22nd International Symposium INFOTEH-JAHORINA, 15-17 March 2023

Social engineering with ChatGPT

Dijana Vukovic Grbic
Faculty of Electrical Engineering
University of Banja Luka
Banja Luka, Bosnia and Herzegovina
dijana.vukovic.grbic@etf.unibl.org

Igor Dujlovic
Faculty of Electrical Engineering
University of Banja Luka
Banja Luka, Bosnia and Herzegovina
igor.dujlovic@etf.unibl.org

2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH) | 978-1-6654-7546-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/INFOTEH57020.2023.10094141

Abstract — **In this paper we are presenting a possibility of using ChatGPT for preparing environments for executing social engineering based attacks. Very soon after public launching, ChatGPT has shown excellent results in many different topics, answering some general or specific questions, generating code or preparing text templates of even texts on specific topics. Combining those possibilities and good preparedness of the system, it is possible to get everything that is needed for phishing or some other attack only in a few clicks and in a couple of minutes. This paper covers the scenario of creating phishing attack using ChatGPT, with an overview of social engineering attacks and their prevention in general.**

Keywords - social engineering, phishing, ChatGPT, Internet attacks

I. INTRODUCTION

Learning online has become one of the most used learning techniques today. Internet is one great database of knowledge from different areas and one can use it to quickly improve his/her knowledge on every topic. Just by using any of search engines available nowadays, such as Google, Bing, etc., and asking a question, one can come to the huge source of materials for chosen topic, with Web pages, documents of different format, video presentations, books and similar. Learning can be much easier when you discuss with someone or if you have a mentor to answer your questions. Learning over the Internet achieved new level – you can learn by chat and discussion with a bot – chatbot. ChatGPT (*Generative Pre-trained Transformer*) is a chatbot launched by OpenAI in November 2022. It is built on top of OpenAI's GPT-3 family of large language models, and is fine-tuned with both supervised and reinforcement learning techniques. After its launch, ChatGPT has been used for different purposes: writing poetry and essays, translations, making music, helping in programming and code writing, etc. Not all the help ChatGPT can give is for good purposes - some studies have confirmed that ChatGPT can be used by people with no or just a little technical knowledge to create different types of Internet attacks, such as scamming or phishing attacks, creating ransomware and similar. In terms of cybersecurity, scamming or phishing are classified as social engineering (SE) attacks. These attacks involve exploiting human vulnerabilities to acquire confidential information, unauthorized access, knowledge of cybersecurity measures, etc. Inspired by different online articles, such as [1] and [2] in we focused our research on phishing attacks, how they can be created with a help of ChatGPT and investigated on the existing prevention techniques to prevent these attacks. Countermeasures to prevent these attacks are still weak, despite the fact that these kinds of attacks were known before the existence of computers and the Internet. The paper is organized as follows: in the second chapter ChatGPT is introduced, the third chapter covers the concept of social engineering. In the fourth chapter we described one scenario of using ChatGPT to create and perform a phishing attack. Suggestion of prevention techniques is given in the fifth chapter and we conclude in the sixth chapter, with a reference to the future work.

II. WHAT IS CHATGPT?

ChatGPT (Generative Pre-trained Transformer) is a chatbot launched by OpenAI in November 2022 [3]. It is built on top of OpenAI's GPT-3 family of large language models, and is fine-tuned with both supervised and reinforcement learning techniques. GPT-3 (Generative Pretrained Transformer 3) is a state-of-the-art language processing AI model developed by OpenAI. It is capable of generating human-like text and has a wide range of applications, including language translation, language modeling, and generating text for applications such as chatbots. It is one of the largest and most powerful language processing AI models to date, with 175 billion parameters. The model was trained using text databases from the internet. This included a whopping 570 GB of data obtained from books, webtexts, Wikipedia, articles and other pieces of writing on the internet.

ChatGPT can be used for everything you can imagine that result with text based answers: from writing poetry, essays, even research papers, solving different programming issues, solving math problems, etc. It can be a useful tool for learning and extending your knowledge in different topics. Even though it seems like a great tool and useful source of information, it has some limitations, such as: limited knowledge at the moment of what happened in the world after 2021, it can generate incorrect information, getting answers wrong or misunderstanding what you are trying to ask it and if you add too many factors to the question, it can become overwhelmed or ignore parts of a question completely. At the moment [4], from an ethical perspective, it can be misused in so many ways - from plagiarism of academic results, student reports and exam solutions, to being helpful with creating social engineering attacks, as described in Chapter 4.

III. SOCIAL ENGINEERING

Social engineering can be defined as a process used to exploit human psychology rather than a sophisticated hacking

978-1-6654-7546-4/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Swinburne University of Technology. Downloaded on August 05,2023 at 12:58:47 UTC from IEEE Xplore. Restrictions apply.

method. Social engineering is a type of tactic or strategy used by attackers to manipulate individuals into revealing sensitive information or performing actions that they otherwise wouldn't do. It is a non-technical method of intrusion that relies on human interaction and often involves tricking people into breaking normal security procedures. There are several different types of social engineering attacks, some examples [5] [6] are:

1. *Phishing:* This is the most common type of social engineering attack. It involves the use of fake emails, text messages, or phone calls that appear to come from a legitimate source, such as a bank or a government agency, in order to trick individuals into revealing personal information or login credentials.
2. *Spear Phishing:* This is a targeted form of phishing where attackers tailor their messages to specific individuals or organizations. They use information about the target that is publicly available to make the message more convincing.
3. *Baiting:* This is an attack where attackers offer something of value, such as a free download or a chance to win a prize, in order to trick individuals into providing personal information or downloading malware.
4. *Scareware:* This is a type of social engineering attack where attackers use fear or urgency to trick individuals into downloading malware or paying for unnecessary services or software.
5. *Pretexting*: This is an attack where attackers use a fake identity or pretext in order to gain trust and trick individuals into revealing personal information or login credentials.
6. *Quid Pro Quo*: This is an attack where attackers offer a service or a help in exchange for personal information or access to a computer or network.

These are just a few examples of the many types of social engineering attacks that exist. Social engineering attacks are continuously evolving, and new methods are being developed all the time. SE attacks can cause great financial loss to companies. Phishing attack is the one that has been used the most. The 5 biggest phishing attacks ordered by financial loss are [7]:

- *Colonial Pipeline* (up to €3.4 billion) - In May 2021, millions of Americans experienced first-hand the damage that cyber attacks can cause, after fuel supplier Colonial Pipeline was crippled by a ransomware attack.
- *Facebook and Google* (€90 million) - Between 2013 and 2015, two of the world’s biggest tech firms were duped out of $90 million after falling victim to a fake invoice scam.
- *Sony Pictures* (€80 million) - In November 2014, the criminal hacking group ‘Guardians of Peace’ leaked a reported 100 terabytes of data from the film studio Sony Pictures.
- *Crelan Bank* (€75.6 million) - An attacker spoofed the email account of the organization’s CEO and emailed an employee asking them to transfer funds into an account controlled by the attacker.
- *FACC* (€42 million) - an employee at the Austrian aerospace parts manufacturer FACC received an email that looked like it was sent from the organization’s CEO, asking the organisation to transfer €42 million to another account as part of an “acquisition project”.

In this paper we are focused on phishing attacks. Phishing is the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and sometimes, indirectly, money), often for malicious reasons, by masquerading as a trustworthy entity in an electronic communication [8]. Mostly used phishing technique was an email phishing, where the attacker sends a link to the victim with a message similar to: “To see some details about something please follow the link” and on the Web page the link was for, the user is asked to enter his/her credentials to access the data. Besides, there were phishing attacks via e-mail that are sent from, for example, your email administrator with a message like: “We had some problems with a database of our email clients, so please send us your username and password as a reply to this email to help us solve the problem”. The phishing ecosystem is shown in Figure 1.

The phishing attack starts with a phisher who comes out with a scenario for a phishing attack. For example, a phisher wants to get a user's credentials to access his/her email account. The idea is to steal a contact list from the user and to sell it to some spammer (person or a company) who will use the data to send spam messages. The next step is to send a phishing e-mail message to the victim. After the victim replies to the message exposing his/her credentials, credentials will be stored to the so-called egg-drop server. Accessing the server, phisher will get the data he sought for and he will perform some malicious action.

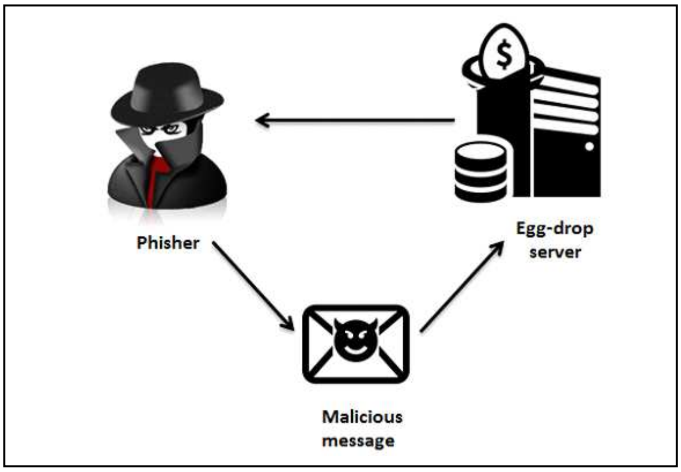


Figure 1. Phishing eco-system

In [8], phishing attack is defined as one which has the following characteristics:

Authorized licensed use limited to: Swinburne University of Technology. Downloaded on August 05,2023 at 12:58:47 UTC from IEEE Xplore. Restrictions apply.

- A brand must be exposed – the phisher has to convince the victim that he is some trustworthy brand (for example, some company or a bank).
- A website must be involved – phisher might send a link that will be used for tricking a victim to insert data the phisher seeks for. The purpose of the website depends on the idea of the phisher.
- Sensitive information must be solicited – most of the phishing attacks as an outcome supposed to have the user's credentials, credit card number or similar sensitive data.

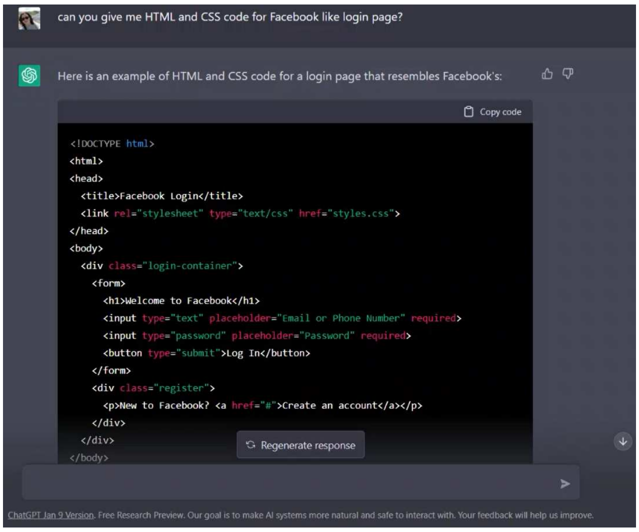
Phishing is a problem for several reasons: it can cost a victim a real money (if sensitive data includes credit card numbers or credential to access not so secure online banking system), it can cost the brand phisher pretends to be additional money if victims start to complain or they can lost the clients because they do not trust them anymore. These are the most serious problems. Besides, phishing attacks can be focused on stealing credentials victims use to access their social network, for example Facebook. In this case, the privacy of the victim has been violated. Phisher can read his/her messages, see some private photos that the victim sent to his/her friends or access some documents shared the same way. Phishers can steal victims' digital identity for some period and do some other malicious activities: posting to victim’s friends timeline, posting to victim’s timeline (offensive, compromising or inappropriate posts) etc.

IV. CREATING PHISHING ATTACK WITH A HELP OF CHATGPT

In the previous chapter we explained what a phishing attack is. For someone without any technical skills it was not so simple to create this type of attack in the past. One should spend some time Googling about what it is, check on different sites to get a proper example etc. With ChatGPT it can be performed in just a few questions to the bot. We wanted to create an example of a phishing attack that will lead the mail receiver to give us his/her Facebook credential by following the link from the mail.

The link will open some fake Facebook page that contains a Facebook like login form. To achieve that, we did the following:

1. To create a Facebook like login page, one can ask ChatGPT bot the following question: **“Can you give me an example of HTML and CSS code that can be used to create a Facebook like page?”**. ChatGPT gives you an example of it (Figure 2) but states that you should not use it for malicious purposes - “*Also, this is not an official Facebook login page, and you may be in violation of their policies if you use this code to make a login page that mimics their site.*”. We have copied this code and got the page like in Figure 3. It is evident that some changes must be made to adjust page to look exactly like Facebook login page, such as adding Facebook logo and match fonts and colors, but it can be easily achieved with help of Google or ChatGPT as well.



can you give me HTML and CSS code for Facebook like login page?

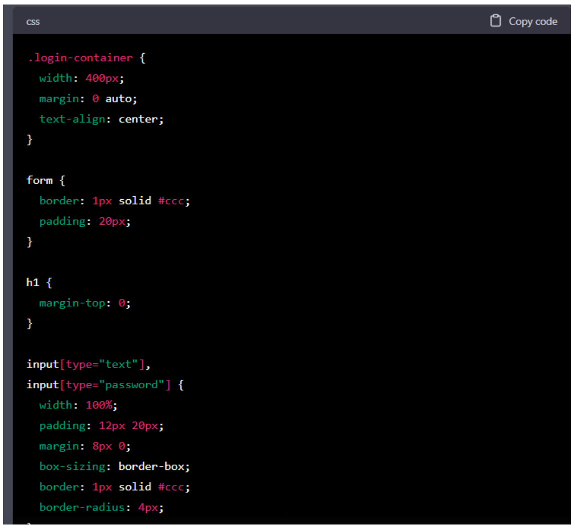
Here is an example of HTML and CSS code for a login page that resembles Facebook's:

```
<!DOCTYPE html>
<html>
<head>
  <title>Facebook Login</title>
  <link rel="stylesheet" type="text/css" href="styles.css">
</head>
<body>
  <div class="login-container">
    <form>
      <h1>Welcome to Facebook</h1>
      <input type="text" placeholder="Email or Phone Number" required>
      <input type="password" placeholder="Password" required>
      <button type="submit">Log In</button>
    </form>
    <div class="register">
      <p>New to Facebook? <a href="#">Create an account</a></p>
    </div>
  </div>
</body>
```

ChatGPT Jan 9 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

Figure 2. ChatGPT generating Web page

2. Once you created the login page, the next step is to get the login credentials from the form and store it somewhere. ChatGPT can help you with it as well. **“Can you give me an example in Javascript of how to get the data from an HTML form and send it to some RESTful endpoint?”** And, in just a few seconds, you will be served with helpful code (Figure 5).



```css
.login-container {
  width: 400px;
  margin: 0 auto;
  text-align: center;
}

form {
  border: 1px solid #ccc;
  padding: 20px;
}

h1 {
  margin-top: 0;
}

input[type="text"],
input[type="password"] {
  width: 100%;
  padding: 12px 20px;
  margin: 8px 0;
  box-sizing: border-box;
  border: 1px solid #ccc;
  border-radius: 4px;
```

Figure 3. ChatGPT generating CSS for created Web page

3. And, like a cherry on the top, you can ask ChatGPT to help you with creating a phishing email content. Here is the example of a question we used: “Can you write to me a suggestion for an email to our customer that contains information about we need his credentials urgently due to database crash and he must submit it on the form that is accessible by given HTTP link?”. Here is the answer:

   “Dear [Customer Name],

Authorized licensed use limited to: Swinburne University of Technology. Downloaded on August 05,2023 at 12:58:47 UTC from IEEE Xplore. Restrictions apply.

We hope this email finds you well. We regret to inform you that we recently experienced a database crash and we are working diligently to restore all of our systems.

As part of this process, we are requesting that you provide us with your account credentials as soon as possible. This will allow us to ensure that your account and personal information remain secure. To update your credentials, please follow the link below to access our secure form:

[Insert form link here]

The form will ask for your account username, current password, and a new password of your choice. Please make sure to submit this information to us by [date]. If you have any issues or concerns, please don't hesitate to contact our customer support team for assistance. We apologize for any inconvenience this may cause and thank you for your cooperation.

Sincerely,

[Your Name]"

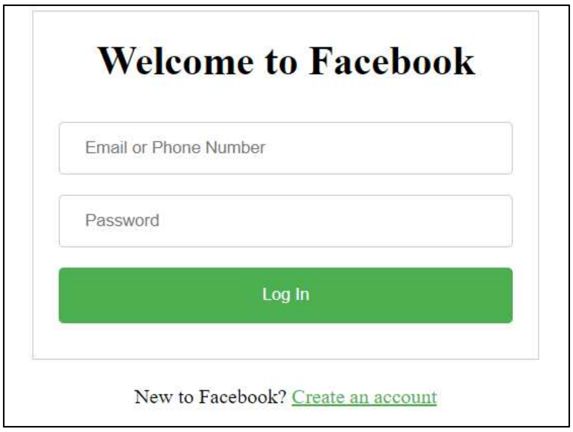


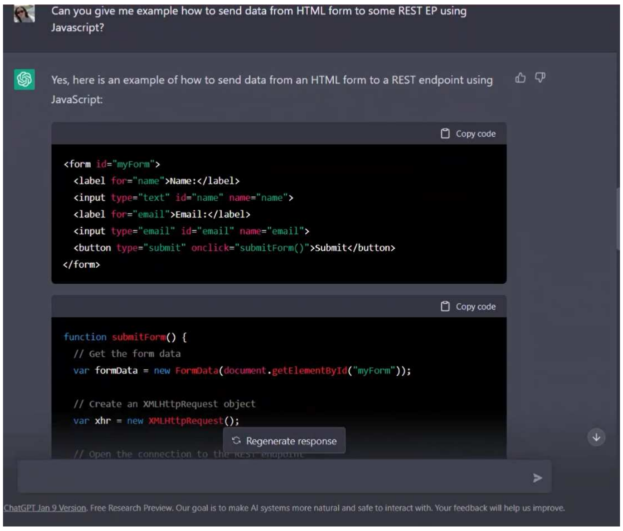
Figure 4. ChatGPT generating web page

Besides the email, where you can see clearly what has to be updated, you got the note as well, with detailed explanation about what you have to change *"In this email, you should make sure to include the specific date by which you need the customer to update their credentials and also provide the customer with a contact information of your customer support team. Also, you should remind them to keep their account credentials safe and use a strong password. Additionally, you could mention any steps taken to prevent this type of crashes in the future.*

Also, make sure the link you provide is secure and only accessible by the intended recipients."

Successful phishing attack gives you username and password of some Facebook user in this particular attack. When you come into possession of Facebook credentials of one person (if two-factor authentication is not used), you can access his/her account, and do whatever you want. You can change passwords, write some inappropriate Facebook status, send messages to the friends list, or you can use the friends list to gain their email addresses and repeat the attack.

You can even improve your phishing attack and make it more personal, if you take the identity of the person you stole credentials from and send a message in his/her name to some friend to donate you some money or something.



Can you give me example how to send data from HTML form to some REST EP using Javascript?

Yes, here is an example of how to send data from an HTML form to a REST endpoint using JavaScript:

```
<form id="myForm">
  <label for="name">Name:</label>
  <input type="text" id="name" name="name">
  <label for="email">Email:</label>
  <input type="email" id="email" name="email">
  <button type="submit" onclick="submitForm()">Submit</button>
</form>
```

```
function submitForm() {
  // Get the form data
  var formData = new FormData(document.getElementById("myForm"));

  // Create an XMLHttpRequest object
  var xhr = new XMLHttpRequest();

  // Open the connection to the
```

ChatGPT Jan 9 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

Figure 5. ChatGPT generating JavaScript code

This is just one example of how ChatGPT can be used. One can vary the questions to the bot to create a login page that looks like some other page, to get code for managing form data in some different language, to write phishing email content in the different language etc. There is a great set of possibilities.

V. SOCIAL ENGINEERING ATTACK PREVENTION

There are a number of best practices that can help keep your email safe from social engineering attacks such as phishing [9] [10]:

1. *Be cautious of unsolicited emails:* Be wary of any emails that come from unknown senders, or that ask for personal information or login credentials.
2. *Verify the sender's identity:* Before clicking on any links or downloading any attachments from an email, make sure to verify the sender's identity. This can be done by checking the email address or by contacting the sender directly.
3. *Look out for suspicious links:* If an email contains a link, hover your mouse over it to see where it leads before clicking on it. Be wary of links that lead to unfamiliar or suspicious-looking websites.
4. *Be careful with attachments:* Be cautious when opening attachments, especially if they come from unknown senders or if they have suspicious file names.
5. *Keep your anti-virus and anti-malware software updated:* Make sure to keep your anti-virus and anti-malware software up to date so that it can detect and

Authorized licensed use limited to: Swinburne University of Technology. Downloaded on August 05,2023 at 12:58:47 UTC from IEEE Xplore. Restrictions apply.

remove any malicious software that may be present in an email.

6. *Use two-factor authentication:* Two-factor authentication adds an extra layer of security to your email account by requiring a second form of authentication, such as a fingerprint scan or a code sent to your phone, in addition to a password.
7. *Be aware of Phishing Scam and Spear Phishing Scam:* Phishing is a scam that uses emails, text messages, or phone calls to trick you into providing personal information, such as passwords, credit card numbers, and social security numbers. Spear phishing is a type of phishing attack that targets specific individuals or organizations.
8. *Be aware of the message:* Be aware of the message that the email carries and be sure to verify it.

By following these best practices, you can help protect your email from social engineering attacks such as phishing. Remember to always be vigilant and to trust your instinct if something seems suspicious. Besides best practices, special tools that provide protection from phishing attacks can be used as well, such as: IRONSCALES, Avanar, Tristifi and Microsoft Defender for Office 365 [11]. Most of the tools are commercial, and not affordable for regular users. Some of them have a free trial, so it can be used during the trial period to make the user more aware of the phishing attacks and how they look like.

VI. CONCLUSION AND FURTHER WORK

Social engineering attacks become easier to apply than ever. Even without technical skills people can create some attacks such as phishing and come into possession of someone's credentials, credit card numbers, PINs or other sensitive information and use it to make harm from financial or personal point of view just in a few queries by using ChatGPT. ChatGPT will provide answers to all questions that users ask, with excellent results with generating code and page layouts and template messages. It is possible to get very quality replicates of the any popular web site, code for processing malicious requests and messages and email text that is realistic and similar to official notation. That way it is very easy to quickly prepare everything that is needed for a successful social engineering attack. ChatGPT will provide warnings about using generated resources in malicious purposes but that won't stop potential attacker to use it.

With presence of AI solutions, and mainly ChatGPT that is easily accessible to many people, it is possible that number of social engineering attacks will increase. With that fact, it is important to learn how to defend yourself from these attacks, to be careful when answering all requests and hope that users won't use ChatGPT or other intelligent systems like that for bad purposes. To the best of our knowledge, similar research on this topic is not published yet related to this topic, so we will continue investigating in order to compare, expand and improve our work.In the future it will be interesting to check whether ChatGPT will be more aware of misuses of its answers and add some extra layer of security. Also, ChatGPT could provide some solutions for prevention of social engineering attacks. Phishing attacks can be prevented by following best practices for protection or using appropriate tools. Since most of the tools are commercial, our future work will be focused on development of one free tool that will prevent phishing attacks.

REFERENCES

[1] "How ChatGPT will change cybersecurity", available online at: https://www.kaspersky.com/blog/chatgpt-cybersecurity/46959/

[2] R. D. Caballar, "Cybercrime Meets ChatGPT: Look Out, World", published 27th January 2023, available online at: https://spectrum.ieee.org/chatgpt-and-cybercrime

[3] A. Hughes, "ChatGPT: Everything you need to know about OpenAI's GPT-3 tool", published 16th January 2023., available online at: https://www.sciencefocus.com/future-technology/gpt-3/

[4] B. Gordijn., H. Have, "ChatGPT: evolution or revolution?", Med Health Care and Philos (2023), available online at: https://doi.org/10.1007/s11019-023-10136-0

[5] F. Salahdine, N. Kaabouch, "Social Engineering Attacks: A Survey", Future Internet 11, no. 4: 89, available on https://www.mdpi.com/1999-5903/11/4/89.

[6] K. Chetioui, B. Bah, A. Ouali Alami, A. Bahnasse, "Overview of Social Engineering Attacks on Social Networks", Procedia Computer Science,Volume 198,2022, Pages 656-661, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.12.302.

[7] L. Irwin, "The 5 Biggest Phishing Scams of All Time", published 22nd October 2022, available online at: https://www.itgovernance.eu/blog/en/the-5-biggest-phishing-scams-of-all-time

[8] R., Zulfikar ,"Phishing attacks and countermeasures", In Stamp, Mark & Stavroulakis, Peter.Handbook of Information and Communication Security. Springer 2010, ISBN 9783642041174

[9] A. Kumar Jain, B.B. Gupta, "A survey of phishing attack techniques, defence mechanisms and open research challenges", Journal of Enterprise Information Systems, vol. 16, pages 527-565, 2022, available online at: https://doi.org/10.1080/17517575.2021.1896786.

[10] B. Gupta, N. Arachchilage, K. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions", Telecommun Syst 67, 247–267. https://doi.org/10.1007/s11235-017-0334-z, 2018.

[11] C. Jones, "The Top 10 Phishing Protection Solutions", published in January 2023, available online at: https://expertinsights.com/insights/top-10-phishing-protection-solutions/.

Authorized licensed use limited to: Swinburne University of Technology. Downloaded on August 05,2023 at 12:58:47 UTC from IEEE Xplore. Restrictions apply.