Volume 47 Number 4 (178) 2015

DOI: 10.5604/17318157.1200176

THE SECURITY OF IT RESOURCES PROTECTED BY PASSWORDS

Artur SZLESZYŃSKI[*], Anna WOJACZEK[**]

[*] *Institute of Command, Military Academy of Land Forces*
e-mail: a.szleszynski@wso.wroc.pl
[**] *ICT Department, Military Academy of Land Forces*
e-mail: a.wojaczek@wso.wroc.pl

Received on April 12th 2015; accepted after revision in September 2015

Abstract:

This paper presents the protection effectiveness assessment of the contents of files created using an office suite. The protection effectiveness is understood as the time required for obtaining a password blocking the access to IT resources. IT resources are files containing data of varying degrees of sensitivity. The study analysed the time it takes to crack a password for different types of information resources. It has been shown that passwords consisting of five or a smaller number of characters, and can be found in less than 1.5 hours. Such a short time needed to find the password does not guarantee the effective protection of the contents of the file. This, in turn, makes it possible to breach of confidentiality attribute of an IT resource.

Keywords:

IT resource protection, a password, password recovery

INTRODUCTION

In the information systems the information used by users or devices operating systems, in which information is collected, processed or transmitted, is stored in files. Therefore, they should be treated as a kind of information resources used by an organisation in its activities. Files may contain data of a particular structure or do not have a strict structure defining the location of the data. Files having a fixed data structure are for example: databases, spreadsheets, etc. Examples of files that do not have fixed data structures are graphic files, mp3s, those created using a text editor, etc. These

types of resources are of a specific value for an organisation and they require protection. The easiest way to protect files against the disclosure of their contents to unauthorised persons is to make the access to it password-protected. This possibility exists in office suites e.g. MS Office, Open Office, etc.

The described method of the file content protection has been implemented by producers of office suites[1] and constitutes a sufficiently effective means of protecting their contents. The appropriate protection effectiveness is related to the construction of a password used to encrypt the content of an IT resource.

The presented method of the content protection enables the exchange of files of varying degrees of sensitivity[2] between users located in different places and using data transmission channels, such as wireless access points, electronic mail, etc.

When analysing the available literature it can be seen that the length of a password used has an impact on the level of a file content protection [5]. The password length affects the resistance to brute force consisting in guessing it, which is presented in a statement published on the website of the producer of the software used to passwords cracking. The table shows the time required to find a password that make it possible for an attacker to reach the source code information stored in an encrypted file (Table 1).

The statement was developed in 2006. The author of the study did not include information on the technical parameters of the computer and the software used by which the attempts were made to find a password to protect the file content. The lack of the data mentioned above makes it impossible to verify the results. It should be remembered that the performance of microprocessors used in mobile and fixed devices have increased since the publication was issued. This means that a successful attack on passwords protecting the file content can be carried out without any specialised hardware needed.

Knowledge of the effectiveness of the protection of files content using passwords to protect them from being broken can be used to develop file encryption procedures within an organisation. This is important for people working outside the organisation's office and using mobile devices. For the IT department or persons responsible for the organisation's information security it will be a source of information used to estimate the possibility of a breach of confidentiality attribute of an information resource [1,6]

1. DEFINITION OF A RESEARCH PROBLEM

On the basis of the information contained in the literature the authors decided to check the password traceability in order to reach the body of the message contained in a protected file. Knowledge on the possibility of a breach of the information resource confidentiality attribute will be benefited from so as to develop rules for the creation

[1] Protection is available in commercial packages and distributed under Open Source license.

[2] The degree of sensitivity is understood as the importance of the information contained in a file for a particular organisation. The larger the value of the information contained the higher sensitivity of the file See: A. Bialas Ochrona informacji i usług...

and distribution of keys for information resources transferred through open-access telecommunications channels. This is the case of the information exchange between users, which will take place across various locations. They cannot be regarded safe places due to the fact that there is a danger that unauthorised persons will become familiar with the content of the resource information. These locations do not provide physical protection of a device (or devices) that transmits information between users. The situation described above is typical for mobile workers exchanging data with the information system of the organisation.

Table 1. Estimated time required for cracking a password by means of brute force with the use of a single computer source: [5]

Type of characters	Allowable number of characters to create a password	3 character password	6 character password	8 character password	12 character password
		Time necessary to decrypt a password	Time necessary to decrypt a password	Time necessary to decrypt a password	Time necessary to decrypt a password
Lower case letters only – the Latin alphabet	26	0,02 s	5 min	58 h	3000 years
Lower case letters only – the Latin alphabet and digits	36	0,04 s	36 min	32 days	150000 years
Upper and lower case letters – the Latin alphabet and digits	62	0,2 s	15 h	7 years	100 million years
Upper and lower case letters – the Latin alphabet, digits, special characters	94	1 s	8 days	193 years	Longer than the Earth exists

Common accessibility of the Internet allows the convenient data exchange. Implementing solutions based on cloud computing technology as carriers of data-storage enhances users' comfort. However, the advantages described, under certain specific conditions, may constitute the vulnerability in the security system. Their skilful use may constitute a source of incidents consisting in the breach of security attributes of files to be transferred. The problems that need to be answered can be expressed as the following questions:

- How effective is the protection of the file content provided by passwords used to prevent the file from being opened?
- What password needs to be created that decoding it requires the longest possible time?

The aim of the study was to verify the effectiveness of the file content protection done by a password on the basis of which the file content is encrypted. The effectiveness of the content protection is understood as the time that an attacker would require for

intercepting the encrypted file content. It is irrelevant how the attacker gained the physical access to the file.

As stated in the introduction, such a password is sought that it will take a long time for a test program to break it. A long time of reaching the source form of the information contained in the file content means the period of time not less than two hours. This assumption is based on the fact that during this period information may become "obsolete" and its usefulness for an attacker will be smaller or zero. Lengthy waiting times to find the password can deter an attacker, which allows retaining the file contents confidential. The issue of setting out the minimum period of time necessary to reach the source file form depends on the type of the IT resource and should be selected in line with its sensitivity. The higher the sensitivity of the information resource, the longer time needed for decoding it.

The basis for the verification of the data was a research experiment. In the experiment there was used a computer equipped with a dual-core microprocessor Intel Core 2 T9600, 4 GB operational memory and 250 GB hard drive. The Office Password Unlocker in the demo version was taken to find the passwords that protect content files prepared using MS Word text editor. This program is available on the producer's website[3], so an intruder after accessing the files may download it, install it on his /her computer and then try to access the contents of a file or files. What is more, the program was chosen since not all of the other applications found were fully functional and free. Producers charge for some of them.

52 processes allocating 2079 MB of operational memory worked in the operating system environment in which the experiment was performed. 2068 MB of operational memory remained available to the program. In the password cracking process the parallel computing with the graphic processing units GPUs was not performed.

The described method of protecting the file content is not the only possible. The file content can be encrypted using special software in the operating system[4] or using the program AxCrypt. It was decided to choose the use of a password protecting the contents of files against decryption as it requires no additional software installation. Moreover, it provides effective protection, provided that the password used is not trivial, for example 12345 or ABCDE.

2. THE MEASUREMENT OF THE TIME REQUIRED FOR FINDING PASSWORDS THAT PROTECT A DOCUMENT

The RC4 stream cipher with the 40[5]-bit encryption key is used in the Microsoft Office suite [2,5]. The password entered by the producer of the document serves to create the key encryption. In the opinion of the package makers, this system should

[3] http://www.passwordunlocker.com

[4] The software to encrypt the contents of files and folders as a system component is present in solutions designed for servers, e.g. Windows Server 2012.

[5] This refers to MS Office files to version 2003. Since the 2007 package release a new encryption algorithm with longer keys has been introduced thus enhancing the security of protected files.

ensure an adequate level of protection of the document contents. Data shown in Table 1 contradict the assumption as it shows that a certain group of passwords does not constitute effective protection of the document content. The question should be raised: should the entire content of the file or only its selected fragments be subject to encryption? In order to answer the question the number of bytes changed in the protected file after using encryption must be determined. The number of changed bytes would be calculated by comparing the contents of two files [6]. For measurement purposes it is necessary to have a point of reference that is a file on the basis of which bytes that have changed in a protected file would be specified [6]. With this aim in view, the relationship shown in formula (1) should be introduced [6].

$$dbm = \frac{n_{erb}}{n_{cf}} \tag{1}$$

where:

dbm – a coefficient determining the number of changed bytes,

n_{erb} *– the number of changed bytes in a file compared,*

n_{cf} *– the number of bytes in reference information.*

For the purposes of the experiment, a *reference* file and a set of passwords were developed and then used to secure the file content, as shown in Table 2. The adopted structures of passwords helped to create several spaces of possible passwords calculated according to the formula (2).

$$S = L^n \tag{2}$$

where:

S – the number of possible passwords,

L – the number of possible characters used to create a password,

n –the number of characters in a password.

Thus, in the case of digits only in a three-character password, the range of passwords sought will be 1000. In the case of 26 characters of the Latin alphabet and the same password length, the number of possible passwords will be 17576. And in the case of three-character password being a combination of letters and digits (36 characters) the number of possible passwords is 46656. For the five-character passwords the number of possible combinations would be: 100000 (digits only), 1881376 (letters of the Latin alphabet) and 60466176 (the combination of letters and digits). Thus, the longer the password and the larger the set of possible characters to use, the larger range of possible combinations the program must search.

The reference file, after saving on the hard drive, had a size of 26624 B. Then, the file was saved using a password to protect it from cracking. After applying a password to protect the file contents its size changed compared to the unencrypted file. The comparison of sizes of the files and the number of the modified bytes is shown in Table 2. A characteristic feature of all the files protected by passwords is to increase the size of

the encrypted file of 512 B in relation to the reference file, which is 1.92%. Increasing the file size as a result of encrypting its content is low (Table 2). The difference between the largest and the smallest value of the coefficient of the number of changed bytes was 0.00105, which is 0.21% of the average test sample (Table 3). The parameters of descriptive statistics of the coefficient determining the number of changed bytes are shown in Table 3.

Table 2. The parameters of files protected by passwords against cracking

File size [B]	Number of bytes changed in the protected file [B]	Value of dbm	Password
27136	13028	0,48933	123
27136	13015	0,48884	Abc
27136	13020	0,48903	1b3
27136	13015	0,48884	a2c
27136	13015	0,48884	1234
27136	13018	0,48896	Abcd
27136	13017	0,48892	a2c4
27136	13012	0,48873	1b3d
27136	13040	0,48978	12345
27136	13037	0,48967	abcde
27136	13036	0,48963	1b3d5
27136	13025	0,48922	a2c4d

source: own study

Table 3. The parameters of descriptive statistics of the coefficient dBm determined for the test sample

Parameter	Value
Average	0,48915
Median	0,48899
Dominant	0,48884
Standard deviation	0,00037
Sample variance	$1{,}36673*10^{-7}$
Kurtosis	-1,08268

Parameter	Value
Skewness	0,71150
Minimum	0,48873
Maximum	0,48978

source: own study

The test sample has a small size equal to 12. When analysing the diagram of the coefficient distribution of test sample's changed bytes, it can be assumed that it is not possible to identify the structure and the length of the password on its basis. The diagram of the coefficient distribution of changed bytes is shown in Figure 1.

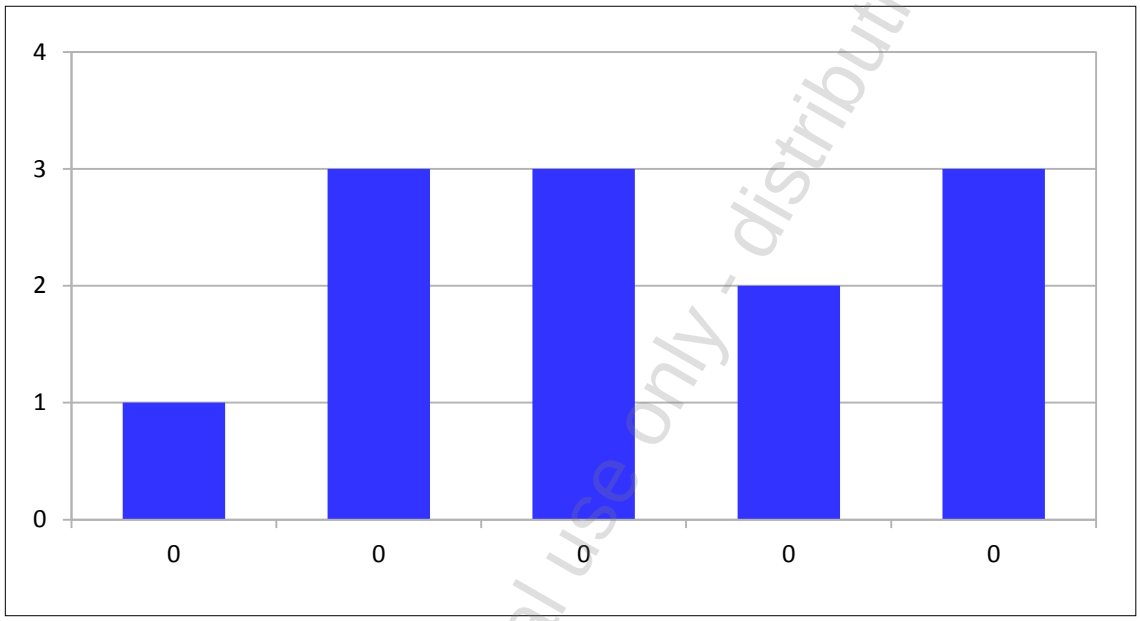

Fig.1. The distribution of the coefficient value of changed bytes (dbm) for the test sample

source: own study

The entire content of a file is not subject to encryption, but only selected parts of it. The method was adopted by virtue of the structure of files created by the software included in office suites [3,4]. Encrypting the content and leaving the file header unaltered allows its opening by a tool[6] contained in the office suite, and displaying information on the need to enter a password by a user in order to open and present its contents. Finding a password needed to open the file content means less elements for programs to check. The search for passwords used to protect the contents of files requires downloading the file fragments that were encrypted. If it is possible to find the contents of the encrypted portion of a file when substituting successive passwords, the task becomes easier and faster. Such a solution allows checking a large number[7] of

[6] A tool is understood as a text editor, a spreadsheet, a database management system, etc.

[7] From several to several hundreds thousand of possible passwords.

possible passwords within 1s. The described method of finding a password is an example of a "brute force"[8] attack, which consists in generating successive passwords, their substitution to the encryption mechanism, and then an attempt to decode the file fragment. This is repeated until the right key enabling the file content decryption is found. This method is time-consuming as shown by the results in Table 1. However, it should be remembered that the computing power of computer devices constantly grows, therefore, it cannot be ruled out that it will be possible to find simple passwords, for example used in the experiment, by the mobile devices.

3. RESEARCH RESULTS

The Office Password Unlocker software in the demo version was used to conduct research. The program can be downloaded from the producer's website. The program offered the full functionality of the commercial version. There was no other limit than the time of its exploitation. There was no restriction on the number of files the passwords of which the program cracked. Test files were protected against opening by passwords presented in Table 4. After entering the file to such a program, the time needed to crack a password was measured. Results from the research are shown in Table 4.

Table 4. Time required for cracking a password protecting the file content

Time for cracking a password by a test program[s]	Password	Type of characters used in a password	Number of characters in a password
1	123	Digits	3
45§	1234	Digits	4
4286	12345	Digits	5
1	1b3	digits -letters	3
44	1b3d	digits - letters	4
4233	1b3d5	digits - letters	5
1	abc	Letters	3
2	abcd	Letters	4
154	abcde	Letters	5
1	a2c	letters - digits	3

[8] Polish: brutalna siła.

Time for cracking a password by a test program[s]	Password	Type of characters used in a password	Number of characters in a password
2	a2c4	letters - digits	4
197	a2c4d	letters digits	5

source: own study

The test program needed most time to find passwords having only numbers in its structure. The reason was that the program began checking possible combinations of passwords from those beginning with the letters of the Latin alphabet. The adopted procedure is typical for the behaviour of users - often a password protecting the file content is the name of a given person or his/her child. In the case of passwords beginning with digits, the time required to break the password increased more than 22 times compared to passwords formed only of letters. It can therefore be recommended to create a password starting with a digit instead of a letter. Another observation arising from the study is that in the case of passwords of comparable length but consisting of the digits and letters combination, the time of cracking them was shorter than breaking passwords consisting of digits only (Table 4).

Passwords that were made up of special characters[9] and characters typical for a given language were not studied in the experiment. The use of special characters, for example W - W increases the number of possible combinations, which expands the search range. Since the programs for finding passwords available on the Web used the English-language communication interface, it can be assumed that they would have difficulty cracking passwords that contain characters typical for the Polish alphabet. Coding the national characters, according to the UTF-8 standard, requires the use of two bytes, which means that the program that finds passwords would interpret it as a password longer by one character.

Based on the results shown in Table 4 it can be concluded that none of the passwords used provided the desired long time to reach the file content. Finding passwords allowing access to the content of the protected file took the program less than 2 hours. This means that applying 5-character or shorter passwords (even if they are combinations of digits and letters) does not provide the expected level of protection for the file content. An attacker can gain access to the file content in time shorter than assumed, which may infringe the file confidentiality attribute [1,6]. The confidentiality attribute is considered the most important of the three security attributes of IT resources [1].

In the experiment there was used the demo version of the password cracking program. It can be assumed that commercial programs of this type will operate more efficiently than the applied one. Some of them will use parallel computing based on GPUs. These processors are parts of Nvidia graphics cards. In the case of taking advantage of paral-

[9] Special characters are understood as signs such as: &,^,#, ?, etc.

lel processing the time required for finding a correct password is shorter. Reducing the time will be proportional to the algorithm used to create the combination of passwords and the division of the set into particular processors in accordance with the principle of the uniformly distributed loading.

It is important to point out that in the case of producers of password-cracking programs it was emphasised that they should be used only with the aim of gaining access to files created by the user. It was informed that they should not be used to gain access to files of other persons and entities. This means that the program should not be used as a tool in criminal activities.

CONCLUSION

The work presents the results of measuring the time required for breaking a password protecting the file produced in the office suite against opening. On the basis of the measurement data obtained from the program cracking passwords it was found that the use of passwords shorter than 5 characters, beginning with the letters, did not provide an adequate level of protection for the file content. Passwords of this type are the fastest to be found. The program needed slightly more time to find passwords that are combinations of letters and digits. The longest time was needed to crack passwords being combinations of digits and letters or consisting only of digits. What is interesting, passwords beginning with a digit were found after a period of time longer than passwords beginning with a letter. This was the effect of the algorithm for searching for a password used in the program.

Only half of the file protected by a password was encrypted as demonstrated by the data in Table 2. Such a solution allows faster decryption of the file content, but at the same time its vulnerability is increased, which can be exploited by an attacker.

The discrepancies that can be observed between the results shown in Table 1 and Table 4 for the time needed to crack the password protecting the file content stem from two factors. The first of them is the method for measuring the time needed to find a password. The results in Table 4, concerning time for finding passwords were obtained from the program used in the experiment. The data in Table 1 can be obtained from the operating system where the system functions measure time accurate to milliseconds. Another factor affecting the differences are algorithms used to break passwords used in both programs. The algorithm used in the test program checked passwords starting with letters, next those consisting of letters or special characters. After checking all the possible combinations starting with a letter (at a given password length), passwords that begin with digits were created.

Knowledge of the time needed to crack a password used will help to develop an intra-organisational system of creating and distributing passwords. This is an important issue as users tend to underestimate risks related to access to information contained in files exchanged through the Internet. In various publications it is often emphasised that passwords protecting access to important IT resources should be sufficiently long and frequently changed. Why, then, do passwords similar to those used in the study continue to apply? This is mainly due to human habits – it is easier to remember the pass-

word "abcdefgh" rather than "1b#4E6&8". Shorter time is required for cracking the first password than the second one. This is an example of the search for a point of balance between a user's convenience and security of an information resource.

REFERENCES

1. Białas A., *Bezpieczeństwo informacji i usług we współczesnej firmie i organizacji*, WNT, Warszawa 2006.
2. Karbowski M., *Podstawy kryptografii*, Helion, Gliwice 2008.
3. Open Office.org's *Documentation file. Microsoft Excel File Format. Excel versions 2,3,4,5,95,2000,XP,2003*.
4. [Word Doc]: Word .doc Binary File Format, Microsoft.
5. Semjonov P., *Password Recovery/Cracking FAQ*, [online]. [available: 03.02.2014]. Available on the Internet: http://www.password-crackers.com/en/articles/12/#3.2.
6. Szleszyński A., *Pomiar bezpieczeństwa informacji w zarządzaniu bezpieczeństwem w systemie teleinformatycznym*, [in:] „Zeszyty Naukowe Politechniki Śląskiej", Organizacja i zarządzanie, no. 74, Gliwice 2014.

BIOGRAPHICAL NOTES

Artur SZLESZYŃSKI, M.A. Eng. – a graduate of the Faculty of Electronics at the Military University of Technology, a holder of the MCP title. The main areas of his academic interests include information security, reliability of ICT systems as well as artificial intelligence methods.

Anna WOJACZEK, M.A. Eng. – a graduate of the Silesian University of Technology in the fields of computer science and mathematics. The main areas of her academic interests include the security of information systems and cryptography. She specialises in the administration of computer networks. She has professional experience in the management and monitoring of systems and the network infrastructure, as well as work on the 1st and 2nd line of the helpdesk.

HOW TO CITE THIS PAPER

Szleszyński A., Wojaczek A. (2015). The security of resources protected. Zeszyty Naukowe Wyższa Szkoła Oficerska Wojsk Lądowych im. gen. Tadeusza Kościuszki Journal of Science of the gen. Tadeusz Kosciuszko Military Academy of Land Forces, 47 (4), pp. 109-119. http://dx.doi.org/10.5604/17318157.1200176