Investigation of the method of determination of password resistance to brute force based on an artificial neural network

K. E. Izrailov¹, P. E. Zhukovskaya², P. A. Kurta³ and A. A. Chechulin⁴

¹ Candidate of Technical Sciences, Associate professor of the Bonch-Bruevich Saint-Petersburg state university of telecommunications, Saint-Petersburg, Russia
² Undergraduate of the Bonch-Bruevich Saint-Petersburg state university of telecommunications, Saint-Petersburg, Russia
³ Engineer of the Bonch-Bruevich Saint-Petersburg state university of telecommunications, Saint-Petersburg, Russia
⁴ Candidate of Technical Sciences, Associate professor of the Bonch-Bruevich Saint-Petersburg state university of telecommunications, Saint-Petersburg, Russia

E-mail: konstantin.izrailov@mail.ru

Abstract. The paper deals with a topical subject of information system access security through development of a polling system aimed at forecasting the resistance of the user’s passwords against brute force. A method of development of such a system, which is based on computer-aided learning, is suggested. This method is based on the following hypothesis: “Social and psychological features of the password system users have a bearing on the complexity of created passwords.” The method is elaborated in the form of the following 6 steps: development of a group of questions which reflect the preferences of the password system user; implementation of a polling system which requests the user’s typical password; selection of a mechanism of independent estimation of password complexity; application of the polling system to generate implicit interrelations of specific features of users and their passwords; development and training of a neural network to solve a regression problem based on the results of the polling system use; and estimation of the neural network by a test set of users. The experiment results are presented in a form of a diagram. The results of the experiment are discussed and conclusions are drawn about the method development.

1. Introduction

Ensuring security of access to information systems is one of the fundamental problems in the area of cybersecurity [1-3]. While hardware and software methods can provide a high level of security, the human factor still remains the main source of threats. In particular, users’ irresponsible attitude towards choosing complex passwords makes it possible for an intruder to use brute force resulting in unauthorized access to information. If these users are known, it will be possible either to prevent their access to critical information resources, or take additional measures to improve their “cybersecurity
culture”. Therefore, the detection of “weak” users in terms of password systems is a crucial IT task. One of the methods aimed at changing the situation for the better can be the one suggested in our paper, that is, the use of a polling system for predicting the complexity of passwords and their resistance to brute force. In contrast to the existing methods which are focused mainly on the detection of insiders [4-8] and regulation of password policies, this one addresses a more complex qualitative problem: checking and improving personnel’s diligence, thereby decreasing the impact of the human factor on cybersecurity.

The most challenging thing in the solution is the necessity to combine qualitatively different aspects – social and psychological factors influencing the passwords set by users [9], as well as probabilistic and statistical factors determining the complexity of brute force used by an intruder [10]. In this case, taking into consideration that finding the conformity of aspects is another difficult problem, we propose using the computer-aided learning, as far as the artificial neural network (frequently used in cybersecurity [11]) is concerned. It will allow for classification or regression of answers to the user’s questions by the complexity of his/her passwords.

2. Method description

Let us describe a hypothesis that forms the basis of the proposed method. Let us assume that there are two users of the password system. The first one (for example, Alice) likes reading fiction, has a pet and a romantic mindset. The second one (for example, Bob) is fully engrossed in mathematical sciences, fond of programming and computer games, and also is an absolutely practical person. Without getting into the specifics of different considerations, and also stemming from the preliminary investigation of similar users, it is reasonable to assume the following. It is likely that Alice’s password will have to do with the Humanities, presumably, the password will include the names of literature characters, titles of artistic works and suchlike; note that the authors’ research has partly confirmed this guess. On the other hand, Bob, who has always been very knowledgeable in IT and understands the algorithmic procedures of brute force (using the alphabet or a dictionary), will offer a more complicated password, perhaps, using buzz words or phrases commonly used only in his “digital community”. Hence, we may assume that is password is much more difficult to guess. Obviously, these considerations are trivial, partial and cannot underpin the general logic of determining the degree of password complexity; for instance, infinite complication of password will result in a situation when the user cannot keep it in mind and therefore it turns to be useless. Nevertheless, the gist of the hypothesis stemming from our previous considerations is that “social and psychological specific features of the password system user have a bearing on the complexity of created passwords”.

Our aim is to reveal weak users, so let us outline our method.

2.1. Step 1. Development of a group of questions

It is essential to compile a list of questions reflecting the user’s personality comprehensively. In this context, it would be absolutely wrong to make any assumptions about the influence of the answers on the properties of the final password, because, as was already mentioned, these interrelations present an individual intractable problem requiring a lot of research. Thus, it makes sense to select questions from different areas in such a way that one could “catch” even lightweight connections between the user’s features and his/her passwords.

2.2. Step 2. Implementation of a polling system

It is necessary to develop a polling system which, apart from accepting answers to questions, would request a typical password from the user. From the user’s point of view, the system can have functionality only related to the questions and answers. It is also advisable to ask the password at the beginning and at the end of the polling process to check whether it is kept in the user’s mind and therefore reflects his/her specific features.
2.3. **Step 3. Selection of a mechanism for determining the passwords complexity**

It is essential to select a mechanism for determining the complexity of the password. To do this, different online resources may be used. For example, Kaspersky Password Check site (https://password.kaspersky.com/) presents the time of selecting the entered password, which can be considered as its complexity. In the future, it makes sense to develop a special technique for password complexity evaluation which will be adapted to specific password-based access control system.

2.4. **Step 4. Application of the polling system**

It is necessary to select a group of unrelated users and test them with the developed system. It is obvious that the dataset should contain various types of users, e.g. by the age, social groups, etc. It would result in a table of answers to questions and the complexities of their typical passwords.

2.5. **Step 5. Development and training of a neural network**

An artificial neural network should be constructed and learned to solve a regression problem with questions as factors, answers as input data, and password complexity as output data. It goes without saying that implementation of the password complexity evaluation technique is not limited to neural networks. Genetic algorithms that represent a malefactor who tries to brute-force the password can also implement it, for instance.

2.6. **Step 6. Estimation of the neural network**

The trained artificial neural network should be tested for new input data. The results received on online resources must be compared to those obtained by the network. Later, the neural network can require the adjustment, for instance, by introducing additional layers, retraining, etc.

2.7. **Typical Password**

The term “typical password” has to be clarified. It implies a common password which is easily kept in mind, reflects the user’s social and psychological specific features and presents a certain basis for passwords used in real life. The hypothesis about some consistent patterns present among all created passwords was based on the authors’ previous study.

For example, it has been noted that people with humanitarian turn of mind, who are fond of reading fiction, frequently use the names of famous literary characters, phrases from classic literature, or historical events in parts of their passwords. Scientists can use the names of physical installations, constant values and the dates of great scientific discoveries. The adherents of nonconformist youth subcultures use slang terms, repeated symbols and substitute letters for visually compatible signs. Naturally, such consistent patterns are not a validated knowledge; however, they allow for stating that a concept of “typical password” has the right to exist. Checkups at the Step 2 are meant to support what has been said above. Their successful results will confirm that the password is really typical for the user because it was created at the beginning of polling and accurately reproduced in the end. Obviously, the system places reliance on the user keeping his/her created password in mind, and not, for example, in a data storage source.

Since the system suggests inventing a new, easy-to-recall password, rather than entering an already existing one, the confidentiality of user information will never be disclosed, either in processing or in sending it to an online resource (Step 3).

Graphical representation of the scheme is presented in the Fig. 1.

According to the Fig.1, the polling system is the key part of the method. The system gets the group of questions as the input and uses a specified mechanism for determining the passwords complexity, trains the neural network by one group of users and with the neural network assessing the complexity of other users’ passwords.

On completion of these steps, an assessment of a degree of success will be made, when using the artificial neural network for forecasting the complexity of users’ passwords based on their responses to questions.
3. The results of the experiment

In order to confirm the hypothesis and check the efficiency of the method, a prototype polling system was developed using Google Form for polling, Kaspersky Password Check for assessing the password complexity (normalized from 0 to 1), and Python software for implementing an artificial neural network. The polling consisted of 37 questions such as “What is your gender?”, “Have you got any pets?”, “Do you like taking photos?” etc. To simplify the implementation, the system called for the password only at the end of polling (i.e. once, not twice) without checking how much it resonated with the user. Of course, it randomizes the accuracy of answers, but this is not critical.

Altogether, questions were asked to 83 users; 67 of them were used for the artificial neural network training and 16 for testing. After that, the complexity of passwords was compared between the tested users; actual complexity was assessed using the site, and predicted complexity – using the artificial neural network. The results are shown in Fig. 2.

Fig. 2. Results of comparison of real and predicted complexity of user’s password.
The results of the polling system testing (Fig. 2) were good: the average dispersion of real and predicted password complexity was as high as ~17%.

4. Discussion
Naturally, the obtained results cannot be taken as a guaranteed certainty of the efficiency of the method and the prototype system, since the latter was simplified and the total sample of users was rather small and most likely did not possess a required quality. Nevertheless, it may be stated even now that there are certain consistent patterns existing between the social and psychological characteristics of a person and the degree of complexity of his/her passwords.

5. Conclusions
In order to verify our hypothesis, a polling system for detecting weak users (in terms of password systems) was developed and investigated by a preventive method, i.e. before the immediate operation. The first results proved the theoretical efficiency of the method. Nonetheless, in order to assess the efficiency of the proposed method, the range of verified questions (including those related to interface elements of the system [12]) has to be enlarged, a more independent sample of users has to be obtained, a full-scale testing should be performed involving a greater number of users, and different metrics of the results quality need to be estimated. The authors are planning to deal with all these tasks in their future research.

*The research is conducted with the support from the Ministry of Education and Science of the Russian Federation under Agreement No. 05.607.21.0322 (ID RFMEFI60719X0322).*

References

[1] Buynevich M.V., Vasil'yeva I.N., Vorob'ev T.M., Gnidenko I.G., Yegorova I.V. etc. Zashchita informatsii v kom'puternykh sistemakh (Information security in computer systems): monograph, Saint-Petersburg: Sankt-Peterburgskiy gosudarstvennyy ekonomicheskiy universitet, 2017. – 163 p.

[2] Utkin O.V., Vlasov D.S., Il'in A.V., Yefremenkov Ye.YU. Metodika otsenki deyatelnosti dolzhnostnogo litsa TSUKS MCHS Rossii (Methods for assessing the activities of an official of the Central Control Center of the Ministry of Emergency Situations of Russia) // Podgotovka kadrov v sisteme preduprezhdeniya i likvidatsii posledstviy chrezvyachaynykh situatsiy: materialy mezhdunarodnoy nauchno-prakticheskoy konferentsii. – 2017. – pp. 227-228.

[3] Antyukhov V.I., Sugak V.P., Yaroshenko A.YU., Ostudin N.V. Modelirovaniye protsessa obolecheniya bezopasnosti informatssii v podrazdeleniyakh MCHS Rossii (Modeling the process of ensuring the security of information in the units of the Russian Emergencies Ministry) // Servis bezopasnosti v Rossii: opyt, problemy, perspektivy. Obolecheniye bezopasnosti pri chrezvyachaynykh situatsiyakh: materialy VII Mezhdunarodnoy nauchno-prakticheskoy konferentsii. – 2015. – pp. 71.

[4] Zhuk R.V., Vlasenko A.V., Titenko Ye.A. Sistemy protivodeystviya insayderam (Systems of counteraction to inside) // Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. – 2012. – № 6 (45). – pp. 030-033.

[5] Davletkhanov M.G., Stolyarov N.V. Nadezhnaya zashchita ot insayderov (Reliable protection from insiders) // Zashchita informatssii. Insayd. – 2008. – № 5 (23). – pp. 44-49.

[6] Alyabushev I.B., Babushkin V.V. Metodiki zashchity baz dannikh ot vnutennikh zloonyshlennikov (Methods for protecting databases from internal intruders) // Zashchita informatssii. Insayd. – 2006. – № 6 (12). – pp. 58-63.

[7] Shugayev V.A., Alekseyenko S.P. Klassifikatsiya insayderskikh ugroz informatssii (Classification of insider threats to information) // Vestnik Voronezhskogo instituta MVD Rossii. – 2020. – № 2. – pp. 143-153.
[8] Buynevich M.V., Vlasov D.S. Sravnatel'nyy obzor sposobov vyyavleniya insayderov v informatsionnykh sistemakh (Comparative review of methods for identifying insiders in information systems) // Informatizatsiya i svyaz'. – 2019. – № 2. – pp. 83-91.

[9] Voyskunskiy A.Ye., Naftul'yev A.I. Aktual'nyye psikhologicheskiye problemy kiber-etiki (Actual psychological problems of cyber ethics) // Gumanitarnaya informatika. – 2007. – № 3. – pp. 31-39.

[10] Blinov A.S., Stepanenko M.A. Otsenka dostatochnoy slozhnosti parolya dlya bezopasnogo ispol'zovaniya na veb-resursakh (Assessment of the sufficient complexity of a password for secure use on web resources) // Issledovaniya v oblasti yestestvennykh nauk. – 2014. – № 8 (32). – pp. 24-27.

[11] Buynevich M.V., Izrailov K.E. Obobshchennaya model' staticeskogo analiza programmnogo koda na baze mashinnogo obucheniya primenitel'no k zadache poiska uyazvimostey (A generalized model of static analysis of program code based on machine learning as applied to the problem of finding vulnerabilities) // Informatizatsiya i svyaz'. – 2020. – № 2. – pp. 143-152.

[12] Akhunova D.G., Vostrykh A.V., Kurta P.A. Otsenka pol'zovatel'skogo interfeysa informatsionnykh sistem posredstvam modeley kachestva programmnogo obespecheniya (Assessment of the user interface of information systems by means of software quality models) // Informatizatsiya i svyaz'. – 2020. – № 2. – pp. 127-135.