Security and Safety Concerns: Username and Password Paradigm

Rehan Ullah Khan and Waleed Albattah

Information Technology Department, Qassim University, KSA

Summary
Usernames and password-based login are one of the widely used approaches to authentication for accessing information resources. In this paper, we analyze millions of usernames and passwords, by investigating common words, density, numbers, special characters, strength and society related parameters. The results shed valuable light on the way we select passwords and that we ignore the fact that our passwords can be easily cracked or guessed by foes or hacker. As a contribution to this area, it educates the masses of how a hacker could easily predict and possibly crack the passwords. It also enables users to be more vigilant while using the online resources and cloud services based on usernames and password authentication. By studying and analyzing common words, density, numbers, special characters, strength, and society related parameters, we believe that the in-depth analysis provides sufficient useful information related to passwords selection and thus millions of minds and individual behaviors in online and offline passwords based systems. Thus, the results and the recommendations are a valuable contribution of this article and augment the state-of-the-art.

Keywords:
Password, Authentication, Security, Online resources

1. Introduction

Authentication and authorization are the processes of confirming that the identity of the entity is valid and that the entity has the right to be serviced for a particular resource. Consequently, the rate of security attacks on the resource providing sources increases over time. These attacks cause high financial losses and yet, the degree of our reliance on these systems is growing exponentially. Users face serious problems when passwords are stolen or misused. If we look at our everyday activities such as checking email, checking account balance, they are protected by combinations of usernames and passwords. The level of safety and privacy, or in other words, the level of security is evaluated by the strength of such combinations. Organizations normally have usernames and passwords policies. These policies include rules about the way usernames and passwords are selected. For example, how the password is formed and what characters must be included, and how often the password should be reset? These kind of policies are very important and they have been improved over time to increase their efficiency. However, another important factor is the end user experience with these policies, and the way users understand and deal with them. If the end user doesn't understand the goal behind the password policy, they will end up with a weak or poor password that actually (semantically) does not follow what is stated in the policy. This leads to shedding 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, 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, username and password paradigm is still the common method [1][2]. Some reasons for that include cost-effectiveness or administration, simple and popular concept, and user-friendliness [3]. Because of the popularity of using passwords as an authentication method, it has increasingly been subjected to increasing attacks, especially weak passwords (i.e., popular and common words, movie names, cell phone numbers, etc.). This type of weak passwords are more exposed and easily can be predicted [4]. Another reason that makes predicting or guessing passwords possible is the password data leakage from popular web systems such as Facebook, Google, LinkedIn, Twitter, Yahoo, and others [5].

In this paper, we analyze the username and password paradigm from a practical usage scenario point of view and thus find weaknesses associated with the usernames and passwords selection. From the (10 + 2) millions of usernames and passwords dataset, we investigate common words, density, numbers, special characters, strength and society related parameters. We believe that such an analysis benefits the society in many ways. It sheds light on the concept that the most secure systems which we use online are basically as strong as the strength of the password we use. Secondly, the usernames passwords

Manuscript received October 5, 2017
Manuscript revised October 20, 2017

study and research educate the masses of how a hacker could easily predict and possibly crack the passwords. By studying and analyzing common words, density, numbers, special characters, strength, and society related parameters, we believe that the in-depth analysis provides sufficient information to the millions of usernames and passwords combination and thus millions of minds and individual behaviors in online and offline passwords based systems. We believe that the parameters list investigated in this paper is adequate and further additions of the parameters will be carried as an extension of this work.

The rest of this paper is organized as follows. In Section 2, we introduce some related work. Section 3 defines the datasets used in the experiments and Section 4 presents the experimental analysis. We conclude in Section 5.

2. Related Work

IT systems rely on password-based authentication for secure access to resources. The information systems that allow users to avail web-based services and/or perform certain specific service-oriented actions on behalf of the user, the system typically needs an authorization and authentication steps [6][7]. The authors in [6] develop a benchmark which assesses the authentication approaches used in web-based service-oriented systems. Zhao et al. [8] show that without using strict evaluation metric for ideal ciphers, the security in an ideal cipher is limited. In [9], the authors point out that the authentication and privacy of Tso's protocol can be compromised by using offline guessing attacks on the passwords. The work in [10] sheds light on the password based authentication in detail. The work in [5] defines authentication 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" [10]. This authentication mechanism is analyzed and studied thoroughly for many years [11][12][13][14][15] 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, even identified as early as dated back to 1980's [12][13][16]. Many other studies show the weaknesses in username and passwords paradigm and the tricks to use effective and strong password [17][18][19][20]. The authors in [21] demonstrate a concept based theoretical, implementable design using memory aides for password security to be used for multiple systems that are connected by legitimate user's actions.

In [22], the authors conducted experiments to see the influence of passwords rules and meters on the selection of the passwords. Password meters are an evaluation that hints at the strength of the passwords. In [6], the authors define three measures for authentication. These measures are password strength requirements, password usage methods, and password reset requirements. In [23], the authors analyze the alternative approaches to password-based authentication. The results show that many users are willing to adopt new methods and are aware of the password related problems.

The authors in [24] 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. According to [8], a password which is ideal should be able to secure against attacks. A lot of people studied the problems based on password-based protocols. In [25], authors investigated the password based problems. They used an encrypted public key to safeguarding against offline passwords guessing attacks. The authors in [26] present the concept of Encrypted Key Exchange (EKE). This EKE became the base for many studies which came afterward [27]. According to [8], people should be extremely cautious when forming password-based protocols with some provable security in cipher models which are ideal. According to [10], there are three major aspects of effective authentication. They include authentication through knowledge (That is something that they know, authentication through ownership (that is something they have) and lastly authentication by characteristic (refers to something that they are).

According to Eichin [28], all data even the encrypted data needs to be authenticated since it is subject to catalog attacks. Purdy [29] believes that interception is not the only problem likely to compromise the identification and authentication of data. Manber [11] believed that guessing was not the only risk involved with passwords but also the risk that people would gather a list of encrypted passwords and spread the list to other people. UNIX was faced with a lot of attacks most of which were caused by grabbing the password file [30]. Manber [11] checking the passwords on a regular basis. There have been continuous updates to the dictionaries, whereby more words, numbers, and phrases are added [31]. The authors in [11] came up with a scheme which made passwords more random for everyone without people having to remember the strings which were random. According to [32], among the key elements in information security is confidentiality and authentication.

According to [17] most people don't usually secure procedures for the construction of passwords. The users who were required to change their passwords were found to set passwords which were less secure and also revealed them frequently. Sharing of passwords by groups was found to be very insecure [33]. Many modern systems have

taken up password methods which are simple [34]. The system usually has the view that the password is easy to the user but difficult for an intruder. Better selection of passwords helps reduce the number of breaches [31]. The attacks on the security systems are divided into three: social engineering, discovery and technical [35]. Among the things that designers have come up with to counter these problems are password rules and system rules. According to [36], the password choice of the user usually has a significant effect on the security level of the system.

3. Dataset

We use two datasets. First dataset (DS1) is approximately 10 million usernames and passwords dataset, provided by Mark Burnett [37]. As the usernames and passwords of individuals are their very personal and secret entity, therefore, we believe that the whole responsibility for misuse of the data and any complaints as such direct to [37]. For analysis, we represent it as (DS1) in the experimental setup. The second dataset (DS2) has about two million passwords (without usernames) and it is obtained from [38]. This dataset is made available by Vincent Granville [39].

4. Experimental Analysis

Our analysis of passwords paradigm is based on the theoretical assumptions made in the state of the art and the follow through of the many years of research based on psychological and social impacts of the password paradigm. In the following sub-sections, we discuss the analytical parameters that are being analyzed in this research. We believe that the list is adequate and further additions of the parameters will/can be carried as an extension of this work. The parameters we analyze experimentally have far more outreaching benefits compare to the theoretical study alone mostly done in the state-of-the-art.

4.1 Common Words in Usernames and Passwords

In this experimental analysis, we use the two datasets: DS1 and DS2. For passwords analysis, we combine the two data sets DS1 and DS2 to jointly process (10+2) 12 million passwords. However, as the DS2 does not contain usernames, we limit common words analysis of usernames to DS1 (10 million) only.

Firstly, we gather repository of 9915 most common English words based on the Peter Norvig's resources [40]. This repository contains approximately 9915 most common English words in the order of increasing frequency of usage. During analysis, we search every word in every password and username. This search has also taken into consideration the possibility of the partial or full presence of the particular word. Figure 1 shows the password (alphabets) count and its presence (count) in passwords.

From the experiments, the three alphabets words start to show the presence of full words combined with digits and special characters. The most used three alphabets words are “man” repeated 126023 times in 12 million passwords. 95% of the times, it was used as solo combined with other characters. Second three alphabets words come out to be “and” which is repeated 113351 times. However, the three alphabets words have many passwords that represent another complete word for example “pay” word in password “papayas” and “maxpayne”.

For four alphabets words, one of the most used words is the “love” word which is used 52923 times in 12 million passwords. The second most four alphabets word is “pass” used 49622 times and “word” which is used 31603 in 12 million passwords.

Five alphabets words “sword” and “angel” repeated 29394 and 14143 times in 12 million passwords. Six alphabets words “master” and “dragon” are repeated 11198 and 10827 correspondingly. Seven alphabet words “Michael” is repeated 6750 times and the “mustang” is repeated 5840 times. Eight alphabets words “password is repeated 27222 times and “football” 5685 times. It was interesting to see that not only the nine alphabets words frequency decreases in 12 million passwords, but also it decreases in the 100000 words list of English vocabulary. Nine alphabets words like “Liverpool” is repeated 1145 times and “alexander” repeated 896 times.

Ten alphabets famous words in passwords are found to be “basketball” and “Manchester”. Most of the eleven alphabets words were not present in the 12 million passwords. Eleven alphabets examples are “Christopher” and the “playstation”. Twelve alphabets words are “professional” repeated 78 times and “masturbation” found 51 times. Thirteen character words “administrator” is found in 108 passwords and “international” in 51 passwords. Fourteen character words (“administrators” 8 times) and (“administration” 6 times) are found correspondingly. Fifteen character words are very rare and “congratulations” is found 1 time only. Sixteen character words and beyond are never used and never encountered in the 12 million passwords.

Figure 2 shows the word alphabet count and its presence count in usernames. With reference to the common words in usernames, the Usernames like “alex”, “chris” and

"master" tops the list with counts of 36823, 20508, and 11113 correspondingly. The trend of length vs count (Figure 2) in passwords is almost similar to the passwords trend (Figure 1). The trend of using lengthy usernames decreases with the increase of the length of words.

From the alphabets of passwords and their length statistics, we construct a useful count flow graph. Figure 1 shows the human choice in a graphical manner for words length and words selection. Human likes to have small passwords. Small words selection combined with some digits or character is the choice for many people as of the normal system requirements for passwords selection nowadays require the user to have letters and digits. However, it was not strictly required in legacy systems. The Figure 1 shows that the people with smaller words have more repetition than higher words length. However, an interesting pattern is found in Figure 1. As the real words (nouns, verbs, and adjectives) that are less mixed as a part of other words starting with 5 and 6 character length, the trend in Figure 1 decreases from four to six and then increases slowly between six and eight alphabets. We believe that the real words (nouns, verbs, and adjectives) in combination with the other characters' combination starts at 5 and beyond. We believe that the passwords in this range are easier to remember and manage.

Figure 2 shows the flow of the length of the words vs its count presence in usernames. Compare to the passwords in Figure 1, the flow is slightly smooth and we believe that people tend to like smaller usernames combined with other characters. Password spike in Figure 1 may also be due limitations of the systems for which the passwords are intended to be used.

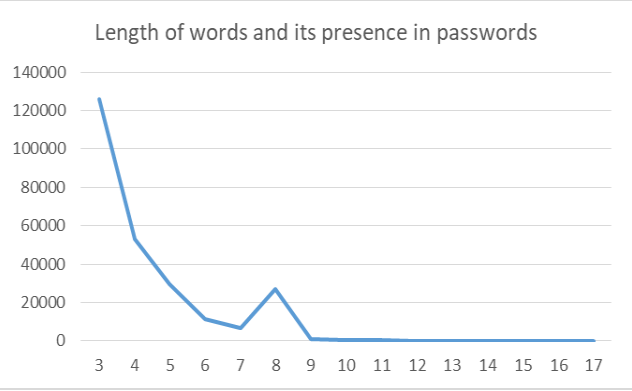

Fig. 1 Word length (alphabets count) and its presence in total number of passwords (X-axis: word length, Y-axis: the frequency of the words in 12 million passwords)

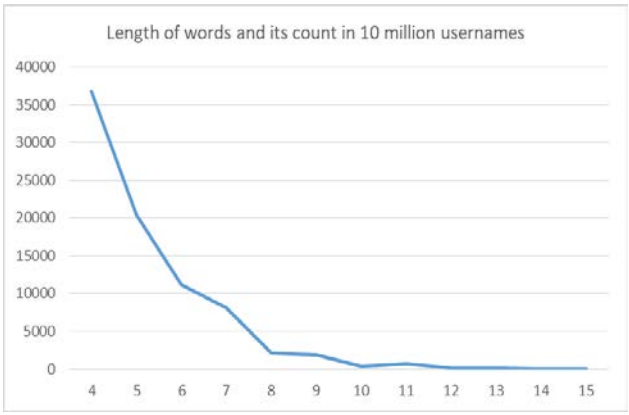

Fig. 2 Words with increasing length (number of alphabets) and its presence (count) in 10 million datasets (DS1) usernames.

4.2 Density Analysis:

In this paper, the density analysis refers to the over-all length statistics of the usernames and passwords. Figure 3 shows the (sampled) spread of the username length plotted against password length. The blue line (dotted) shows the length of the usernames, Orange line shows the length of the passwords, and the Grayline (dashed) shows the difference of the length of the username and the password. This difference statistic is of key importance and gives a hint about the nature of the passwords and usernames combinations. Overall in Figure 3, the Grayline stays low. A Gray line with 1 (on Y-axis and X-axis) in Figure 3 shows that username length is 6, the password length is 5 and the difference is 1.

With reference to the average length statistics, it is found that users like to have (or by chance) a similar character usernames and passwords. The total average difference of 10 million passwords comes out to be 1.23. The average usernames 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 high.

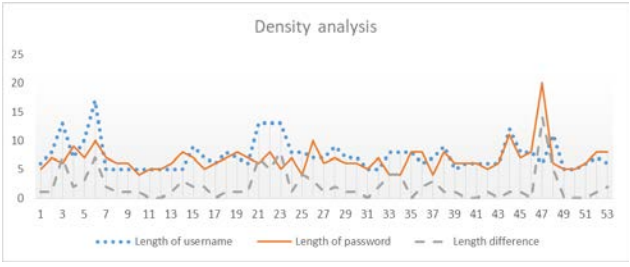

Fig. 3 Password length vs usernames length sampled

4.3 Numbers in usernames and passwords

Numbers are of great importance in usernames and passwords. They not only add the strength to the passwords but also helps 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. The most used digits at the beginning of the password are 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 for digits used at the end of the passwords. The digit 1 and 2 is mostly used at the end of the passwords in 10 million.

Figure 4 shows the average digits repeated in passwords and shows a smooth pattern starting from 1 at peak and slowly decreasing towards 7, finally, increasing for 8 digits, 9 and 0. It also confirms that users on average like to use first few digits (1, 2, 3) and last digits (8, 9, 0). For usernames, 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 show almost same characteristics for the digits at the end of the passwords with 1 being the most used digit at the end of the usernames. 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 digits (8 9 and 0). This can be contributed to the fact that it is easier to remember these digits combination as compared to the digits in between (4, 5, 6 and 7).

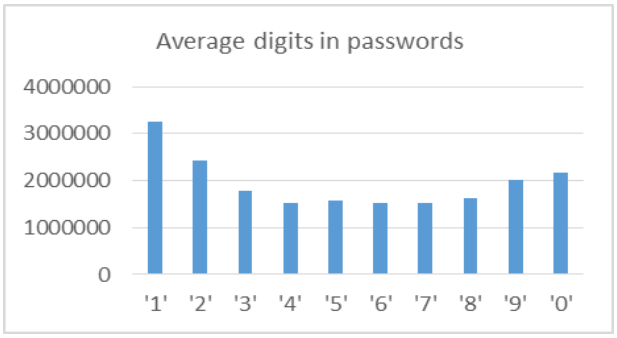

Fig. 4 Distributions of digits (0-9) in passwords

4.4 Special characters

Like numerical digits, the special characters are of key importance for the not only uniqueness of usernames passwords 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: ['_' '.''-' '%' ':' '*' '!' '"' '$' '&' '+' '#' ';' '^' '/' '[' ']' '}' '\' '`' '~' '|' ')' '>' '?' '{' '<' '@' '(' '"' ',' '=']. Figure 5 shows the distributions and the count of these special characters in 10 million usernames and passwords. In Figure 5, we observe that the special character "_" is the highest in terms of the usage in passwords with total 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 "-". One interesting pattern found that users are likely to use the "_" and "-" as special characters in usernames and passwords compare to other special characters.

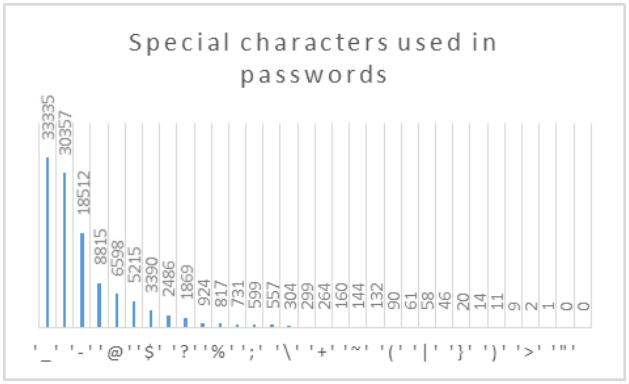

Fig. 5 Special characters used in passwords

4.5 Strength, Length, Society

The strength of the passwords has many parameters. However, in this experimental setup, we use the applicable approach that depends on the common restrictions and conditions of using passwords in many modern systems nowadays. We check the presence of four parameters only which are:

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

Based on these criteria, we find a very small numbers of passwords satisfying it, and declared as strong passwords. Only about 0.15 % of the passwords are strong and 99.8 % of the passwords are weaker based on this simple set of criteria. If we increase the parameters in criteria by adding just one extra parameter of the presence of at least one

lower character in the password, the number of strong passwords further decreases. The addition of one restriction (at least one lower character presence), further reduces the number of strong passwords present in the dataset. The numbers reduce from 15460 to 14026. The online resources must strict limitations on these criterions and the users should not be allowed to proceed before satisfying the criterion.

Society, linguistics, and interactions in society have a deep impact on the passwords as well. Rather than a survey or direct human interaction regarding the passwords and the impact of social variables, we take a practical different approach. We analyze separate variables by searching these variables in passwords and demonstrate that the society has a deep impact on the password selection. We start with the names (obtained from [41] and [42]) in passwords.

The family names in passwords are found to be common and “John” and “Andre” were the most used family names. The “Daniel” and “Roger” family names are also heavily used in the passwords selection. Regarding female names in the password, “Ana”, “Mari”, and “Angel” are mostly used female names, with “Ana” count equal to 58259. For male names, “Alex” and “Jack” are used 16311 and 14225 respectively.

We also performed sentiment analysis (sentiment data obtained from [43]) of the passwords. Figure 6 shows the spread of the positive sentiments. Total positive sentiments are found to be 546849 which is 4.5% of the total passwords. Negative sentiments are more of violent feelings fueled by the current environment and actions. To our astonishment, the negative sentiments are found to be 1155092 which is 9.6% of the total passwords. We believe that this represents the general public using slang language to express feelings and therefore, this behavior is quite common in societies.

For personality analysis, we use the personality traits data from [44]. The analysis of passwords shows that terminologies like “Happy”, “Frank”, and “Nice” are used 5094, 4635, and 3259 times correspondingly and sheds interesting light on this aspect. Compared to the state-of-the-art, our statistics are directly related to the words used by individuals, which we hope expresses his/her personality and our work augments the similar approaches. Our work demonstrates that personality and the personality terminologies effect the user’s selection of passwords.

4.6 Recommendations

Based on different sections in the experimental analysis, we argue that besides the standard restrictions/ suggestions on password selection, the following points should be taken into considerations when selecting passwords.

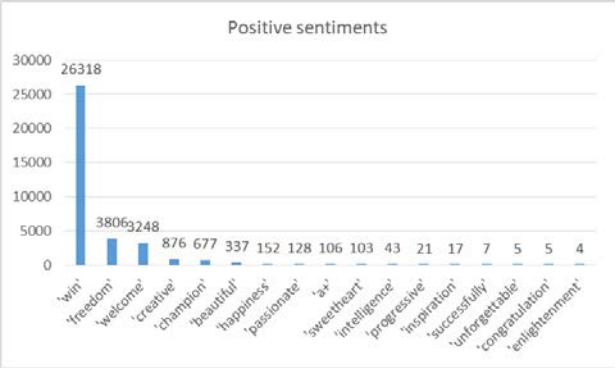

Fig. 6 Positive sentiments in passwords

- The passwords should not be based on common words used in societies.
- If a password has to be picked from social words, it should be broken down into pieces and digits and special characters should be inserted in between and the end.
- If numbers are used in the passwords, the flow should not be in ascending order
- If numbers are used the numbers should not represent years and birthdays
- Recommendation of the numbers is to use numbers in between two words and at the end of the word
- Special characters like "_" and "-" should be avoided as it has been used excessively and the hackers are aware of this fact.
- Passwords should not be famous family names from societies
- Passwords should not be names of individuals
- Passwords should not contain feeling related words
- Passwords selection should be planned and not chosen at a particular instant during registration. This help in avoiding selecting passwords based on immediate feelings

5. Conclusions

In this paper, we analyzed usernames and passwords for finding the weakest link in the human perception of password security. With this useful study and analysis of millions of usernames and passwords, our results 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 hacker. We believe that this research paper benefits the society in many ways and educate the masses of how a hacker could easily predict and possibly crack the passwords. The study enables us to be more vigilant while using the online resources and cloud services based on usernames and password authentication. By studying and analyzing these parameters, we believe that the in-depth analysis provides sufficient information to the millions of usernames and passwords and thus millions of minds and individual behaviors in online and offline passwords based systems. We believe the recommendations add a valuable contribution and augment the state-of-the-art.

Acknowledgments

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

References

[1] C. Herley and P. Van Oorschot, “A Research Agenda Acknowledging the Persistence of Passwords,” IEEE Secur. Priv. Mag., vol. 10, no. 1, pp. 28–36, Jan. 2012.

[2] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz, “Quantifying the security of graphical passwords,” in Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS ’13, 2013, pp. 161–172.

[3] W. E. Burr, D. F. Dodson, and W. T. Polk, “Electronic Authentication Guideline,” Spec. Publ. (NIST SP) - 800-63 Ver 1.0.2, 2006.

[4] S. Ji, S. Yang, T. Wang, C. Liu, W.-H. Lee, and R. Beyah, “PARS: A Uniform and Open-source Password Analysis and Research System.”

[5] “Over 2 million stolen Facebook, LinkedIn and Google passwords leaked online - National | Globalnews.ca.” [Online]. Available: https://globalnews.ca/news/1009800/stolen-facebook-linkedin-google-passwords-leaked-online/. [Accessed: 28-Sep-2017].

[6] H. Mattord, Y. Levy, and S. Furnell, “Factors of Password-based Authentication,” in Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois, August 15-17, 2013, 2013, no. 1996, pp. 1–9.

[7] B. Lampson, M. Abadi, M. Burrows, and E. Wobber, “Authentication in distributed systems: theory and practice,” ACM Trans. Comput. Syst., vol. 10, no. 4, pp. 265–310, Nov. 1992.

[8] Z. Zhao, Z. Dong, and Y. Wang, “Security analysis of a password-based authentication protocol proposed to IEEE 1363,” Theor. Comput. Sci., vol. 352, no. 1–3, pp. 280–287, Mar. 2006.

[9] M. S. Farash and M. A. Attari, “An efficient client–client password-based authentication scheme with provable security,” J. Supercomput., vol. 70, no. 2, pp. 1002–1022, Nov. 2014.

[10] R. Anderson, J. P., & Vaughn, A Guide to Understanding Identification and Authentication in Trusted Systems. National Computer Security Center, 1991.

[11] U. Manber, “A simple scheme to make passwords based on one-way functions much harder to crack,” Comput. Secur., vol. 15, no. 2, pp. 171–176, Jan. 1996.

[12] B. Menkus and Belden, “Understanding the use of passwords,” Comput. Secur., vol. 7, no. 2, pp. 132–136, Apr. 1988.

[13] B. L. Riddle, M. S. Miron, and J. A. Semo, “Passwords in use in a university timesharing environment,” Comput. Secur., vol. 8, no. 7, pp. 569–579, Nov. 1989.

[14] D. Santhi Jeslet, G. Sivaraman, M. Uma, K. Thangadurai, and M. Punithavalli, “Survey on Awareness and Security Issues in Password Management Strategies,” IJCSNS Int. J. Comput. Sci. Netw. Secur., vol. 10, no. 4, 2010.

[15] W. Ma, J. Campbell, D. Tran, and D. Kleeman, “A Conceptual Framework for Assessing Password Quality,” IJCSNS Int. J. Comput. Sci. Netw. Secur., vol. 7, no. 1, 2007.

[16] D. L. Jobusch and A. E. Oldehoeft, “A survey of password mechanisms: Weaknesses and potential improvements. Part 2,” Comput. Secur., vol. 8, no. 8, pp. 675–689, 1989.

[17] P. Adams, Anne and Sasse, Martina Angela and Lunt, “Making Passwords Secure and Usable,” in Proceedings of HCI on People and Computers XII, 1997, pp. 1–19.

[18] R. Fagin, M. Naor, and P. Winkler, “Comparing information without leaking it,” Commun. ACM, vol. 39, no. 5, pp. 77–85, May 1996.

[19] R. Hauser, P. Janson, G. Tsudik, E. Van Herreweghen, and R. Molva, “Robust and secure password and key exchange method,” J. Comput. Secur., vol. 4, no. 1, pp. 97–111, 1996.

[20] D. P. Jablon and D. P., “Strong password-only authenticated key exchange,” ACM SIGCOMM Comput. Commun. Rev., vol. 26, no. 5, pp. 5–26, Oct. 1996.

[21] A. Conklin, G. Dietrich, and D. Walz, “Password-based authentication: a system perspective,” Syst. Sci. 2004., vol. 0, no. C, pp. 1–10, 2004.

[22] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley, “Does my password go up to eleven? The impact of password meters on password selection.,” Proc. SIGCHI Conf. Hum. Factors Comput. Syst., pp. 2379–2388, 2013.

[23] S. M. Furnell, P. S. Dowland, H. M. Illingworth, and P. L. Reynolds, “Authentication and Supervision: A Survey of User Attitudes,” Comput. Secur., vol. 19, no. 6, pp. 529–539, 2000.

[24] S. Halevi and H. Krawczyk, "Public-key cryptography and password protocols," ACM Trans. Inf. Syst. Secur., vol. 2, no. 3, pp. 230–268, Aug. 1999.
[25] L. Gong, T. Mark, A. Lomas, R. M. Needham, and J. H. Saltzer, "Protecting Poorly Chosen Secrets from Guessing Attacks," 1993.
[26] S. M. Bellovin, S. M. Bellovin, and M. Merritt, "Encrypted Key Exchange: Password-Based Protocols Secure Against Dictionary Attacks," IEEE Symp. Res. Secur. Priv., pp. 72--84, 1992.
[27] M. Bellare, D. Pointcheval, and P. Rogaway, "Authenticated Key Exchange Secure Against Dictionary Attacks," in International Conference on the Theory and Applications of Cryptographic Techniques, 2000, pp. 139–155.
[28] M. W. Eichin and J. A. Rochlis, "With Microscope and Tweezers: An Analysis of the Internet Virus of November 1988."
[29] G. B. Purdy and G. B., "A high security log-in procedure," Commun. ACM, vol. 17, no. 8, pp. 442–445, Aug. 1974.
[30] C. Stoll, The cuckoo's egg : tracking a spy through the maze of computer espionage. Pocket Books, 2000.
[31] E. H. Spafford, "OPUS: Preventing weak password choices," Comput. Secur., vol. 11, no. 3, pp. 273–278, May 1992.
[32] D. B. Parker, "Restating the Foundation of Information Security," in Proceedings of the IFIP TC11, Eigth International Conference on Information Security: IT Security: The Need for International Cooperation, 1992, pp. 139–151.
[33] FIPS, "Password Usage," 1985. [Online]. Available: http://csrc.nist.gov/publications/PubsFIPSArch.html. [Accessed: 18-May-2017].
[34] C. Cachin, "Modeling Complexity in Secure Distributed Computing," Springer Berlin Heidelberg, 2003, pp. 57–61.
[35] J. D 'arcy and A. Hovav, "IS Security Research: An Analysis and Integrative Framework for Future Work," in The Annual Security Conference, 2003.
[36] A. Conklin, G. Dietrich, and D. Walz, "Password-based authentication: A System Perspective," in 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, 2004, vol. 0, no. C, pp. 1–10.
[37] Mark Burnett, "Today I Am Releasing Ten Million Passwords – xato: security," 2015. [Online]. Available: https://xato.net/today-i-am-releasing-ten-million-passwords-b6278bbe7495. [Accessed: 14-Jun-2017].
[38] Vincent Granville, "Password and hijacked email dataset for you to test your data science skills - Data Science Central," 2012. [Online]. Available: http://www.datasciencecentral.com/forum/topics/password-dataset-for-you-to-test-your-data-science-skills. [Accessed: 14-Jun-2017].
[39] Vincent Granville, "Vincent Granville's Page - Data Science Central." [Online]. Available: http://www.datasciencecentral.com/profile/VincentGranville. [Accessed: 14-Jun-2017].
[40] Peter Norvig, "Natural Language Corpus Data: Beautiful Data." [Online]. Available: http://norvig.com/ngrams/. [Accessed: 14-Jun-2017].
[41] Mark Kantrowitz, "Package: areas/nlp/corpora/names/." [Online]. Available: http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/nlp/corpora/names/0.html. [Accessed: 17-Oct-2017].
[42] "Top Male First Names in America [Alphabetical List]." [Online]. Available: https://names.mongabay.com/male_names_alpha.htm. [Accessed: 17-Oct-2017].
[43] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation," Comput. Linguist., vol. 37, no. 1, pp. 9–27, Mar. 2011.
[44] "Personality types - English Vocabulary Word List | Learner's Dictionary." [Online]. Available: http://learnersdictionary.com/3000-words/topic/personality-types/1. [Accessed: 17-Oct-2017].

Rehanullah Khan graduated from the University of Engineering and Technology Peshawar, with a BSc. degree (Information Systems) in 2004 and MSc (Information Systems) in 2006. He obtained PhD degree in 2011 from the Vienna University of Technology, Austria. He is currently an Assistant Professor at the IT Department, CoC, Qassim University, KSA. His current research interests include, security, segmentation, machine learning and recognition.

Waleed Albattah received his Ph.D. from Kent State University, Ohio, USA. Dr. Albattah is a faculty member at the Information Technology Department, Qassim University, Saudi Arabia. His research interests are software engineering, software measurements, software design and agile software development, and software quality. Recently, he has been working in Big data and cloud computing security projects. He is the dean of College of Computer at Qassim University, and he is a member in ACM Society SIGSOFT.