The Password Doesn't Fall Far: How Service Influences Password Choice

Miranda Wei University of Chicago Chicago, United States weim@uchicago.edu Maximilian Golla Ruhr-University Bochum Bochum, Germany maximilian.golla@rub.de Blase Ur University of Chicago Chicago, United States blase@uchicago.edu

ABSTRACT

Users often create passwords based on familiar words or things they like, using these passwords across many web services. But does the type of web service influence how users construct their password? In this paper, we observe how and how often passwords are specific to the services for which they were created. We analyze leaked passwords from five web services. We find that passwords from each service reflect the category of the service, often by including the name or semantic theme of the service. Through a qualitative analysis of passwords, we further identify unique characteristics of the passwords created for each service. Service-specific passwords can reveal other shared interests or demographics of that service's userbase. This contextual perspective on password creation suggests improvements for site-specific blacklists and password-strength meters.

1. INTRODUCTION

It is well-known that users craft passwords around common themes like names [14], dates [17], phrases [8], and more [18]. Much work has also been done to understand the content and structure of passwords that users create [7], as well as the strength of passwords against adversarial attacks [13].

When users differ from each other in what they choose to use for their passwords, it may be because they choose passwords more carefully for accounts they care more about [14], or because stricter password composition policies force them to do so [5]. Password choices may also be correlated with demographic factors [9]. However, are these factors enough to explain all variation in passwords? What else might cause users to choose passwords differently? In this paper, we study how the category of a web service influences the passwords chosen for that service. While most users choose and reuse meaningful concepts for their passwords, no matter the account they are for, we find that other users choose service-specific concepts for their passwords. Focusing on the web service's category, we aim to highlight the service as an important contextual factor in password creation.

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

Who Are You?! Adventures in Authentication (WAY) 2018. August 12, 2018, Baltimore, MD, USA.

We use qualitative methods to understand how passwords are related to the type of service they are created for and find that passwords are service-specific in two distinct ways. First, users include the name of a service in their passwords for that service. For example, variations on the name LinkedIn account for six out of the top ten LinkedIn-specific passwords. Second, users include terms related to the semantic theme of a service in their passwords for that service. Passwords for adult-content services include explicit terms for adult content, while passwords for online gaming services include competitive and violent terms. Our study also reveals that passwords often reflect how users utilize or feel about a service. The password, *jobsearch*, corroborates the claim that users search for jobs on LinkedIn, while the password, *freemusic*, indicates the value that users get from such a music-streaming service.

Despite the large literature on passwords, little work has been done to quantify the ways in which passwords are related to the services for which they were created. Our paper provides a lower bound for this phenomenon by analyzing passwords sampled from the password leaks of five web services: Battlefield Heroes, Brazzers, Last.fm, LinkedIn, and Mate1. We find that passwords related to the service by name or semantic theme vary by category, accounting for 2-7% of the top 1,000 passwords specific to that service.

Our findings underscore the importance of adjusting and augmenting blacklists for each service. As attackers can leverage contextual information about a service to improve their password guessing, defenders must have an empirical understanding of how users choose passwords to implement the most effective defenses. For example, popular passwords for adult-content websites also included references to cars, trucks, and other vehicles. Sometimes, the passwords reveal characteristics of the userbase, as passwords for a game produced by a Swedish company included many Swedish terms. This paper contributes to a fuller contextual understanding of passwords, in hopes of informing the implementation and expansion of blacklists and other related password defenses.

In Section 2, we review related work on password creation, as well as efforts to improve users' password creation. We describe the methodology of our study in Section 3 and present the results in Section 4. We conclude with a discussion the implications of our findings in Section 5.

2. RELATED WORK

In this section, we review research on users' strategies for password creation, as well as efforts to improve passwords. Users often create passwords from a small set of dictionary words, potentially using a small set of substitutions [7]. Studying these passwords by their semantic theme reveals that these words often relate to pets' names, people's names, or dates [17,18]. Other frequent terms include sports teams, geographic locations, or song lyrics [8]. Keyboard walks or patterns (e.g., zxcvbn) are also common passwords [1].

Efforts to improve password creation emphasize intervening during password creation, as there are many discrepancies between users' perceptions of a password's security and reality [15]. Creating passwords is often frustrating and unrewarding for users [14], so many users are unlikely to change their password (especially to something more complex) without being required to. Password meters are well-placed to intervene, especially if they are designed with usability concerns in mind [13]. Successful meters can provide users with advice about their potential passwords [3, 12, 16]

Blacklists, or lists of frequently used passwords that are banned, play an important role in preventing users from choosing the most vulnerable passwords [4, 6]. Through proactive password-checking, blacklists can be an efficient way to encourage more secure passwords [20], provided that the blacklists do not significantly impede usability [9].

3. METHODOLOGY

In this section, we describe the methodology of our study.

3.1 Leak Selection

For our study, we analyzed passwords leaked from five web services: Battlefield Heroes, Brazzers, Last.fm, LinkedIn, and Mate1. We chose these datasets due to the salience of the services from which they were leaked. Each service provides distinct and unique value for its users: Battlefield Heroes was an online third-person shooter game, whereas Brazzers is an adult-content production company and content-hosting website. LinkedIn provides a platform for professional social networking, Last.fm provides free music streaming, and Mate1 is an online dating site. We thus specified a category for each of these leaks: *gaming, adult, music, social,* and *dating.*

3.2 Analysis Methods and Metrics

Although Brazzers and Mate1 passwords were stored in plaintext, those from Battlefield Heroes, Last.fm, and LinkedIn were hashed, albeit with weak hash functions. We use the 99.01%, 98.17%, and 98.68%, respectively, of plaintext passwords we were able to recover from the hashes. As we are interested in the most common passwords from these services, which are unlikely to be the (potentially strong) unrecovered passwords, we anticipate that excluding these passwords will have little effect on our results.

We conducted our analysis on the top 1,000 passwords of our five chosen services. In order to focus on passwords unique to a given service, for each service we filtered out passwords that were also among the top 1,000 passwords of any of the other four services. We chose not to use any of the widely distributed, pre-produced lists of common passwords because doing so would risk excluding desired passwords. If a password from a given service was included on a widely distributed list of common passwords precisely because it was popular on that exact service, it should not be removed. After filtering passwords from the top 1,000 passwords in each of the five services, about a quarter remained. Specifically, there were 349 passwords remaining from the Battlefield Heroes, 242 from Brazzers, 77 from Last.fm, 250 from LinkedIn, and 214 from Mate1. All percentages reported in the results, save for the first percentage given in each section, are calculated within the top 1,000 passwords.

A member of the research team used two initial criteria to identify passwords that were related to the service:

- 1. Is it related to the name of the service provider?
- 2. Is it related to the semantic theme of the service or the service's category?

We then performed open-coding on the remaining passwords to identify themes in their semantics. The same researcher iteratively updated a codebook with themes until no new themes emerged. Themes were not mutually exclusive, as a password could express multiple themes, or none. Themes were exhaustive, but only themes accounting for more than 0.5 % of the top 1,000 passwords are reported. These codings were verified by another member of our team.

3.3 Ethical Considerations

While the data for our study concerns real-world passwords, users were not harmed in the process of our research. Data was collected from publicly-available sources, and we stored and analyzed passwords independent of other identifying information. The analysis for our study was conducted on secure, encrypted computer systems.

4. **RESULTS**

Overall, we find that the name and semantic theme of a service influence the passwords created for that service. In all but one category, passwords related by name or semantics account for the greatest proportion of the top 1,000 passwords; in the remaining category, related passwords were the second greatest proportion.

A summary of the top ten service-specific passwords for our five selected services is presented in Table 1.

4.1 Battlefield Heroes (Gaming)

The top 1,000 Battlefield Heroes passwords accounted for 44,076 within the total 548,774 accounts contained in the Battlefield Heroes leak, or 8.0% of all Battlefield Heroes passwords. After filtering, 349 passwords from the top 1,000 remained (cf. Table 2).

As one might expect, *battlefield* was the most common password for this service. Seven other passwords were directly related to the name of the service (e.g., bf2). We further found that 2.7% of passwords were semantically-related to the Battlefield Heroes game. Most of these passwords referred to concepts within games, such as roles within the game (e.g., *trooper*, *killer123*, *commander*) and weapons or violence (e.g., *headshot*, *pirate* pistol). Interestingly, some passwords also reflected aspirational desires for game performance, such as *iamthebest* and *master1*.

A plurality of passwords referred to broader gaming-related terms, and in particular, demonstrate other games that Battlefield Heroes users like, e.g., *runescape*, *halflife2*, and *callofduty.* 3.2% of passwords reflected terminology unique to the online gaming culture, such as translations of *your mom*

Table 1: The top ten passwords per service after removing passwords that were among the top 1,000 passwords for any of the other four services. Each password is shown with its relative percentage of total passwords for that service. The sums of the top ten passwords for that service are shown at the bottom.

Battlefield (Gaming)		Brazzers (Adult)		Last.fm (Music)		LinkedIn (Social)		Mate1 (Dating)	
Password	Of Total	Password	Of Total	Password	Of Total	Password	Of Total	Password	Of Total
battlefield	0.053%	brazzers	0.064%	lastfm	0.150%	linkedin	0.120%	sexy	0.053%
lol123	0.028%	211211	0.022%	music	0.063%	linked	0.019%	mate1	0.050%
xbox360	0.028%	giants	0.019%	abcdefg123	0.049%	Linkedin	0.012%	promise	0.033%
warhammer	0.017%	titties	0.019%	last.fm	0.030%	linkedin1	0.011%	love123	0.024%
starwars1	0.016%	bigboobs	0.018%	foxpass	0.025%	ZZZZZZZZ	0.011%	looking	0.023%
runescape	0.015%	pornstar	0.017%	musica	0.024%	krishna	0.010%	olamide	0.017%
fp2241	0.014%	patriots	0.013%	qqww1122	0.013%	sairam	0.009%	money6	0.016%
4815162342	0.014%	braves	0.012%	ahov	0.011%	super123	0.009%	kissme	0.015%
bfheroes	0.013%	iverson	0.011%	A123456	0.009%	linkedin123	0.008%	damilola	0.015%
hejsan	0.012%	hooters	0.011%	ahovwpib	0.009%	LinkedIn	0.008%	lovingyou	0.015%
Sum	0.210%	Sum	0.205%	Sum	0.383%	Sum	0.217%	Sum	0.260%

(*jemoeder* in Dutch and *dinmamma* in Swedish) commonly used as insults, as well as *ownage*, *pwnage*, and *godlike*, referring to success. This also included 2.2% of passwords that mentioned gaming equipment or technologies used for gaming, such as *xboxlive*, *nintendo64*, *geforce*, and *logitech1*. 1.1% of passwords were keyboard walks, such as *qazqaz* and *qwerasdf*. Other themes were culture, e.g., movies and TV shows (1.8%), and sports (0.9%).

Interestingly, the set of Battlefield Heroes-specific passwords included a high percentage of foreign-language words; 0.8% were Dutch and 0.7% were German. Swedish accounted for another 0.6%, with passwords like *hejsan* (hello), *hejhej* (hi) and *bajskor* (poo); this may be because Battlefield Heroes was developed by EA Digital Illusions CE AB, a Swedish video game company. The 1.4% of passwords that were common given and last names reflected these origins, e.g., *markus, fabian, johannes*, and most are traditionally male.

4.2 Brazzers (Adult)

The top 1,000 Brazzers passwords accounted for 103,902 within the total 928,072 accounts in the Brazzers leak, or 11.2% of all Brazzers passwords. After filtering, 242 passwords from the top 1,000 remained (cf. Table 2).

Unsurprisingly, *brazzers* was the most common password for this service, and another two passwords were related to the name of the service (*brazzers1* and *brazzer*). Further, 4.6 % of passwords were directly related to Brazzers' explicit nature, including *titties* and *pornstar*. Some of these explicit passwords may have been chosen by users because they reflected the users' line of thinking when accessing their accounts: *enjoyporn* and *iloveporn*.

Sports were popular among Brazzers passwords, accounting for 7.0%. These passwords included names of favored teams and athletes (e.g., *iverson*). Team sports (e.g., *hockey1*) were mixed with individual sports like *bowling* and *surfing*. 2.1% of passwords mentioned vehicles, ranging from Honda *accords* to *silverado* and *peterbilt* trucks.

 $1.8\,\%$ of passwords were cultural references, e.g., the band *chevelle*, the movie *godfather*, or character *kramer* from the TV show Seinfeld. Of the $1.6\,\%$ of passwords that were

places, all but four (*alabama*, *texas*, *minnesota*, *maryland*) were cities in the United States, such as *detroit* and *houston*.

0.9% of passwords referred to food and drink, including *pepsi1*, *bigmac*, and (topically) *hooters*. Military references accounted for 0.8% of passwords, such as *airforce* and *infantry*. Finally, 0.7% of passwords were given names, all of which are traditionally male, such as *james1* and *billybob*.

4.3 Last.fm (Music)

The top 1,000 Last.fm passwords accounted for 3,570,290 of the total 43,157,173 accounts contained in the Last.fm leak, or 8.3% of all Last.fm passwords. After filtering, 77 passwords from the top 1,000 remained (cf. Table 2). After our filtering process, Last.fm had the lowest number of remaining passwords out of all five services, indicating that passwords from Last.fm were the least service-specific. This may be because terms relating to music are frequently selected for passwords, regardless of category [8].

As expected, *lastfm* was the most common password for this service; another five passwords were related to the name of the service. The most common theme for Last.fm passwords (1.5%) referred to the value of Last.fm, i.e., music streaming. Passwords such as *mymusic*, *rockon*, and *music4me* demonstrate users' conception of their password as an access point for their music-streaming services.

1.1% of passwords referred to popular musical artists, such as *radiohead*, *linkinpark*, and *rammstein*. Common given names, like *callum* and *laura* accounted for another 0.8%. Finally, references to pop culture (*hellokitty*, *peterpan*), sports (team *westham*, player Wayne *rooney*), and food or drink (*skittles*, *pizza*) were another 1.5% of passwords.

4.4 LinkedIn (Social)

The top 1,000 LinkedIn passwords accounted for 9,676,896 within the total 160,947,194 accounts in the LinkedIn leak, or 6.0% of the total LinkedIn passwords. After filtering, 250 passwords from the top 1,000 remained (cf. Table 2).

linkedin was the most common service-specific password, with 16 other variations on the name in the remaining passwords set. The high number of LinkedIn name-related pass-

Table 2: The result of qualitative coding passwords from each service. Themes are shown with the percentage of passwords expressing that theme out of the top 1,000 passwords for that service. Passwords *filtered* as being common on other services was the most common category. We have bolded service-specific themes (the service's semantic theme and the service's name).

Battlefield (Gaming)		Brazzers (Adult)		Last.fm (Music)		LinkedIn (Social)		Mate1 (Dating)	
Category	Of Top	Category	Of Top	Category	Of Top	Category	Of Top	Category	Of Top
(Filtered)	65.1%	(Filtered)	75.8%	(Filtered)	92.3%	(Filtered)	75.0%	(Filtered)	78.6%
Other Games	3.4%	Sports	7.0%	Service Theme	1.5%	Names	10.2%	Service Theme	6.4%
Gaming Culture	3.2%	Service Theme	4.6%	Music	1.1%	Service Theme	3.5%	Names	3.1%
Service Theme	2.7%	Vehicles	2.1%	Names	0.8%	Religious	1.9%	Religious	0.9%
Culture	1.8%	Culture	1.8%	Service Name	0.6%	Service Name	1.7%	Years	0.9%
Names	1.4%	Places	1.6%	Culture	0.5%	Places	1.1%	Service Name	0.2%
Keyboard	1.1%	Food/Drink	0.9%	Sports	0.5%			—	
Sports	0.9%	Military	0.8%	Food/Drink	0.5%	—		—	
Service Name	0.8%	Names	0.7%	—		—		—	
Dutch	0.8%	Service Name	0.3%	—				—	
German	0.7%	—		—		—		—	
Swedish	0.6%	—							

words suggests that "LinkedIn" is an especially salient description of the service for users.

A significant portion (10.2%) of passwords were common given or last names, such as *isabelle* and *florence*. These names came from a wide variety of ethnic origins, including Hebrew (*abigail*), Portuguese (*catalina*), Greek (*penelope*), and Indian (*rajesh*).

3.5% of passwords reflected the services that LinkedIn provides. Common passwords included *networking* and *marketing*, as well as *professional* and *career*. In particular, it appears that many users are hopeful that their LinkedIn password will provide for job changes, considering the common passwords, *jobsearch* and *newjob*. Further, many users may have set their LinkedIn passwords with a *carpediem* (Latin for "seize the day") or *hariom* (Sanskirt mantra for "removes suffering") mindset towards their job aspirations.

Religious themes accounted for another 1.9% of passwords, with a range of religions represented. Christian passwords commemorated *jesuschrist* and *john316*. The Hindu deities *ganesha* and *aditya*, and Sikh for God, *waheguru*, were also popular as passwords. Interestingly, the Hindu mantras *omsairam* and *jaimatadi* were very popular, meaning "that which saves" and "victory to the mother," respectively. There were also general religious statements, such as *godisgood*, *godisgreat*, and *godislove*. As religious themes are not commonly found in the Battlefield Heroes, Brazzers, or Last.fm service-specific passwords, this suggests that users may feel that they would benefit from the intervention of higher powers in their careers.

Finally, geographic places, like countries and cities, comprised 1.1% of passwords; most were in South or Southeast Asia (*bangalore, singapore, thailand*) or Europe (*france*), but there were some in North America (*toronto*) and Africa (*casablanca*) as well.

4.5 Matel (Dating)

The top 1,000 Matel passwords accounted for 3,746,719 within the total 27,403,958 accounts contained in the leak, or 13.7% of the total Matel passwords. After filtering, 214 passwords from the top 1,000 remained (cf. Table 2).

Common service-specific passwords included both mate1 and

mate1.com. 6.4% of passwords were identified as semantically related to Mate1's status as a dating platform, such as *dating* and *promise*. Some indicated that users were hoping to find love, e.g., *lovingyou*, *honesty*, *loving*, and *icare123*. As with many dating services, there were also explicit passwords: *ilovepussy* and *sexybitch*.

3.1% of passwords were common given or last names, such as *joe* and *michael*. Notably, Nigerian names accounted for about half of all names; some referred to Nigerian celebrities, such as the Nigerian hip hop artist *olamide* or the Nigerian actress *opeyemi* Ayeola.

Religious concepts accounted for another 0.9% of passwords. Some of these passwords were simply religious statements, e.g., *ilovegod, jesusislord*, while others may have been more hopeful or wishful, e.g.,*godhelpme, ingodwetrust*. Similar to the LinkedIn passwords, this may indicate that users are praying to higher powers in their Matel dating situations.

Interestingly, another 0.9% of passwords were years in the late twentieth century, ranging from 1969 to 1987, potentially indicating common birth years of Matel's users.

5. DISCUSSION

In this study, we observed the ways in which the category of a web service influences the passwords created for that service. We qualitatively analyzed the top 1,000 passwords leaked from five popular web services, each of which represented different categories. We find that a minimum of 3-6% of accounts' passwords could be easily guessed by attackers by trying variations on the service's name or semantic theme. Additionally, attackers would be successful in making targeted guessing attempts based on characteristics of the service's userbase, such as shared interests or demographics. For example, at least 6.5% of passwords for one gaming service reflected topics salient to gamers, such as other games and gaming equipment.

Our findings provide empirical grounding for the implementation of blacklists. From an attacker's perspective, using service-related information to improve guessing attacks is nothing new; many password guessing models and algorithms support custom strings for improving guesses [2, 19]. Our work emphasizes the need for defenders to think broadly in preempting such attacks. As has been previously suggested, system administrators should prohibit passwords composed of the service's name at a minimum [21]. We further argue that terms related to the semantic theme of the service should also be blacklisted. While previous work has found that users may create weaker passwords when blacklists prevent them from using their desired password [6], it is unclear whether this holds equally true across categories of web services. Future work should investigate how users perceive the category of a service and conduct related user studies of their thinking during password creation.

A categorical understanding of how passwords vary by web service also stands to improve any trained password strength meters, such as PCFG-, Markov model-, or RNN-based [16] meters. The popular *zxcvbn* meter allows custom input of forbidden strings [22]. To use Mate1 as an example, password strength estimated by *zxcvbn*'s guess number would drop from 74, 416 to 3 if configured with relevant strings.

An in-depth understanding of a service's passwords may not always be possible (and ideally, if password breaches never occurred, would not be). Thus, the results of our paper support previous work by Schechter et al. [10,11] on popularitybased password-composition policies. While it is not inherently insecure for users' passwords to be service-specific, security risks increase when those passwords become popular among that userbase. Popularity-based policies could dynamically adjust for topics that are salient for one userbase but not another, defining a flexible approach that could be less frustrating for users than static blacklisting.

6. **REFERENCES**

- J. Bonneau and E. Shutova. Linguistic Properties of Multi-word Passphrases. In Workshop on Usable Security, pages 1–12. Springer, 2012.
- [2] C. Castelluccia et al. When Privacy Meets Security: Leveraging Personal Information for Password Cracking. *CoRR*, abs/1304.6584:1–16, 2013.
- [3] X. de Carné de Carnavalet and M. Mannan. From Very Weak to Very Strong: Analyzing Password-Strength Meters. In Symposium on Network and Distributed System Security. ISOC, 2014.
- [4] D. Florêncio et al. An Administrator's Guide to Internet Password Research. In Large Installation System Administration Conference, pages 44–61. USENIX, 2014.
- [5] D. Florêncio and C. Herley. A Large-scale Study of Web Password Habits. In *Conference on World Wide* Web, pages 657–666. ACM, 2007.
- [6] H. Habib et al. Password Creation in the Presence of Blacklists. In Workshop on Usable Security. Internet Society, 2017.
- [7] M. Jakobsson and M. Dhiman. The Benefits of Understanding Passwords. In Workshop on Hot Topics in Security. USENIX, 2012.

- [8] C. Kuo et al. Human Selection of Mnemonic Phrase-based Passwords. In Symposium on Usable Privacy and Security, pages 67–78. ACM, 2006.
- [9] M. L. Mazurek et al. Measuring Password Guessability for an Entire University. In Conference on Computer and Communications Security, pages 173–186. ACM, 2013.
- [10] S. Schechter et al. Popularity Is Everything: A New Approach to Protecting Passwords from Statistical-Guessing Attacks. In Workshop on Hot Topics in Security. USENIX, 2010.
- [11] S. M. Segreti et al. Diversify to Survive: Making Passwords Stronger with Adaptive Policies. In Symposium on Usable Privacy and Security, pages 1–12. USENIX, 2017.
- [12] R. Shay et al. A Spoonful of Sugar?: The Impact of Guidance and Feedback on Password-Creation Behavior. In *Conference on Human Factors in Computing Systems*, pages 2903–2912. ACM, 2015.
- [13] R. Shay et al. Designing Password Policies for Strength and Usability. Transactions on Information and System Security, 18(4):13:1–13:34, 2016.
- [14] B. Ur et al. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In Symposium on Usable Privacy and Security, pages 123–140. USENIX, 2015.
- [15] B. Ur et al. Do Users' Perceptions of Password Security Match Reality? In Conference on Human Factors in Computing Systems, pages 3748–3760. ACM, 2016.
- [16] B. Ur et al. Design and Evaluation of a Data-Driven Password Meter. In *Conference on Human Factors in Computing Systems*, pages 3775–3786. ACM, 2017.
- [17] R. Veras et al. Visualizing Semantics in Passwords: The Role of Dates. In Symposium on Visualization for Cyber Security, pages 88–95. ACM, 2012.
- [18] R. Veras et al. On the Semantic Patterns of Passwords and their Security Impact. In Symposium on Network and Distributed System Security. ISOC, 2014.
- [19] D. Wang et al. Targeted Online Password Guessing: An Underestimated Threat. In Conference on Computer and Communications Security, pages 1242–1254. ACM, 2016.
- [20] M. Weir et al. Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. In Conference on Computer and Communications Security, pages 162–175. ACM, 2010.
- [21] D. L. Wheeler. zxcvbn: Realistic Password Strength Estimation, Apr. 2012. https://blogs.dropbox.com/tech/2012/04/ zxcvbn-realistic-password-strength-estimation/, as of June 25, 2018.
- [22] D. L. Wheeler. zxcvbn: Low-Budget Password Strength Estimation. In USENIX Security Symposium, pages 157–173. USENIX, 2016.