ISBN 978-981-11-7861-0
Proceedings of 2018 the 8th International Workshop on Computer Science and Engineering (WCSE 2018)
Bangkok, 28-30 June, 2018, pp. 521-525
doi: 10.18178/wcse.2018.06.089

Do Students with High Grade Point Averages Create Better Passwords?

Josephine M. de la Cuesta[1, 2 +]

[1] Asia Pacific College, Philippines

[2] College of Computer Studies, De La Salle University, Philippines

Abstract. Technological advances made possible different ways to secure data and information. Passwords though remain to be the popular choice to authenticate users for its ease of use and, for the organization, it is inexpensive to administer. Studies show however that passwords have limitations, which is addressed by defining stricter and more complex password creation policies. User reactions to such policies vary, from delaying conformance to opting for ease and convenience when creating passwords, with little consideration for the resulting security risks. Studies have likewise been made regarding the factors that affect password creation, such factors mostly being behavioral in nature. This study intends to determine if Asia Pacific College students' academic performance positively or negatively impacts password creation. In compliance with the National Institute of Standards and Technology's (NIST) Digital Identity Guidelines, the password of each student logging in to the information system was captured and recorded over a duration that included peak periods of system activity. These passwords were hashed and compared with a corpus of 320 million breached passwords, to identify which were compromised and not compromised. To this list, the grade point average (GPA) of each student was added. The resulting data set, with student information, password status and GPA data were analyzed. Statistical analysis shows that there is a significant difference between the compromised and not compromised group in terms of the mean of the GPA. Academic performance may positively impact password creation. The study can be expanded to further test this conclusion.

Keywords: password authentication, breached passwords, password strength, password creation

1. Introduction

Developments in information and communications technology (ICT) have made available different tools that can be used in day-to-day activities. From students who make use of tools for school work to office workers who take advantage of available tools to be more efficient and productive, ICT is ever present and has made possible almost every task imaginable. Improved connectivity gave rise to the Internet, where everyone and every device is connected or online. Online meant that one can capture data, process it into information, store, then share the information. Online electronic devices became permanent fixtures in the workplace across industries, including education, used by children and adults alike. Over the years, developments in software have likewise made available different tools that may be used in support of basic office tasks, training and education, finance and accounting, graphics design, animation, and even games, by both individual and team workers.

Electronic mail, short message sending, and shared directories, to name a few, enable workers to collaborate and share their ideas and knowledge towards achieving a goal [1]. With the many areas for collaboration in organizations, more and more tools are being developed although the question is whether organizations are able to take advantage of such tools. Equally important is the question of whether organizations are able to safeguard the data and information captured and generated.

+ Corresponding author. Tel.: +639176319183; fax: +6328530535
E-mail address: joannec@apc.edu.ph

A key area of consideration is access to these tools and systems. Advancements in technology have made possible different methods for securing tools and systems, although, passwords remain to be the most used option for securing access [2]. Passwords are theoretically simple to create, and from an information technology standpoint, easier to manage; and, since passwords are the common first choice for protecting accounts, it remains to be a main vulnerability and is prone to hacking [3]. System administrators and web site owners are tasked to ensure that the systems they are administering provide the necessary guidelines for users with regard to creating passwords, but users should also follow these guidelines correctly [4] [5]. A review of literature shows many factors affect the creation of passwords, including user behaviour [6]. One of the issues is that most users are more inclined to go for convenience – easy to create and recall – rather than ensuring their accounts (and therefore, the tool or system) are secure [7].

In Asia Pacific College (APC), several tools are in use to aid both faculty and students in day-to-day operations and academic work. The information system, learning management system and electronic mail system, to name a few, are web-based and accessible both from within and without APC. The Information Technology Resource Office (ITRO) is tasked to ensure these systems are secure. One of the ITRO's tasks is to routinely check the passwords of users against known compromised passwords, per the National Institute of Standards and Technology (NIST) Digital Identity Guidelines [2].

These guidelines, released in June 2017, include updated standards for managing and accepting user passwords with one of the new guidelines stating that user passwords should be compared to a list of known breached/compromised passwords. A breached or compromised password refers to a password that is included in the list (such as the list curated by security researcher Troy Hunt, https://haveibeenpwned.com) of known passwords, compiled from various data 'breaches.' These passwords (e.g. password12345, mypassword123) are considered compromised as these are the first passwords that hackers will try, before using more complicated strategies. Using a compromised password will make a user's account easier to compromise.

Using a corpus of 320 million breached passwords, obtained through security researcher Troy Hunt, it was determined that a significant number of students in APC were using compromised passwords.

Several studies have been made on the different behavioural factors that affect password creation. Given the available data as a result of the routine check of ITRO, and since there is limited literature available, this study aims to evaluate if a student's academic rating has likewise an impact on password creation. The results of the study may be used as a basis for defining APC's policy on information security. The paper is organized as follows: the research design and methodology will be discussed followed by the analysis of the results. Finally, the conclusion and recommendations for further research will be presented.

2. Literature Review

Technology advances have made available different ways to secure systems. Aside from using *something you know* (e.g. passwords or PINs), different options are available for *something you have* (e.g. physical devices), or a combination of something *you have* and *something you are* (e.g. biometrics) [2] [8]. However, studies show that passwords remain to be the popular choice owing to its being easy to use and manage [8] [9]. This is despite the fact that passwords have limitations, which organizations addressed by defining more complicated and stricter requirements for password creation [7].

These additional requirements are to ensure the password is both easy to remember and hard to guess [8]. Easy to remember passwords resulted in mostly numeric passwords (which are easier to crack as opposed to alphanumeric passwords) [3], dictionary words, or the most obvious (e.g. 123456, password, qwerty) [5]. To create hard to guess passwords, password composition policies were defined (e.g. password must be a combination of numbers, letters, symbols, etc.) [6]. These policies however had negative impact on the productivity of the employee [10]. Additionally, these complex password policies drove users to write down their 'hard to guess' passwords or, use the same password for their different accounts [6] [8] [11] or, share their passwords with others [8] [11]. In some instances, users resist complying with, or hold off on implementing, the password policies (without consideration for, in effect, being the weak link in security) [10]. If users are 'forced' to comply with policies, they find their way around implementing it [9].

Technology, obviously, is not the only factor. User behaviour, getting the people to properly implement such policies, is a major consideration [4]. Studies show that with proper guidance on password creation, users are able to create strong passwords [12].

Studies have also been conducted to evaluate the different factors that impact password creation, as well as different organizations' password policies and use. Aside from behaviour and sentiments [10], age, gender, company policies and academic preparation [6] [7] are considered as contributing factors to password creation. It was also posited that people from different geographic locations may create passwords differently [13].

These different studies show that many factors influence users in password creation, albeit majority of the studies are about user behaviour. While one study considered academic preparation, testing Computer Science or Information Technology students' behaviour, it did not extend towards the same students' academic performance. This paper hopes to contribute to existing literature by presenting the results of the study on academic performance as a factor that impact password creation.

3. Research Design and Methodology

To facilitate checking the passwords of users in APC, the corpus of 320 million breached passwords was obtained through security researcher Troy Hunt (https://haveibeenpwned.com/Passwords), an Australian Microsoft Regional Director and Microsoft Most Valuable Professional for Developer Security. Hunt released this list of breached passwords, which contains only Secure Hash Algorithm 1 (SHA-1) hashes of plaintext passwords, after the NIST 2017 guidelines was published. SHA-1 processes a message to produce a condensed version called a message digest and enables the ascertainment of its integrity [15].

As part of its security maintenance activities, the development team in ITRO initiated steps that would enable the comparison of the passwords of each user logging in to the APC Information System with the corpus of breached passwords. This special software infrastructure, composed of short PHP codes and an SQLite3 database, was left running and recording results for three (3) months, capturing peak periods of system activity during midterms and finals weeks. For the purpose of this study, only the students' accounts were recorded. Since the corpus of breached passwords contained only SHA-1 hashes, the passwords of the users were similarly hashed before searching through the corpus for a match. A special database table was created that recorded whether the users' passwords were compromised or not compromised. Users are not recorded twice. In the event that a previously recorded user logs in again, the software simply updates the user's existing record. At the end of three (3) months, 91.52% (1,252 individual students) of currently enrolled College students were logged by the software. It was expected that a few students would not be recorded since they may not need to log in to the information system. Grade Point Average (GPA) data was next added to the data set of 1,252 students with password status (TRUE – compromised password, FALSE – not compromised password). The resulting data set, with student information, compromised password status, and GPA data was then analysed.

The data set was divided into two, one set contains all FALSE and another set all TRUE, and descriptive statistics were obtained (refer to figure 1). The mean refers to the average of the data set; standard deviation evaluates how distributed the numbers are from the center of the data set; and, P-value refers to the probability of finding the observed results when the null hypothesis is true [16]. The F-test was used to check whether the two sets have equal variances and to determine the type of t-test to be used. Furthermore, since the data collected is greater than 30 [14], it is assumed that the values are normally distributed which is another important assumption in order to use t-test. The standard significance value of 0.05 [16] was used.

4. Discussion of Results

The data was analysed using z-test to determine if the students with the higher GPA tend to have passwords which are not compromised. At 0.05 level of significance, the result of the test is -2.183498 (refer to figure 1), which falls within the critical region. This means we reject the null hypothesis that there is no significant difference in the means of each sample. There is a significant difference between the means of each group (TRUE = 2.527581 and FALSE = 2.601418). The mean of the average GPA of TRUE

(compromised passwords) is less than the mean of the average GPA of FALSE (non-compromised passwords). This indicates that academic performance does impact password creation.

TRUE	
Average Grade	
Mean	2.527581395
Standard Error	0.030801911
Median	2.46
Mode	2.22
Standard Deviation	0.45164467
Sample Variance	0.203982908
Kurtosis	-0.540914979
Skewness	0.354805056
Range	2.22
Minimum	1.42
Maximum	3.64
Sum	543.43
Count	215

FALSE	
Average Grade	
Mean	2.601417551
Standard Error	0.013954674
Median	2.56
Mode	2.24
Standard Deviation	0.449375173
Sample Variance	0.201938047
Kurtosis	-0.46939118
Skewness	0.291001522
Range	2.27
Minimum	1.5
Maximum	3.77
Sum	2697.67
Count	1037

	TRUE	FALSE
Mean	2.527581	2.601418
Known Variance	0.203983	0.201938
Observations	215.000000	1037.000000
Hypothesized Mean Difference	0.000000	
z	-2.183498	
P(Z<=z) one-tail	0.014500	
z Critical one-tail	1.644854	
P(Z<=z) two-tail	0.028999	
z Critical two-tail	1.959964	

Fig. 1: Descriptive statistics, z-test.

5. Conclusion and Recommendations

This study was done using one (1) password per user, in consideration of data privacy as passwords are sensitive data. Based on the statistical analysis of the data, there is a possibility that a student's academic performance can influence password creation. The study can be expanded to include the users' other passwords, to further test that academic performance positively impacts password creation. Additionally, the correlation of the GPA with password creation may be determined, using a larger number of users and longer period of capturing and recording data.

6. Acknowledgements

The author would like to acknowledge Dr. Jaime Cempron of De La Salle University for whom the concept paper (submitted as a course requirement) this research was based on was developed. Additionally, the author sends her heartfelt thanks to Asia Pacific College for the generous support in the development of this research. Lastly, expressions of gratitude are extended to Mr. JV Roig and Ms. Roselle Wednesday Gardon, both from Asia Pacific College, for their invaluable assistance in gathering and analysing the data used in the research.

7. References

[1] Pillet, J., Carillo, K. (2016). Email-free collaboration: An exploratory study on the formation of new work habits among knowledge workers. International Journal of Information Management, 36(1), 113–125. https://doi.org/10.1016/j.ijinfomgt.2015.11.001

[2] Grassi, P., Perlner, R., Fenton, J., Burr, W., Richer, J. (2017). Digital Identity Guidelines: Authentication and Lifecycle Management. NIST Special Publication 800-63b, 1-79. http://doi.org/10.6028/NIST.SP.800-63b

[3] Shen, C., Yu, T., Xu, H., Yang, G., & Guan, X. (2016). User practice in password security: An empirical study of real-life passwords in the wild. Computers and Security, 61, 130–141. https://doi.org/10.1016/j.cose.2016.05.007

[4] Furnell, S. (2014). Password practices on leading websites - Revisited. Computer Fraud and Security, 2014(12), 5–11. https://doi.org/10.1016/S1361-3723(14)70555-X

[5] Furnell, S., & Esmael, R. (2017). Evaluating the effect of guidance and feedback upon password compliance. Computer Fraud and Security, 2017(1), 5–10. https://doi.org/10.1016/S1361-3723(17)30005-2

[6] Hussain, T., Atta, K., Bawany, N. Z., & Qamar, T. (2018). Passwords and User Behavior. Journal of Computers, 13(6), 692–704. https://doi.org/10.17706/jcp.13.6.692-704

[7] Mujeye, S., Levy, Y., Mattord, H., Li, W. (2016). Empirical results of an experimental study on the role of password strength and cognitive load on employee productivity. Online Journal of Applied Knowledge Management, 4(1), 99-116.

[8] Haque, S., Wright, M., Scielzo, S. (2014). Hierarchy of users' web passwords: Perceptions, practices and susceptibilities. International Journal of Human-Computer Studies, 72(2014), 860-874. http://dx.doi.org/10.1016/j.ijhcs.2014.07.007

[9] Wang, D., He, D., Cheng, H., Wang, P. (2016). fuzzyPSM: A New Password Strength Meter Using Fuzzy Probabilistic Context-Free Grammars, Proceedings of the 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, 595-606. http://dx.doi.org/10.1109/DSN.2016.60

[10] Belanger, F., Collignon, S., Enget, K., Negangard, E. (2016). Determinants of early conformance with information security policies, Information & Management, 54(2017), 887-901. http://dx.doi.org/10.1016/j.im.2017.01.003

[11] Jenkins, J., Grimes, M., Proudfoot, J., Lowry, P. (2014). 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, 20(2), 196-213. http://dx.doi.org/10.1080/02681102.2013.814040

[12] Furnell, S., Khern-am-nuai, W., Emsael, R., Yang, W., Li, N. (2018). Enhancing security behaviour by supporting the user. Computers & Security, 1-19. https://doi.org/10.1016/j.cose.2018.01.016

[13] Kim, H., Jun, H. (2012). PIN selection policies: Are they really effective? Computers & Security, 31(2012), 484-496. http://doi.org/10.1016/j.cose.2012.02.003

[14] Pandit, V. (2014). A Study on Statistical "Z Test" to Analyse Behavioural Finance Using Psychological Theories. Quest Journals, 2(1), 1-7.

[15] Secure Hash Standard (SHS). (2015). *Federal Information Processing Standards Publication*. Retrieved from NIST online http://dx.doi.org/10.6028/NIST.FIPS.180-4

[16] Frost, J. (2015, March 19). Understanding Hypothesis Tests: Significance Levels (Alpha) and P values in Statistics. The Minitab Blog. Retrieved from http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-hypothesis-tests%3A-significance-levels-alpha-and-p-values-in-statistics