УДК 81’33:004

Е.А. Юрковская[1], Е.А. Юрковский[2], Н.С. Балданова[1]

[1]Иркутский государственный университет путей сообщения
[2]Иркутский национальный исследовательский технический университет

ЛИНГВИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ НАДЕЖНОЙ ПАРОЛЬНОЙ ФРАЗЫ

Аннотация. *Аутентификация по паролю остается одним из наиболее популярных методов авторизации пользователей в корпоративных системах, социальных сетях, различных приложениях. В статье затрагивается актуальная проблема обеспечения безопасности пользовательских паролей, способных противостоять современным хакерским методикам взлома паролей, в том числе, методу брутфорс («грубого перебора»). Наиболее надежным типом пароля признается парольная фраза и анализируются лингвистические приемы оформления парольной фразы, способные увеличить ее информационную энтропию. К таким приемам относятся сочетание слов, не имеющих очевидного общего семантического контекста, несоблюдение правил английской грамматики, использование слов из разных языков и транслитераций.*

Ключевые слова: *парольная фраза, энтропия парольной фразы, лексический состав парольной фразы, синтаксис парольной фразы, мнемоника парольной фразы.*

E.A. Iurkovskaia[1], E.A. Iurkovskii[2], N.S. Baldanova[1]

[1]Irkutsk State Transport University
[2]Irkutsk National Research Technical University

LINGUISTIC FEATURES OF A SECURE PASSPHRASE

Abstract. *Password authentication remains one of the most popular methods of user authorization in corporate systems, social networks, and various applications. The article touches upon the urgent problem of ensuring the reliability of user passwords capable of resisting modern hacker methods of password cracking, including the brute force attack technique. A passphrase is considered to be the most reliable type of password and linguistic techniques of structuring a secure passphrase that can increase its information entropy are analyzed. These techniques include combining words having no evidently common semantic context, not observing the rules of English Grammar, using words from different languages and transliterations.*

Keywords: *passphrase, passphrase entropy, passphrase lexical composition, passphrase syntax, passphrase mnemonics.*

Introduction

Passwords remain the most widely used mode of user authentication that is why the password problem is one of the most researched issues in the world’s IT community.

The password problem refers to the challenges and vulnerabilities associated with creating, managing, and securing passwords, which often leads to weak passwords and increased security risks [1].

A password is a string of characters used to authenticate a user and provide access to a system or application [2]. Passwords started with the Compatible Time-Sharing System (CTSS), an operating system introduced at MIT in 1961. It was the first computer system to implement a password login. Although the increase in data breaches, social engineering attacks, and cybercrimes has done harm to the reputation of passwords, people are still using them as a medium to ensure security on their platforms [3].

Common Types of Passwords

The classical and most-widely used types of passwords are

1. alphabetic that consist of the letters a-z in the upper and lower cases;
2. numeric that are composed of numbers 0-9;
3. alphanumeric, a more complex string of letters, numbers and special characters.

With this type anything from a to z and 0 to 9 counts and you might be required to mix uppercase and lowercase versions. Unusual symbols from dashes to dollar signs to parenthesis are included.

Information security specialists specify that protection of information systems and infrastructure are most easily compromised by bad password management. Since users have to create their own passwords, it is highly likely that they won't create a secure password. It might be because users want to have a password that is easy to remember, use their name or birthdate in their passwords or reuse similar passwords across different networks and systems which makes their passwords vulnerable to the top password security risks.

Common Password Security Risks and Solutions

One of the easiest ways to get access to someone's password is to have them tell you. Through this method, hackers can even bypass the password authentication process by

- tricking users into typing their passwords into malicious websites they control (known as phishing);
- infiltrating insecure, unencrypted wireless or wired network (commonly known as sniffing),
- installing a keylogger (software or hardware) on a computer;

Systems that allow users to recover or reset their password if they have forgotten it can also let malicious actors do the same. Cybercriminals can mimic users and attempt to gain access to users' accounts by trying to reset the password.

More sophisticated methods use special software or automated tools to generate billions of passwords and trying each one of them to access the user's account and data until the right password is discovered. Password-cracking programs create variations from a dictionary of commonly known passwords, or attempt every possible combination using a method called a brute force attack [4].

A brute force attack or cracking is one in which an attacker will try all combinations of letters, numbers, and symbols according to the password rules, until they find the one that works. Billions of passwords can be tried per second.

In an effort to contribute to making authentication more secure, Mark Burnett, an infosec consultant and the author of the book "Perfect Passwords", revealed 10 million leaked username/password combinations that he had collected from the Web. This is one of the largest data collections of this sort available to the public. Among the 10 million passwords in the dataset, only about one-half are alphanumeric and unique as the top 20 most popular passwords are preferably either alphabetic such as *password, qwerty, letmein*; or numeric, e.g. *123456, 111111, 696969* [5].

Alphanumeric passwords proved to be much stronger and less vulnerable to cracking than alphabetic and numeric and this type is the one suggested by state-of-the-art Random Password Generators like LastPass [6], 1Password [7] and others.

The main advantage of an alphanumeric password is much entropy. Entropy is defined as unpredictable randomness [8]. Entropy can be mathematically calculated and measures a password's strength in a unit called bits. The stronger a password is, the more bits of entropy it has.

Password entropy also categorizes passwords into four types depending on their strength - very weak, weak, strong, and very strong [9].

Entropy results from the following features of a password:

1. it is long, it uses more than 10 characters (a random 13-digit string of characters would be strong enough to resist the most advanced cracking programs of today [4]);
2. it is difficult for someone else to guess, it uses both uppercase and lowercase letters, numbers, and symbols and includes no obvious personal information or common words;
3. it is unique as it cannot be found in any dictionaries of leaked or compromised passwords.

But such strong high-entropy passwords lead to another problem. They might be too complicated to remember. Randall Monroe, an American programmer, engineer and cartoonist muses, "we've successfully trained everyone to use passwords that are hard for humans to remember, but easy for computers to guess." [10].

With his famous cartoon [11] Monroe suggests a solution and uses math to form a password that is easy to remember but has enough entropy that it will take centuries for a computer hacking program to guess. His ideas are described in Figure 1.

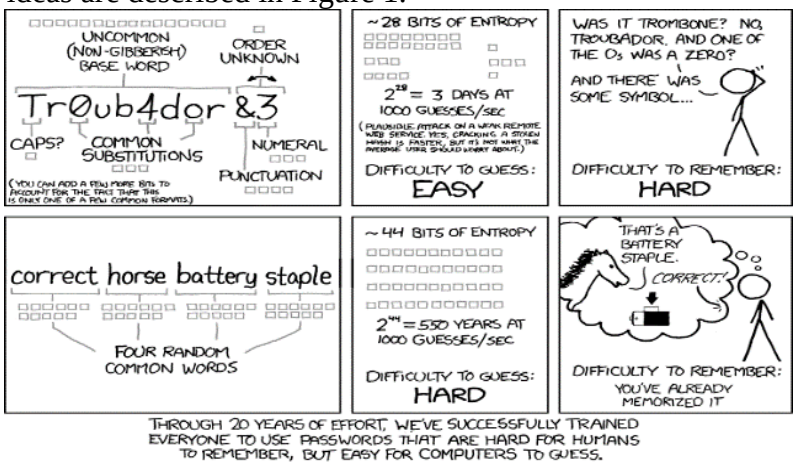

Figure 1. Monroe's cartoon on password security

Passphrases

The solution Monroe suggests is to use passphrases instead of passwords. A passphrase is typically longer and contains spaces. A passphrase can also contain symbols, and it does not need to be grammatically correct [12].

Monroes suggests using four common yet unrelated words, and remembering a situation that involves all four of them. His example *correcthorsebatterystaple* is a much stronger password (16 bits of entropy more) than the alphanumeric *Tr0ub4dor&3* even if it seems like it wouldn't be. But the most significant advantage of a random passphrase is that it provides the best combination of memorability and security.

Linguistic Features of a Secure Passphrase

A passphrase is the point where Linguistics can assist Cyber Security. To contribute to the idea, some linguistic features of passphrases can be suggested to maximize entropy of a passphrase. They are meant to make it look as much different from a grammatically correct and meaningful sentence as practicable. Such passphrases are not expected to be found in brute force attackers' dictionaries as they commonly include "common words or phrases found in dictionaries" [13].

1. Words in a passphrase mustn't belong to the same thematic group based on common contextual associations.

So, a bad passphrase would be *studentstudyhomeworkathome* as all the words are associated with the same semantic context of a learning environment.

A good passphrase would be *riverstudylaw-abidingon Monday*. All the words demonstrate different semantic references therefore it will be problematic for a cracking program to get them together in one passphrase.

2. Words in a passphrase must be different parts of speech, both content (nouns, verbs, adjectives, adverbs, numerals) and function words (prepositions, conjunctions, modal verbs). It can also be of various grammar types (singular and plural nouns, possessive nouns, verb tense forms and verbals, comparative and superlative adjectives).

A bad passphrase would be *rivertablegirldog* as all the words are countable singular nouns.

A good passphrase would be *riverstudylaw-abidingon Monday*. All the words in the passphrase belong to different parts of speech – *river* is a noun, *study* is a verb, *law-abiding* is a compound adjective, and *on Monday* is a prepositional phrase.

In order to make a brute forcer's task even more difficult, it is a good idea to use the words in various grammar forms, e.g. *riversstudiedlaw-abidingon Monday* where *rivers* is a plural noun, *studied* is a Past Simple verb.

3. The syntax of a passphrase must be incorrect. Word order is one of the fundamental aspects of English syntax. It determines how words are arranged in a sentence to convey the intended meaning clearly and effectively. The model syntax of an English sentence is

(adjective) + Subject noun + (adverb) + Verb + ((adjective) + object noun) + (adverb)

A confusing passphrase can start with a verb or an adverb and finish with an adjective. The wrong choice of prepositions ore use on articles may also be a useful factor.

A good passphrase would be *studiedlaw-abidingon Mondayrivers*. The sentence starts with a verb and a prepositional phrase precedes a noun which is a syntactical violation.

An even better passphrase could be *studiedthelaw-abidingin Mondayrivers*. In the phrase the adjective *law-abiding* is wrongly preceded by the article, and the incorrect preposition *in* is used before *Monday*.

4. A passphrase can include words from different languages both spelt in English and, if the system allows, in another language.

Thus, the good passphrase *studiedthelaw-abidingin Mondayrivers* can be transformed to an even better passphrase *studiedthelaw-abidingy понедельникreki.* The English wrong prepositional phrase *in Monday* was changed to the equally wrong *y понедельник* and the word *rivers* is transliterated to *reki.*

The final step in creating a passphrase should be take care to memories it. It is necessary to think of a situation where all these randomly selected words get together, it is expected to be absurd and funny enough to serve as a mnemonic device. We can easily imagine a situation like this: in the morning (*y* is the first letter of this word translated into Russian) on Monday (*понедельник* in Russian) some *law-abiding* people *studied* Russian rivers (*reki*). The order of the elements can be practiced by asking and answering questions about this situation:

What did people do? – *studied*

What kind of people were they? – *thelaw-abiding*

When did they study? – *y понедельник*

What did they study? – *reki*

The linguistic transformations described above are meant to provide the passphrase as much entropy as possible which can be tested with password entropy calculators. The passphrase *studiedrekithelaw-abidingy понедельник.* was tested with 2 calculators estimating results in accordance with the two basic entropy indications – level of strength and entropy value. As Figure 2 demonstrates, both of them proved the highest level of strength – the 4th level (very strong) out of 4 possible [14]; and 222.60 bits of entropy [15] with the sufficient value of 78 bits [9].

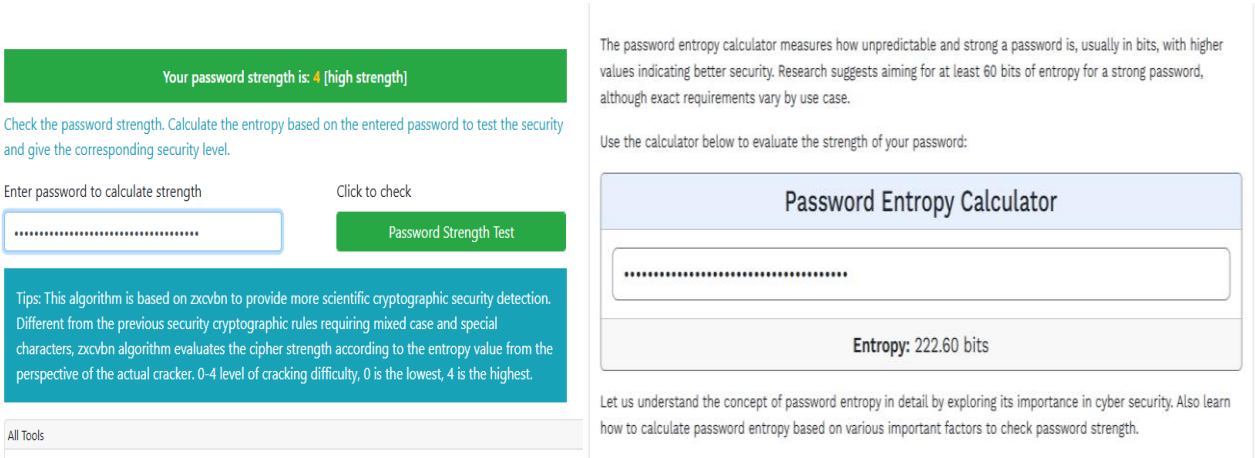

Figure 2. Results of entropy value tests with password strength calculators [14,15]

Conclusion

The art and science of Cyber Security is currently evolving and security specialist are developing new reliable authentication methods such as one-time passwords, using multi-factor authentication, biometrics, hardware keys and others [16]. Nevertheless, when it comes to having to

create a new password, a passphrase made according to the features described in the article may be a good idea.

An ideal passphrase should demonstrate certain linguistic features. It will be difficult to guess due to its high entropy if it includes words having no evidently common semantic context; it does not observe the rules of English Grammar, it is double lingual or uses transliterated words.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Understanding Alpha Characters in Passwords for Better Cyber Security // PassBits. URL: https://www.passbits.com/blogs/what-is-an-alpha-character-in-a-password.html (дата обращения: 30.04.2025).
2. Types of Passwords: What to Know & How to Pick the Best One // Password Clinic. URL: https://passwordclinic.com/creating-the-safest-password/types-of-passwords-what-to-know-how-to-pick-the-best-one/ (дата обращения: 30.04.2025).
3. Richard M. 5 reasons why passwords are no more safe – What's next? // Shufti. URL: https://shuftipro.com/blog/5-reasons-why-passwords-are-no-more-safe-whats-next/ (дата обращения: 30.04.2025).
4. How Password Cracking Work // Keeper. URL: https://www.keepersecurity.com/blog/2016/09/28/how-password-crackers-work/ (дата обращения: 30.04.2025).
5. Burnett M. Perfect passwords: selection, protection, authentication. Rockland, Mass.: Syngress, 2006. 202 p.
6. LastPass : online password manager. URL: https://www.lastpass.com/ (дата обращения: 30.04.2025).
7. 1Password: online password manager. URL: https://1password.com/ (дата обращения: 30.04.2025).
8. What is Entropy In Cryptography And Encryption? // Quside. URL: https://quside.com/what-is-entropy-in-cryptography-and-encryption/ (дата обращения: 30.04.2025).
9. What Is Password Entropy? // IDStrong. URL: https://www.idstrong.com/sentinel/what-is-password-entropy/ (дата обращения: 30.04.2025).
10. Johnsons D. Get Better Security with Plain English Passwords // CBS News. URL: https://www.cbsnews.com/news/get-better-security-with-plain-english-passwords/ (дата обращения: 30.04.2025).
11. Password Strength // xkcd. URL: https://xkcd.com/936/ (дата обращения: 30.04.2025).
12. Password vs. Passphrase: Differences Defined & Which Is Better? // Okta. URL: https://www.okta.com/identity-101/password-vs-passphrase/ (дата обращения: 30.04.2025).
13. Unlocking the Power of Brute Force: A Dictionary Approach // Metaversum. URL: https://metaversum.it/unlocking-the-power-of-brute-force-a-dictionary-approach/ (дата обращения: 30.04.2025).
14. Password Test: password strength calculator // ToolPie. URL: https://strength.toolpie.com/; (дата обращения: 30.04.2025).
15. Password Entropy Calculator // Insecure. URL: https://www.insecure.in/tools/password-entropy-calculator (дата обращения: 30.04.2025).
16. Glover R. 2024's least and most secure authentication methods // 1Password Blog. URL: https://blog.1password.com/authentication-methods/ (дата обращения: 30.04.2025).

REFERENCES

1. Understanding Alpha Characters in Passwords for Better Cyber Security. *PassBits*. Available at https://www.passbits.com/blogs/what-is-an-alpha-character-in-a-password.html (accessed 30 April 2025).

2. Types of Passwords: What to Know & How to Pick the Best One. *Password Clinic*. Available at https://passwordclinic.com/creating-the-safest-password/types-of-passwords-what-to-know-how-to-pick-the-best-one/ (accessed 30 April 2025).
3. Richard M. 5 reasons why passwords are no more safe – What's next? *Shufti*. Available at https://shuftipro.com/blog/5-reasons-why-passwords-are-no-more-safe-whats-next/ (accessed 30 April 2025).
4. How Password Cracking Work. *Keeper*. Available at https://www.keepersecurity.com/blog/2016/09/28/how-password-crackers-work/ (accessed 30 April 2025).
5. Burnett M. *Perfect passwords: selection, protection, authentication*. Rockland, Mass.: Syngress, 2006, 202 p.
6. *LastPass*: online password manager. Available at https://www.lastpass.com/ (accessed 30 April 2025).
7. *1Password*: online password manager. Available at https://1password.com/ (accessed 30 April 2025).
8. What is Entropy In Cryptography And Encryption? *Quside*. Available at https://quside.com/what-is-entropy-in-cryptography-and-encryption/ (accessed 30 April 2025).
9. What Is Password Entropy? *IDStrong*. Available at https://www.idstrong.com/sentinel/what-is-password-entropy/ (accessed 30 April 2025).
10. Johnsons D. Get Better Security with Plain English Passwords. *CBS News*. Available at https://www.cbsnews.com/news/get-better-security-with-plain-english-passwords/ (accessed 30 April 2025).
11. Password Strength. *xkcd*. Available at https://xkcd.com/936/ (accessed 30 April 2025).
12. Password vs. Passphrase: Differences Defined & Which Is Better? *Okta*. Available at https://www.okta.com/identity-101/password-vs-passphrase/ (accessed 30 April 2025).
13. Unlocking the Power of Brute Force: A Dictionary Approach. *Metaversum*. Available at https://metaversum.it/unlocking-the-power-of-brute-force-a-dictionary-approach/ (accessed 30 April 2025).
14. Password Test: password strength calculator. *ToolPie*. Available at https://strength.toolpie.com/ (accessed 30 April 2025).
15. Password Entropy Calculator. *Insecure*. Available at https://www.insecure.in/tools/password-entropy-calculator (accessed 30 April 2025).
16. Glover R. 2024's least and most secure authentication methods. *1Password Blog*. Available at https://blog.1password.com/authentication-methods/ (accessed 30 April 2025).

Информация об авторах

Юрковская Елена Александровна – канд. филол. наук, доцент кафедры «Иностранные языки», Иркутский государственный университет путей сообщений, г. Иркутск, e-mail: eayur@mail.ru.

Юрковский Егор Алексеевич – студент группы ИСТб-21-1, Институт информационных технологий и анализа данны, Иркутский национальный исследовательский технический университет, г. Иркутск, e-mail: egoriurkovsky@yandex.ru.

Балданова Наталья Сергеевна – студентка группы БИ.4-24-1, Иркутский государственный университет путей сообщений, г. Иркутск, e-mail: baldanovanatasa417@gmail.com.

Information about the authors

Iurkovskaia Elena Aleksandrovna – Ph.D. in Philology, Associate Professor, the Subdepartment of Foreign Languages, Irkutsk State Transport University, Irkutsk, e-mail: eayur@mail.ru.

Iurkovskii Egor Alekseevich – student of Group ISTb-21-1, Information Technology and Data Analysis Institute, Irkutsk National Research Technical University, Irkutsk, e-mail: egoriurkovsky@yandex.ru.

Baldanova Natalia Sergeevna – student of Group BI.4-24-1, Irkutsk State Transport University, Irkutsk, e-mail: baldanovanatasa417@gmail.com.