

ASSESSING THE DEGREE OF THE SOCIAL MEDIA USER'S OPENNESS USING AN EXPERT MODEL BASED ON THE BAYESIAN NETWORK

M. S. Frolova
St. Petersburg State University
m.f.26@mail.ru

A. A. Korepanova
St. Petersburg Federal Research Center of the Russian
Academy of Sciences
(SPC RAS)
aak@dscs.pro

M.V. Abramov
St. Petersburg Federal Research Center of the Russian
Academy of Sciences
(SPC RAS)
mva@dscs.pro

Abstract. One of the most topical problems of information security is the prevention of successful incidents of social engineering attacks — psychological manipulation of people to get access to confidential information. It is necessary to assess personal characteristics of an information system user to estimate his vulnerability to some type of attack. Assessment the degree of openness of a social media user is perspective this way (openness is considered according to the factor of "openness to experience", which is part of the personality model "The Big Five"). In this paper, we consider an approach to assessing the openness of a online social network user using the bayesian network. Based on the five-factor personality questionnaire, the characteristics that can be extracted from the user's page, potentially affecting his openness to accepting information, are considered. The structure of the Bayesian network for solving this problem is proposed.

Key words: social engineering attacks, social media, Bayesian networks.

I. INTRODUCTION

Currently, in various areas of manufacture and in the service sector, important documents are stored electronically. Attackers can gain access to this information from the outside either by trying to hack the information security system of the enterprise, or through employees of this enterprise who have access to the necessary data. Therefore, social engineering attacks — a set of applied psychological and analytical techniques for manipulating a person's mental consciousness, which attackers use to covertly motivate users of a public or corporate network to violate established rules and policies — is a major threat to the company's security [31].

Social engineering attacks are improving year by year, requiring constant development of existing methods of protection [27], focusing on the system user's personal characteristics.

User's vulnerability profile, that is a set of his psychological and behavioral characteristics potentially affecting the outcome of social engineering attacks, directed at him, is a convenient tool for the software development to prevent successful incidents of social engineering attack [7]. This work is devoted to improving the method of compiling a user's vulnerability profile to social engineering attacks by assessing his openness. The study is based on The Big Five model [20]. The five-factor model is a personality model that allows you to describe a person using 5 traits, or, in other words, factors. The scale of "openness to new experiences" (we consider this factor in the paper) shows how much a person is interested in expanding his horizons, learning new things, getting new information. Thus, this work is devoted to assessing the degree of a user's openness in the context of the eponymous factor of the Big Five.

Many studies show a positive correlation between the user's openness to experience and vulnerability to various types of manipulation and social engineering attacks [39, 15]. Therefore, obtaining this assessment is a promising task in the software development for protecting information systems from harmful threats. It is convenient to use the information that the user places himself on his pages in social media, as well as data of his reactions to posts in the news feed, to make the necessary assessment. This information is easily accessible, so it can potentially be used by an attacker. Data in online social networks are characterized by fuzziness and incompleteness, so in this paper, Bayesian networks are used to solve the estimation problem. It is a convenient tool for developing models with causal relationships to answer probabilistic questions, and also allow working with incomplete data.

The results obtained in this paper are useful for further research on improving the security of information systems

from social engineering attack, for developing protection methods against it based on the user's vulnerability profile.

II. LITERATURE REVIEW

This work is a part of a general study devoted to the analysis of the information systems user's vulnerability to social engineering attacks in order to develop software to prevent successful incidents [6]. In [12], a scheme of a social engineering attack development is proposed, the dependencies of the attack outcomes on the actions of intruders and user responses are considered, and a hierarchy of information models is proposed, including information objects (documents), users, and possible acts of the attack [17].

One of the most important directions for solving the problem of increasing the security of information system users from social engineering attacks is building a profile of user's vulnerabilities. The works [8, 9] underline the importance of building a psychological profile that reflects the degree of severity of the psychological characteristics of the user, and present developments in this area. There are various approaches to analyzing the severity of the user's psychological characteristics, for example, psychological testing or analysis of personal information on the user's page in a social media [14, 16, 18, 19]. In paper [14], based on the digital footprint of a social media user, his personal characteristics are evaluated, and as a result of comparing the "human" and "computer" assessment of psychological characteristics, the "computer" assessment has a higher accuracy.

One of the most complex and detailed systems for assessing a person's characteristics, including the "openness to new experiences" required for this study, is the "Big Five". In [31], the author reveals in detail the validity of using this particular test to obtain the most accurate knowledge about a person's behavior and inclinations, and also draws a relationship between the obtained indicators for certain test factors and a person's tendency to be more open to information from the outside.

There are studies showing a link between openness and different characteristics of user behavior in a social media [38]. The study [5] assessed the correlation between the big five factors and user activity on social media. According to this study, openness to experience was positively correlated with the number of posts on a user's wall and the number of their friends on the social media.

A convenient mathematical model for working with the received information is Bayesian networks [4], which have many applications, including solving problems in the field of information security and social engineering attacks, for example, to assess the intensity of user behavior in a social network on the example of posting [10], [11]. In paper [13], approaches to modeling social engineering attacks on the user based on Bayesian network are proposed. The paper [29] investigates the possibility of using Bayesian network to build a diagnostic device for social anxiety disorder. The authors conclude that this mathematical tool is very adaptive, allows you to present information in a simple and convenient form,

significantly reduces the computational complexity, and allows you to achieve high accuracy of calculations.

III. PROBLEM STATEMENT

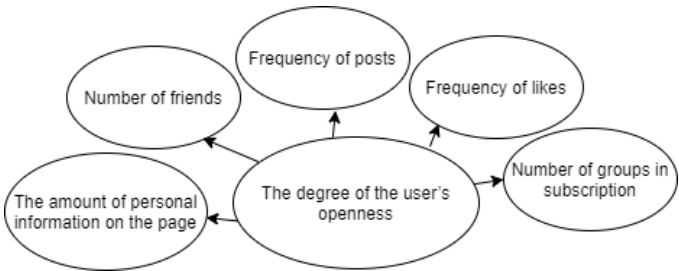
The purpose of this work is to develop a method for assessing the social media user's openness based on a five-factor personality model and automate it. To achieve this goal, we need to get data from the user's social media profile: the number of groups in his subscriptions, the number of his friends; the presence of information in the columns "status", "about yourself", "classes", "books", "quotes", "interests", "inspiration", "the main thing in people", "the main thing in life". It is necessary to identify a series of criteria to make the assessment, and, based on a five-factor questionnaire, create a mathematical model that reflects the results of the study. In this paper, it was decided to use the Bayesian network — a graphical structure to represent probabilistic relationships between a large number of variables and to implement probabilistic inference based on these variables [3]. This model allows you to work with incomplete data. To denote the developed mathematical model, it is necessary to write a parser program and collect data about VKontakte users, on the basis of which to calculate the average indicators for each factor in the model.

IV. MODEL DESCRIPTION

As a result of the work, a prototype of the model was proposed, the structure of which is described below. This model is designed to assess the user's openness by analyzing the data available in social media. In the work, the estimation is based on some numerical characteristics that can be directly obtained from the user's page in the social media. To assess the user's openness, the characteristics, reflecting the activity of his interaction with the social media, were selected: the frequency of "likes", that is the number of "likes" marks, by which the user evaluates the content consumed over a certain period of time; the frequency of posting — this includes the frequency of posting his own posts on the page and the frequency of reposts from various communities and other users; the number of friends; the amount of personal information in the profile; the number of groups in subscriptions. To assess the amount of personal information on the page, we used data of filling in/not filling the fields from the "Interests" and "Life Position" sections by the VKontakte user, namely: status, about yourself, activities, interests, favorite books, favorite quotes, sources of inspiration, the main thing in life, the main thing in people. The presence of the above fields was analyzed, as, presumably, they reflect information about the personal characteristics of the user deeper, in contrast to the sections "Contacts", "Education" or "Career". That is, filling them out may indicate the desire of this user to tell others about his interests and life priorities, from the presence of which it is possible to draw an indirect conclusion about his psychological characteristics (extroversion/introversion) [32]. Variables in the Bayesian network correspond to these characteristics. Each variable can take one of the following values: "Low", "High". In addition to numerical characteristics, it may be useful to analyze the user's areas of interest by the content consumed by him. According to the assumption, the presence of a large number

of groups with news content, advertisements (as well as the presence of similar reposts on the user's page) may indicate a high degree of openness to new experiences, while a small number of groups with narrowly focused specifics indicates that the user is closed to information that is out of his circle of interests. In this paper, the practical evaluation uses only information about the number of groups in the user's subscriptions, and in the future it is also planned to analyze its content.

The proposed model of the degree of user openness is shown in picture 1.



Pic. 1. The degree of the user's openness

The above model is universal for making the necessary assessment based on the user's profile in any of the existing social media (VKontakte, Instagram, Facebook, Twitter, Odnoklassniki, etc.). In this paper, the analysis is based on the pages of users in the social media VKontakte.

V. DENOTING VARIABLES

The proposed network model is binary, i. e. each node can take one of two values. To perform the calculations, it is necessary to match the numerical indicators to the binary indicators in the network nodes. In order to match numerical characteristics to binary, the average numerical values of users of the VKontakte social media were used as thresholds for each characteristic. These values were calculated based on a dataset of 100 users with less than 150 friends and 100 groups in subscriptions, in order to simplify data collection. In order to get a more average result for each user, the frequency of "likes" was calculated as follows: the number of "likes" that user left on the posts of his friends and groups in his subscriptions for 3 days was divided by 3. A time interval of a week, selected experimentally, is used to calculate the frequency of posting. The correspondence between numerical and binary values is shown in table 1.

TABLE I. CONVERTING NUMERICAL INDICATORS TO THE BINARY

	Low	Hight
Number of friends	<= 71 people	> 71 people
Frequency of posts	1 post per week	> 1 post per week

Frequency of likes	<= 4 "likes" per day	> 4 "likes"per day
The amount of personal information on the page	<= 2 units	> 2 units
Number of groups in subscription	<= 52 groups	> 52groups

a. Converting numerical indicators to the binary

At this stage, the conditional probabilities in the network set by experts. Based on the proposed model, using the obtained average indicators for each criterion, an application was written in C#, which collects the necessary data using the VKontakte user Id and uses the Smile library to work with Bayesian networks to perform the necessary evaluation.

Table 2 shows an example of evaluation based on information about three users:

TABLE II. THE DEGREE OF VKONTAKTE NETWORK USER'S OPENNESS

	Numb. of friends	Freq. of posts	Freq. of likes	The amount of personal informati on on the page	Numb of groups in subscri ption	The degree of the user's openness
1	148	0	0	0	68	High: 0,262 Low: 0,738
2	51	0	1	0	53	High: 0,092 Low: 0,908
3	132	0	10	2	64	High: 0,768 Low: 0,232

a. The degree of VKontakte network user's openness

VI. CONCLUSION

This paper presents a prototype model for assessing the user's openness to new experiences according to the five-factor personality model. As a result of the work, a set of criteria were identified for assessing the degree of social media user's openness, and a model was proposed to make the necessary analysis. In future work, it is planned to improve the mathematical model to obtain a more accurate estimate: train the model on a dataset instead of expert evaluation of conditional probabilities, and add to the model not only numerical characteristics, but also an analysis of the content consumed and published by the user in the social media.

he work made as a part of the project under the state task of SPIIRAN No. 0073-2019-0003, with the financial support of the RFBR, projects No. 20-07-00839.

REFERENCES

- [1] How to protect yourself in social media.. Expert recommendations. Available at: <https://www.m24.ru/news/obshchestvo/16072020/125353> (accessed 20.09.2020)
- [2] Approaches to the development of models for analyzing text information in social media profiles in order to build a user's vulnerability profile. Available at: <https://elibrary.ru/item.asp?id=29850652> (accessed 20.09.2020)
- [3] Learning Bayesian Networks Neapolitan R.E., 2004
- [4] Tulup'ev A. L. Algebraic Bayesian networks: local logical-probabilistic inference. - SPb SUT, 2019. C. 2-12. (In Russian)
- [5] S. A. Shchebetenko Big five and usage of the "VK" online social network - Bulletin of the south ural state university. Series: psychology, 2013. Vol. 6, no. 4, pp. 73-83.
- [6] Khlobystova, A.O., Abramov, M.V., Tulupyev, A.L., Zolotin, A.A. Search for the shortest trajectory of a social engineering attack between a pair of users in a graph with transition probabilities // Informatsionno-Upravliaushchie Sistemy. 2018. No. 6. Pp. 74-81
- [7] Abramov, M.V., Azarov, A.A. Identifying user's of social networks psychological features on the basis of their musical preferences // Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017. 2017. Pp. 90-92
- [8] E. D. Fraustein, S. Flowerday Susceptibility to phishing on social network sites: A personality information processing model - Elsevier Ltd, 2017.
- [9] M. Gratian, S. Bandi, M. Cukier, J. Dykstra, A. Ginther Correlating Human Traits and Cybersecurity Behavior Intentions - Elsevier Ltd, 2017.
- [10] Kalashnikov I. D. Statistical models for social media data and the use of Bayesian methodology - Proceedings of the International Conference on Soft Computing and Measurement (SCM-2016). St. Petersburg 2019. vol. 1, pp 406-409. (In Russian)
- [11] Toropova A. V., Tulupyeva T. V. Bayesian belief network as a behavior intensity rate model on the example of posting in a social network Proceedings of the International Conference on Soft Computing and Measurement (SCM-2016). St. Petersburg 2020. vol 1, pp 20-22. (In Russian)
- [12] Tulupyev A. L., Paschenko A. E., Azarov A. A. Information model of the user, who may be under the threat of socioengineering attack - "Trudy SPIIRAN", St. Petersburg, 2010. No. 2, pp 143-155. (In Russian)
- [13] Abramov, M. V., Azarov, A. A. Modeling of social engineering attacks using Bayesian networks. Proceedings of the International Conference on Soft Computing and Measurement (SCM-2016). St. Petersburg, 2016. pp. 71-74. (In Russian)
- [14] Youyou W., Kosinski M., Stillwell D. Computer-based personality judgments are more accurate than those made by humans // Proceedings of the National Academy of Sciences of the United States of America. 2015. Vol. 112 (4). Pp. 1036-1040.
- [15] C. Wolverton, D. Stevens The impact of personality in recognizing disinformation - University of Louisiana at Lafayette, Lafayette, Louisiana, USA. C 181-191.
- [16] Shalamov R. A., Abramov M. V., Tulupyev A. L., Azarov A. A. Automation of the assessment of the degree of severity of the psychological characteristics of users for making personnel decisions: testing in a social media. In the materials of the conference "Information Technologies in Management" (ITM-2018). Central Research Institute "Electropribor". 2018. pp. 497-500. (In Russian)
- [17] Abramov M. V., Azarov A. A., Tulupyeva T. V., Tulupyev A. L. Model of malefactor competencies profile for analyzing information system personnel security from social engineering attacks - "Information and control systems", St. Petersburg, 2016. No. 4, pp. 77-84. (In Russian)
- [18] Jan-Erik Lönqvist, Juha V.A. Itkonen, Markku Verkasalo, Panu Poutvaara The Five-Factor Model of personality and Degree and Transitivity of Facebook social networks - Journal of Research in Personality, volume 50, June 2014. C 98-101.
- [19] Abramov MV, Azarov AA. Identifying user's of social networks psychological features on the basis of their musical preferences. In the proceedings of the International Conference on Soft Computing and Measurement (SCM-2017). Saint-Petersburg, 2017. Vol. 1, pp. 130-133. (In Russian)
- [20] Khromov A. B. Five-factor personality questionnaire - Ministry of education of the russian federation, Kurgan state university, 2000. (In Russian)
- [21] J.J. Voronkova, O.M. Radyuk, I.V. Basinskaya The "big five" or five-factor model of personality - Khabarovsk: Publishing House of the Pacific State University, 2017. - p. 39-45. (In Russian)
- [22] Goldberg L. R. An alternative «description of personality»: The Big-Five factor structure - Journal of Personality and Social Psychology, 1990. T. 59, № 6, C. 1216-1229.
- [23] Goldberg L. Language and individual differences: the search for universals in personality lexicons - Review of Personality and Social Psychology, 1981. T. 2, C. 141-165.
- [24] Zhao H. The big five personality dimensions and entrepreneurial status: a meta-analytical review - The Journal of applied psychology, 2006. T. 91, № 2, C. 259-71.
- [25] Burlachuk L. F. Dictionary-handbook of psychodiagnostics, 2007. C 689-690. (In Russian)
- [26] A five-factor personality questionnaire. Available at: <http://www.syntone-spb.ru/library/news/content/283.html> (accessed 08.12.2020)
- [27] The number of attacks using social engineering increased by 147% in 2020. Available at: <https://www.securitylab.ru/news/515178.php> (accessed 11.03.2021)
- [28] Stepanova M. M., Khaykin V. D. Analysis of the content of social media for the purpose of diagnosing a psychological portrait of a person - Prospects for the development of science and education: a collection of scientific papers based on the materials of the international scientific and practical conference on September 28 2012 Part 12. Tambov: LLC "Consulting Company Ucom", 2012. pp. 136-137. (In Russian)
- [29] Zakiyeh S. E., Mohammad M. R. K., Farnaz J. M., Simin S., Zahra T., Shima M. M. & Koosha S. O. Bayesian network modeling for diagnosis of social anxiety using some cognitive-behavioral factors - Network Modeling Analysis in Health Informatics and Bioinformatics volume 2, 2013. C. 257-265.
- [30] Hulya Olmus, S.Oral Erbas Determining the conditional probabilities in bayesian networks - Hacettepe Journal of Mathematics and Statistics vol. 33, 2004. C 69-76.
- [31] Azarov, A. A., Tulup'eva, T. V., Suvorova, A.V., Tulup'ev, A. L., Abramov, M. V., & Yusupov, R. M. (2016). Social Engineering attacks. Problems of analysis. Nauka Publishing House., 2016. pp. 352. (In Russian)
- [32] Stankevich M. A., Ignatiev N. A., Smirnov I. V., Kiselnikova N. V. Personality Traits Prediction from VKontakte Social Media - "Voprosy kiberbezopasnosti", 2019, No. 4. pp. 80-87. (In Russian)
- [33] Sapon I. V., Ledenev D. E. Virtual personality: the network effect of self-disclosure. Scientific result. Sociology and Management. 2020. Vol. 6, no. 2. pp. 36-50. DOI: 10.18413/2408-9338-2020-6-2-0-3 (In Russian)
- [34] Sapon I. V., Ledenev D. E. Motives for using the social media as a factor of self-disclosure in the profile - Siberian State University of Telecommunications and Informatics, Russia, Novosibirsk, 2020. No 61-11. pp 75-78. DOI: 10.18411/lj-05-2020-248 (In Russian)
- [35] Sapon I. V., ledenev D. E. User's media activity as a self-disclosure factor in the sns profile - "Trends in the development of science and education" 2020. pp. 322-235. (In Russian)
- [36] Bezogova M. S., Iontseva M. V. Socio-psychological aspects of user interaction in virtual social networks. Online magazine "World of Science" 2016, Vol. 4, No. 5. Available at: <http://mir-nauki.com/PDF/43PSMN516.pdf>
- [37] Zlokazov K. V., Kolmykova T. I., Rybakova E. A. Extroversion, neuroticism and psychoticism influence on perception of potential threat to health, reputation and social status - УрГПУ, Педагогическое образование в России, 5, 2017. C 75-83. (In Russian)
- [38] Shchebetenko S.A. (2016) Online Social Network Users' Attitudes toward Personality Traits Predict. (In Russian)
- [39] Behavior of their Friends. National Psychological Journal, [Natsional'nyy psikhologicheskii zhurnal], 4, 34-44. DOI: 10.11621/npj.2016.0404. (In Russian)
- [40] Cusack, B., & Adedokun, K. (2018). The impact of personality traits on user's susceptibility to social engineering attacks. DOI: <https://doi.org/10.25958/5c528ffa66693> (In Russian)