. No numeric passwords are found in the policy-based or meter-based category.

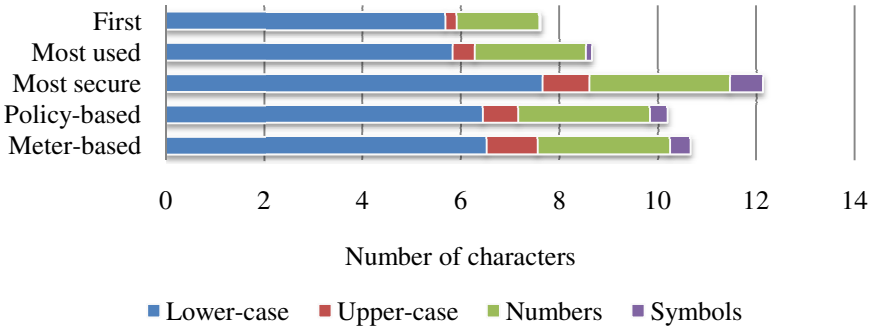

Fig. 1. Password composition by categories (see Table 1) comparing the amount of lower-case letters, upper-case letters, numbers and symbols

3.3 Usage and Behavior

45% of our participants still use their first password. While 30% use it only to authenticate with old services, 70% of them still use their first password to register with new services. This indicates that users rarely change passwords they once created and password reuse is common.

Fig. 2 gives an overview of the mean Levenshtein distances of four comparisons. The distance test reveals that four (10%) participants still use their first password as the most secure password (dist. = 0). In addition, 15% use their unchanged (dist. = 0) first password or a slightly changed (dist. = 1-3) version most often.

A one-way repeated-measures ANOVA comparing the means of the calculated distances reveals a significant main effect of password ($F_{3,87} = 4.61, p < 0.05$), Post-hoc tests show that this effect is caused by the significantly larger distance between the first password and the most secure password (all = $p < 0.05$). The analysis indicates that people often reuse passwords for multiple services: 40.5% of our participants use an unchanged (dist. = 0) or slightly changed (dist. = 1-3) policy-based password most often. This can be explained by the fact that 67.5% of the participants respond to password policies by adopting an already used password.

A similar conclusion can be drawn when comparing meter-based passwords to most often used passwords: 31.3% of our participants use an unchanged (dist. = 0) or only slightly changed (dist. = 1-3) meter-based password most often. Comparing the distance of the most secure and the most used passwords reveals that 17.5% of our participants use their most secure password most often (dist. = 0). In addition, 5% of the most secure passwords show only minor differences (dist. = 1-3) compared to the most used passwords.

The qualitative data supports the quantitative results as 51.4% stated to reuse passwords that they once created. According to the participants, this behavior is necessary since otherwise, they would have to memorize too many passwords. Even with reuse, 42.5% write down their passwords. A dependent t-test comparing the mean numbers of frequently used passwords of participants, who write down

passwords (M = 5.5, SE = 0.7) with those, who do not write down their passwords (M = 3.8, SE = 2.3) reveals that the latter group also uses (and therefore has to memorize) significantly less passwords, $t_{37} = -2.06, p < 0.05, r = 0.32$. This indicates that the number of used passwords has a medium effect on the behavior of writing them down. One participant, who stated to use 28 out of 150 passwords frequently, was excluded from the analysis as this amount was outside of the doubled standard deviation (SD = 4.6) and therefore was assessed as an outlier. This participant mentioned that he uses a master password protected keychain to manage his passwords.

Finally, the results indicate that the influence of password policies on passwords was stronger than the influence of password meters as 92.5% stated that policies did influence their passwords, but only 57.5% stated the same for password meters. This might be based on the fact that password-meters are pure recommender systems, while password-policies often force users to follow specific guidelines.

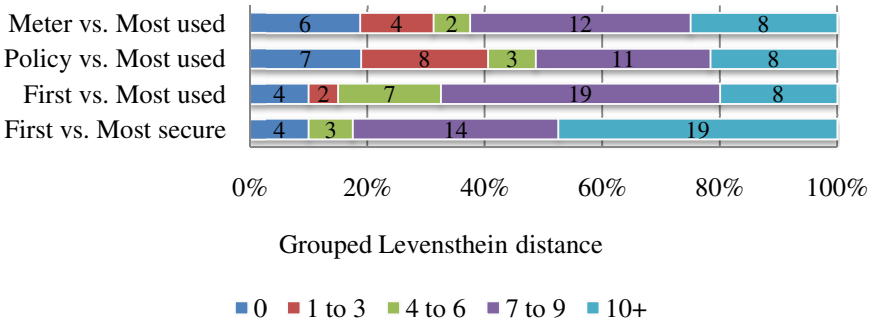

Fig. 2. Grouping of the calculated Levenshtein distances. The biggest differences can be found between the first passwords and the most secure passwords.

4 Limitations

Since the analysis is based on self-reported data, it is possible that the participants did not report their real passwords. Reasons might be that they did not want to expose them or could not perfectly remember some of their passwords. To check the validity of the reported data, we added several control questions into the questionnaire. In addition, we instructed our participants to report any memory gaps. Based on the results of this monitoring, we have no reason to assume that our participants reported contrived passwords and argue that our data is valid.

The complexity analysis is based on 40 participants and can therefore, in terms of generalization, not compete with analyses based on thousands of real user passwords. We like to mention that our goal was not to give an overview of currently used passwords, but to combine quantitative and qualitative data to be able to trace the individual-related evolution of passwords. Therefore, we are confident that the sample size meets the requirements for this approach.

5 Conclusion and Future Work

We conducted one-on-one interviews and used password analysis tools to gather valuable insights into the complexity of different user passwords. We could trace the evolution of our participants' passwords and the factors that influenced this process.

The results indicate that, in contrast to their first passwords, users today know how to build more secure passwords and thus some current passwords are based on significantly more characters. However, most people still rely on weak (e.g. short) passwords for most authentications, especially, when services are not rated sensitive. In addition, recommender systems and password guidelines had only marginal effects on the password strength and the reuse of old-established passwords is still common.

This is a serious security flaw as attackers could start by finding out a password of a low-sensitivity service, just by guessing or by using some deficiencies of the implementation. The reuse of passwords and the small distances to more secure passwords could consequently enable access to more passwords and more sensitive data. As the growth of web-based services will demand memorizing even more passwords in the future, we argue that usable alternatives to alphanumeric authentication have to be found.

Another point for further investigation is the potential influence of long-term mobile device use on password selection. Text input on such devices is cumbersome, which might, in the long run, negatively influence the security of such passwords.

Acknowledgments. This work was partially funded by a Google Research Award.

References

1. Adams, A., Sasse, M.A., Lunt, P.: Making passwords secure and usable. In: Proc. HCI 1997, pp. 1–19. Springer, London (1997)
2. Adams, A., Sasse, M.A.: Users are not the enemy. Commun. ACM 42, 40–46 (1999)
3. Riley, S.: Password Security: What Users Know and What They Actually Do. Usability News 8, 1 (2006)
4. Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: user attitudes and behaviors. In: Proc. SOUPS 2010, pp. 2:1–2:20. ACM, New York (2010)
5. Komanduri, S., Shay, R., Kelley, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F., Egelman, S.: Of passwords and people: measuring the effect of password-composition policies. In: Proc. CHI 2011, pp. 2595–2604. ACM, New York (2011)
6. Proctor, R., Lien, M.-C., Vu, K.-P., Schultz, E., Salvendy, G.: Improving computer security for authentication of users: Influence of proactive password restrictions. Behavior Research Methods 34, 163–169 (2002)
7. Weir, M., Aggarwal, S., Collins, M., Stern, H.: Testing metrics for password creation policies by attacking large sets of revealed passwords. In: Proc. CCS 2010, pp. 162–175. ACM, New York (2010)
8. Bonneau, J.: The science of guessing: analyzing an anonymized corpus of 70 million passwords. In: Proc. SP 2012, pp. 538–552. IEEE Computer Society, Washington, DC (2012)