Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Analysis of passwords: Towards understanding of strengths and weaknesses

Waleed Albattah *

Information Technology Department, College of Computer, Qassim University, Saudi Arabia

ARTICLE INFO

Article history:
Received 14 May 2018
Received in revised form
25 August 2018
Accepted 10 September 2018

Keywords:
Password strength
Password analysis
Security
Password prediction

ABSTRACT

In this paper, we analyze the passwords' strength from real-world data; perform an in-depth analysis, and extract useful information related to the millions of usernames and passwords being utilized. This useful information thus represents the millions of minds and the individual behaviors in online and offline passwords based information systems. From the twelve million usernames and passwords, we investigate density, numbers in usernames and passwords, special characters, and strength analysis of the usernames and passwords. To the best of our knowledge, this work is unique based on the selected parameters and the amount of processed data. With the extensive analysis, we seek the weak link in the username and password paradigm. With density analysis, it can be deduced that users like to have (or by chance use) similar character usernames and passwords. From the digits analysis in passwords, it is found that users like to use the first few digits (1, 2, and 3) and the last digits (8, 9, and 0). With the special character analysis, we found that "_" is the most widely used character. With the strength analysis, we determined that it is better to use non-popular English vocabulary words and the inclusion of the special characters, lower, upper and digits are in between different words. Also, if a word can be converted to other languages and used as a password, it will be extremely robust. Most users use their username partly or fully as passwords. This opens doors for hackers. The extensive experimentation and results in the appropriate sections provide useful contributions.

1. Introduction

Organizations normally have a username and password policy. This policy will include rules about the way a username and password are selected. For example, how the password is formed and what characters must be included, and how often the password should be reset. This kind of policy is very important, and they have been improved over time to increase their efficiency. However, another important factor is also so important, which is the end user experience in this kind of policy and the way they understand and deal with it. If the end user does not understand the goal behind the password policy, they will end up with a weak or poor password that actually follows what is stated in the policy. This leads to shedding a light on what is called the usability of password policy in organizations. Consequently, the password policy can be unusable and as a result, insecure or vulnerable if the end user experience is neglected. For example, a regular change requirement of the password is a good policy, however, forgetting passwords or repeating the previous passwords is an unwanted user practice. Without a good user experience, the password policy may be unusable. Although the literature has a number of authentication mechanisms, a username and password paradigm is still the common method (Herley and Van Oorschot, 2012; Uellenbeck et al., 2013). Some reasons for that include cost-effective or administration, simple and popular concept, and user-friendly. Because of the popularity of using passwords as an authentication method, it has been increasingly subjected to a larger number of attacks, especially weak passwords (i.e., popular and common words, movie names, cell phone numbers, etc.). These types of weak passwords are more exposed and can easily be predicted (Ji et al., 2015). Another reason that makes predicting or guessing passwords possible is the password leakage of

* Corresponding Author.
Email Address: w.albattah@qu.edu.sa
https://doi.org/10.21833/ijaas.2018.11.007
2313-626X/© 2018 The Authors. Published by IASE.

popular web systems such as Facebook, Google, LinkedIn, Twitter, Yahoo, and others.

In this paper, we continue our work in Khan and Albattah (2017). We found the results in Khan and Albattah (2017) are promising and encouraging for further investigation to analyze the username and password paradigm from a practical usage point of view and thus find weaknesses and strengths associated with the usernames and passwords paradigm, which this paper tries to present. We analyze millions of usernames and passwords to see the weakest link in the human perception of password security. With this useful study and analysis of millions of usernames and passwords, we hope that the results will shed valuable light on the way we chose passwords and that we ignore the fact that our passwords can be easily cracked or guessed by foes or hackers. From the two datasets of (10 + 2) million usernames and passwords, we investigate: A) Density, B) Numbers in usernames and passwords, C) Special characters, and D) Strength analysis. With this extensive analysis, we seek the weak link in the username and password paradigm. With density analysis, it can be deduced that users like to have (or by chance select) similar character usernames and passwords. From digit analysis in passwords, it is found that users like to use the first few digits (1, 2, and 3) and the last few digits (8, 9, and 0). With special character analysis, we found that "_" is the most widely used character. With the strength analysis, it is better to use non-popular English vocabulary words and the inclusion of the special characters, lower, upper, and digits are in between different words. Also, if a word can be converted to other languages and used as a password, it will be extremely robust. Most users use a username partly or fully as passwords. This opens doors for hackers. By studying and analyzing these parameters, we believe that the in-depth analysis provides sufficient information related to the millions of usernames and passwords and thus millions of minds and individual behaviors in online and offline password based systems.

2. State of the art

Information Technology (IT) systems rely on password-based authentication for secure access. The information facility systems that are allowing users to avail themselves of web-based services and or perform certain specific service-oriented actions on behalf of the user, typically need authorization and authentication steps (Mattord et al., 2013; Lampson et al., 1992). The authors in Mattord et al. (2013) developed a benchmark which assesses the authentication approaches used in web-based service-oriented systems. Mattord et al. (2013) focused on three distinct areas: First, the requirements for a strong password, secondly, how to effectively use passwords, and finally, the requirements for resetting the passwords. Zhao et al. (2006) showed that without using a strict evaluation metric for ideal ciphers, the security in an ideal cipher is very limited. Farash and Attari (2014) pointed out that the authentication and privacy of Tso's protocol can be compromised by using offline guessing attacks on the passwords. Password-based authentication is as old as computer usage itself; Anderson and Vaughn (1991) shed light on this authentication in detail. Authentication is defined as a step that proves that the request of a service is being generated from a valid (allowed) entity. In the simplest form, it is the user ID and the secret code "password" (Anderson and Vaughn, 1991). This authentication mechanism has been analyzed and studied thoroughly for many years (Manber, 1996; Menkus, 1988; Riddle et al., 1989) and is still used in almost all the distributed and cloud services. However, there are many threats associated with the use of username and passwords authentications, identified even as early as dated back to the 1980's (Menkus, 1988; Riddle et al., 1989; Jobusch and Oldehoeft, 1989). Many other studies show the weaknesses in the username and passwords paradigm and the tricks to using an effective and strong password (Adams et al., 1997; Fagin et al., 1996; Hauser et al., 1996; Jablon, 1996). Conklin et al. (2004) demonstrated a concept based theoretical, implementable design using memory aides for password security to be used for multiple systems that are connected by a legitimate user's actions.

Egelman et al. (2013) conducted experiments to see the influence of password rules and meters on the selection of the passwords. Password meters are an evaluation that hints at the strength of the passwords. Generally, this strength is classified as weak or strong. The authors conclude that the meters force the users to select stronger passwords. However, we believe that stronger password selection has other drawbacks. One drawback comes in the form of memorability and retrieve-ability. The user is more likely to easily forget the password in this case.

In Furnell et al. (2000), the authors discuss that user authentication is mostly done by passwords for accessing IT resources. However, this password-based authentication has serious issues. The biggest issue is that it can be compromised. There are many other methods as well, but the problem is the user acceptance of these solutions. The authors conducted a detailed survey that sheds light on different approaches of authentication and the user acceptance of these methods. The results show that many users are willing to adopt new methods and are aware of the password related problems. Halevi and Krawczyk (1999) came up with the concept of security for password authentication. They gave a list of attacks that a protocol which was password based would guard against. Gong et al. (1993) made a research study on the problems faced by the password based problem. They used an encrypted public key to safeguarding against offline password guessing attacks. Bellovin and Merritt (1992) also came up with an Encrypted Key Exchange (EKE). This EKE became the basis for many studies which came afterward. Other authors who dealt with the

password-based protocol problem were Bellare et al. (2000). They came up with a model for the problem, and claimed that the model could deal with the problem of password guessing. According to Eichin and Rochlis (1989), all data even the encrypted data, needs to be authenticated since it is subject to catalog attacks. Purdy (1974) believed that interception is not the only problem likely to compromise the identification and authentication of data. He believes that things such as mishandling of data versions which are offline such as the backup files and the fault induced system dump also compromise identification and authentication of data. According to Stoll (2005), many users usually tend to choose passwords which are easy to guess. He found that about eighty-five percent of a user's passwords were guessable. There have been continuous updates to the dictionaries, whereby more words, numbers, and phrases are added to passwords (Spafford, 1992). According to Parker (1992), among the key elements in information security is confidentiality and authentication that are the major mechanisms, and that authentication has two stages: user identification and a user authentication stage. Parker (1992) said that one of the factors which cause hacking of passwords is a lack of knowledge on security. Parker (1992) then identified a major doctrine in password security as the need to know principle. By this, he means that only the people who need to know should be informed about the security mechanisms. The users who were required to change their passwords were found to set passwords which were less secure and also revealed them frequently. The Federal Information Processing Standards (FIPS) were of the opinion that ownership of passwords which was individual increased the accountability and also decreased the illegal use of passwords. This is because of the possibility of audit trailing, which is a byproduct of authentication. Sharing of passwords by groups was found to be very insecure. Group passwords would only be used when they refer to a team of people who work together.

3. Analysis and evaluation

3.1. Datasets

We use two datasets: The first one (DS1) is approximately 10 million provided by Burnett (2015). The second dataset (DS2) having approximately two million passwords only (without usernames) is obtained from Granville (2012).

3.2. Analysis setup

Our analysis of the passwords paradigm is based on the theoretical assumptions in the state of the art and many years of research based on psychological and social impacts of the paradigm. Table 1 includes the analytical parameters we studied in this research. We still believe that the list can have further additions of the parameters essential in the future. However, the analytical parameters we discuss and experimentally visualize have far more outreaching benefits compared to the psychological study.

Table 1: Analytical parameters

Density	The difference between the length of the username and passwords
	On average, how much is the length of the username people use
	The length of the password on average people use
Numbers in usernames and passwords	Digits at the end of the passwords
	Digits at the start of the passwords
	Average numerical digits in passwords
	The number of digits in usernames
Special characters	The special characters in passwords
	The special characters in usernames
	The matching special characters between usernames and passwords
Discussion and strength analysis	Impact of length on passwords
	The strength analysis of the passwords

3.3. Density analysis

Density analysis here refers to the over-all length statistics of the usernames and passwords in this research context. Fig. 1 shows the sampled spread of the username length plotted against password length. The blue line (dotted) shows the length of the usernames, the red line shows the length of the passwords, and the green line (dashed) shows the difference of the length of the username and the password. We believe that this difference statistic is also of key importance and can provide a hint about the nature of the password and username combinations. Overall in Fig. 1, the green line stays low. A green line with digit one in Fig. 1 shows that username length is 6, the password length is 5 and the difference is 1. Fig. 2 shows the average username length, the average password length, and the average difference of usernames and passwords in 10 million datasets. From Fig. 2, it can be deduced that users like to have (or by chance select) a similar character username and password. The total average difference of 10 million passwords comes out to be 1.229. The average username length is 8.82 and the average password length is 7.59. We argue that this information can be useful for hackers, as the hacker can start with a seed length close to that of the username length. We believe that for stronger password and username combinations, this difference should be higher. The maximum length password we obtained from the dataset is "band*** otmneworleanslouisishaza_cunl**ve96" comprising

of 42 characters (some characters replaced due to maintaining its secrecy).

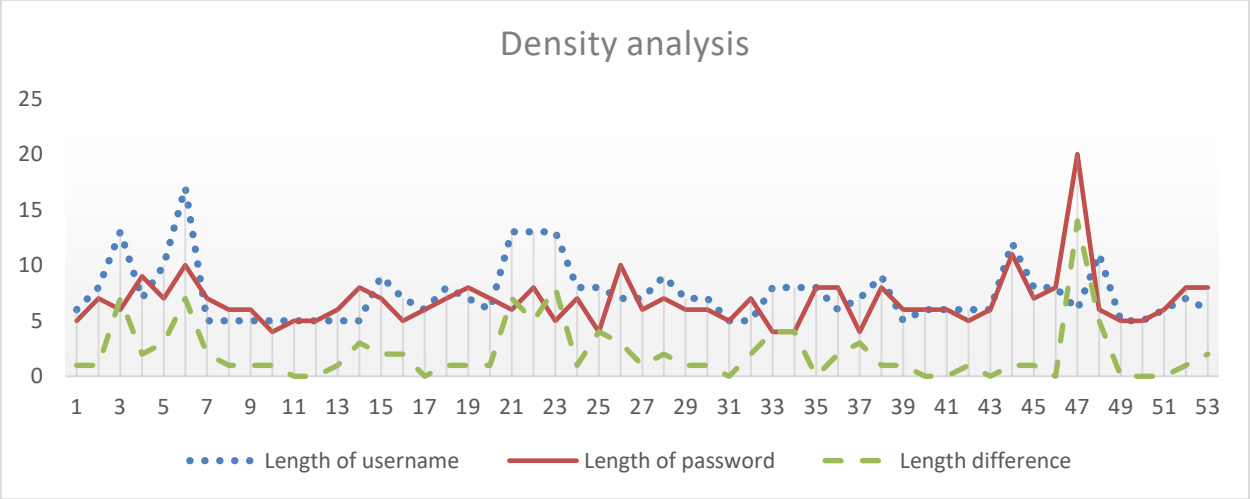


Fig. 1: Password length vs username length sampled

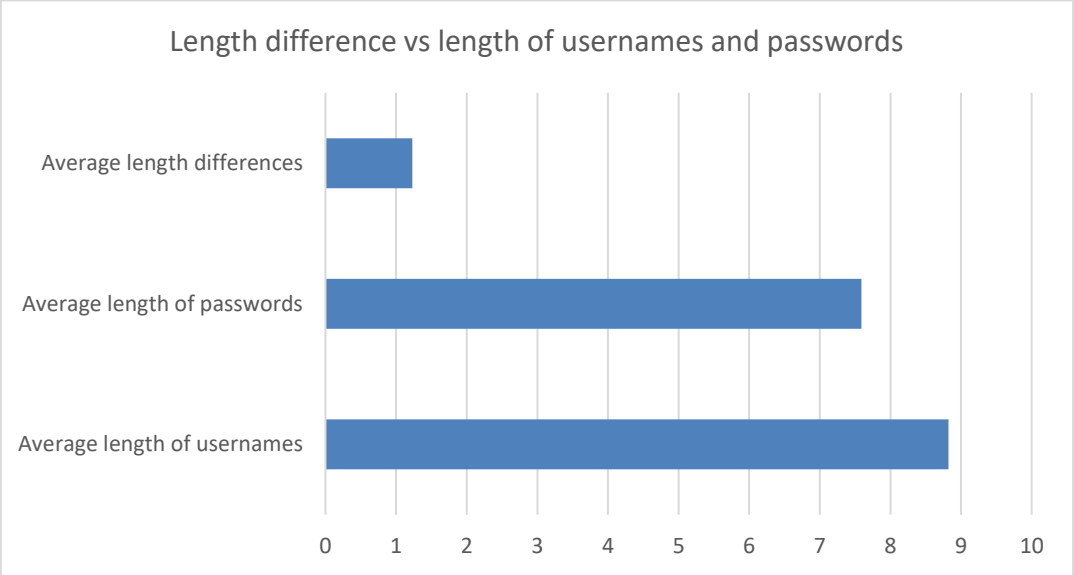


Fig. 2: Average length usernames vs. passwords

3.4. Numbers in usernames and passwords

Numbers are of great importance in usernames and passwords. They not only add strength to the passwords but also help in memorability of the passwords. Also, the online resources motivate the addition of digits not only to passwords but also to usernames as to uniquely construct the combination. Fig. 3 shows the presence distributions of numerical digits from 0 to 9 in 10 million usernames and passwords. In Fig. 3, we arrange numbers starting at 1 and ending at 0 for visualization purpose only. Regarding passwords, in Fig. 3a, the most used digits at the beginning of the password is 1 followed by the 2 and 0. The least used digit at the beginning is 9, 6, and 4 correspondingly. The almost similar trend is found in Fig. 3b for digits used at the end of the passwords. The digit 1 and 2 is mostly used at the end of the passwords on 10 million. Average digits usage in Fig. 3c, shows a smooth pattern starting from 1 at peak and slowly decreasing towards the 7, then finally increasing for 8 digits, 9 and 0. It also confirms that users on average like to use first few digits (1, 2, and 3) and the last few digits (8, 9, and 0). For usernames, Fig. 3d, the pattern is even more interesting than with passwords. We find the smooth flow of digits count from 1, 2, and 3 and all the way to the digit 0. Digits at the end of the usernames (Fig. 3e) show almost the same characteristics for the digits at the end of the passwords, with 1 being the most used digit at the end of the usernames. Another interesting pattern found in these statistics is that of the average digits used in usernames (Fig. 3f). The average digits in usernames follow the similar trend of the average digits in passwords, with 1 being the most used, followed by 2, and slowly decreasing. Similar to passwords, an increasing trend is observed at the end of the digits (8, 9, and 0). We deduce an interesting result from this analysis. Users like to use the first few digits (1, 2, and 3) and last few digits (8, 9, and 0). This can be attributed to the fact that it is easier to remember these digit combinations as a reference compared to the digits in between (4, 5, 6, and 7).

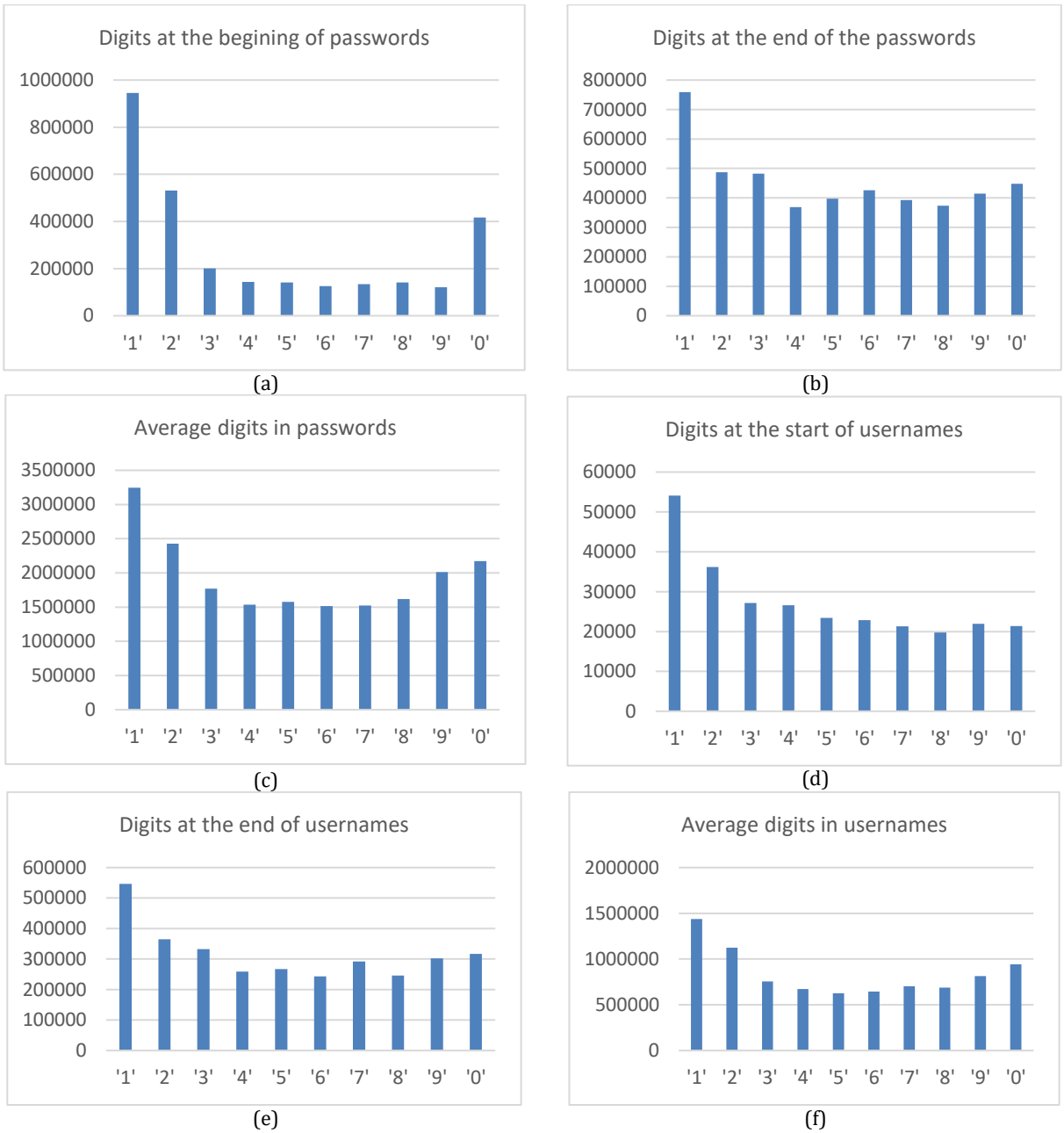


Fig. 3: Distributions of digits (0-9) in 10 million usernames and passwords

3.5. Special characters

Like numerical digits, the special characters are of key importance for not only the uniqueness of username/password combinations but also adds strength to the corresponding combinations in terms of password cracking times. We analyzed the presence of 32 special characters, as follows: ['_' '.' '-' '%' ':' '*' '!' '"' '$' 'and' '+' '#' ';' '^' '/' '[' ']' '}' '\' '`' '~' '|' ')' '>' '?' '{' '<' '@' '(' '"' ',' '=']. Fig. 4 shows the distributions and the count of these special characters in 10 million usernames and passwords. Special characters are of key importance, and the latest online resources nowadays have restrictions on using them in passwords, i.e., there must be at least one special character in the password. In Fig. 4, we observe that the special character "_" is the highest in terms of the usage in passwords, with a number of 33,335 passwords 33335 in which this character is used. Followed by special character "." and "-" special characters. In usernames, the initial trend is similar as well, the highest used special character is "_" followed by "." and then "-" in 10 million usernames and passwords. Fig. 5 shows the matching behavior of users. In 3,243 username and password combinations, the users used "_" in both the usernames and the passwords. In 2,509 occasions, the user's use "." in both the usernames and passwords. Finally, "-" is shared between 903 usernames and passwords. One interesting pattern found in these statistics is the usage of the three special characters. Users feel easier with the "_" and "-" as special characters in usernames and passwords compared to other special characters, and this could be because of the need to separate two words, which eventually makes it easy to remember.

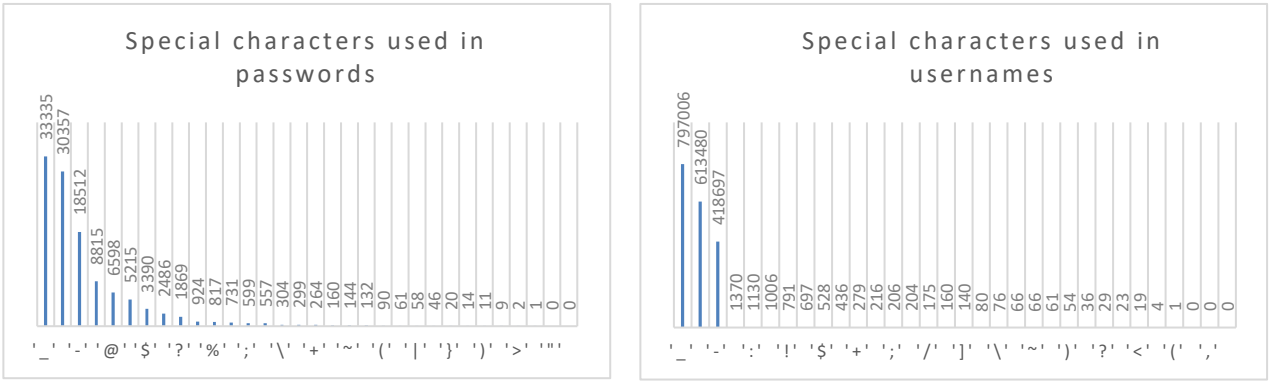


Fig. 4: Special characters used in usernames and passwords for the 10 million dataset

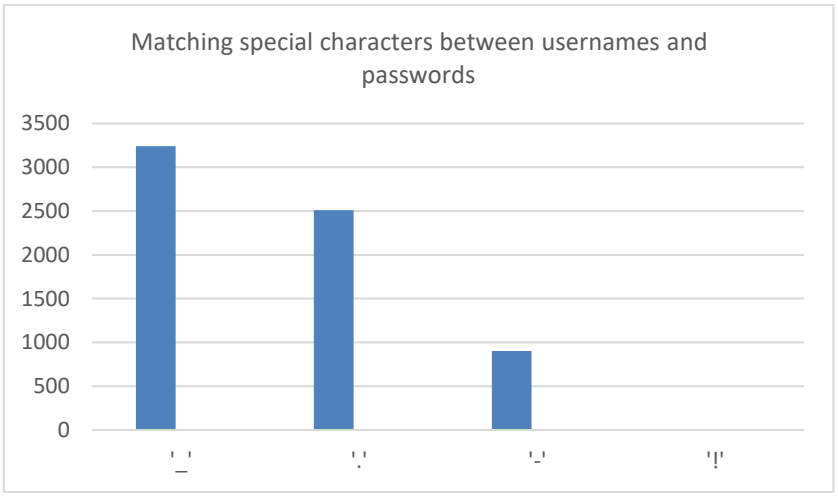


Fig. 5: Graph showing the count of special characters that are matched between usernames and its passwords in the 10 million datasets

3.6. Length and strength

Table 2 shows the average length of the usernames, passwords, and the difference of the 10 million datasets. Table 2 shows that on average, users select similar passwords length as of their usernames. Another interesting statistic would be the presence of usernames in passwords. This is a common practice many of users do. This practice has many reasons. One of the major ones is the memorability of the passwords. However, this makes it extremely easy for a hacker to crack passwords that have a matching part of the username in the password.

Fig. 6 shows the distribution of the presence of the usernames in passwords. The total number of users who used the username as passwords comes out to be 331,033. We believe that this is a high number (3.3 users out of 100) and the masses need to be educated about not using the username in passwords as a part of or combined with other characters.

Table 2: Average length of usernames and passwords and their length difference

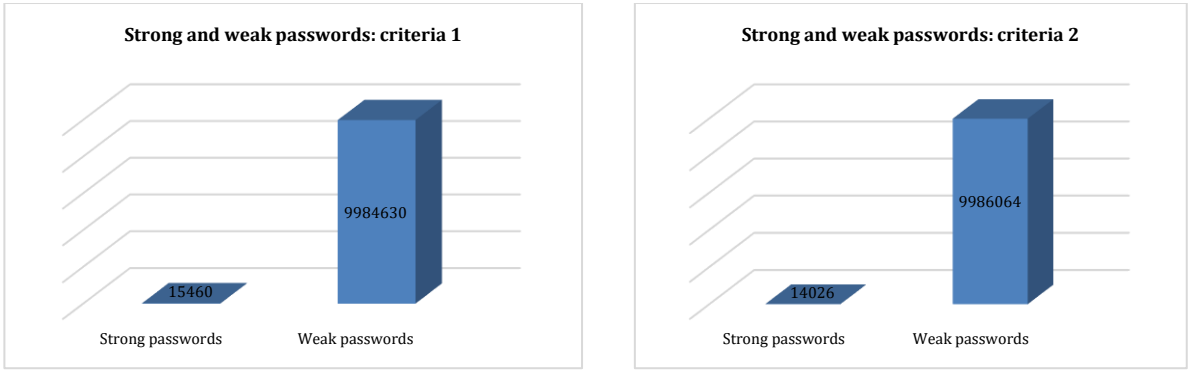
Average length of usernames	Average length of passwords	Length differences
8.82	7.6	1.23

Fig. 6: Presence of usernames in passwords

We conducted simple yet important experiments to check the strength of the passwords in a 10-million dataset. The strength of the passwords has many parameters. However, in this experimental setup, we use the applicable approach that depends on the common conditions of using passwords in many systems nowadays. We check the presence of four parameters only which are (criteria 1):

- Passwords should be greater than six characters
- At least one upper case alphabetical character is used in passwords
- At least one special character is used in passwords
- At least one numerical digit is used in passwords

Fig. 7a shows the results of a strength check based on criteria 1. Based on criteria 1, the statistics that show a very small numbers of passwords can be declared as strong. Only about 0.15% of the passwords are strong and 99.8 % of the passwords are weaker. If we increase the parameters in criteria 1 by adding just one extra parameter of the presence of at least one lower character in the password (thus, criteria 2), the number of strong passwords further decreases. Fig. 7b shows that the addition of one restriction (at least one lower character presence) to criteria 1 further reduces the number of strong passwords present in the dataset. The number reduces from 15,460 to 14,026. The online resources must impose strict limitations on these criterions and the user should not be allowed to proceed before satisfying criteria 1 and 2 for passwords.



(a) (b)

Fig. 7: Visualization of strong vs weak passwords used on criteria 1 and criteria 2 in the dataset

4. Discussion and strength analysis

For password strength analysis from practical perspectives, and based on criterion 2 of the previous section, we select three passwords (g8NJr*QxcCkF, Raubmaus=09=09, M.a.x.641114) to analyze their strengths. We use the well-known resource of the Kaspersky (2017) study to check the strength of passwords. The passwords seem strong, but an analysis of Fig. 8 shows that though it is impossible to crack those using normal computers, however, using high-end resources, the cracking is limited to a few hours or days.

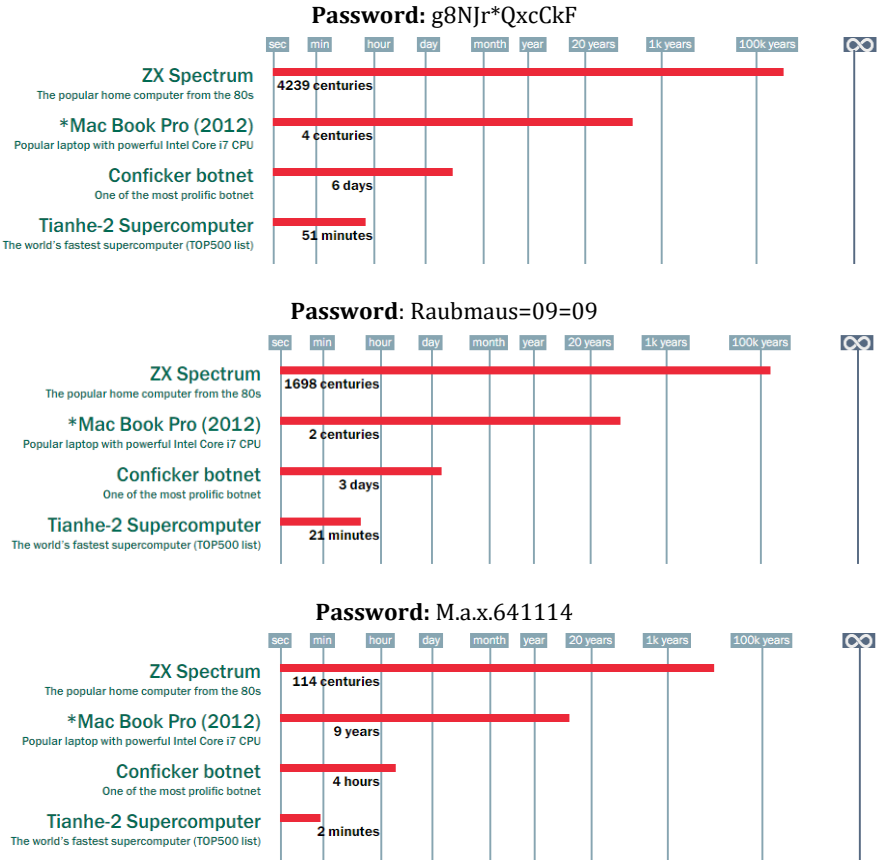


Fig. 8: Three selected passwords based on criteria 2 and the estimated time to crack using different systems. The resource used to check the cracking times is Kaspersky Labs

This information is helpful in designing passwords. The plainer the password, the easier it is to crack. The more complex combinations of special characters, lower, and upper characters make the password stronger and difficult to crack. As in Fig. 8, the password g8NJr*QxcCkF is difficult to crack due to the systematic use of lower, upper, special characters, and digits used.

Also, we analyzed that if the words belong to the English dictionary, the cracking is easier. We analyzed and experimented with this phenomenon in detail. Also, the length of the passwords adds to the cracking time in most cases. We tested a scenario; we starting with plain passwords (passwords without characters, uppercase, numbers). We tried to use non-English words (words in other languages pronounced in the English alphabet). We believe that this way, the cracking will be difficult as compared to English dictionary words. We keep increasing its length by adding extra character at a time. Then we also add special characters and digits. Table 3 shows the flow of this scenario. The password selected is non-English "salam" meaning "peace" in the Arabic language. We test both the English version and the Arabic versions. In its entirety, the password "salam" with the length of 5 is easier to crack then adding further alphabetical characters with it. In Fig. 8, when we add a few alphabetical characters to advance the length to 10, the cracking time increases to 2 months. With the addition of a digit and a special character, the cracking time reaches a very long time (47 years). We compare the same word translated to English as peace and carried on the experiments. Table 3 shows that compared to non-English words, the English words are considerably easy to crack. The password peace takes nine seconds to crack, compared to the password "salam" (Arabic word with same meaning) which takes three minutes. If the length is increased in the peace password case, the time increases. However, there is a problem with this, if the second word is a well-known English word, the cracking time will decrease. And in that case, the length cannot benefit the strength until the lower alphabets, upper alphabets, digits, or special characters are used. As can be seen in Table 3, the inclusion of a digit increases the time of cracking to three months. When a special character is used, it is then we get a very strong password taking roughly 300 years to crack.

From these statistics and experimentation, we derive useful results. It is better to use non-popular English vocabulary words and the inclusion of the special characters, lower, upper, and digits are in between different words. Also, if a word can be converted to another language and used as a password, it will be extremely robust. Most users use a username partly or fully as passwords. We have carried out this analysis using the distance of usernames and passwords based on Navarro (2001).

Fig. 9 shows that flow of distance with the combined length of the username and passwords. The higher the distance of usernames and password, the better is the combination.

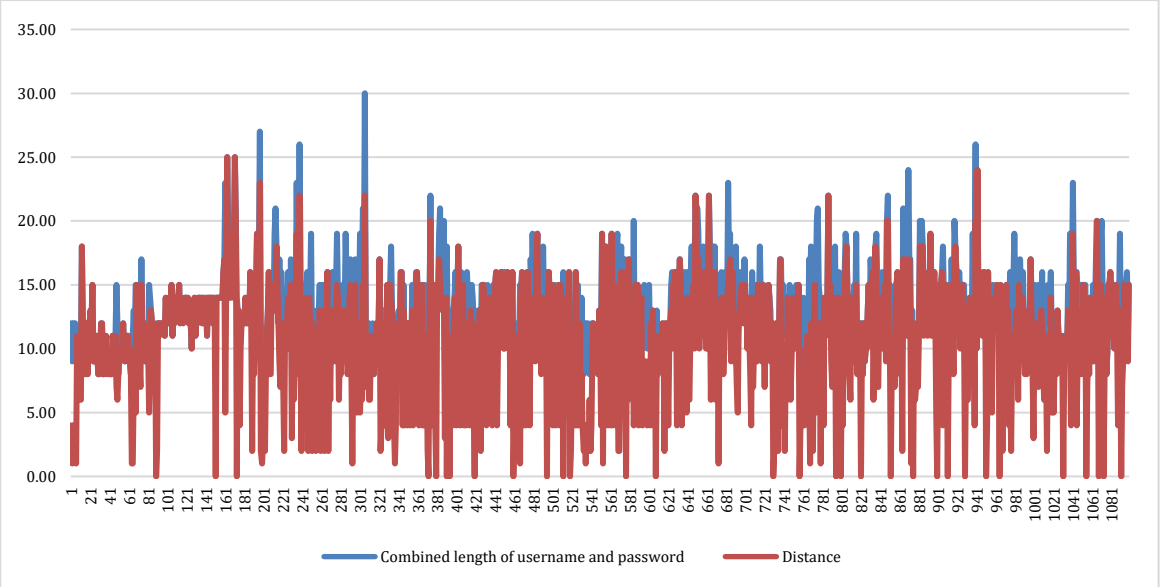


Fig. 9: Levenshtein distance (Navarro, 2001) of usernames and passwords (sampled). A smaller value of distance shows that the password matches the username. The higher the distance, the better the combination

Ideally, in Fig. 9, the distance must be equal to the combined length (blue peaks). However, many scenarios are found where the distance is reaching zero. A zero distance means that the user has exactly used the username as a password. In the statistics, in Fig. 10, the total of 36,081 cases are found where the users have used exactly the username as the password. A distance of value 1 means that the user has selected a password that is only one character different than the username. More than 8,000 passwords are different from the usernames by just one character or alphabet. A distance of two passwords is 10,753, meaning that users have only two digits different from their passwords. We

believe that hackers are aware of this useful information.

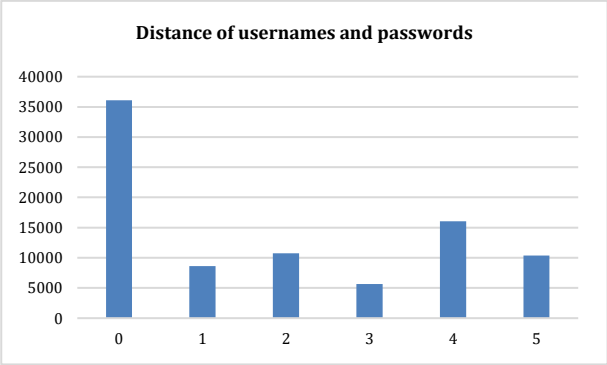


Fig. 10: First six distances of usernames and passwords

In ideal conditions, the password must be totally different than the usernames. We also believe that a distance higher than five is acceptable for most systems, as apparent from the density experiments.

5. Conclusion

In this article, we analyzed passwords using two datasets. We investigated A) Density, B) Numbers in usernames and passwords, C) Special characters, and D) Strength analysis. With density analysis, it is deduced that users like to have (or by chance select) similar character usernames and passwords. From digit analysis in passwords, it is found that users like to use the first few digits (1, 2, and 3) and last few digits (8, 9, and 0). With special character analysis, we found that "_" is the most widely used character. With the strength analysis, it is better to use non-popular English vocabulary words and the inclusion of the special characters, lower, upper, and digits are in between different words.

Also, if a word can be converted to another language and used as a password, it will be extremely robust. Most users use a username partly or fully as passwords. This opens doors for hackers. By studying and analyzing these parameters, we believe that the in-depth analysis provides sufficient information related to the millions of usernames and passwords and thus millions of minds and individual behaviors in online and offline passwords based systems. The article thus enables us to be more vigilant while using the online resources and cloud services based on usernames and password authentication.

Table 3: The impact of the word selection, length, special characters, and digits on cracking of example passwords in increasing length and complexity. Cracking time is estimated by an average home computer using the resource

Length	Password	Time to crack	Modality
6	sakoon	2	hours
7	sakoonm	9	hours
8	sakoonma	12	days
9	sakoonmat	3	months
10	sakoonmats	4	years
11	sakoonmatsh	33	years
12	sakoonmatsho	400	years
	Adding special characters and digit		
13	sakoonmatsho1	3300	years
14	sakoonmatsho1$	32700	years
	We start with English words and keep adding characters		
5	peace	9	seconds
6	peaced	5	minutes
7	peacede	16	minutes
8	peacedes	12	days
9	peacedest	14	days
10	peacedestr	4	years
11	peacedestro	3	days
12	peacedestroy	4	hours
13	peacedestroye	2	days
14	peacedestroyed	5	hours
	Adding special characters and digit		
15	peacedestroyed1	19	days
15	peacedestroyed_	3	months
17	peace_destroyed#1	300	years

Acknowledgment

The work in this paper is funded in its entirety by the Deanship of Scientific Research (SRD), Project number: 1313-coc-2016-1-12-S at Qassim University, Kingdom of Saudi Arabia.

References

Adams A, Sasse MA, and Lunt P (1997). Making passwords secure and usable. In: Thimbleby H, O'Connaill B, and Thomas PJ (Eds.), People and computers XII: 1-19. Springer, London, UK.

Anderson JP and Vaughn R (1991). A guide to understanding identification and authentication in trusted systems. No. NCSC-TG-017. National Computer Security Center Fort George G Meade Md. Available online at: https://fas.org/irp/nsa/rainbow/tg017.htm

Bellare M, Pointcheval D, and Rogaway P (2000). Authenticated key exchange secure against dictionary attacks. In the International Conference on the Theory and Applications of Cryptographic Techniques, Springer, Berlin, Germany: 139-155.

Bellovin SM and Merritt M (1992). Encrypted key exchange: Password-based protocols secure against dictionary attacks.

In the IEEE Computer Society Symposium on Research in Security and Privacy, IEEE, Oakland, USA: 72-84.

Burnett M (2015). Today I am releasing ten million passwords. Available online at: https://xato.net/passwords/ten-million-passwords/

Conklin A, Dietrich G, and Walz D (2004). Password-based authentication: A system perspective. In the 37th Annual Hawaii International Conference on System Sciences, IEEE, Big Island, USA: 1-10.

Egelman S, Sotirakopoulos A, Muslukhov I, Beznosov K, and Herley C (2013). Does my password go up to eleven?: the impact of password meters on password selection. In the SIGCHI Conference on Human Factors in Computing Systems, ACM, Paris, France: 2379-2388.

Eichin MW and Rochlis JA (1989). With microscope and tweezers: An analysis of the internet virus of November 1988. In the IEEE Symposium on Security and Privacy, IEEE, Oakland, USA: 326-343.

Fagin R, Naor M, and Winkler P (1996). Comparing information without leaking it. Communications of the ACM, 39(5): 77-85.

Farash MS and Attari MA (2014). An efficient client–client password-based authentication scheme with provable security. The Journal of Supercomputing, 70(2): 1002-1022.

Furnell SM, Dowland PS, Illingworth HM, and Reynolds PL (2000). Authentication and supervision: A survey of user attitudes. Computers and Security, 19(6): 529-539.

Gong L, Lomas MA, Needham RM, and Saltzer JH (1993). Protecting poorly chosen secrets from guessing attacks. IEEE Journal on Selected Areas in Communications, 11(5): 648-656.

Granville V (2012). Password and hijacked email dataset for you to test your data science skills - Data science central. Available online at: https://www.datasciencecentral.com/forum/topics/password-dataset-for-you-to-test-your-data-science-skills

Halevi S and Krawczyk H (1999). Public-key cryptography and password protocols. ACM Transactions on Information and System Security, 2(3): 230-268.

Hauser R, Janson P, Tsudik G, Van Herreweghen E, and Molva R (1996). Robust and secure password and key change method. Journal of Computer Security, 4(1): 97-111.

Herley C and Van Oorschot P (2012). A research agenda acknowledging the persistence of passwords. IEEE Security and Privacy, 10(1): 28-36.

Jablon DP (1996). Strong password-only authenticated key exchange. ACM SIGCOMM Computer Communication Review, 26(5): 5-26.

Ji S, Yang S, Wang T, Liu C, Lee WH, and Beyah R (2015). Pars: A uniform and open-source password analysis and research system. In the 31st Annual Computer Security Applications Conference, ACM, Los Angeles, USA: 321-330.

Jobusch DL and Oldehoeft AE (1989). A survey of password mechanisms: Weaknesses and potential improvements. Part 1. Computers and Security, 8(7): 587-604.

Kaspersky (2017). Password strength checking we use words to save the world. Available online at: https://password.kaspersky.com/

Khan RU and Albattah W (2017). Security and safety concerns: Username and password paradigm. International Journal of Computer Science and Network Security, 17(10): 145-152.

Lampson B, Abadi M, Burrows M, and Wobber E (1992). Authentication in distributed systems: Theory and practice. ACM Transactions on Computer Systems (TOCS), 10(4): 265-310.

Manber U (1996). A simple scheme to make passwords based on one-way functions much harder to crack. Computers and Security, 15(2): 171-176.

Mattord HJ, Levy Y, and Furnell S (2013). Factors of password-based authentication. International Journal of Computer Science and Network Security, 17(10): 145-152.

Menkus B (1988). Special feature: Understanding the use of passwords. Computers and Security, 7(2): 132-136.

Navarro G (2001). A guided tour to approximate string matching. ACM Computing Surveys, 33(1): 31-88.

Parker DB (1992). Restating the foundation of information security. In the 8th International Conference on Information Security: IT Security: The Need for International Cooperation, North-Holland Publishing Co., Amsterdam, Netherlands: 139-151.

Purdy GB (1974). A high security log-in procedure. Communications of the ACM, 17(8): 442-445.

Riddle BL, Miron MS, and Semo JA (1989). Passwords in use in a university timesharing environment. Computers and Security, 8(7): 569-579.

Spafford EH (1992). OPUS: Preventing weak password choices. Computers and Security, 11(3): 273-278.

Stoll C (2005). The cuckoo's egg: tracking a spy through the maze of computer espionage. Simon and Schuster, New York, USA.

Uellenbeck S, Dürmuth M, Wolf C, and Holz T (2013). Quantifying the security of graphical passwords: the case of android unlock patterns. In the 2013 ACM SIGSAC Conference on Computer and Communications Security, ACM, Berlin, Germany: 161-172.

Zhao Z, Dong Z, and Wang Y (2006). Security analysis of a password-based authentication protocol proposed to IEEE 1363. Theoretical Computer Science, 352(1-3): 280-287.