Understanding the Propensity of Password Creation: Comparative Analysis of Three Language Spheres

Keika Mori
Waseda University
3-4-1 Okubo, Shinjuku, Tokyo, Japan
keika@nsl.cs.waseda.ac.jp

Yunao Zhou
Waseda University
3-4-1 Okubo, Shinjuku, Tokyo, Japan
zhouyunao@nsl.cs.waseda.ac.jp

Tatsuya Mori
Waseda University
3-4-1 Okubo, Shinjuku, Tokyo, Japan
mori@nsl.cs.waseda.ac.jp

Abstract
This work aims to understand the propensity of password creation through the lens of *language spheres*. Namely, we focus on three different languages, Chinese, English, and Japanese, each of which is rooted in different cultures. To recruit the participants who live in these language spheres, we used three crowdsourcing services, Sojump for recruiting Chinese participants, Amazon Mechanical Turk for recruiting English participants, and Lancers for recruiting Japanese participants, respectively. Our survey revealed that although how schools and companies have people be aware of password security is different among countries, over 80% of users in all three countries create passwords by themselves without using a password generator. Regarding the people who create a password by themselves, there are intrinsic differences in password creation strategies among the language spheres.

Author Keywords
password creation; online survey; comparative analysis;

ACM Classification Keywords
K.6.5 [Authentication]: Security and Protection

Introduction
Despite having several security risks such as password cracking or massive password breaches, password authen-

Table 1: What kind of computing devices do you use? (Multiple choices allowed.)

	CN	JA	EN
PC (%)	92.4	94.2	96.6
Smartphone (%)	98.3	68.8	93.1
Tablet (%)	47.1	18.6	46.6

Table 2: Do you have a degree in computer science or information security? Are you taking a degree in them currently?

	CN	JA	EN
Yes. (%)	17.6	9.88	18.1
No. (%)	81.9	90.1	81.9
Prefer not to say. (%)	0.42	0.0	0.0

Table 3: Have you received training on how to manage your passwords at work or school? (Multiple choice allowed.)

	CN	JA	EN
Yes, at school. (%)	12.4	28.1	18.1
Yes, at work. (%)	15.1	46.5	26.7
No I haven't. (%)	79.4	38.0	62.1

Table 4: Have you received any information about the risks of not doing password management? (Multiple choice allowed.)

	CN	JA	EN
Yes, at school. (%)	15.3	15.1	19.0
Yes, at work. (%)	22.3	26.0	31.0
No I haven't. (%)	70.4	64.7	56.0

tication still is a primary authentication mechanism widely used in the diverse services, because of its simplicity and user-friendliness. There have been several prior studies that analyzed large corpora of leaked passwords [5, 7, 4] with the aim of assessing the risks of password cracking. One of the commonly obtained findings was that there were a set of frequently used passwords. Also, Li et al. [5] studied the characteristics of passwords leaked from Chinese web services and reported that the observed passwords had different characteristics, compared to the ones leaked from English-based services. There has been another type of studies that aimed at performing in-depth analysis of password creation propensity through the user study approach [6, 8].

While prior studies on password creation, except Ref. [5], have studied the characteristics of passwords created by English speakers, there have been little studies that focused on the passwords created by people whose native language is *not* English. While passwords are usually composed of alphanumeric letters[1], many languages use other letters, such as Chinese characters, Korean Hangul, or Japanese Hiragana. We deem that such linguistic difference, as well as the cultural difference, may strongly affect the password creation processes and resulting password properties.

Given this background in mind, we aim to understand the propensity of password creation through the lens of language spheres. We perform an online survey to the users from three different language spheres —Chinese, Japanese, and English. To recruit the participants from each language sphere without introducing possible bias factors, we use three crowdsourcing services that operate in each country; i.e., Sojump [3] for recruiting Chinese participants, Lancers [2] for recruiting Japanese participants, and Amazon Mechanical Turk [1] for recruiting English participants. Because we want to highlight the characteristics of passwords created by people whose native languages are different, we made our questionnaires in the three languages so that we can ensure their native languages are actually the ones they reported to us. We note that this approach has some limitations, however, this approach increases the opportunities to recruit the participants in a specific language sphere.

We carried out user surveys through crowdsourcing services and compare the propensity of password creation processes such as the use of particular types of words and their languages, use of random letters, use of password generator, and management strategies of Chinese, English, and Japanese users.

Related Work

Analysis of non-English-based passwords

Li et al. [5] analyzed a large corpus of Chinese web passwords. They reported that Chinese prefer digits and include Pinyin in their passwords.

User Survey on Password Creation

Riley et al. [6] conducted a user survey to understand users' practices of password creation and storing. They asked participants' habit with the Internet, real strategies to create a password, and practices they think are safe. They revealed that users did not employ the best practice they knew. Ur et al. [8] interviewed 49 participants about their password creation strategies. In his study, he asked them to create passwords for three websites (banking, email, news website). Not only did he identify their misconceptions of strong passwords but he also found out that their opinion of the

[1] Although we are aware of some exceptions on this assumption, we omit the issue due to the space limitation.

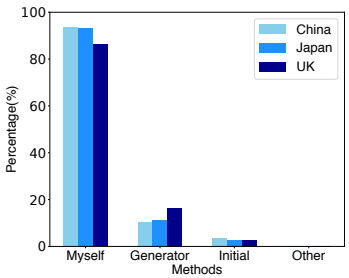

Figure 1: How do you create passwords? (Multiple choices allowed.)

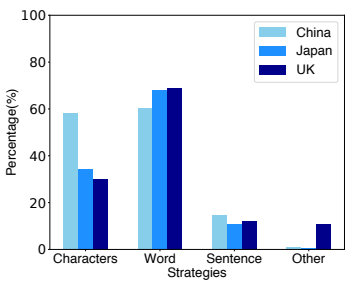

Figure 2: How do you think of passwords? (Multiple choices allowed.)

value of each account was different from that of the security community.

Methodology

Recruitment

To have participants from three language spheres, we recruited 525 participants from China through Sojump, 529 participants from Japan through Lancers and 165 participants from the UK through Amazon Mechanical Turk. We required participants that their first language should be either Chinese, Japanese, or English. For each participant, we provide 0.4–1.0 USD. The amount of payment was adjusted based on the income level of each country.

Survey Design

Our survey consists of five parts, informed consent, demographics, knowledge of password security, password creation, and password management. The survey was designed to take 10-15 minutes. Our questions were about password creation strategies and management. At the beginning of the survey, we clarify our purpose and the usage of the answers. Next, we asked participants demographic questions and their knowledge of passwords, and then asked how they create passwords. Finally, we asked them how they manage their passwords; e.g., "How many accounts/passwords do you have?" and "Do you use password management software?" Due to the space limitation, we omit the results for the last question.

Results

We received 525 responses from China, 561 from Japan, and 143 from the UK, respectively. We removed invalid responses and investigated 116, 476, and 516 responses from each country. Demographics of the participants are shown in Table 5. While the ages of participants were widespread for China and Japan, the ages of participants for the UK

Table 5: Demographics of the participants.

	Gender # participants	Age (Years) –19 / 20–29 / 30–39 / 40–49 / 50–59 / 60–
CN	F: 270 M: 205 O: 1	1.47 / 36.8 / 51.3 / 7.77 / 2.31 / 0.42 (%)
JP	F: 217 M: 298 O: 1	2.52 / 22.1 / 33.9 / 29.4 / 9.11 / 2.91 (%)
EN	F: 32 M: 83 O: 1	11.2 / 88.8 / 0.00 / 0.00 / 0.00 / 0.00 (%)

were biased toward young generation for some reasons that we were not able to control. The description of devices they use is shown in Table 1. We see that majority of the participants use PC across the three countries.

Educational Background and Knowledge about Passwords

We first detail their educational background and knowledge of password management in Table 2, 3, and 4. Table 3 presents that training of password management was most common in Japan out of the three countries; in fact, the Chi-Square test revealed that the differences were significant. Also, the factions of participants who have learned the information about the risks of weak password management had differences among the countries.

Password Creation

We first ask how a user creates passwords. Figure 1 presents the results. We see that "Thinking by themselves," which is prone to be cracked in most cases, was the most common password creation method in all three language spheres. It is interesting that the differences in the educational background did not seem to affect the convention of self-creation of passwords. The use of password generator has not been a primary method for password creation in the three countries; among them, the UK had the highest adoption rate of 16.4%. We also notice that Chinese and Japanese methods of creating passwords were similar with each other, while English ones were different, which was proved to be statistically significant with the Chi-square test (significance level of 1%).

Table 6: What words or numbers do you use? (Multiple choices allowed.)

	CN (%)	JA (%)	EN (%)
Personal words			
First name	**43.5**	**23.1**	2.90
Last name	**37.5**	**16.6**	0.0
Nickname	**36.1**	**24.3**	13.0
Birthday	**43.9**	**17.2**	5.80
Phone number	**29.4**	1.85	7.25
Credit card	6.69	1.54	0.0
Person you love	**19.0**	6.77	7.25
Important date	**38.3**	12.6	10.1
Generic words			
Famous person	**19.3**	13.2	7.25
Website name	9.29	**11.4**	1.45
Place name	11.5	8.92	11.6
Love word	**19.0**	0.615	2.90
Music word	14.5	9.85	**20.3**
Sport word	8.92	5.23	5.80
Animal word	10.4	7.08	**15.9**

Table 7: If you create passwords from base words or sentences, what language do you use? (Multiple choices allowed.)

	CN (%)	JA (%)	EN (%)
Prefer CN	72.6	0.0	0.0
Prefer JA	0.0	70.6	0.0
Prefer EN	61.2	60.7	93.6
Prefer others	1.57	6.93	12.8

Next, to those who answered that they create passwords by themselves, we asked their thinking processes. The result is shown in Figure 2. Here, we see clear differences among the language spheres. While "choosing characters randomly" was not a common strategy in English and Japanese, it was as common as "choosing words" in Chinese.

For those who answered that they create passwords from base words or numbers, we asked what word they use. Table 6 summarizes the results. Again, we see intrinsic differences among the language spheres. While generic words such as music words or animal words are preferred by English participants, personal words such as names or birthdays were preferred by Chinese and Japanese participants. We note that these differences were statistically significant with the significance level of 1%.

We also studied the differences in the use of mangling rule, which is a technique to transform a dictionary word into an obfuscated word; e.g., "Donald Trump" may be transformed into "d0n4ld 7rump" by using the "Leet" technique (which is one of the mangling rules). Adding '!' or '123' at the end of passwords is another example of mangling rule. Figure 3 shows the results. We see the differences of mangling rules among the language spheres. Japanese participants preferred to connect words or mix them. English and Chinese participants are fond of mixing words or replacing certain characters with other characters such as Leet.

Finally, we asked the language they use when creating passwords by themselves. The result is shown in Table 7. As expected, English participants mostly use English and some other languages such as Arabic, Norwegian, Tamil, and so on. In contrast, Chinese and Japanese participants use both English and their first languages. Such differences may impact the strategy of selecting effective dictionary when cracking passwords.

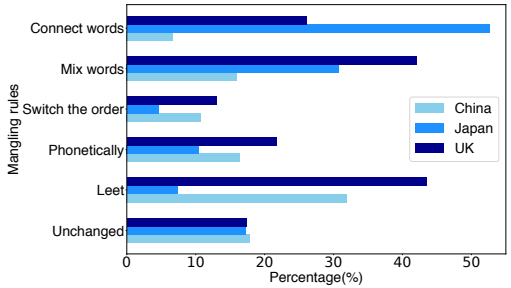

Figure 3: How do you randomize the words or the numbers? (While multiple choices were allowed for British and Japanese, single choice was allowed for Chinese participants by accident.)

Discussion and Future Work

As the ages of English participants from the UK were all under 30, the derived results may have some bias. We need to look for effective ways to recruit representative participants from each language sphere. While this work focused on the creation of passwords, extending the study to other topics such as the management of passwords is the next step toward establishing better password practices on the basis of language sphere. We also note that users' actual behavior and self-reports may have inconsistency. To address the issue, we are planning to perform user study where we ask participants to create passwords in a realistic environment. The user study will be followed by the semi-structured interview. We also plan to correlate the results of user study and analysis of large-scale password corpora.

REFERENCES

1. 2018. Amazon Mechanical Turk. (04 May 2018). `https://www.mturk.com/`.

2. 2018. Lancers. (03 May 2018). https://www.lancers.jp/.

3. 2018. Sojump. (08 May 2018). https://www.wjx.cn/.

4. keeper. 2016. The Most Common Passwords of 2016. (2016). https://keepersecurity.com/public/Most-Common-Passwords-of-2016-Keeper-Security-Study.pdf.

5. Zhigong Li, Weili Han, and Wenyuan Xu. 2014. A Large-Scale Empirical Analysis of Chinese Web Passwords. In *Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22, 2014.* 559–574. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/li_zhigong

6. Shannon Riley. 2006. Password Security: What Users Know and What They Actually Do. In *Usability News, Volume 8, Issue 1*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.597.5846&rep=rep1&type=pdf

7. splashdata. 2017. WORST PASSWORDS OF 2017 Top 100. (2017). https://s13639.pcdn.co/wp-content/uploads/2017/12/Top-100-Worst-Passwords-of-2017a.pdf.

8. Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In *Eleventh Symposium On Usable Privacy and Security, SOUPS 2015, Ottawa, Canada, July 22-24, 2015.* 123–140. https://www.usenix.org/conference/soups2015/proceedings/presentation/ur