DOI: 10.3969/j. issn. 1673-5188. 2016. S0. 006
http://www.cnki.net/kcms/detail/34.1294.TN.20160630.1830.002.html, published online June 30, 2016

Password Pattern and Vulnerability Analysis for Web and Mobile Applications

LI Shancang, Imed Romdhani, and William Buchanan
(School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, Scotland, UK)

Abstract

Text-based passwords are heavily used to defense for many web and mobile applications. In this paper, we investigated the patterns and vulnerabilities for both web and mobile applications based on conditions of the Shannon entropy, Guessing entropy and Minimum entropy. We show how to substantially improve upon the strength of passwords based on the analysis of text-password entropies. By analyzing the passwords datasets of Rockyou and 163.com, we believe strong password can be designed based on good usability, deployability, rememberbility, and security entropies.

Keywords

password strength; security entropies; password vulnerabilities

1 Introduction

Although receiving plenty of criticism, the text-based passwords are still heavily used for authenticating web and mobile application users [1], [2]. Many research efforts have been made to protect user's password against attacks [3]. In recent, many password managers have been developed to help people to create/manage secure passwords with enough strength and easy to remember (e.g. Dashlane, Keepass, Lastpass). However, when using a password manager, at least a master password needs creating and remembering [4], [5].

A number of websites have recently been hacked and millions of user credentials were leaked online [6], [7]. In 2012, six millions of LinkedIn users' credentials were leaked. Actually, it was reported that in this case over 117 million user credentials were leaked on the Dark Web [8]. In 2015, the Sony Pictures was hacked and many confidential data were leaked, including 173,000 emails and 30,000 separate documents [9]. It was reported that in China 130,000 users' data were leaked via China's train ticketing site 12306 [10] in Dec. 2014. In 2015, a large number of websites (including 163.com, CSDN, TianYa, Duduniu, 7k7k, 178.com, Rockyou, and Yahoo) were hacked and over 100 million of user credentials were leaked online. We believe that investigating the leaked passwords will be helpful in improving the strength of passwords in real data sources.

Most of users have various passwords for different web or mobile application accounts. However, it is difficult to remember so many passwords for a user. Although mobile devices have increasingly been used, it is still difficult to run a password manager over mobile devices. Besides, unfriendly on-screen keyboards make it more challenging or inconvenient to type passwords with special symbols or mixed-case characters [4]. Many websites and mobile applications (apps) require users to choose complicate passwords (e.g., mixed-case letters, digits, special characters) and the authentication of passwords becomes more complicated. In this case, the text-password input interfaces (e.g. touchscreen virtual keyboards) are applied to protect users' passwords from malwares [4].

In this paper, we will investigate these leaked passwords to comprehensively identify the strength of passwords. Basically, we will focus on four features of the passwords: 1) Length of passwords, 2) variety of character types in a password, 3) the randomness of passwords, and 4) uniqueness of passwords. Mathematically, we will analyze the password entropy, guessing entropy, and Minimum entropy for passwords in the leaked passwords lists. A number of password analyzing tools, including John the Ripper, Hashcat, and the password analysis and cracking toolkit (PACT) will be used to analyze the password lists for password length, password entropies, character types, pattern detection of masks, and other password features.

2 Background and Previous Work

There have been many research efforts made for helping users to choose passwords. Measuring the strength of passwords is an important topic. In [1], a method for calculating password entropy was proposed, which is based on the summarization of the distributions of passwords length, placements of character, and number of each character types. Yan et al. [2] filtered weak passwords by improving dictionary-based checking with 7 character alphanumeric passwords. The password quality indicator (PQI) was proposed to measure or evaluate the quality of passwords [3], [4].

2.1 Typical Hash-Based Login Systems

A typical website or mobile application login system contains the following four basic steps: 1) A user registers and uses an assigned password; 2) the password is hashed and stored in the database (but the plain-text password should never be written to the database); 3) when logging the system, the hash of the entered password will be checked against the hash saved in the database; 4) if the hashes match, the user will be granted access. The weakness in this system is that the passwords can often be guessed, forgotten, or revealed. To overcome this weakness, a stronger password should be created or additional authentication factors be used, such as physical token, digital certificates, one-time access code, etc. In many mobile app authentication systems, the mobile devices are increasingly used to enhance login security [11].

A lot of concerns have been raised on the security of web and mobile application login systems, however, we should bring awareness to the inherent weakness of such systems that are vulnerable to passwords cracking by considering the following situations:

- Most authentication of websites and mobile applications have not moved to stronger hash type. Many weak hash types, such as Unix DES, NTLM and single-round salted SHA1 are still in use.
- Existing password policies have led to exploitable predictability.
- Authentication systems with design flaws are vulnerable to pass-the-hash attacks, for example the websites like 163com, CSDN, Tianya (before 2014).
- Power graphics processing unit (GPU) can significantly speed up brute force attacks to weak hashes, or even for long password length.
- In some applications, the username and password combination are inadequate for strong authentication.

2.2 Password Attacks

A hacker can break passwords with many ways, out of which the following attacks are widely used:

1) Dictionary Attack

A hacker may use a password cracking dictionary (such as wordlist, dictionary, and password database leak) to find a password. The password dictionary is a very large text files that includes millions of generic passwords. The hacker may use higher performance computer or game graphics cards to try each of these passwords until find the right one [12].

2) Brute-Force Attack

The brute-force attack, or exhaustive password attack, is still one of the most popular password cracking methods. It tries every possible combination until it gets the password. In practice, the password space for all possible combinations might be huge, which makes the brute force attack very difficult to carry out.

3) Man-in-the-Middle Attack

The man-in-the-middle (MITM) attacks may be used for regular web/mobile apps logins. It is possible for an attacker to find out authentication requests/responses from the recorded network traffic, and then capture candidates for password related contents. Passwords attack dictionary can be built by choosing login information over highly popular websites.

Traditionally, strong passwords are created by the following methods:

1) Proper length of passwords. The length of passwords should be properly selected by balancing the user convenience and security. An eight-character password with a mix of numbers, symbols, and uppercase and lowercase can take at most months or years to crack. There is no minimum password length everyone agrees on, but in generally a password is required to be a minimum of 8 to 14 characters in length. A longer password would be even better [13].
2) Mix of numbers, symbols, uppercase and lowercase. It is very difficult to crack such password; the only techniques are to try huge number of combinations until find the right one. For an eight-character password there are 83^8 possible combinations and need 10 days and 2 hours to crack [13], [14].
3) Salt password and avoiding passwords listed in password cracking dictionary. The dictionary, wordlist, and password database are widely used in password cracking. In [12], a password cracking dictionary with a size of 15 GB has been released, which was used to successfully crack 49.98% of a password list with 373,000 passwords. To create a safer password, a better salting scheme is needed.
4) One-size-fits-all password. Most websites or mobile apps apply the one-size-fits-all approach to ensure that users choose strong passwords.
5) Outsourcing the security. It is a trend for most websites or mobile apps to outsource their security systems. This trend will seemingly continue.

In summary, it is not very difficult to create a strong password with proper length and a mix of many different types of characters. It is hard to guess such passwords due to its randomness. However, memorizing such a strong password is a problem. It is very difficult for most users to memorize a strong password created with a random password generator of websites and mobile applications. In creating a strong and memorable password, we need to think about how to avoid using something obvious with dictionary characters. For example, we can create a strong password based on a simple sentence like “I live in 20 Colinton Road at Edinburgh. The rent is $500 each month." We can easily turn this simple sentence into a strong password by using the first letter or digit of each word, as "Ili20CR.Edi $5em", which is a memorable and strong password with mix of numbers, characters, symbols, and uppercase letter and lowercase. It may be hacked in at least 420,805,123, 888,006 years [14].

3 Password Strength Metrics and Evaluation

Password strength measurements can help to warn users away from highly vulnerable passwords [15]. Many authentication systems of websites and mobile applications require passwords must be able to resist eavesdroppers and off-line analysis of authentication protocols run. In general, the security of passwords can be measured with password strength. Password strength is defined in terms of probability of a determined attacker discovering a selected users' password by an inline attack. The password strength is also a function of both the entropy of the password and the way unsuccessful trials are limited. Entropy is believed as a standard measure of security [5].

3.1 Password Entropy

Shannon entropy is a popular method to evaluate the security strength of a password, which is also used as password entropy. Assuming a finite variable X corresponds to n passwords set (p_1, p_2 ,..., p_n), the password entropy can be modeled with Shannon entropy as $H(X)$

$$H(X) = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i) \tag{1}$$

where p_i denotes the occurrence probability of ith possible outcome. A password using lowercase characters can be represented as $log_2(26) \approx 4.7$ bits of entropy per character. For a password "iliveinedinburgh" would have an entropy value of about $4.7 \times 16 \approx 75$ bits.

The Shannon entropy is commonly used to measure the passwords. Some variants of entropy have recently been proposed to measure other features of passwords such as guessing entropy, Minimum Entropy, and relative entropy.

3.2 Guessing Entropy

The ability of passwords that resists against complete off-line attacks can be measured with Guessing entropy. Guessing entropy is a measure of the difficulty to guess the passwords in a login system [6]. If the values of $Y = sort_d(X)$ are sorted with decreasing probability, the guessing entropy of Y can be defined as

$$G(Y) = \sum_{i=1}^{n} i \cdot p_i \tag{2}$$

The guessing entropy is closely related to the average size of passwords. If a password has n bits guessing entropy, an attacker has as much difficulty in guessing the average password as in guessing an n bits random quantity [16].

3.3 Minmum Entropy

Since in some cases, the password strength cannot warn users away from reusing the same password because they are usually based on heuristics (e.g., numbers, password length, upper/lowercase, symbols). Minimum entropy is a way to estimate the strength of a password, which is defined as

$$H_{\min}(X) = -\min \log_2(p_i) \tag{3}$$

For example, a low strength password p_b has low minimum entropy ($H_{min}(p_b) = 1$). High minimum entropy ($H_{min}(X) = \alpha$) guarantees that with high probability the adversary will always need to use around 2^{α} guesses to recover the users' passwords. The Minimum entropy shows the resistance of offline password cracking attacks with high probability.

3.4 Password over Mobile Applications

Many mobile applications require password input and the authentication task over mobile platforms is more complicated by using full-size key-board. In some mobile applications, the inconveniences caused by an unfriendly interface can affect users to create/use strong passwords. An example is that more than 80% of mobile device users are using digit-only passwords [6]. In recent, a number of password generation methods have been developed for mobile applications. For example, the object-based password (ObPwd) has been implemented over Android platform for generating password from a user-selected object (e.g., pictures) [7].

4 Analysis of the passwords

In this section, we investigated the way people create their passwords from five aspects: length, character types, randomness, complexity, and uniqueness. We analyzed over 100 million leaked and publicly available passwords from several popular websites (Rockyou, CSDN, TianYa, 163com).

4.1 Password Length

Most of the passwords have length of 6 to 10 characters as shown in **Tables 1** and **2** that specify the percentage of the total analyzed passwords.

▼**Table 1. Analysis of password length (Rockyou.txt)**

Length	Percentage (%)	Number of items
8	20	2,966,037
7	17	2,506,271
9	15	2,191,039
10	14	2,013,695
6	13	1,947,798

▼Table 2. Analysis of password length (163com.txt)

Length	Percentage (%)	Number of items
8	23	1,159,984
7	17	973,951
6	17	870,857
9	16	822,077
10	11	572,185
11	7	399,021
12	2	141,935
13	1	69,776
14	1	58,601

▼Table 3. Analysis of password character-sets (Rockyou.txt)

Character-set	Percentage (%)	Number of items
loweralphanum	88	4,720,183
upperalphanum	06	325,942
mixedalphanum	05	293,432

▼Table 4. Analysis of password character-sets (163com.txt)

Character-set	Percentage (%)	Number of items
Numeric	58	2,931,867
loweralphanum	30	1,527,719
loweralpha	08	450,746
loweralphaspecialnum	00	38,913
mixedalphanum	00	26,097
upperalphanum	00	23,905
specialnum	00	15,614
loweralphaspecial	00	4830
mixedalpha	00	4353
upperalpha	00	3142
All	00	2172
upperalphaspecialnum	00	1722
mixedalphaspecial	00	550
Special	00	164
upperalphaspecial	00	133

▼Table 5. Analysis of password advanced masks (Rockyou.txt)

Advanced Masks	Percentage (%)	Number of items
?l?l?l?l?l?l?d?d	07	420,318
?l?l?l?l?l?d?d	05	292,306
?l?l?l?l?l?l?l?d?d	05	273,624
?l?l?l?l?d?d?d?d	04	235,360
?l?l?l?l?d?d	04	215,074

From both Tables 1 and 2, we can find that 85% of passwords are between 8 to 10 characters long, which is pretty predictable. Around 50% of the passwords both in Rockyou and 163.com lists are less than eight characters. Few passwords have a length greater than 13. The main reason is that most websites and mobile apps require a maximum length of 8 and long passwords are difficult to be remembered.

4.2 Character Types

The diversity of the character types in passwords can be categorized into the following sets: number, uppercase, lowercase, and special-case.

The character-sets in Rockyou.txt and 163com.txt are shown in **Tables 3** and **4**.

The character-set analysis helps us understand the usability and security of passwords. It is good to consider three or more character types when creating a password. More than 80% Rockyou passwords had only one-character type (lowercase). In 163com, more than 88% of the passwords had only numeric passwords.

4.3 Randomness

In our investigation, we found that many of the usual culprits are used such as "password", "123456", "abc123", and city names. We also found that many passwords were apparently related to a combination: part of user names, city names, country names, etc. A few of these are very specific but there may be context to this in the sign up process.

We analyzed the mask of passwords in both Rockyou.txt and 163com.txt (**Tables 5** and **6**). Table 6 shows only 1% of all the passwords have the patterns matching the advanced masks and the majority is "string-digit" passwords that consist of a string with two or four digits.

4.4 Uniqueness

This uniqueness is about password sharing for different accounts. According to the analysis of many Chinese websites (12306, 163.com, 126.com, Tianya, CSDN, etc.), we found that many users are sharing the same passwords among their accounts in these websites.

It is believed that the leakage of such websites as 12306 and Tianya are caused by the hit-the-library attack, in which leakage of users' privacy data is more like that by a hacker hitting the library behavior. The hit-the-library attack is used by hackers to collect username, password, and other private information. After generating the corresponding dictionary with collected information, that attacker is able to attempt another batch landing sites. By this way, the hacker can deal with almost any website login systems. If a user uses the same username and password as the master key to log on different sites, he facilitates himself but also provides convenience for hackers.

In the Sony leakage case, 92% of passwords were reused across both in "Beauty" and "Delboca login systems. Only 8% of identical passwords are used. In internet web and mobile ap-

Table 6. Analysis of password advanced masks (163com.txt)

Advanced Masks	Percentage (%)	Number of items
?d?d?d?d?d?d	14	727,942
?d?d?d?d?d?d?d?d	13	701,557
?d?d?d?d?d?d?d	13	692,425
?d?d?d?d?d?d?d?d?d	06	348,786
?d?d?d?d?d?d?d?d?d?d?d	04	244,521
?d?d?d?d?d?d?d?d?d?d	03	162,921
?l?l?l?l?l?l?l?l	02	117,516
?l?l?l?d?d?d?d?d?d	01	90,281
?l?l?l?l?l?l?l?l?l	01	78,441
?l?l?l?l?l?l	01	74,678
?l?l?d?d?d?d?d?d	01	71,890
?l?l?l?l?l?l?l	01	67,221
?l?l?l?l?l?l?l?l?l?l	01	66,316
?l?l?d?d?d?d?d?d?d	01	65,743
?l?l?l?d?d?d?d?d?d?d	01	61,386
?l?d?d?d?d?d?d?d	01	53,065

plications, many users are using the same emails as their login usernames, which increases the risks of password sharing.

5 Conclusion

In this paper, we analyze the strength of passwords and investigate the password leakages cases from the viewpoints of length, character types, randomness, complexity and uniqueness, which is expected to warn users away from highly vulnerable passwords.

References

[1] W. E. Burr and D. F. Dodson. (2016, May 11). *Electronic Authentication Guideline: Recommendations of the National Institute of Standards and Technology, 1.0.2 ed.* [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-63/SP80063V1.pdf

[2] J. Campbell, D. Kleeman, and W. Ma, "Password composition policy: does enforcement lead to better password choices? " in *ACIS 2006*, Adelaide, Australia, 2006, Paper 60.

[3] J. Campbell, D. Kleeman, and W. Ma, "The good and not so good of enforcing password composition rules," *Information Systems Security*, vol. 16, no. 1, pp. 2–8, 2007. doi: 10.1080/10658980601051375.

[4] R. Cisneros, D. Bliss, M. Garcia, "Password auditing applications," *Journal of Computing in Colleges*, vol. 21, no. 4, pp. 196–202, 2006.

[5] M.-H. Lim and P. C. Yuen, "Entropy measurement for biometric verification systems," *IEEE Transactions on Cybernetics*, vol. 46, no. 5, pp. 1065–1077, May 2016. doi: 10.1109/TCYB.2015.2423271.

[6] M. Jakobsson, E. Shi, P. Golle, and R. Chow, "Implicit authentication for mobile devices," in *4th USENIX Conference on Hot Topics in Security*, Montreal, Canada, Aug. 2009, pp. 9–9.

[7] M. Mannan and P. C. van Oorschot. (2016, May 11). *Passwords for both mobile and desktop computers ObPwd for Firefox and Android* [Online]. Available: https://www.usenix.org/system/files/login/articles/mannan.pdf

[8] MakeUseOf. (2016, May 11). *What you need to know about the massive Linkedin accounts leak* [Online]. Available: http://www.makeuseof.com/tag/need- know-massive-linkedin-accounts-leak

[9] S. Fitz-Gerald. (2016, May 11). *Everything that happened in the Sony leak scandal* [Online]. Available: http://www.makeuseof.com/tag/need-knowmassive-linkedin-accounts-leak

[10] E. Yu. (2016, May 11). *130K users' data leaked via China's train ticketing site* [Online]. Available: http://www.zdnet.com/article/130k-users-data-leaked-via-chinas-train-ticketing-site

[11] M. Sarrel. (2016, May 11). *Authentication via Mobile Phone Enhances Login Security* [Online]. Available: http://www.informationweek.com/applications/authentication-via-mobile-phone-enhances-login-security/d/d-id/1103017?

[12] E. Escobar. (2016, May 11). *How long to hack my password* [Online]. Available: http://www.quickanddirtytips.com/tech/computers/how-to-crack-a-password-likea-hacker?page=1

[13] B. Buchanan. (2016, May 11). *Encryption* [Online]. Available: http://asecuritysite.com/encryption/passes

[14] Random-ize. (2016, May 11). *How long to hack my password* [Online]. Available: http://random-ize.com/how-long-to-hack-pass

[15] J. Blocki. (2016, May 11). *Password strength meters* [Online]. Available: http://www.cs.cmu.edu/ jblocki/entropyAndMinimumEntropy.htm

[16] NIST. (2016, May 11). *Electronic authentication guideline (NIST Special Publication 800-63)* [Online]. Available: http://itlaw.wikia.com/wiki/NIST-Special-Publication-800-63

Manuscript received: 2016-05-21

Biographies

LI Shancang (s.li@napier.ac.uk), PhD, is a lecturer in Network Forensics in School of Computing at Edinburgh Napier University, UK. Over the last few years, he has been working on a few research projects funded by EU, EPSRC, Academic Expertise for Business (A4B), Technology Strategy Board (TSB), and industry. Based on these research projects, dozens of papers have been published. His current research interests include network forensics, security, wireless sensor networks, the Internet of Things (IoT), and lightweight cryptography over IoT.

Imed Romdhani (i.romdhani@napier.ac.uk) is an associate professor in computer networking at Edinburgh Napier University, UK. He received his PhD from the University of Technology of Compiegne (UTC), France in May 2005, and an engineering and a master degree in networking obtained respectively in 1998 and 2001 from the National School of Computing (ENSI, Tunisia) and Louis Pasteur University of Strasbourg (ULP, France). He worked extensively with Motorola Research Labs in Paris and authored 4 patents in the field of IPv6, multicast mobility and IoT.

William Buchanan (w.buchanan@napier.ac.uk) is a professor in the School of Computing at Edinburgh Napier University, UK, and a fellow of the BCS and the IET. He currently leads the Centre for Distributed Computing, Networks, and Security and The Cyber Academy, and works in the areas of security, cloud security, web-based infrastructures, e-crime, cryptography, triage, intrusion detection systems, digital forensics, mobile computing, agent-based systems, and security risk.