Studying Passwords to Create Domain-Specific Blacklists

Kentrell Owens
Carnegie Mellon University
Pittsburgh, PA 15213, USA
kentrelo@cmu.edu

Ziheng Ni
Carnegie Mellon University
Pittsburgh, PA 15213, USA
zni@andrew.cmu.edu

Mengchen Yong
Carnegie Mellon University
Pittsburgh, PA 15213, USA
myong@andrew.cmu.edu

Josh Tan
Carnegie Mellon University
Pittsburgh, PA 15213, USA
jstan@cmu.edu

Neha Sridhar
Carnegie Mellon University
Pittsburgh, PA 15213, USA
nksridha@andrew.cmu.edu

Lorrie Cranor
Carnegie Mellon University
Pittsburgh, PA 15213, USA
lorrie@cmu.edu

USENIX Symposium on Usable Privacy and Security (SOUPS), August 11-13, 2019
Santa Clara, CA, USA

Abstract

Password policies can incorporate a blacklist check to ensure that created passwords do not contain predictable patterns, dictionary words, or previously leaked passwords. In addition, some organizations also incorporate organization- or site-specific terms into their blacklist (e.g. Carnegie Mellon (CMU) might choose to prevent passwords containing the word "tartan"). Currently, National Institute of Standards and Technology (NIST) recommends that passwords containing "context-specific words, such as the name of the service, the username, and derivatives thereof" be blacklisted [5]. In this paper, we investigated what kind of domain-specific information users include in passwords they create when they are prompted with a new website. Our study collected data from 680 Amazon Mechanical Turk (MTurk) participants who were randomly presented with one of three imitation websites and asked to create a password. We analyzed the passwords created in our study and found that almost 10% of the passwords contained domain-specific information. This information included text found on the websites, visual features of the website, topics related to the content of the website, and words specific to the recruiting platform we used for the experiment.

Author Keywords

Passwords; usable security; blacklists

Introduction

The approach of the end of passwords has been declared by experts in the security community for decades now [12]. Contrary to these predictions, passwords remain the authentication method that achieves the best balance of deployability, usability, and security [1] and will continue to be used for the foreseeable future. Previous research on password creation has found that an effective blacklist reduces the risk of a user having their password guessed [6, 16]. Blacklists contain words that users are not allowed to include in their passwords. It is important to identify how website administrators can improve their password policies to better protect their users. We're interested in investigating what types of domain-specific information users might include in passwords. Studying this domain-specific information could lead to more robust blacklists and consequently stronger passwords for users.

We hypothesized that users would attempt to create passwords based on a website's domain name, words associated with the website, images found on the website, and other identifiable information found on the website, including transformations (or "mangling") of domain-specific information.

Related Work

Wei et al [15] studied the "service-specific" information found in the top 1000 leaked passwords from five web services. They found "that passwords from each service reflect the category of the service, often by including the name or semantic theme of the service" [15]. Our study aimed to demonstrate the similar results to Wei et al. using passwords created in a user study and instead of leaked passwords from a field study. We also examined all of the passwords collected in the study instead of the most common ones.

Password studies have shown that users include the website domain name or service name in the passwords, either directly as words or phrases, or through some predictable transformations and mangling [9, 10, 15]. In addition, users may use words or content associated with the website or users' purpose of visiting the website [7, 11]. For instance, passwords like "+Money369" for a bank account use words/phrases closely related to the general type of website/service. This can be highly predictable and make the passwords for this particular service more guessable [13]. Therefore, we suggest that a domain-specific blacklist, if well targeted, could increase the overall strength and security of passwords on a certain domain. However usability and security must be balanced; if a blacklist is too large its use becomes impractical [4].

Methodology

The data in this study was collected using Qualtrics, an online survey tool. Recruitment occurred on MTurk and was restricted to individuals who live in the United States of America and are 18 years of age or older. We paid participants $0.55 for completing Part 1 and $0.70 bonus for completing Part 2. Out of our 680 participants in Part 1, 75.0% of them were able to complete Part 2 of the survey. 49% of users identified as women, 49% as men, and 1% as trans/non-binary. 75.4% of participants said that they are not "majoring in or...have a degree or job in computer science, computer engineering, information technology, or a related field."

Part 1

After examining the Alexa Top 50 websites, we determined that social networking websites would be good candidates for imitation due to their popularity; we chose to imitate Twitter (Panddar), Tinder (Torch), and WhatsApp(HowYoDoin). Users were presented with one of three imitation websites

in which they were asked to create a password (following a *comprehensive8* password policy) that was not one of their own real-world passwords and asked to remember this password using the method they would normally use. Participants then took a survey (about their password creation process and demographics) and were asked to return in 48 hours for Part 2.

Part 2
After a 48-hour waiting period, users were sent a recall survey with the same website they had seen in the initial survey and asked to recall their password. After completing a series of questions about how they recalled their password, participants were compensated for completing the recall.

Analysis
We had two qualitative coders manually inspected each password and determine if it contained domain-specific information. Their inter-coder reliability metric was satisfactory (Cohen's Kappa = 0.71).

Our primary security metric was guessability. We calculated the guessability of each password using CMU's Password Guessability Service (PGS) [14].[1] This service generates a *guess number* for passwords using up to five different password guessing methods. To analyze the strength of the passwords collected in this study, we used a Cox Proportional-Hazards regression [2].

Our usability metric was the number of attempts it took users to enter their password correctly in the Part 2. We used Pearson's Chi-squared test to determine if the number of recall attempts (0-3) was different for passwords flagged as containing domain-specific information and non-flagged passwords (α = 0.05). In analyzing the recall attempts, we only included users who said they entered their password from memory and said they did not reuse their study password for another account.

[1] https://pgs.ece.cmu.edu/

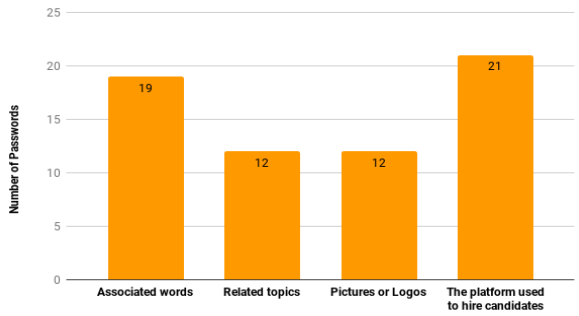

Figure 1: Number of Domain-Specific Passwords in Different Groups

Results

Frequency of domain-specific passwords
We collected 228 passwords for HowYoDoin, 234 for Torch, and 218 for Panddar. 9.4% of the passwords created contained domain-specific information. We divide these domain-contextual passwords into four groups according to the information contained: associated words (Panddar1$, TOrched1!), related topics (LoveBugs56$, Sexy!1337), pictures or logos (Fire-2019, 9Hands%%), and the recruitment platform (mTurkpassw0rd!, Mturk01!). The distribution of the passwords containing domain-specific information into these categories is shown in Figure 1. Torch was the website for which users created the highest percentage (14.5%) of domain specific passwords, followed by Panddar (7.3%) and HowYoDoin (6.1%).

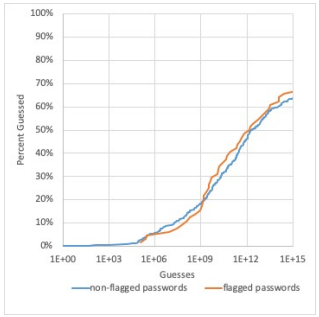

Figure 2: Guessability of passwords flagged as containing domain-specific information and non-flagged passwords.

Guessability of domain-specific passwords
As evident from Table 1, none of the factors identified had any statistically significant impact on password guessability. Neither if a password contained domain-specific information nor which website was presented to users affected the guessability of the passwords created significantly. Figure 2 shows that the percent of passwords guessed does not differ much between the two sets of passwords.

Usability of domain-specific passwords
We found no statistically significant difference in the number of recall attempts needed by a user who made a flagged password vs users who made non-flagged passwords $\chi^2(df = 3, N = 160) = 0.82548, p = 0.84$.

Factor	Coef.	Exp(coef)	SE	*p*-value
flagged	-0.09490	0.90946	0.18995	0.617
website	-0.10348	0.90169	0.06922	0.135

Table 1: Final Cox Regression output for all passwords

Discussion
The password analysis shows that our hypotheses about the type of domain-specific information we'd find was confirmed. The Torch website had the highest percentage of domain specific passwords (14.5%). This could be because users found the topic of this website (dating) to be more compelling than the other websites' topics (messaging and social media).

We observed no significant difference in the security or usability of passwords containing domain-specific information. Since PGS was not trained to include domain-specific information in their password guesses (as an attacker would), the password guess numbers are higher than they actually should be.

Of the domain-specific passwords, 33% contained information about the recruitment platform. This may indicate that the content of passwords collected from crowd-sourcing platforms such as MTurk may not be representative of real-life passwords (although they are similar in strength [3, 8]).

Limitations
The passwords in this study were created in a simulated environment in which users were presented with a photograph of a fake website and asked to create a password. The differences between this method and real-life passwords creation include: 1) A photograph of a website instead of interaction with the actual website, 2) the website was a fake one and 3) there was no risk of users losing personal information if their passwords were compromised. For these reasons, we cannot be sure that users created passwords as they would in a real-life scenario. Although prior work [3, 8] has shown that passwords created in this environment are useful for studying real passwords, we have not found any literature discussing whether the authenticity of a website or how that website is presented to a user impacts password creation.

Conclusion and Future Work
We analyzed 680 passwords created by MTurk participants for imitation websites during a two part password creation study. We identified domain-specific information in the created passwords (e.g. website name, design elements) and analyzed the strength of these passwords as well as their usability. We found that 10% of passwords used some form of domain-specific information. A password blacklist targeting these domain-specific words/phrases could be useful to some extent, and future work should attempt to have a much larger sample size to determine if there is statistical significance in changes to password security.

Acknowledgements

We would like to thank Hana Habib for her advice and valuable help with designing our study.

REFERENCES

1. Joseph Bonneau, Cormac Herley, Paul C Van Oorschot, and Frank Stajano. 2012. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy*. IEEE, 553–567.
2. David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.
3. Sascha Fahl, Marian Harbach, Yasemin Acar, and Matthew Smith. 2013. On the Ecological Validity of a Password Study. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS '13)*. ACM, New York, NY, USA, Article 13, 13 pages. DOI:http://dx.doi.org/10.1145/2501604.2501617
4. Dinei Florencio, Cormac Herley, and Paul van Oorschot. 2014. An administrator's guide to internet password research. *Proceedings of the 28th Large Installation System Administration Conference (LISA14)* (2014).
5. PA Grassi, JL Fenton, EM Newton, RA Perlner, AR Regenscheid, WE Burr, JP Richer, NB Lefkovitz, JM Danker, Yee-Yin Choong, and others. 2017. NIST Special Publication 800-63b: Digital Identity Guidelines. (2017).
6. Hana Habib, Jessica Colnago, William Melicher, Blase Ur, Sean Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Cranor. 2017. Password creation in the presence of blacklists. *Proc. USEC* (2017), 50.
7. Zhigong Li, Weili Han, and Wenyuan Xu. 2014. A large-scale empirical analysis of Chinese web passwords. *Proceedings of the 23rd USENIX Security Symposium* (2014).
8. Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. 2013. Measuring Password Guessability for an Entire University. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS '13)*. ACM, New York, NY, USA, 173–186. DOI: http://dx.doi.org/10.1145/2508859.2516726
9. Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. 2017. Lets Go in for a Closer Look: Observing Passwords in Their Natural Habitat. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS 17* (2017). DOI: http://dx.doi.org/10.1145/3133956.3133973

10. Richard Shay, Lorrie Faith Cranor, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and et al. 2016. Designing Password Policies for Strength and Usability. *ACM Transactions on Information and System Security* 18, 4 (2016), 1âĂŞ34. DOI:http://dx.doi.org/10.1145/2891411
11. Elizabeth Stobert and Robert Biddle. 2014. The password life cycle: user behavior in managing passwords. *Tenth Symposium On Usable Privacy and Security* (2014).
12. Daniel Terdiman. 2013. Google security exec: 'Passwords are dead'. (Sep 2013). https://www.cnet.com/news/google-security-exec-passwords-are-dead/
13. Blase Ur, F Noma, J Bees, S M Segreti, R Shay, L Bauer, N Christin, and L F Cranor. 2015a. âĂİI added âĂŹ!âĂŹ at the end to make it secureâĂİ: Observing password creation in the lab. *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)* (2015), 123âĂŞ140.
14. Blase Ur, Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Saranga Komanduri, Darya Kurilova, Michelle L. Mazurek, William Melicher, and Richard Shay. 2015b. Measuring Real-World Accuracies and Biases in Modeling Password Guessability. In *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, Washington, D.C., 463–481. https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/ur
15. Miranda Wei, Maximilian Golla, and Blase Ur. 2018. The Password Doesn't Fall Far: How Service Influences Password Choice. *Proceedings of the 5th SOUPS Who Are You?! Adventures in Authentication Workshop (WAY)* (2018).
16. Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. 2010. Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. In *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS '10)*. ACM, New York, NY, USA, 162–175. DOI: http://dx.doi.org/10.1145/1866307.1866327