ANALYSING PASSWORD STRENGTH FOR SOPHOMORES

Omar Saad Almousa

Computer Science Dept., Jordan University of Science and Technology, Irbid, Jordan

ABSTRACT

Passwords are ubiquitous and this will continue for long. Strong passwords are a necessity to protect sensitive information. However, users not only tend to pick weak passwords, but also reuse them over several authentication systems. The existence of weak passwords in a system not only jeopardize that system, but also other systems with overlapping users because of password reuse phenomena. Investigating users' behaviour in password creation leads to finding ways to avoid weak passwords. One aspect of that is to study the very passwords. In this study we analyse 662 passwords created by fresh students in our faculty. The students picked their passwords to authenticate themselves to a platform for programming practice and assignment solving. Our analysis relied on basic structural parameters such as password length, constructing characters, and entropy. To that end, we coined two definitions for weak and strong passwords. One is alphabet-based, and the other is entropy based. Accordingly, we found that majority of students do not tend to create strong passwords. We believe that this is due to the lack of enforcement of a strong password policy.

KEYWORDS

passwords, analysis, weak password, strong password.

1. INTRODUCTION

Passwords remain the most dominant authentication type despite the advancement in other types [1]. This domination over other authentication types is expected to continue [2]. This is basically due to their low-cost and usability [3].

Service providers, and thus users, rely on passwords to protect the privacy of their sensitive data on different applications [4]. Applications that vary from social networking, business accounts, personal emails, online banking, and e-shopping. Having the importance of such sensitive data in mind, one may rightly assume that users pick strong passwords to secure their accounts. However, practice has proven many times that this is not the case. Many users still pick weak passwords unless forced by a password policy. Picking a weak password is mainly due to the fact that users prefer easy to remember passwords [5].

The threat of creating weak passwords is persistent even though a robust password policy is enforced. This is because cyber threats rely not only on the poor quality of passwords, but also on other factors of users' behavior. From these behaviors that demolish security efforts we have passwords lending, using personal information in creating passwords, and password reuse. The authors of [6] statistically found that 53% of the users reuse the same passwords through several systems. They also found that 21% of them lend their passwords, and 17% wrote down their

David C. Wyld et al. (Eds): CCSIT, NLPCL, AISC, ITE, NCWMC, DaKM, BIGML, SIPP, SOEN, PDCTA – 2024
pp. 01-11, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141701

passwords in order to remember them once needed.

A crucial issue of secure passwords is reusing the same password for multiple accounts. Therefore, leaking a password from one account entails that all the other accounts are accessible as well [7]. Additionally, the authors of [8] found that almost 60% of all users use the same password through multiple systems, whereas [9] noticed that most users are not aware of the best practices in creating strong passwords.

The existence of weak password in a system is not only harmful for the account owner [10]. It is also risky for the whole system. Attackers may exploit a weak password to jeopardize the whole system [11]. Therefore, password strengthening procedures must be rather than being left to users' responsibility.

Most of the systems enforce a strong policy to pick a password. They enforce users to create complicated and diverse passwords. Such policies although sound beneficiary, they have certain drawbacks. They make it cumbersome to users for users to memorize and thus retrieve their passwords [12][13][14]. This is becoming an issue especially with the vast amount of accounts for each user [14]. This explains the aforementioned habits of users such as password reuse and writing down passwords.

In this study, we aim at assessing the level of security awareness for our fresh students at the faculty. We assume that security awareness of students is reflected by their password creation. The rest of this paper is organized as follows: in Section 2 we briefly discuss some related studies. We explain our methodology in section 3. We then show our results in section 4, and in section 5 we give some recommendations. Finally, we conclude in section 6.

2. RELATED STUDIES

Several researchers studied passwords lists for different purposes. These studies aimed at either understanding users' behavior in password creation or developing password metrics. To that end, some researchers used huge lists of passwords such as [15] and [16]. Others used relatively smaller ones such as [17].

The authors of [11] investigated the behaviors of users when creating password and how it affects its strength. Several attempts to understand user-chosen passwords have been performed by researches in several studies. For example, the researchers of [18] detailed the main reasons behind the problems related to text-based authentication. They consider that creating memorable passwords is a perplexing matter. Therefore, they proposed a method based on the psychological dynamics of the user to encourage creating memorable passwords and yet strong.

Analysis of users' habits over a Chinese network was performed by the authors of [19]. They studied the strength of passwords through a comprehensive analysis of password different parameters such as password length, characters type, and other parameters. They reported the existence of some repeated patterns in the analyzed passwords.

3. METHODOLOGY

Now we explain our data set source and how we processed it. We also explain the parameters we extract from our data set.

3.1. Data Set

We use an obsolete password list for fresh students in computer and information technology faculty at the Jordan university of science and technology. The students created their accounts and thus passwords on a programming system that was developed in the university to serve a programming course. Students were supposed to use this system during the whole semester, therefore we can assume that they picked their passwords realistically. The system was deployed for two semesters in the university before replaced by another system and thus the "obsoleteness". This obsoleteness is what encouraged us to study its password list.

The password list consists of 671 anonymized passwords. the anonymization is performed by simply deleting user ids from the list. After anonymization, we performed data cleansing by eliminating default passwords and blanks. The remaining dataset consisted of 662 passwords.

3.2. Data Processing

For the total of 662 passwords in the list, we advised an algorithm that reads each password and computes a list of simple parameters such as the length of a password, the existence of capital letters, and similar parameters. We explain the complete list in Table 1.

Table 1: Password Parameters Explained

Parameter	**Type**	**Explanation**
Length	Integer	The length of a password
hasSmall	Boolean	If it has small letters
#Small	Integer	Number of small letters
hasCaps	Boolean	If it has capital letters
#Caps	Integer	Number of capital letters
has SpecialChars	Boolean	If it has special carachters
#SpecialChars	Integer	Number of special carachters
hasDigits	Boolean	If it has digits
#digits	Integer	Number of digits
hasSpaces	Boolean	If it has spaces
#spaces	Integer	Number of spaces

3.3. Password Entropy

Our algorithm also computes an additional parameter. This additional parameter is nothing other than the infamous entropy. An entropy of a password measures attacker's uncertainty when trying to find a secret (password) [4]. The entropy of a password (H) is computed based on its length (L) and the alphabet size (N) using formula (1). Entropy is measured in bits.

$$H = \log_2 N^L \quad (1)$$

For example, a password of seven characters (L= 7) and contains only small letters (A=26) has the entropy of 32.9 bits. Whereas a password with the same length but with mixed capital and small letters has an entropy of 39.9 bits.

4. RESULTS AND DISCUSSION

In this section we present our results.

4.1. Alphabet Distribution

For the lack of a password policy, students picked their passwords without any restrictions. Therefore, 29.3% of the passwords were containing only digits.

In Table 2 we display the distribution of different password classes. We classify the passwords on the basis of their alphabets. By password alphabet we mean the set from which a user (student) picked his/her password. The first column in this table gives the alphabet name. Each row represents a class (set) of passwords. Next we clarify each of those classes.

Table 2: Alphabet distribution

class#	Class alphabet	Size	Percentage	Count
1	Digits Only	10	29.3%	194
2	One case letters	26	1.8%	12
3	One case letters+ digits	36	15.4%	102
4	Digits + symbols	42	0.3%	2
5	Two case letters	52	0.5%	3
6	One case letters+ symbols	58	0.3%	2
7	Two case letters+ digits	62	10.6%	70
8	One case letters+ digits + symbols	68	3.6%	24
9	Two case letters+ symbols	84	0.3%	2
10	Two case letters+ digits + symbols	94	37.9%	251

The first column is simply a numbering for classes. We have 10 different classes. The second column describes the alphabet of each class. “Digits Only” - as name indicates – means that the alphabet consists of number digits only, i.e., [0-9]. For example, the password “123456”. We have also “One case letters” that means the alphabet is either small letters [a-z] or capital letters [A-Z] exclusively but never both. For instance, “omar”, and “OPPO”. “One case letter+ digits” includes the passwords that contain digits and either small or capital letters. For example, “ali5” and “7SAM”. “Digits + symbols” contains passwords with digits and printable symbols in ASCII table other than letters or digits. Symbols here are also called “special characters” elsewhere. For instance, “678$” and “2%3”. The fifth class: “Two case letters” includes passwords with both cases present (small and capital letters), i.e., passwords belong to the language {a-z, A-Z}+. For example, a password like “Omar” belongs to “Two case letters”, where “saad” or “ALMOUSA” belong to “One case letters”.

The sixth class: “One case letters+ symbols” contains passwords that has one case letters (small or capital exclusively) and digits (at least one). For example, “ALAN%” and “@sami”. The next class “Two case letters+ digits” includes passwords with small and capital letters along with a digit at least. For example, “Ali5”, and “7sAm”. The eighth class “One case letters+ digits + symbols” contains passwords that have only one case of letters along with digits and symbols (at least one of each). Passwords like “&jason8” and “88HALA*” are examples of this class. The next class #9 that is “Two case letters+ symbols” represents the passwords with letters from both cases merged with a symbol or more. For instance, “Ahmad_)”, “Sarah&”, and “U$er”. The last class: “Two case letters+ digits + symbols” includes the passwords that have it all!, i.e., passwords with small and capital letters, digits and symbols. For example “Ja%9” and “$uH49_ak”.

The third column in Table 2 gives the alphabet size. Recall that we have: 10 digits, 26 small letters, 26 capital letters, and 32 symbols in ASCII table. The sum of the previous sets is 94 and thus the last row of the table.

The last two columns of Table 2 show the percentage and count of users for each of the 10 classes. As one can notice, 37.9% of the students freely picked a password with "Two case letters + digits + symbols". Recall that this was not a forced policy on them. This indicates that they have committed themselves to pick a strong password as recommended by several web services (including the university web site). On the other hand, 29.3% of them created a "digits only" password. This may be due to their belief that no one will be interested in hacking into their programming training tool. We noticed that several passwords of this category have a mobile number pattern. This also points to the students' willingness to reuse a well-remembered easy password rather than creating a memory challenging password.

Now we classify the passwords into two categories based on their alphabets. Recall that our alphabets are constructed from four different sets, namely: digits, small-letters, capital letters, and special characters (symbols). We call the two categories "alphabet-Weak" and "alphabet-Strong". For abbreviation, α-weak and α-strong. An α-weak password is a password that is generated from at most two sets of characters. An α-strong password is a password that is generated from at least three sets of characters. For example, a password that is constructed from small letters and digits is an α-weak password. Whereas a password that is constructed from small letters, digits, and special characters is an α-strong password.

Based on that, we have a new distribution for our passwords that we call α-distribution shown in Table 3. Note that Table 3 is an abstraction for Table 2, i.e., rows 1-6 in Table 2 corresponds to row i in Table 3, and rows 7-10 in Table 2 correspond to row ii in Table 3.

Table 3: α-distribution

#	**α-DISTRIBUTION**	**Percentage**	**Count**
i	α-weak password	47.3%	313
ii	α-strong password	52.7%	349

From Table 3, we notice that our sophomores pick strong passwords more often than weak passwords. However, the percentage of weak passwords is not ignorable. A 47.3% is almost a half. Keep in mind that there is password policy that directed student to strengthen their passwords. Additionally, we believe that many students had not considered their programming tasks and assignments as a sensitive information that must be protected by a strong password.

4.2. Password Length Distribution

Now we present and discuss the distribution of passwords on basis of their lengths. As shown in Figure 1, the passwords have at least 3 characters and at most 24. Only 2 passwords have the minimum length that is 3 characters, and another 2 have the maximum length of 24 characters. The average password length is 9.68 characters, where the mode and the median equal 9.

Moreover, 131 passwords were of 9 characters' length. This is a positive indication of students' password creation habits. Especially if combined with the fact that more than 73.5% of them picked a password of 8 and more characters.

4.3. Small Letters Frequency

Figure 2 shows that frequency of passwords that contains at least one small letter. It is obvious that 205 passwords have no small letters with a percentage of 33.0%. for all passwords, the average of small letters is 3.3, where the maximum number of small letters in a password is 15. The mode is 4 and the median is 2. However, if we exclude the passwords that have no small letters, the median becomes 5.

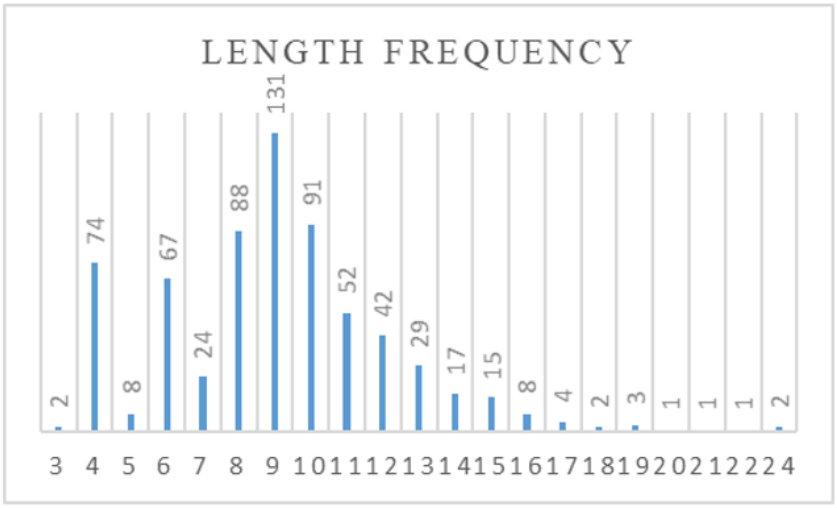


FIGURE 1: PASSWORD LENGTH FREQUENCY

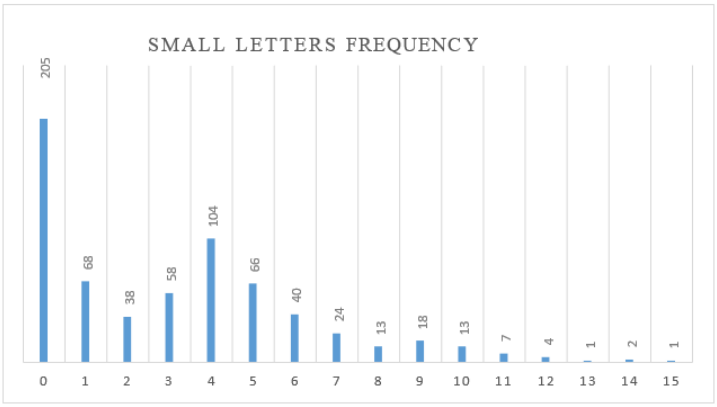


FIGURE 2: SMALL LETTERS FREQUENCY

4.4. Capital Letters Frequency

The presence of capital letter is far less than of that of small letters. As Figure 3 depicts, almost half (49.4%) the passwords have no capital letters. Moreover, most (89.3%) of the other half have only one capital letter. However, we can say that more than half of the students use at least one capital letter in their passwords.

The maximum number of used capital letters in a single password is 11, with an average of 0.67 per password. The mode and median equal 1 regardless the inclusion of the passwords with no capital letters.

4.5. Special Characters' Frequency

In Figure 4 we present the frequencies of special characters in our passwords. A majority of 57.5% (381) passwords have no special characters. Less than 7% contain 2 or more special characters. The maximum number of special characters in a password is 4, this occurred only in 3 passwords. this shows the scarce use of special characters by our somphores.

On average, a password has 0.52 special characters, with 0 as a mode and a median. If we exclude the passwords with no special characters, the median and mode become 1.

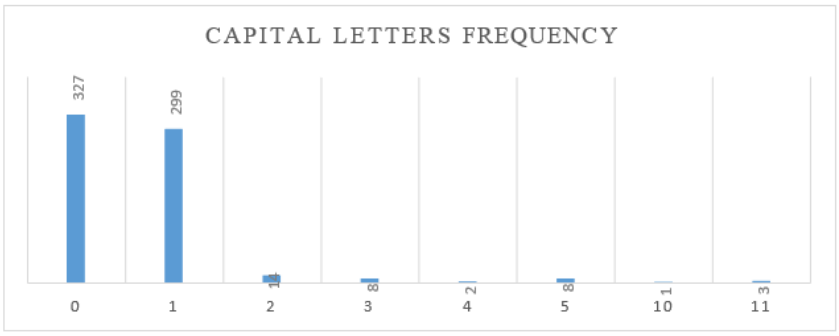


FIGURE 3: CAPITAL LETTERS FREQUENCY

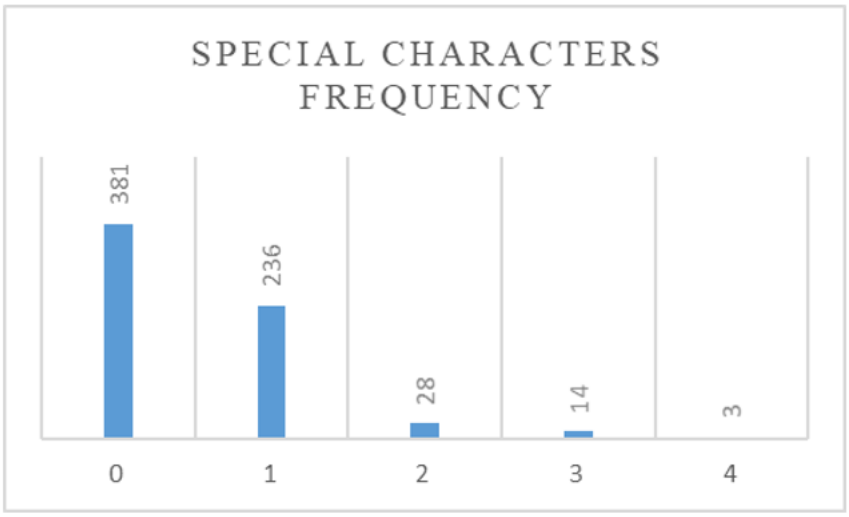


FIGURE 4: SPECIAL CHARACTERS FREQUENCY

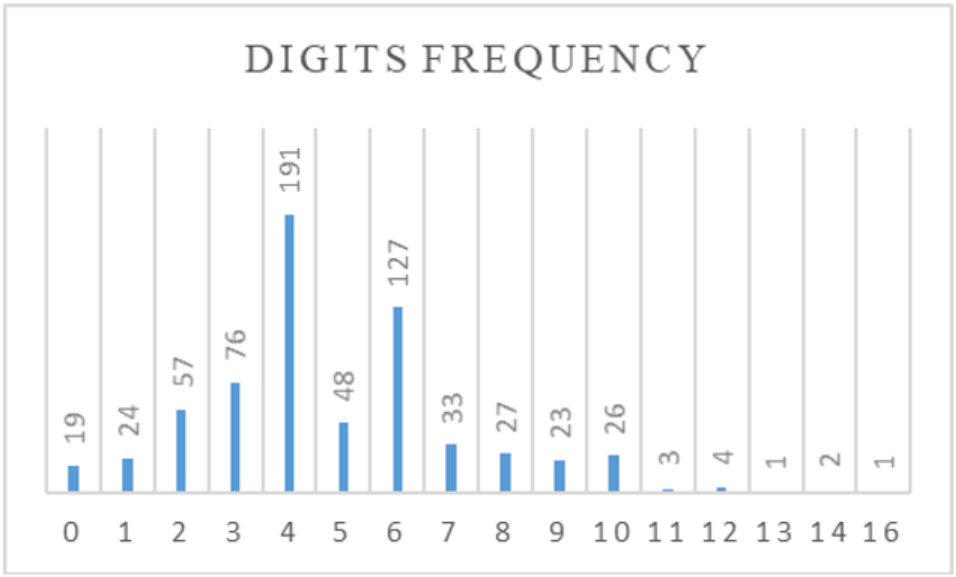


FIGURE 5: DIGITS FREQUENCY

4.6. Digits' Frequency

Figure 5 shows the digits' frequency in study passwords. Unlike special characters and capital letters, the vast majority of passwords contain digits. In fact, as we discussed earlier, a third of the passwords contain only digits. Only 19 passwords have no digits (almost 3%).

The average number of digits per password is 4.8, with a mode and median of 4 digits in a password. This is regardless of the passwords with no digits.

4.7. Entropy Related Results

Our sophomores' passwords have an average of 50 bits of entropy. The maximum is 118 bits and the minimum is 30 bits. Figure 6 depicts the entropy frequency for all passwords.

For a better presentation, we use NIST [4] cut point of 80 bits of entropy to put our passwords into two classes that we call "Entropy-weak" and "Entropy-Strong" passwords. For abbreviation we use "η-weak" and "η-strong" correspondingly. An η-weak password is a password with less than 80 bits of entropy. Whereas, an η-strong password is a password with 80 bits of entropy and more. Figure 7 shows the frequencies on the new distribution that we call η-distribution.

As the Figure 6 shows, we have over 90% of weak passwords. we carry back our previous explanation for this high percentage of weak passwords that is two folded. First, the lack of a password policy, and second, the students' disregarding of a password protecting their programming tasks.

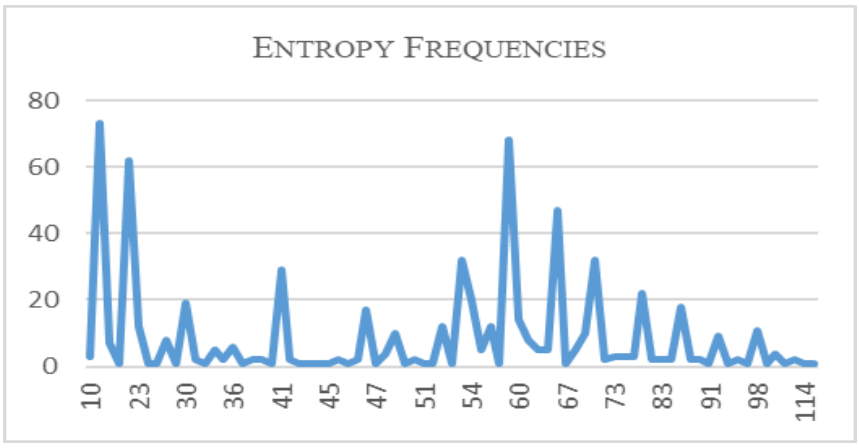
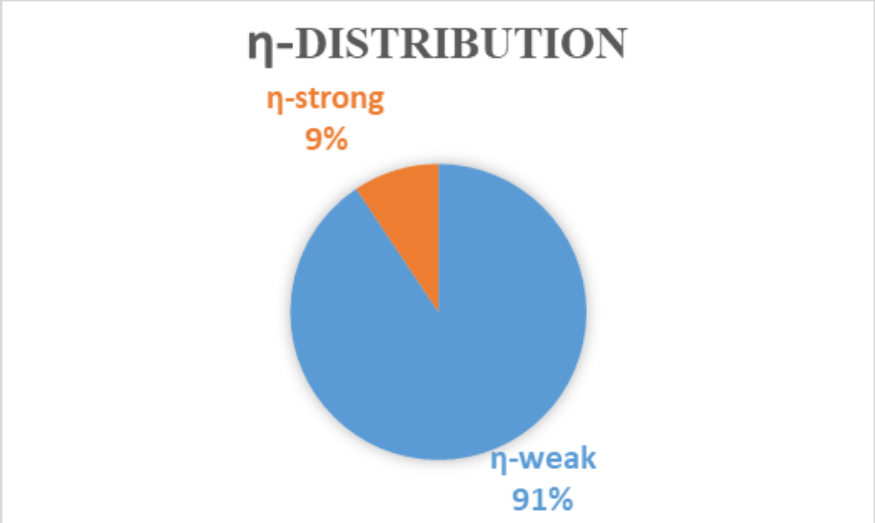


FIGURE 6: ENTROPY FREQUENCY



5. RECOMMENDATIONS

FIGURE 7 η-DISTRIBUTION

First of all, we recommend the enforcement of a strong password policy no matter the level of information sensitivity of a system. If a system will use password authentication, then the existence of a password policy is a must. This is because users tend to reuse passwords all the time. Thus, the existence of reused passwords that protect sensitive data on other accounts side by side with weak passwords in the absence of a password policy is of great risk.

We also recommend following a secure method to store passwords. Storing passwords in plain format must be avoided even prevented. Hashing salted and peppered password might be a good start.

Finally, we recommend system designers and developers to endorse a secure design approach in the development life cycle. This will prevent the aforementioned issues.

6. CONCLUSION

We collected 667 passwords created by fresh students in the faculty of computer and information technology at the Jordan university of science and technology. The students created their passwords on a programming platform as a requirement in a programming course. The programming platform was developed by the university. The system is now obsolete, and thus we see no much risk of studying its passwords.

After data cleansing, we end up with a list of 662 passwords. We analysed the list by password length, the existence and frequencies of lower-case letters, upper-case letters, digits, and special characters. We also compute the entropy for each password in the list and analysed the results.

We coined two definitions for weak and strong passwords. One is alphabet-based that we called α-distribution that classify passwords into α-strong and α-weak according to the number of character sets a password is generated from. The other definition is entropy based that we called η-distribution that classify passwords into η-strong and η-weak according to the NIST cut point of 80 bits of entropy for strong passwords.

Accordingly, we found that majority of students do not tend to create strong passwords on both of our definitions. More precisely, almost a half of them pick an α-weak password. While over 90% of them pick an η-weak password. We believe that this is due to the lack of strong password policy and the students' consideration of their programming tasks as non-sensitive information that needs to lay behind a strong password. We presented some recommendations to address this issue.

For future work, we plan to conduct a more rigorous investigation on passwords' quality. Such an investigation can be performed using a suite of metrics and tools.

REFERENCES

[1] Melicher, W., Ur, B., Segreti, S. M., Komanduri, S., Bauer, L., Christin, N., & Cranor, L. F. (2016). Fast, lean, and accurate: Modeling password guessability using neural networks. In 25th USENIX Security Symposium (USENIX Security 16) (pp. 175-191).

[2] K. Sivapriya, L. R. Deepthi, "Password strength analyzer using segmentation algorithms," 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 605-611, 2020.

[3] M. Dell'Amico, P. Michiardi, & Y. Roudier. (2010, March). Password strength: An empirical analysis. In 2010 Proceedings IEEE INFOCOM (pp. 1-9). IEEE.

[4] W. Burr, D. F. Dodson, W. Polk, "Electronic authentication guideline," NIST Special Pub 800- 63, 2006.

[5] D. Florencio and C. Herley. 2007. A large-scale study of web password habits. In Proceedings of the 16th international conference on World Wide Web (WWW '07). Association for Computing Machinery, New York, NY, USA, 657–666.

[6] K. Solic, H. Ocevcic, D. Blazevic, "Survey on password quality and confidentiality," Automatika, Vol. 56, No. 1, pp. 69-75, 2015.

[7] M. Yildirim, I. Mackie, "Encouraging users to improve password security and memorability," International Journal of Information Security, Vol. 18, No. 6, pp. 741–759, 2019.

[8] M. Grimes, J. Proudfoot, P. Benjamin, "Improving password cybersecurity through inexpensive and minimally invasive means: detecting and deterring password reuse through keystroke-dynamics monitoring and just-in-time fear appeals," Information Technology for Development, Vol. 20, No. 2, pp. 196-213, 2013.

[9] E. F. Gehringer, "Choosing passwords: security and human factors," IEEE 2002 International Symposium on Technology and Society (ISTAS'02), Raleigh, NC, USA, pp. 369-373, 2002.

[10] Dastane DO. The effect of bad password habits on personal data breach. International Journal of

Emerging Trends in Engineering Research. 2020 Oct 22;8(10).

[11] J. Darby, J. Phelan, P. Sholander, B. Smith, A. Walter and G. Wyss, "Evidence-Based Techniques for Evaluating Cyber Protection Systems for Critical Infrastructures," MILCOM 2006 - 2006 IEEE Military Communications conference, Washington, DC, USA, 2006, pp. 1-10, doi: 10.1109/MILCOM.2006.302504

[12] S. Komanduri, R. Shay, P. G. Kelley, M. Mazurek, L. Bauer, N. Christin, L. F. Cranor, S. Egelman, "Of passwords and people: measuring the effect of password-composition policies," The ACM CHI Conference on Human Factors in Computing Systems, Vancouver, Canada, 2011.

[13] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, "Correct horse battery staple: exploring the usability of system-assigned passphrases," The eighth Symposium on Usable Privacy and Security (SOUPS), Washington DC, USA, 2012.

[14] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, "Encountering stronger password requirements: user attitudes and behaviors," The sixth Symposium on Usable Privacy and Security (SOUPS), Redmon, USA, 2010.

[15] J. Bonneau, "The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords," 2012 IEEE Symposium on Security and Privacy, San Francisco, CA, USA, 2012, pp. 538-552, doi: 10.1109/SP.2012.49.

[16] O.S.. Almousa, H. Migdady, (2020). Introducing a machine learning password metric based on EFKM clustering algorithm. International Journal of Communication Networks and Information Security, 12(3), 394-400.

[17] M. Awad, Z. Al-Qudah, S. Idwan, & A. Jallad, (2016, December). Password security: Password behavior analysis at a small university. In 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA) (pp. 1-4). IEEE.

[18] Z. Zheng, H. Cheng, Z. Zhang, Y. Zhao, P. Wang, "An alternative method for understanding user-chosen passwords," Security and Communication Networks, Vol. 2018, p.p. 1-12, 2018.

[19] K. Siau, Y. Ma, N. Twyman, "Cybersecurity: personal information and password setup," MWAIS Conference, St. Louis, Missouri, USA, pp. 1-6, 2018.