Mail Service Password Security

George Markov
Information Security Department
Bauman Moscow State Technical University
Moscow, Russia
gm@cnpo.ru

Vladislav Sharunov
Department of Computing and IS
University of Greenwich
London, Great Britain
mrvo788@gmail.com

***Abstract*—As the first protection line for a computer system, the authentication system is of critical importance in the information security area. Despite the steady development of information security mechanisms, the password is the most commonly used authentication tool. The key vulnerability of such a protection mechanism is selecting an insecure password. The period of 2014-2017 saw major Internet companies suffer a number of password database leaks, which followed by a study of real password security. It should be noted that password system protection has not advanced much over the past years; mainly, there has been a clear tendency for imposing stricter requirements on the password entry interface. This being the case, there is still a question, yet to be answered, which passwords can be considered secure and which cannot. The work offers examples of password system assessment and reviews leaked passwords for their security under the current requirements. Security was verified using metrics (password security indices). These metrics provided the basis for defining objective requirements for password system security.**

Keywords— identification, authentication, password metric, information security, information protection, password system

I. Introduction

Even though the problems of password system security are subject to permanent studies, this issue remains yet to be dealt with in practice for subjective reasons. For example, some software system developers treat the password system security issue in different ways, while users often fail to meet in full the authentication system security policy.

Over the last two years, there have been a number of major leaks from password databases of communication Internet services (Twitter, Hotmail, Yandex, Google, Mail, Dropbox etc.). This allowed password security to be investigated using informal and formal indices.

II. Password Security Strength Concept

Let's consider the password mining probability formula:

$$P = \frac{V \cdot T}{|A|^n},$$

where V is attacker's password brute-forcing rate, T is password age, $|A|^n$ is password space capacity, n is password length.

The above formula allows a conclusion that password security is largely affected by password change frequency and password space capacity characterized by length and the alphabet used to create a password.

The above makes it possible to formulate simple substantive criteria whereby a password is considered secure:

- Password length must be at least 8 characters;
- Characters of various cases must be taken into account;
- Numeral must be used;
- Special characters must be used;
- Password must not be based on a word;
- Password must not include words relating to the password owner.

Meanwhile, there is a discussion underway in literature as to formal requirements for password systems [1-5, 7-11].

III. Regulatory Requirements

As is well known, the password requirements were set out in the Orange Book and were actually reduced to password length (6 and 8 characters depending on the system protection class). Current documents define the availability, as a minimum, of a password protection policy at an organization or in a computer system. Some requirements can be made more specific, while others are left in the hands of system administrators (Table 1).

IV. Password Security Metrics

Listed below are some of the best known password security index classes:

- Numerical metrics (e.g. Orange Book);
- Probabilistic metrics [2, 7, 8];
- Shannon informational entropy [6];
- Heuristic entropy modifications [3, 10, 11];
- Probabilistic entropy modifications [9].

Table 1. Information security regulatory documents

Document title	Minimum password length requirement	Password updating frequency/ security control requirement
Payment Card Industry Data Security Standard	7 characters	+ (90 days)
Australian Government Information Security Manual. Controls (Australia)	-	+ (90 days)
The IT-Grundschutz Catalogues (Germany)	-	+
Cyber Essentials Scheme Requirements for basic technical protection from cyber-attacks (Great Britain)	8 characters	+ (60 days, 3 months, 6 months)
Information Security Provisions in Federal Information Systems (Russia)	6, 8 characters	+ (180, 120, 90, 60 days)
Requirements for Information Security in Process Control Systems… (Russia)	6, 8 characters	+ (180, 120, 90, 60 days)
NIST SP 800-53/ NIST SP 800-63B (USA)	-	+
Information Assurance Implementation. Department of Defense Instruction 85002.2 2003 (USA)	8 characters	+ (3, 6 months)

Numerical metrics include password brute forcing time values. Unfortunately, this method takes no account of deliberate brute forcing and guessing.

Probabilistic metrics are based on the available password statistics for specific systems, which is not always practicable.

This work will consider the Shannon entropy and heuristic entropy (recommended by NIST SP 800-22). The methods differ in that the Shannon entropy assumes a password to be generated by a random-number generator, while heuristic entropy implies a human created password.

The Shannon entropy is calculated as follows:

$$H = \log_2|A|^n = n \cdot \log_2|A| = n\frac{\ln|A|}{\ln 2},$$

where $|A|$ is alphabet capacity, n is password length.

The metric suggests that the more complicated the alphabet and the longer the password, the more secure the latter is.

Given below is the Shannon entropy calculation example (Table 2).

Table 2. Entropy calculation example

Alphabet/length	5	6	7	8
Latin	23.5	28.2	32.9	37.6
Numerals	16.6	19.9	23.2	26.5
Latin + uppercase + numerals	29.7	35.7	41.6	47.6
Latin + Cyrillic + uppercase + numerals	35	41.9	48.9	55.9

NIST-recommended password entropy can be calculated by the following formula:

$$S = 4 + \sum_{i=2}^{8} 2 + \sum_{i=9}^{20} 1.5 + \sum_{i=21}^{n} 1 + 6\chi_A,$$

where $i \leq n$, n is password length, χ_A is characteristic function of a password containing non-alphanumeric or uppercase characters.

This formula can be described as follows: the first password character receives a value of 4 bits, each subsequent character from the second to the eighth one receives 2 bits, from the ninth to the twentieth – 1.5 bits and each subsequent – 1 bit. If there are non-alphanumeric or uppercase characters, 6 bits are added to the obtained result.

For these metrics, a password is considered secure if it conforms to the entropy [11]:

- according to Shannon – 56 bits or more;

- on NIST recommendations – 24 bits or more.

The above criteria should be restricted, i.e. if a password is recorded in password brute forcing databases (dictionaries), entropy is reduced to zero.

V. Study Outcomes

Password databases have been very often compromised of late. For example, such password databases became publicly available more recently

Research software was used to process a few password databases made publicly available by hackers on the Internet last year. Presented below is certain statistics obtained for each password database. The outcomes of the study into the compromised password database of Yandex (1,261,809 passwords) are listed in tables 3-5, Mail.ru (45,000) – tables 6-8, Google (4,926,673) – tables 9-11.

Table 3. Password length (Yandex)

Password length	Number of passwords
6	380,732
7	174,782
8	282,641
9	130,676
10	103,926
11	71,948
12	45,387
13	20,127
14	14,950
15	9,895
16	7,646
17	3,487
18	3,104
19	1,747
20	2,660

Table 4. Top 10 repeated passwords (Yandex)

Password	Number of repetitions
123456	39,177
123456789	13,892
111111	9,826
qwerty	7,926
1234567890	5,853
1234567	4,668
7777777	4,606
123321	4,324
000000	3,304
123123	3,031

Table 5. Password alphabet (Yandex)

Used alphabet	Number of passwords
Passwords (PWD) including numerals only	608,125
PWD composed of characters	233,561
PWD composed of lowercase letters only	218,319
PWD composed of uppercase letters only	3,136
PWD similar to mobile phone number	40,980
PWD coinciding with login	1,489
PWD similar to dates	171,906
PWD suitable for secure password informal description	345
PWD suitable as per Shannon security	143,802
PWD suitable as per NIST security	108,951

Table 6. Password length (Mail.ru)

Password length	Number of passwords
6	17,484
7	4,155
8	12,562
9	3,212
10	2,421
11	1,399
12	1,106
13	627
14	438
15	293
16	205
17	12
18	20
19	2
20	15

Table 7. Top 10 repeated passwords (Mail.ru)

Password	Number of repetitions
qwerty	4,291
987654321	1,385
4815162342	661
11111111	615
123123123	578
789456123	448
12341234	408
147852369	380
444444	353
q1w2e3	331

Table 8. Password alphabet (Mail.ru)

Used alphabet	Number of passwords
PWD including numerals only	18,806
PWD composed of characters	14,650
PWD composed of lowercase letters only	13,835
PWD composed of uppercase letters only	53
PWD similar to mobile phone number	138
PWD coinciding with login	3,619
PWD similar to dates	9,287
PWD suitable for secure password informal description	5
PWD suitable as per Shannon security	3,916
PWD suitable as per NIST security	3,274

Table 9. Password length (Google)

Password length	Number of passwords
6	924,154
7	663,510
8	1,422,999
9	683,315
10	682,811
11	152,256
12	93,202
13	42,387
14	24,853
15	14,851
16	7,291
17	2,549
18	1,781
19	1,082
20	1,166

Table 10. Top 10 repeated passwords (Google)

Password	Number of repetitions
123456	47,918
password	11,554
123456789	11,160
12345	8,096
querty	5,918
12345678	5,250
111111	3,521
abc123	3,011
123123	2,972
1234567	2,911

Table 11. Password alphabet (Google)

Used alphabet	Number of passwords
PWD including numerals only	774,669
PWD composed of characters	1,968,873
PWD composed of lowercase letters only	1,968,873
PWD similar to mobile phone number	22,751
PWD coinciding with login	45,010
PWD similar to dates	156,142
PWD suitable for secure password informal description	0
PWD suitable as per Shannon security	290,530
PWD suitable as per NIST security	157,475

VI. Conclusion

Comparative analysis of the obtained and earlier known statistics [12, 13] showed a trend for slight strengthening of password protection. This is because some Internet services defined stricter rules for interfaces, for example, strengthened requirements for password length (at least 6 characters) and the use of a relatively complicated alphabet. The statistics suggests, however, that the above fact does not stop unorganized and careless users from choosing easily hackable passwords, and the number of Top 500 passwords has hardly changed over the years.

In general, the study confirmed that the authentication system remains highly vulnerable (only 10% of passwords can be considered reliable), which prompts the creation of integrated information protection systems and the improvement of information security management systems.

Finally, it should be noted that using entropic metrics instead of verbal descriptions is more practical in defining technical requirements for information security systems, as they are easier to automate and control. Besides, the use of formal indices helps diminish the degree of subjectivity inherent in system security analysis.

References

[1] Bonneau J. Guessing human-chosen secrets. Technical Report UCAM-CL-TR-819. 2012. 161 p.

[2] Bonneau J. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In 2012 IEEE Symposium on Security and Privacy, IEEE, 2012, pp. 538 – 552. DOI: 10.1109/SP.2012.49.

[3] Boothroyd V., Chiasson S. Writing down your password: Does it help? In Proc. of the 2013 Eleventh Annual Conference on Privacy, Security and Trust, IEEE, 2013, pp. 267 – 274. DOI: 10.1109/PST.2013.6596062.

[4] Burnett M. Perfect Password: Selection, Protection, Authentication. Syngress Publishing, 2006.194 p.

[5] Burr W.E. and etc. Electronic Authentication Guideline. NIST Special Publication 800-63-1. 2011. 110 p.

[6] Christiansen M.M., Duffy K.R. Guesswork, Large Deviations, and Shannon Entropy, IEEE Transactions on Information Theory, 2013, volume 59, issue 2, pp. 796 - 802 DOI: 10.1109/TIT.2012.2219036.

[7] Galbally J., Coisel I., Sanchez I. A probabilistic framework for improved password strength metrics. In Proc. of the 2014 International Carnahan Conference on Security Technology (ICCST), IEEE, 2014, pp. 1 – 6. DOI: 10.1109/CCST.2014.6986985.

[8] Galbally J., Coisel I., Sanchez I.A New Multimodal Approach for Password Strength Estimation—Part I: Theory and Algorithms. IEEE Transactions on Information Forensics and Security, 2017, volume 12, issue 12, pp. 2829 – 2844. DOI: DOI: 10.1109/TIFS.2016.2636092.

[9] Groza B. Analysis of a Password Strengthening Technique and Its Practical Use. In Proc. of 2009 Third International Conference on Emerging Security Information, Systems and Technologies, IEEE, 2009, pp. 292 – 297. DOI: 10.1109/SECURWARE.2009.52.

[10] Kelley P.G. and etc. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In proc. of the 2012 IEEE Symposium on Security and Privacy, IEEE, 2012, pp. 523 – 537. DOI: 10.1109/SP.2012.38.

[11] Markov G., Sharunov V. About Information Security of Email Services. Voprosy kiberbezopasnosti *[Cybersecurity issues]*, 2015, No 5 (13), pp. 55-59. DOI: 10.21681/2311-3456-2015-5-55-59.

[12] The Top 500 Worst Passwords of All Time, 2008. URL: http://www.whatsmypass.com/the-top-500-worst-passwords-of-all-time. (Last access: 01/12/2017).

[13] The Top 500 Worst Passwords of All Time. 2010. URL: https://www.symantec.com/connect/blogs/top-500-worst-passwords-all-time. (Last access: 01/12/2017).