Application of modern hybrid technologies in the field of data analysis to struggle with spear phishing

Dmitriy Gordeev[1], *Margarita* Karaseva[1,2], and *Tatiana* Karaseva[3*]

[1]Reshetnev Siberian State University of Science and Technology, Department of System Analysis and Operations Research, 660037, 31, Krasnoyarsky Rabochy av., Krasnoyarsk, Russian Federation
[2]Siberian Federal University, Department of Digital Management Technologies, 660074, Academic Kirensky St., 26a, Krasnoyarsk, Russian Federation
[3]Siberian Federal University, Department of Business Informatics and Business Process Modeling, 660074, Academic Kirensky St., 26a, Krasnoyarsk, Russian Federation

Abstract. To develop techniques to combat spear phishing, it is necessary to constantly develop new ways to detect it. This requires studying the characteristics and factors that indicate spear phishing. The formation of these factors is based on the methods used in mature organizations and on the basis of statistical analysis over past years. This makes it possible to track trends in changes in the tactics and techniques of attackers over time and taking into account the information security systems used. The paper considers the possibility of classical solutions hybridization with neural networks.

1 Introduction

Today IT technologies play a key role in the life of society, the value of information has increased dramatically. This statement is especially important for organizations and legal entities; for them, digitalization has provided a number of advantages. In this regard, the task of protecting information has arisen [1-2].

The main motivation of attackers is to obtain confidential data, this could be user authentication data, important documents, sensitive information about a company, its employees or owners, for use in blackmail. Often, attackers do this for the purpose of financial enrichment [3-4].

Phishing is increasingly becoming one of the most popular methods of attack. This type is successful because it is based on social engineering methods. And the proportion of such attacks in recent years is second only to DDOS attacks and is growing every year [5]. The greatest danger is spear phishing.

Spear phishing is a tool primarily for serious attacks on large enterprises, banks or famous people, because it requires a higher level of attacker skill [6]. While regular phishing requires less skill and is initially designed to fail, spear phishing, like a targeted attack, is more complex and uses more advanced techniques, and requires more preparation in advance.

* Corresponding author: tatyanakarasewa@yandex.ru

Spear phishing often uses other types of phishing, e.g, to obtain sensitive information that attackers can then use. The targets of such attacks are email addresses, full names of employees in departments of interest to the attackers and etc. To obtain this information, various tricks are invented so that the person trusts them and provides data, often depriving the person of time to think about his actions.

The consequence of this is that one of the main challenges of information security is the "constant race" between security specialists and attackers. As part of this "race," both sides are developing new hacking methods for some and ways to counter them for others. Often these methods are based on new technologies and developments; one of these new technologies that has recently taken over the world is artificial neural network (ANN) [7].

Hybridization with the ANN in information security systems makes it possible to automate the system and reduce the required processing power. Reducing processing power will expand the target audience of this solution, since most effective solutions are now complex solutions for which spear phishing is not the main task.

2 Methods

This paper will discuss a method of attack applying emails. In this regard, it is assumed that this information security system should be installed on the mail server to control all incoming letters.

A targeted attack is carried out in several stages:

- Setting attack targets;
- Preliminary study of data of potential victims;
- Selecting victims and carefully studying the data;
- Creation and distribution of a phishing letter based on collected data and social engineering;
- Development of the attack.

Information security using ANN is proposed to be used at the stage of distributing phishing emails [8]. The principle of operation of this should be based on the analysis of the email headers, as well as the contents of the letter itself.

Before forming factors, it is necessary to determine the capabilities and goals of attackers.

Spear phishing assumes that the information security intruder is external, in addition, it requires a lot of effort and money. In this regard, most attacks of this type are carried out by mature groups. Therefore, the chances that a phishing attack was carried out by attackers with basic capabilities are minimal.

The next group of intruders consists of criminal groups and competing organizations. Their capabilities are already sufficient to carry out a serious attack on the system of medium and small organizations. The goals are to disrupt transactions, obtain commercial and other secrets, conduct industrial espionage, and steal finances.

In some cases, actual intruders may have higher capabilities. These could be terrorist groups and foreign special services. These are especially relevant offenders for large and medium-sized organizations engaged in activities in the fields of finance, healthcare, communications services, energy, military and nuclear industries.

First of all, the factors should be based on current methods of implementing threats. To avoid errors of the first type, an organization should introduce regulations for writing letters, where it should also consider the limitation of the mail protocol used within the organization.

One of the ways to implement many threats is to install malware on the victim's workstation. Using malware, an attacker can steal data by transmitting it through the work to his resources. Another option would be to obtain employee authentication data by installing programs such as keyloggers. And in conjunction with malware that can change OS registries, an attacker can be able to connect to a remote desktop.

Moreover, attacker is able to scan the network by using a malware. In order to deliver a malicious program to a victim's computer, an attacker only needs to create a phishing email containing a file or link that allows the user to download and install the malware without the user noticing. According to Positive Technologies, malware is the most widespread among all methods of compromise in 2023 (Fig. 1).

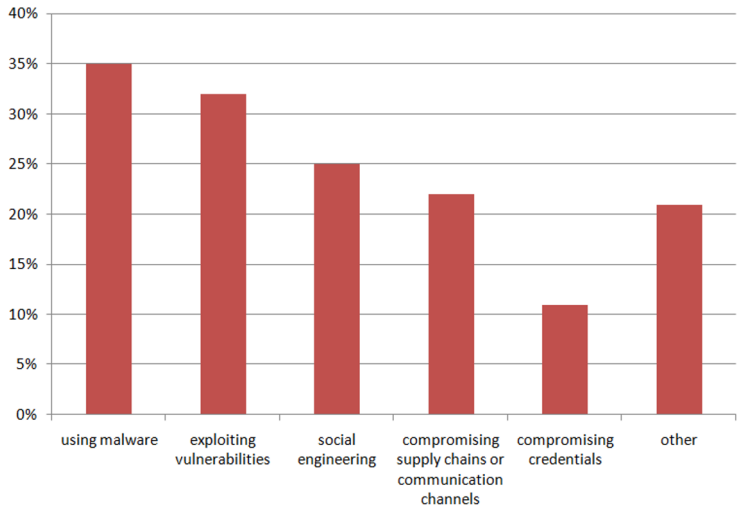

Fig. 1. Methods distribution in 2023 year.

Considering the trends according to the Central Bank of Russian Federation and Positive Technologies, the distribution of malware by class, where the first place in 2019 was occupied by encryptors, and in 2020 spyware. Ransomware ranked first among malware in the financial sector in 2023, followed by downloaders in second place. In 2022, downloaders were in first place with a share of 59%, and ransomware and spyware shared second place with a share of 18%. This trend is justified by the transition of many organizations to a remote and partially remote format due to COVID-19, as well as the reverse transition.

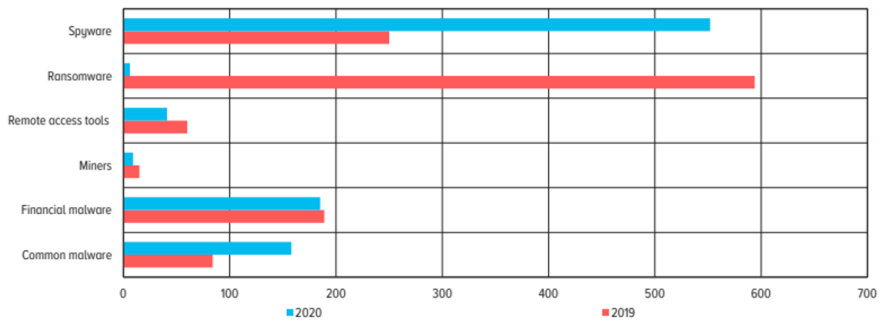

Fig. 2. Malwares distribution in 2019-2020 years.

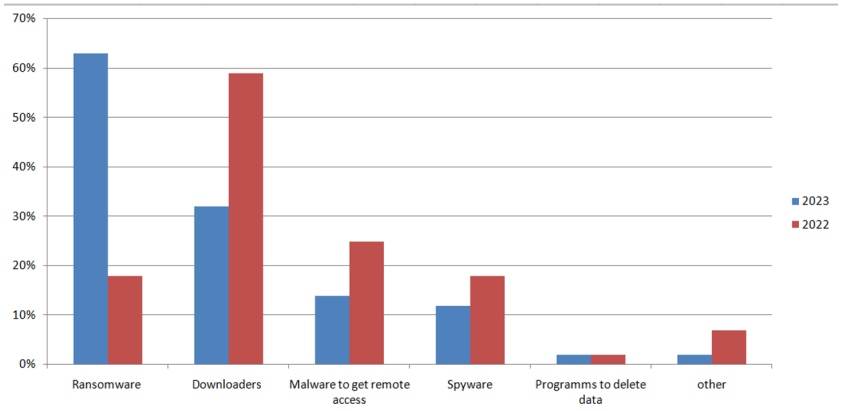

Fig.3. Malwares distribution in 2022-2023 year.

When studying the data collected so far on the types of malware for 2024 in the world, it is very similar to the average for 2023 and 2022 in Russian Federation, but at the same time, the distribution in Russian Federation again changed vector, where spyware began to predominate and ransomware and downloaders are losing their positions.

One of the factors for the ANN is the presence of a header with a file inside the letter and, as a consequence, the format of this file [9]. For some formats, it is worth immediately passing a red marker, for example, the «.sfx» format, and for such «.exe» formats, a yellow marker, for example. For this factor, the security policy should pre-determine legitimate formats and their compliance with a specific color.

When analyzing the text of an email, it is worth focusing on some key phrases. Attackers can use phrases that, in some cases, try to evoke emotions of confusion, guilt, and others in the target of the attack, and then put pressure on the urgency of the task.

Another factor in text analysis can be short links that look very similar to legitimate ones. Their danger lies in the fact that they may use a similar symbol. For example, it is almost impossible for a person to distinguish the English letter “a” from the Russian “a”. In addition to checking only the text for such similarity of characters, you should check email addresses, domain and other headers as a separate factor. It is not uncommon for domain names to be faked by changing just one character. The link should also be checked to ensure that the site has a secure connection.

To simplify the sender header check, you can create a fixed list of all legitimate domains, IP addresses, and internal emails.

While the sender's application may be indicative of the headlines, attackers may use public services to self-test organizations for phishing susceptibility, such as Gophish.

3 Description of tasks

The system is aimed at countering the following threats:

The threat of unauthorized access to authentication information;

It consists in the ability to extract passwords, usernames or other credentials from computer RAM or steal (copy) password files (including those stored in clear text) from computer storage media. This happens by installing malware like keylogger or by creating a copy of a resource website with authentication data that is of interest to the attacker.

The threat of infecting your computer when visiting unreliable sites:

The threat lies in the possibility of violating the security of protected information by malicious programs that are secretly installed when users visit the system from their

workstations (either intentionally or by accidental redirection) to sites with unreliable content and launched with the privileges of discredited users.

This threat is caused by weaknesses in network traffic filtering and anti-virus control mechanisms at the organizational level.

This threat can be realized if system users visit sites with unreliable content from their workstations.

The Phishing Threat:

This threat is caused by insufficient user knowledge about phishing methods and tools.

This threat is relevant for all types of phishing including: spear phishing, vishing and others.

4 Conclusion

The hybrid solution under consideration will simplify the administration of the information security system, as it will reduce the number of tools used to detect spear phishing, and will also increase the efficiency of identifying illegitimate letters by increasing the speed of response and detecting difficult to distinguish factors. It is not enough to just solve a delivery method aimed at blocking to build a mature security system, since no single security measure is capable of providing 100% security. The complex should be aimed at detecting intrusions in case a compromise has occurred, as well as mitigating the consequences after an information security incident. But in addition to building a security package, it is necessary to set requirements for the anti-phishing protection system as well. Also, the ANN must be integrated into the information security system so that the results of the work of each of its elements separately circulate in the system. This will make it possible to make more accurate results of checking emails and, based on all this information, it will be possible to generate reports on information security incidents. It is important to develop regulations for updating the ANN to train it in new tactics and techniques used to implement targeted phishing.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant № 075-15-2022-1121).

References

1. D. P. Lestari, *Factors Affecting Security Information Systems: Information Security, Threats and Cyber Attack, Physical Security, and Information Technology* IJIIS: International Journal of Informatics and Information Systems **7(1)**, 16-21 (2024).
2. V. Skormin, Ja. Moronski, D. Mcgee, D. Summerville, *Biological Approach to System Information Security (BASIS): A Multi-agent Approach to Information Security* Lecture Notes in Computer Science **2691**, 435 (2003).
3. S. G. Govender, E. Kritzinger, M. Loock, *A framework and tool for the assessment of information security risk, the reduction of information security cost and the sustainability of information security culture* Personal and Ubiquitous Computing **25(5)**, 927-940 (2021).
4. P. Y. Leonov, A. V. Vorobyev, A. A. Ezhova and et al, *The main social engineering techniques aimed at hacking information systems* in Proceedings – 2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology, USBEREIT 2021, 13-14 May 2021, Yekaterinburg, pp. 471-473 (2021).
5. K. Mammadova, R. Aslanov, *Installation of integrated intellectual information security systems in open corporate networks – DDoS attack* InterConf **32(151)**, 643-651 (2023).

6. T. Xu, K. Singh, P. Rajivan, *Personalized persuasion: Quantifying susceptibility to information exploitation in spear-phishing attacks* Applied Ergonomics, **108**, 103908 (2023).
7. A. S. Ahmed, S. Kurnaz, A. M. Khaleel, *Evaluation DDoS Attack Detection Through the Application of Machine Learning Techniques on the CICIDS2017 Dataset in the Field of Information Security* Mathematical Modelling of Engineering Problems **10(4)**, 1125-1134 (2023).
8. V. A. Chelukhin, S. E. Tikhonov, Z. A. Piei, *Modern problems of information security in control and access control systems when using neural networks* Journal of Physics: Conference Series **2096,** 012159 (2021).
9. R. A. Antonov, E. V. Karachanskaya, G. V. Khandozhko, *Using Artificial Neural Networks to Estimate the Probability of Information Security Threat Occurrences* Automatic Control and Computer Sciences **55(8)**, 941-948 (2021).