How to Safely Communicate with a Phishing Attacker by Email?

Ladislav Burita, Aneta Coufalikova and Kamil Halouzka
The Department of Informatics and Cyber Operations; Faculty of Military Technology, University of Defence; Brno, Czech Republic

ladislav.burita@unob.cz
aneta.coufalikova@unob.cz
kamil.halouzka@unob.cz

Abstract: The published study is a part of the long-term research of emails with phishing attacks against the article's author. In the previous three years, 3 experiments were carried out to analyze phishing emails. The result is their detailed classification. The subsequent experiment was focused on defense against phishing attacks using the rules of the MS Outlook email client. The last experiment, which is the article's content, is devoted to analyzing communications with phishing attackers. A fake identity was created for the experiment and security rules were set up. A total of 100 phishing emails were answered, with a preference for those whose content was not aimed at fulfilling any request; that was clarified during the communications. The conducted literature search confirmed the assumption that no one is engaged in similar research, so the results of the research may be more interesting for the cybersecurity community. The articles of the literary research are focused on the issue of social engineering from an interdisciplinary perspective. A great deal of attention has also been oriented on the influence of social networks on people information perception or on their exploitation in cyber-attacks. The result of the study is a statistical analysis of communications and a detailed analysis of its content. Out of 100 replies to the phishing email, 32 (32%) were answered by the phisher. The longest communications had 6 cycles. If the phisher insisted aggressively on personal information, the communications was terminated. From the content of the communications, the attacker's procedures and his argumentation to obtain the required information were primarily examined. A detailed analysis of the texts from the communications aimed to answer the question of whether the phisher is a robot or a person. Further considerations are being made within the team on how to continue researching phishing attacks.

Key words: Phishing email, communications with phisher, fake identity, statistics, analysis.

1. Introduction

Over the past few years, there has been a large increase in targeted cyber-attacks on individuals to obtain sensitive, confidential, or personal information of random or specific individuals or organizations. Depending on how this information is obtained, these attacks can be divided into phishing, spear phishing, and whaling in combination with social engineering. A typical phishing attack is targeted on a large number of people, and the attackers hope to catch a few victims. Typically, a group of people receives an email that looks trustworthy and asks the victim to log on to a website and take a certain action. The email always contains a link to something that looks like a trusted website, and the victim are requested to enter their account details. Once these details are entered, the attacker uses them to steal, defraud, or obtain even more valuable information. The tactic of the attack is to target a larger number of potential victims in the hope of getting some to click on the email and fall for it.

While the goal of phishing is to target a large number of victims, spear phishing targets specific individuals with the goal of obtaining specific information. Prior to a spear phishing attack, the attacker collects publicly available information about the potential victim. They then use the information gathered to persuade the victim to provide more information or to perform a task. As spear phishing targets specific individuals, whaling is even narrower and targets only people in high-level jobs. The social engineering consists mainly of psychological manipulation of people with the aim of revealing sensitive information or performing some action. Social engineering is usually motivated by data theft, malware spread, or political aspects.

Three methods can be used to send phishing and answer processing. First, manual processing, which is strenuous, can contain errors and place high demands on the attacker's attention. The attacker must follow the communication flow and follow up on previous messages. If he or she is inattentive and erring, he or she risks being revealed. Second, attackers can use automatic phishing and automatic responses to the victim's initial response. These messages are prearranged and contain predefined information. Third, the use of bots is now common when communicating with customers of online sales portals or customer support lines. There is no reason for attackers to not use this technology to facilitate their actions. Longer communication is needed to assess whether a person or a bot is behind the attack emails, but the outcome is not always clear cut.

The aim of the article is to describe the results of the research on email phishing attacks against the article's author as a part of the long-term research. The first period in the years 2020-2022 was oriented to the analysis

of phishing emails, and the result was their detailed classification into five segments: Business, Charity, Fund, Transfer, and Others. The Business segment is characterized by the cooperation offered on an investment, contract, or project in the recipient's country, as a realization of a business opportunity. The letter in the Charity segment is mostly written in the Christian spirit, offering the recipient the donation to set up a charity fund and getting a reward for this mission. The alleged sender is an old woman, a widow, her deceased husband left millions of USD. The Fund segment included emails that promised the recipient money obtained from a fund, gift or inheritance. The Transfer segment through email requested cooperation for money or other asset transfer. The Others segment contains emails with marginal significance, not included in other segments, with an unspecific offer, repeated contact, undelivered package, etc. These emails were mostly chosen for the currently described experiment.

The following experiment in the phishing research analyzed defense against phishing attacks by applying the filtration rules of the email client. The rules were set up with the keywords characterizing the individual phishing segment carrying out the selected operation. The phishing emails were transferred into the designated directory, and the defense effectiveness reaches 90%.

The last experiment analyzes the author's communications with phishing attackers. A fake identity was created for the experiment and protection rules were set up. The emails were mostly from the Others segment with nonspecific content; a total of 100 phishing emails were answered. The result of the study is the content of the paper. The literature review confirmed the assumption that no found publication is engaged in a similar attempt, so the results of the research are sure interesting for cybersecurity research.

This paper is organized as follows. The "Introduction" provides an overview of our current phishing research and briefly informs about the latest experiment; following the "Literature review". The research methodology, fake identity, and basic security rules of the experiment are described in the section "Research methodology and security rules". The section "Description of the experiment and its statistical analysis" presents the intention of the experiment and its statistical overview; the section "Detailed analysis of the selected cases" deals with interesting communications with phishing attackers and analyzes them in detail. The "Discussion and conclusions" section summarizes the results of the experiment and suggests possibilities for further research.

2. The literature review

It is increasingly possible to see a combination of social engineering and phishing, where in many cases cyber-attacks are not targeted at the organization as a whole, but at the employees of that organization. In this case, specific human characteristics and psychology are used to circumvent the organization's technical security measures. Social engineering is increasingly becoming a more widespread approach used to compromise individuals and organizations, where it is relatively more advantageous to compromise a person compared to exposing a vulnerability in a security system. Social engineering attacks include physical, social, and technical aspects that are used at different stages of the attack. Even if such an attack is initially unsuccessful, any penetration into individual and organizational security processes can be exploited for future attacks.

The issue of social engineering and phishing is addressed in many publications, often focusing on the general characteristics of social engineering and its detailed characterization, such as A Study on the Psychology of Social Engineering-Based Cyberattacks and Existing Countermeasures (Siddiqi, et al., 2022). The study describes cyber-attacks that are based on social engineering, a process that exploits human psychology. It also provides an in-depth analysis of the approaches used to conduct social engineering-based cyberattacks. This study examines the human vulnerabilities that criminals exploit in security breaches. Further, the study highlights existing approaches, including machine learning-based methods that are being used to combat cyber-attacks.

The human factor as a cybersecurity vulnerability is elaborated on in the paper Vulnerabilities of Cyber Security of Technical Intelligentsia in Relation to Social Engineering (Kononovich, et al., 2021). The paper examines the impact of social engineering on human and collective consciousness, and on the human subconscious. The focus is on the influence of social networks and their functioning on information perception and thinking. The analysis mainly focuses on the vulnerability of educated people and describes recommendations to counter social engineering attacks, in particular: general system recommendations for a system of protection against social engineering attacks; methods of self-defense against the attack of cognitive bias "myside bias"; recommendations of non-system methods and means of protection against social engineering.

A different view of social engineering is presented in an interdisciplinary view of social engineering: a call to action for research (Washo, 2021), which examines the topic of social engineering from an interdisciplinary

perspective. The article reviews literature from the fields of information technology, psychology, and business, explaining the interconnectedness of social engineering with these fields and the need to understand it from multiple perspectives. The literature review is followed by an ethical perspective that analyses social engineering research from a philosophical and professional perspective. The proposed framework provides readers with a flexible model that they can use in their studies with an emphasis on a philosophical or practical ethical perspective.

In the field of social engineering, prevention is equally important in terms of preparing the employees of an organization for possible social engineering attacks. One of the articles that deals with this has been published under the title The Role of Employees' Information Security Awareness on the Intention to Resist Social Engineering (Grassegger and Nedbal, 2021) examines individual and organizational factors that influence employees' awareness of information security and the possibilities to protect against social engineering attacks. The paper evaluates a survey of 136 employees and clearly shows the importance of information security education and awareness. It also highlights the importance of the impact of individual risk-taking behavior on employee awareness of information security.

Another paper (Yeng, et al., 2022) conducts a quantitative and qualitative survey of doctors in the field on simulated phishing attacks using SMS. The results show that doctors who prioritized patient care were not as susceptible to those simulated phishing attacks because they did not pay attention to those messages. The paper also highlights the impact of employee security training on overall security in healthcare facilities. The importance of raising awareness of social engineering threats is discussed in (Alghenaim, et al., 2022), (Singh and Singh, 2022).

During the COVID-19 pandemic, there was a large increase in social engineering attacks. The reason for this increase lies in the increase in the amount of time that users, employees, students, and businesses began to use online environments in a big way. This increase in people's online presence is almost never preceded by education about cybersecurity and the different types of attacks to which the everyday Internet user may be exposed. This situation creates an opportunity for cybercriminals to attack, with social engineering attacks being the most common type of attack in a pandemic. The issue of the increase in cyber-attacks using social engineering during the COVID-19 pandemic has been addressed in the following papers (Venkatesha, et al., 2021), (Li and Lalani, 2020), (Kumaran and Lugani, 2020).

3. Research methodology and security rules

The research methodology in the published study is based on experience from previous research on phishing attacks. The data source for the experiment makes up of phishing emails, captured as a result of a previous experiment using email filtering rules in the MS Outlook client.

The basic security measure of the experiment is the creation of a fake identity - an email account of a non-existent person at gmail.com. Selected emails from the phishing email account were gradually copied to the fake identity email account and sent as a response to the phishing attack.

For communication, typical texts were created in advance, which became the content of the answers. If the phisher responded, communication continued from the fake identity email account. All communication was accurately recorded for the subsequent evaluation of the experiment.

Overview of additional safety measures:

- Do not click on URL addresses from phishing emails.
- Communicate outside the university computer network.
- Do not open suspicious email attachments.
- Do not send personal data (even about fake identity).

4. Description of the experiment and its statistical analysis

The experiment took place between May and July 2022 with the aim of finding out how communication with phishing attackers takes place. A total of 100 phishing emails were answered and their statistical evaluation is in Tab. 1. Significant parameters of the communications are the number of responses and the time it takes for the phisher to respond, see Tab. 2.

In the course of communication, some phishers attached arguments so that their identity and demands could

be considered real and true. Photos of the alleged corresponding person (see Fig. 1), their identification cards (see Fig. 2), and certificates confirming the winning of the fund (see Fig. 3) were sent. The patience of the phishers was sometimes admirable, they were not deterred by any negative response or rejection.

From the experiment, a clear experience can be drawn that it is possible to communicate with phishers without risk, provided that proper security measures are observed. A lot of interesting information was obtained, but the decision on whether a phisher is a person or a robot could not be unequivocally confirmed, see the next section.

Table 1: Email communication statistics

Ord	Email communication	Num	%
1	Without an answer	68	68
2	- of which undelivered	8	8
3	Aswer 1 times	10	10
4	Aswer 2 times	8	8
5	Aswer 3 times	11	11
6	Aswer 4 times	1	1
7	Aswer 5 times	1	1
8	Aswer 6 times	1	1
9	Total communicating	32	32
10	Total answers	74	

Table 2: Speed of responses

Ord	Speed of response	Num	%
1	On the same day	13	18
2	In 1 day	12	16
3	In 2 days	6	8
4	In 3 days	15	20
5	In several days	28	38

Figure 1: Example of photo of the alleged corresponding person

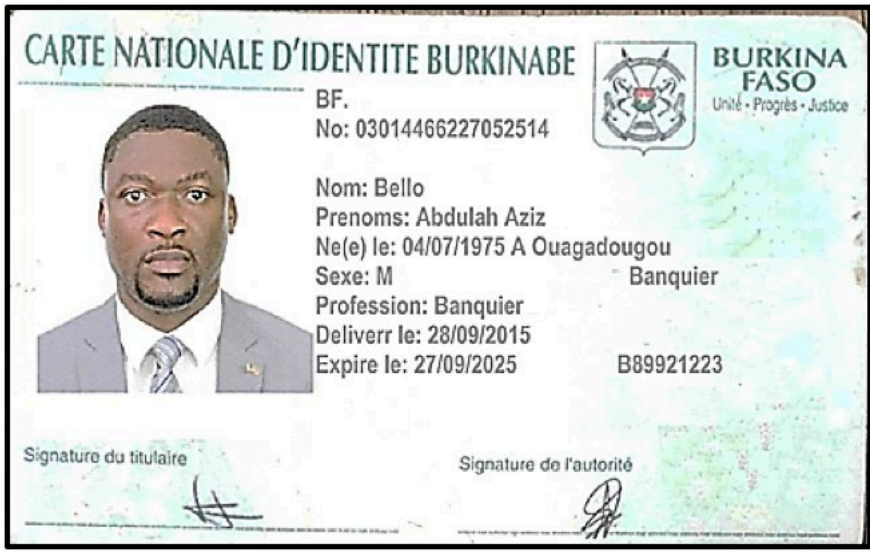

Figure 2: Example of ID-card of the alleged corresponding person

Figure 3: Example of a certificate

5. Detailed analysis of the selected cases

In the section, the analysis of selected communications is provided. The three longest communication is taken into consideration, analyzed, and conclusions drawn. The longest communication has been six rounds before being terminated due to the victim's decision. The main topic is the fund offer from a terminally ill woman. The second communication lasts five rounds, and the topic is a fund offer to help the poor. It is finished by the victim's decision, also. The third communication has four rounds. An attacker acts as an attorney to propose a business offer.

5.1 An offer from a terminally ill person

The longest communication lasted six rounds. The initial phishing email is short, the attacker poses as a terminally ill woman and asks the victim for help in transferring $5.5 million. 30% of this amount can be kept by the victim and 70% is to be used to help the poor.

After the first response from the victim, a comprehensive email is received. There is an emphasis on the urgency of the situation (3x word immediate, 2x word urgent) due to the condition of the patient. Religion is a very distinctive feature of email. The attacker writes about the Bible and about the service to God; the word God appears here 4 times. The reason for contacting the victim is God's will. This is followed by an assurance of the legality of the transfer and a request to send personal information (full name, nationality, occupation, age, picture, mobile phone), which will be used to write the authorization letter for the benefit of the victim. Compassion for orphans and the less privileged is highlighted several times.

After a negative response from the victim who does not want to share any personal data, an even more comprehensive email is received. The attacker again stresses the urgency of the situation, charity, and God's will. The attacker writes that she trusts the victim and would like to get to know the victim better, so she asks for personal information. There is a contradiction here: can one trust someone you do not know? Furthermore, detailed banking information is attached, and a fund release form is attached. The mentioned African Development Bank and its branch in Burkina Faso exist and the international phone code is also correct.

However, the address and phone numbers do not match the official contact details of the branch. The attacker's contact email adb-bf11@accountant.com does not appear to be fraudulent at first glance. Nevertheless, the domain accountant.com is not the real domain of the ADB bank, the real domain is afdb.org.

The victim replies that he cannot make the transaction. The attacker's response is very short, the victim is supposed to follow the bank's instructions. After nearly 2 weeks, the attacker asks if the victim has contacted the bank. The victim responds negatively and expresses concern about the legality of the transaction. The attacker reassures the victim and asks again if the victim has contacted the bank. The victim again expresses concern. The attacker responds after 12 days, reassures the victim again, and insists on contacting the bank. The victim stops the communication here.

The attacker communicates between 5:53 and 18:28 and the time between replies varies between 8 hours and 12 days. Long response times conflict with the urgency that the attacker is trying to evoke in the victim. After the initial response of the victim, automated responses may have been used. There is a general salutation in the response, and nothing from the victim's previous email is used. On the other hand, the response was sent 7 days late, suggesting manual processing. The attacker's motivation to obtain information is at a high level. The attacker responds logically, has an overview, and refers to previous communications. Salutations and goodbyes change slightly, which also indicates manual processing. Longer delays can be caused by the large number of emails being processed.

Language is at a very good level in the initial phase of communication. Later, the language level decreases and contains minor errors. There is a clear transition from the prepared parts of the attack to the unprepared parts, where the attacker's improvisation is needed when communicating with the victim.

5.2 Fund offer for the poor

Delivered phishing contains an offer of funds to help the poor and raise living standards in the recipient's region. The recipient was selected based on a profile from unspecified Microsoft and Google lists. The attacker requests a replay with name, address, country, age, sex, occupation, and phone number. Some of the data should not be requested by the attackers, as it should be known to them from the mentioned profile from Microsoft or Google lists. This makes the email look suspicious at first glance.

Although the victim answers negatively and does not want to share any personal data, the attacker continues his attempt and informs the victim about the funds deposited in the bank ($650,000.00), again requesting a name, address, and phone number. The attacker mentions a 'reference number' and encourages contacting the bank to gain access to the money, which is most likely to increase the credibility of phishing. However, he mentions informationschasebank@gmail.com as the bank's contact email. The address is very suspicious because the gmail.com domain is not used by banks, but they register their own domains. After the victim questions how to access the funds, the attacker responds illogically. In reply, he writes only that the money in the bank is waiting for the transfer. It is only after repeated questioning of the victim that the attacker replies that a fee of 100 euros is needed to transfer the money, and then the money will be transferred immediately. Interestingly, the attacker does not specify or require an account number where the fee or funds should be transferred. The victim responds negatively, no money will be sent, and expresses concern that it is a fraud. The attacker replies again illogically, using the same formulated answer that the money in the bank is waiting to be transferred. After re-emailing information that the victim will not send any money and fears an attack, the attacker sends the same reply for the third time about the waiting money in the bank. The victim terminates the communication.

It is clear from the communication that this is not a professional group, but amateurs. Errors can be considered to require information that the attacker should know, use of a non-bank domain, illogic in replies, and repeated replies. These are errors of a fundamental nature and it is obvious at first glance that this is a fraud. Partial automation is not used here, where after the initial reaction the victim can be automatically sent an email with information about the deposit in the bank, the reference number, and the contact to the representative of the bank. Several-day time delays do not correspond to automation, but rather to manual processing. The time interval of the attacker's responses from 9:37 to 15:40 also corresponds to manual processing. Phisher’s motivation to obtain personal data is not high, due to non-consecutive communication and longtime delays. However, it is possible that the long response time is caused by the large number of emails that the attacker has to manually process. It is also possible that this communication was provided by a robot that had learned the answers and did not modify them in any way. In this case, the time delays could be set deliberately or could be due to a large number of emails. The language level of the attack emails is good. However, it cannot be concluded

that the attackers are linguistically educated, as even the free English translators are at a very good level currently. The constant level of English could also be explained by the use of bots in communication.

5.3 Mutual business proposal

Initialization phishing is very short, the attacker offers mutual business and gives an email address for replies. It is suspicious that the given contact address is different from the sender's one. After a positive response from the victim, a message is sent with detailed information about the business offer. The attacker acts as an attorney for the deceased Mr. Malik, who is looking for his offspring for the settlement of the estate of $23,600,000.00. It is not important that the victim is not biologically related, it is enough that he has the same surname. Unfortunately, this is the last attempt to settle the estate before the entire fund is confiscated. The attacker states that he will split the amount into a 50% ratio for him and 50% for the victim. The attacker wants the victim to act as a relative and the attacker will use his position as a personal attorney for the smooth progress of the transaction. At the end of the email, it is stated that everything is legal. At this point, it is obvious that this is a financial fraud offer. The attacker uses the same surname as the deceased person and the victim, which is supposed to increase credibility and the success of the attack. So far, the attacker does not require any personal data or any information; he just ascertains the possible interest. He also does not induce any time pressure.
A positive response from the victim is followed by a comprehensive email repeating the information from the previous communication, and, in case of interest, the attacker offers the option of filling out a form to release funds and handle other needed documents. Therefore, he requires sending personal information to confirm the identity of the victim (international passport or driver’s license or any valid ID), contact address, occupation, marital status, mobile number, landline, and fax number). To increase credibility, the attacker attaches to the message a notification from the bank about the last chance to claim the funds of late client Engr. Phillip Malik. The attacker creates a slight time pressure.

After a negative response from the victim and a reluctance to share personal information, the attacker responds with pleas, appeals to the partnership, and assurances about the legality of the transaction. He gives the victim his word that he will not misuse personal information and again asks for the required data to be sent. To promote confidence, the attacker encloses a copy of the passport. According to the date of birth on the passport copy, the person is 60 years old, while the passport photo shows a young man. Here, too, the attacker makes another mistake and signs the email with a different name. Again, the victim responds negatively, and the attacker responds with more pleas and assurances. The victim terminates the communication.

The time span of the attacker's replies ranges from 12:27 to 20:15. It could be working hours in a different time zone. However, such a zone does not correspond to the attacker's declared Spain as his place of life. The delay between replies varies from 2 to 9 days. Initially, the attacker responds within 2 days; later, with the victim's unwillingness to share data, the attacker's motivation decreases, and his replies are sent with greater delays. The attacker responds logically, orients himself in the communication, and refers to the previous communication. In the later stages of communication, the attacker makes mistakes. This suggests manual response processing.

The language is at a very good level throughout the communication. Invariable addresses, constant English style, and similar repetitive passages could also be explained by the use of bots in communication. Also, the initialization of phishing and the response to the victim's first positive action could have been processed by automation or bots. Taking the surname and embedding it in the text of the email does not require any activity of a human operator. The parts of the communication where errors occurred are likely to have been processed manually.

6. Discussion and conclusions

Automatic phishing and bots can provide the initial stages of communication. Errors occur in the later stages of communication, where attackers need to react and improvise. This can indicate the limited scope of bots and the manual processing of emails. Bots can operate nonstop, whereas human operators only react at certain time intervals. Thus, it can be deduced from the times of communication whether it could have been a bot or a human. Language level and text style can also be a sign of automatic processing.

In the early stages of communication, where texts and information are defined, automatic responses or bots can be used, using online English language translators, which are currently at a very high level. In subsequent parts of communication, where there is a clear change in style and where errors occur, it is likely that the response

has been processed by a human. From the above, it can be deduced that such indicators can be used to determine whether a human or a machine provides communication.

Further research should be oriented to the best evaluation and use of all the results and conclusions to date for developing defensive abilities and skills of users against phishing attacks. The goal will be both high-quality teaching and training programs, as well as defense methodologies and rules.

Acknowledgments

This article presents the results of the research in cyber security of the Department of Informatics and Cyber Operations, University of Defence, as part of the project KYBERSILY (DZRO-209, 2022).

References

Alghenaim, M. F., Azaliah, N., and Rahim, F. A. (2022). "Exploring the Factors Influencing Employee Awareness of Social Engineering Threats: A Review". Applied Mathematics & Information Sciences, July, pp. 491-500.

DZRO-209. (2022). KYBERSILY: "Cyber forces and assets," Brno, University of Defence, 2021-2025.

Grassegger, T., Nedbal, D. (2021). "The Role of Employees' Information Security Awareness on the Intention to Resist Social Engineering". Procedia Computer Science, 181, pp. 59-66.

Kononovich, V., Kononovych, I., and Shvets, O. (2021). "Vulnerabilities of Cyber Security of Technical Intelligentsia in Vulnerabilities of Cyber Security of Technical Intelligentsia in. Odesa", ISIT 2021, pp. 127-136.

Kumaran, N., Lugani, S. (2020). "Protecting businesses against cyber threats during COVID-19 and beyond". [Online] Available at: https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond.

Li, C., Lalani, F. (2020). "The COVID-19 pandemic has changed education forever. This is how". [Online] Available at: https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/

Siddiqi, M. A., Pak, W., and Siddiqi, M. A. (2022). "A Study on the Psychology of Social Engineering-Based Cyberattacks and Existing Countermeasures". MDPI Journals.

Singh, I., Singh, Y. (2022). "Cyber-Security Knowledge and Practice of Nurses in Private Hospitals in Northern Durban, Kwazulu-Natal". Journal of Theoretical and Applied Information Technology, 15 January, pp. 246-267.

Venkatesha, S., Reddy, R. K., and Chandavarkar, B. (2021). SN Comput Sci., 6 February.

Washo, A. H. (2021). "An interdisciplinary view of social engineering: A call to action for research", Amsterdam: Elsevier.

Yeng, P. K., Fauzi, M. A., Yang, B., and Nimbe, P. (2022)." Investigation into Phishing Risk Behaviour among Healthcare Staff!. MDPI journals.