Information & Security

P. Antonov & N. Georgieva
vol. 47, no. 2 (2020): 276-282
https://doi.org/10.11610/isij.4719

Published since 1998

ISSN 0861-5160 (print), ISSN 1314-2119 (online)

Research Article

Security Analysis of Diceware Passphrases

Petar Antonov and Nikoleta Georgieva (✉)

Nikola Vaptsarov Naval Academy, Varna, Bulgaria, http://nvna.eu/

ABSTRACT:

The purpose of this study is to analyse the security of Diceware passphrases in comparison with various symmetric authenticated encryption schemes against the brute force attack. It proves that the security of these passphrases is deficient and, therefore, its use should end gradually. Additionally, this study offers ways on how Diceware could enhance its security.

ARTICLE INFO:

RECEIVED: 11 JULY 2020
REVISED: 16 AUG 2020
ONLINE: 22 SEP 2020

KEYWORDS:

authentication, security, password, diceware passphrase

Introduction

Currently, the most common form of authentication used to control the access of computer communication systems are passwords. Even though there are various recommendations and requirements of the use and creation of strong passwords, in reality, we can say that there have been multiple occasions where password authentication has failed to provide security. Typically, it is because of the use of weak authentication protocols as well as password storage.[1, 2, 3] The most common causes are: the use of weak passwords, the use of specific lengths based on the authentication schemes (some shorten the passwords while some make them longer), the use of not strong enough hash functions and small size of their hash passwords, the storage, and sharing of passwords in an unprotected way in computer communication systems, etc. In this report, we are going to examine the security of Diceware passphrases compared to other authentication schemes and determine whether Diceware should be used in the future or not.

✉ Corresponding Author: Tel.: +359882416801; E-mail: nikoletha@gmail.com

Methods

The security analyzation of Diceware passphrases is done by comparing Diceware with other common encryption methods and calculating their security by using the most common encryption formulas.

Analysis

In order to examine the security of Diceware passphrases, we must discuss how passwords and passphrases are created and the difference between them. Usually, passwords contain words as well as in some cases phrases and the only difference between them is the number of used symbols. Typically, passwords contain between 8 to 12 symbols and passphrases contain up to 100 and even more, which, therefore, makes passphrases significantly more secure. It is natural for users to choose short and meaningful for them passwords, which can be remembered very easily, but at the same time, these passwords can be cracked in a very short time. This is the reason why, passphrases that contain lengthy keywords with capital and small letters, numbers, and special characters are preferred. These passphrases, however, are very difficult to remember. This is the reason why the best way to protect a system is by the use of passphrases that contain random words and objects that are around the user that can be remembered with ease.

In recent years the most common method for password generation is Diceware passphrase.[3, 9, 10] Diceware uses dice that generate random numbers and in turn, these numbers pick random words from a carefully selected Diceware Word List. The Word List contains 7776 English words with a maximum length of 6 letters and the average length of 4.2 letters. Each word inside the Diceware Word List is linked with a 5-digit number (from 11111 to 66666) and every digit ranges from 1 to 6. In order to generate a new passphrase, one by one dice is thrown to generate a number that picks the word that would be inside the passphrase. If the system only has one dice in order to generate 4 words, it must be thrown 20 times, if the system has 5 dice, they must be thrown four times, etc. Initially, Diceware Word List was only in English, but later it was created in Russian, Spanish, and French, Italian, Japanese, and many more common languages as well as Bulgarian.[6] In July of 2016 the Electronic Frontier Foundation (EFF) introduced a new Diceware Word List, which also contains 7776 words, but with an average length of 7 letters.[6]

In order to test the security of passwords, experts use the formula bellow, where E is the measure of ambiguity and the results is measured in bits. In this formula, L is the length of the passwords (meaning the number of symbols used), M is the total number of the alphabet where this word is chosen from. For example, if the length of the password is 9 symbols and it contains only small English letters, the security of the password with this formula (1) will be E=(9.lg26)/lg2=42.3 bits.

$$E = \log_2 M^L \, [\text{bits}], \qquad (1)$$

The same method is used when evaluating the security of Diceware passphrases. In the document [10] is noted that each extra word in the passphrase increases E with 12.9 bits. A phrase with 5 words is 64.6 bits, passphrases with 6 words are 77.5 bits, etc. Those numbers are calculated from the formula, where M=776 words and L is the number of the words in the passphrase. With L=1, E=(lg7776)/lg2=12.92 bits, with L=5, E=5.12,92=64.6 bits, etc. Further, Reinhold claims that Diceware passphrases with 6 words are secure enough and passphrases with 8 words will be secure up until 2050.[9]

However, this study shows that the security of the passphrases when using formula (1) is too abstract and cannot be used when determining the security of the whole communication computer system when only one part of it is authentication. This is the reason why, the methods in [4, 5] should be used instead because it compares the security of the most popular symmetric encryption schemes against the brute force attack.

If, for example, a basis of comparison is the most popular symmetric cipher AES (with the length of the keys being 128, 192 and 256 bits) and if the key phrase uses randomly chosen words from the Diceware Word List, then the security of the password phrase will be equivalent to the strength of the symmetric cryptographic algorithm when the ratios are used:

$$2^{128}=7776^{L},\ 2^{192}=7776^{L} \text{ и } 2^{256}=7776^{L}, \qquad (2)$$

Where L is the number of the words in the passphrases.

From here we get that $L_{128}=\lceil 9.9 \rceil=10$, which means that if the key phrase contains 10 randomly chosen words from the Diceware Word List, then the sustainability of the passwords against the brute force attack could be compared with the security of AES -128. With AES-192 and AES-256 the result is accordingly $L_{192}=\lceil 14.85 \rceil=15$ и $L_{256}=\lceil 19.8 \rceil=20$. The following is valid:

$$L_{Nks} = \lceil N_{ks} \,.\, lg2/lg7776 \rceil, \qquad (3)$$

Where N_{ks} is the length (the size) in bits of the secret key in the symmetric cipher, which is used for comparison.

In relation to (3), it could be concluded that no matter the size of the secret key N_{ks} the security of a Diceware passphrase with 6 words is:

$$6 = \lceil N_{ks} \,.\, lg2/lg7776 \rceil,\quad N_{ks} = \lceil 6.lg7776/lg2 \rceil = \lceil 77{,}55 \rceil = 78 \text{ bits}, \qquad (4)$$

Currently, the lowest value of N_{ks} that is acceptable is 128 bits and therefore based on this answer, we can conclude:

1. The use of entropy for evaluating passphrase security is inappropriate. The best way to ensure a passphrase's security is to use and compare the security of the most popular symmetric ciphers against the brute force attack.
2. A secure enough Diceware passphrase must contain 10, 15 and even up to 20 words, which is impractical and cannot be remembered or preferred by users.

3. There is a need of developing a way to enhance the security of Diceware passphrases.

The first way to enhance Diceware's security is to increase the size of its Word List and its origin. An organization can create and use its own Word List with the appropriate size and that list can only be available in their organizations' domain. This way of enhancing the security of Diceware could be seen in (4).

In order for this idea to work, the Word List must contain M words which are linked with randomly generated numbers, where the numbers contain K-digits and use a Q counting system. Then $M=Q^{K.}$. The required degree of security of the passphrase could be set according to the value of the N_{ks}, which tests against the brute force attack. In this case, the needed word count L inside the passphrases will be calculated according to this formula.

$$L=\lceil N_{ks}.lg2/lgM \rceil = \lceil N_{ks}.lg2/K.lgQ \rceil. \qquad (5)$$

If a user sets the value of N_{ks}, L would be the needed length of the key phrase depending on the size of the word list. If these words are picked in a completely random way, then the security of the key phrase will be the identical to the security against the brute force attack of the cryptographic algorithm with N_{ks} bits.

It is clear that the needed values of L will gradually decrease depending on the increase of M. The opposite could be said as well. Dictionaries that contain big sizes of M values can easily be created, however they can only be used as a One-Time-Pad and the values of Q and K cannot be related to the values of M and L.

However, there is an analysis that could predetermine the values of M, L, Q and K for every cryptographic scheme and it is the following. First, the value of M is determined and from the value we can create a personal Word List. That personal Word List can be created from an already existing one in a certain language, where the order of the words would be selected from a random number generator. By using the number generator, the organization can set the limits for the minimum and maximum size of the words. When this kind of dictionary/Word List is created, its security is going to depend on the number of the words that it contains L.

For example, if the created dictionary contains $M=Q^K=2^{14}=16384$ words from a language, then the ration (5) can be written in this way:

$$L=\lceil N_{ks}.lg2/lg16384 \rceil = \lceil N_{ks}.lg2/14.lg2 \rceil. \qquad (6)$$

In this case, the number of picked words is 16384 words and the scrambling numbers will be with 14-digit ones from 00000000000000 up to 11111111111111.

In relation to (6), L (words) can be determined in order to generate passphrases from the Word List and can be tested against the brute force attack of the most popular symmetric ciphers with the length of the secret bits $N_{ks:}$

$$N_{ks} = 128 \text{ bits} \qquad L_{128} = \lceil 9.142 \rceil = 10 \text{ words};$$

$$\begin{aligned} N_{ks} &= 168 \text{ bits} \quad L_{168} = \lceil 12.00 \rceil = 12 \text{ words;} \\ N_{ks} &= 192 \text{ bits} \quad L_{192} = \lceil 13.71 \rceil = 14 \text{ words;} \\ N_{ks} &= 256 \text{ bits} \quad L_{256} = \lceil 18.29 \rceil = 19 \text{ words.} \end{aligned} \tag{7}$$

In these calculation, the actual size of the most popular symmetric cipher secret keys are included: N_{ks}=128, 168, 192 и 256 bits. It is shown that from a word dictionary with 16384 words, the size of the secure passphrases is significant, which in real life brings difficulty in their use.

In order to keep the passphrases secure and also lower their word count is to make M higher. For example, if the dictionary contains $M=Q^K=2^{16}=65536$ words, the results of the same calculations would be:

$$\begin{aligned} N_{ks} &= 128 \text{ bits} \quad L_{128} = \lceil 8.00 \rceil = 8 \text{ words;} \\ N_{ks} &= 168 \text{ bits} \quad L_{168} = \lceil 10.5 \rceil = 11 \text{ words;} \\ N_{ks} &= 192 \text{ bits} \quad L_{192} = \lceil 12.0 \rceil = 12 \text{ words;} \\ N_{ks} &= 256 \text{ bits} \quad L_{256} = \lceil 16.0 \rceil = 16 \text{ words.} \end{aligned} \tag{8}$$

Passphrases with a size of 8 words rather than the ones with 10 and more could be accepted and used in real life. From (8) you can see that when increasing the size of the dictionary 4 times (from 16384 to 65536), the secure size of the passphrase is reduced significantly. If the dictionary contains every word in the English language (around 250 000), then the passphrases that correspond to the most popular symmetric ciphers with the length of the key words being N_{ks}=128, 168, 192 и 256 bits, will need to have around L_{128}=8, L_{168}=10, L_{192}=11 и L_{256}=15 words. However, it is worth to mention that if we increase the size of the dictionary, the words that are randomly generated, could be unknown and hard to remember to users since we are going to use the entire dictionary of a single language.

Some research show that typically, users know and use around 8000-10000 in their native languages (Ivan Vazov, a very popular Bulgarian author has used around 31500 unique Bulgarian words in his work, Shakespeare around 29000 English words and Pushkin – 25000 Russian words, however these literature geniuses are the exception and not the rule). Shortened dictionaries (8000÷10000 words) of the main languages are widely being used in authentication, however this makes a certain password very easily broken. If a shortened dictionary is used with a size L of the key phrase, then the security of this cipher would be:

$$\begin{aligned} L_{128} &= \lceil 128.\lg 2/\lg 10000 \rceil = \lceil 9.63 \rceil = 10 \text{ words and} \\ L_{192} &= \lceil 192.\lg 2/\lg 10000 \rceil = \lceil 14.45 \rceil = 15 \text{ words.} \end{aligned} \tag{9}$$

If the length of 10 words could somewhat be acceptable, then the second one cannot.

A second way of enhancing the security of Diceware is suggested in [3] when you create dictionaries from a couple different languages. For example, we create an English Word List with 250 000 words and you combine it with another 150 000 words from a different language, then the secure passphrase with 128 bits, would need to have:

$$L=(128.lg2)/lg400000=6.9 \text{ words}, \quad (10)$$

When the size of the secret key is 192 bits:

$$L=(192.lg2)/lg400000=10.32 \text{ words}, \quad (11)$$

When the size of the secret key is 256 bits:

$$L=(256.lg2)/lg400000=13.76 \text{ words}. \quad (12)$$

It is clear that a passphrase containing too many words is impractical and cannot be memorized by the average users, as well as, the expectation that a user would memorize phrases from two or more languages. The whole purpose of creating the Diceware passphrases mechanism was to make sure users pick longer passwords that are easily remember and are more secure. However, after this analysis we can say that with the current technological advancement, the shortest that a Diceware passphrase can be is eight words from two or more different languages. Therefore, the further use of Diceware passphrases is meaningless.

Conclusions

From the facts and the analysis of this report, we can conclude:

1. The method of passphrase generation Diceware is no longer effective and should carefully be used. If every passphrase contains at minimum 10 words, then it will have the security of a 128 bit secret key of the most popular symmetric authentication ciphers, which as of today provides basic security.
2. In order to increase the security of Diceware, dictionaries bigger that 7776 is a must. As calculated earlier, a dictionary with the size of M=65536 words the length of the passphrase would be 8 in order to satisfy the 128-bit secret key. This dictionary can be generated for a separate organization and could be securely stored.
3. The two methods of increasing the security of passphrases that were discussed, the first being the increase of the size of the dictionaries by combining two languages can be enough to keep the Diceware method secure and active. However, we can conclude that the Diceware passphrase mechanism is limited and its use would gradually be changed to newer authentication schemes.

References

[1] Petar Antonov and Simeon Malchev, *Kriptografia v komputarnite mrezhi* (Varna: TU Varna, 2000), 315.

[2] Petar Antonov, *Naruchnik po criptographia i zashtita na dannite* (Varna: TU Varna, 2014), 27.

[3] Petar Antonov and Valentina Antonova, “Analiz na sigurnosta na parolnite frazi,” *Konferentsiya na tema „Politikata na Evropeĭskiya sŭyuz po zashtitata na informatsiyata*

i lichnite danni", Natsionalen voenen universitet —Vasil Levski - gr. Shumen 12-13 april 2018, 283-286.

[4] Petar Antonov, Valentina Antonova, and D. Razsadov, "An approach and a software system of choosing reliable password phrases for resource access control," In: *Conf. Proceedings of "TEHNONAV 2002", Constanta (Romania), Ovidius University Press,* 2002, pp. 185-187.

[5] Petar Antonov and Valentina Antonova, "Access Control in the Virtual Education Area," In: *Proc. of Fourth Intern. Bulgarian-Greek Conf. Computer Science'08, Kavala, Greece, September 18-19,* 2008, pp. 951-955.

[6] Joseph Bonneau, "Deep Dive: EFF's New Wordlists for Random Passphrases," *Electronic Frontier Foundation*, 2016, https://www.eff.org/deeplinks/2016/07/new-wordlists-random-passphrases.

[7] Diceware™ Randomized Word Lists (Bulgarian): Diceware Randomized Word List 1 (Bulgarian), Diceware Randomized Word List 6 (Bulgarian), http://blog.radunchev.com/2016/06/02/randomized-diceware-word-lists-misc/.

[8] Paul A. Grassi et al., "Digital Identity Guidelines. Authentication and Lifecycle Management," Special Publication (NIST SP) - 800-63B, U.S. Department of Commerce, NIST, June 2017, p. 79, https://doi.org/10.6028/NIST.SP800-63b.

[9] Arnold G. Reinhold**,** "The Diceware Passphrase Home Page," 2020, https://theworld.com/~reinhold/diceware.html.

[10] The Diceware Passphrase FAQ, http://world.std. com/~reinhold/dicewarefaq.html.

About the Authors

Petar **Antonov** has received his BSc, MSc and PhD in the area of Information and Communication Technology from Saint Petersburg Electro technical University in Russia. His scientific research interests have been in computer communication and information security. Since 2016, he has been a professor at the Bulgarian Naval Academy.

Nikoleta **Georgieva** has graduated from the United States Naval Academy in 2020 and has received her BSc in Cyber Operations. Currently she is working as a system administrator at the Bulgarian Naval Academy.