RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

PPIPAE: Protecting Personal Information from Phishing Attacks in E-commerce: *Review paper*

Al-Marhabi Zaid[1] Marhabi2000@gmail.com
Adnan Dahim[2] adnan716381245@gmail.com
Mohammed Hakami[2] mohammed770542264@gmail.com
Mohammed Alaghbari[2] alaghbarim767@gmail.com

1) Department of Cybersecurity and networking, Al-Razi University, Yemen
2) Department of Information Technology, Al-Razi University

I. Abstract

After the widespread spread of the Internet, which now covers all parts of the world, and with the emergence of electronic commerce, most people prefer to buy and sell their products on the Internet. All commercial and banking transactions have shifted from the traditional method to the digital method, making electronic phishing crimes a center of attraction for attackers and violating privacy in the digital space. Phishing is one of the methods used by attackers to obtain personal information by deceiving users by using fake websites similar to legitimate websites or through fake URLs and transferring the user to a scammers' website for various purposes such as obtaining sensitive personal data or bank accounts and passwords. There are many types of phishing but in this scientific paper, we only aim to review the most important techniques used to detect phishing attacks, whether by forging web pages and URLs or through deceptive emails Which is the main points of the discussion of this paper. We also discuss the most important deception methods used by fraudsters on people.

Keywords: *E-Commerce, URLs, Phishing attacks, Fake web pages, Spam messages.*

II. الملخص

بعد الانتشار الواسع لشبكة الإنترنت التي أصبحت تغطي جميع أنحاء العالم، ومع ظهور التجارة الإلكترونية، أصبح معظم الناس يفضلون شراء وبيع منتجاتهم عبر الإنترنت. حيث تحولت جميع المعاملات التجارية والمصرفية من الطريقة التقليدية إلى الطريقة الرقمية، مما جعل جرائم

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

التصيد الإلكتروني مركز جذب للمهاجمين وانتهاك الخصوصية في الفضاء الرقمي. التصيد الاحتيالي هو أحد الأساليب التي يستخدمها المهاجمون للحصول على معلومات شخصية عن طريق خداع المستخدمين باستخدام مواقع ويب مزيفة تشبه المواقع الشرعية أو من خلال عناوين URL مزيفة ونقل المستخدم إلى موقع ويب محتال لأغراض مختلفة مثل الحصول على بيانات شخصية حساسة أو حسابات بنكية و كلمات المرور. هناك أنواع عديدة من التصيد لكننا في هذه الورقة العلمية، نهدف إلى استعراض أهم التقنيات المستخدمة للكشف عن هجمات التصيد، سواء عن طريق تزوير صفحات الويب وعناوين URL أو من خلال رسائل البريد الإلكتروني الخادعة، وهو محور نقاش هذه الورقة كما اننا سنتطرق أيضًا إلى أهم أساليب الخداع التي يستخدمها المحتالون على الأشخاص.

الكلمات المفتاحية: التجارة الالكترونية, URLs, التصيد الاحتيالي, الصفحات المزورة, رسائل البريد المزعجة

III. Introduction

Societal engineering used by Phishers in their attacks to hide/masquerade themselves as official servers. [1] Phishing (brand spoofing or carding all have same meaning for Phishing) Phishing usually uses e-mail messages that always came from official businesses that one might have dealings with, banks such as Citibank; online companies such as PayPal and eBay; Internet service providers such as EarthLink, Yahoo, MSN, and AOL; online retailers also like Best Buy; and insurance company [2]. The messages send to victims may look quite authentic, and also attached with corporate logos and formats similar to the ones used for official messages.

When we wonder how phishers work, most of verification data processes such as email, account numbers or passwords, the phishers exploit personal information, which allow them to use it later in stealing money from the victims' bank accounts or conducting buying and selling on their behalf, or even in Extortion operations victims. [3] because these emails appear to be official and belong to official companies as well. Forrester Research conducted survey in 2005, and the result of that research was that 20% of unexpected recipients of these messages may respond to phishers, and this may make them vulnerable to these fraudulent attacks [4].

Victims of phishing are on the rise up to the present time, especially new or inexperienced computer users As a result the relationship between companies/organizations and their customers facing many societal

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

issues, including a decline in trust or identity theft, reliability in dealing and support, etc.... [5].

In a report issued by RSA, global associations incurred losses amounting to $9 billion due to phishing attacks in 2016. The number of phishing sites in 2019, according to what was published by APWG, reached 266,387 sites [6]. Based on how phishing attacks are carried out, these attacks are classified into three types: social engineering using fake websites and email spoofing, attacks using malicious software such as Malware Phishing (Trojan) or Keylogger, and through the network (free networks), especially networks spread in places Public ones, such as (cafes, hospitals, and markets), which anyone can access for free [7]. Browser creators rely on two methods against phishing, which is to program the blacklist and the whitelist, provided that the whitelist contains legitimate URLs, and the blacklist contains illegitimate URLs. However, this method does not do enough work and is considered a traditional method because it requires a group of people to report the addresses. Here's a list so you can get to know them the ban [8]. Despite blacklisting and intelligent learning to detect phishing in websites and emails, some users still end up being victims of phishing. Phishing is always cheap and easy to deploy. Most vendors use different methods to reduce phishing, however, these solutions cannot keep up. Continuous updates of phishing sites [1-7]. In order for e-commerce to increase in any country, information security must be ensured and security measures must be increased to protect information of vital importance to companies and users, as users still suffer from phishing in attempts to forge web pages as a result of the tricks used by fraudsters or a lack of awareness among users [10]. Fraudsters resort to phishing, especially in e-commerce, because it is an easy and not complicated or expensive process for fraudsters, as it is possible to deceive users and steal sensitive data such as credit cards, login data, and passwords. The most important thing that scammers do is design a fake, physical website similar to the real one, such as Amazon. For example, the attackers then register this website with a specific domain

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

address and send it via emails to a group of users, instructing the recipients to click on the fake link and fill in their data [9]. URL addresses are used to access web pages and Internet resources. These addresses are used to determine a specific site and transfer the user to it. They consist of a set of parts: (protocol, host, port, path). Forged URLs represent a widespread threat to many online companies, even though the World Wide Web has become one of the most important networks in the world. The most famous attacks that use malicious URLs are via spam messages [10]. Phishing in e-commerce is the most common attack in the world, and the issue of protecting personal information from phishing and denial-of-service attacks is an important and necessary issue, as this type of attack aims to obtain sensitive personal information, which causes serious financial damage to users [6]. One common way scammers use to defraud users is by simply changing the name of a particular site such as FACEBOOK and FECEBOOK. In a 2015 McAfee survey, nearly 97% of users were unable to correctly identify phishing emails [8]**.** Phishing is a type of criminal activity that is practiced on users over the Internet in e-commerce, where the fraudster begins broadcasting e-mail messages over the Internet to users around them to enter these sites where the users believe that the mail has arrived from the trusted store as shown. In Figure 1, when they open this email message, they are redirected to another fraudulent website. Users are asked to enter their information so that the fraudster can defraud that data and use it for various purposes [11].

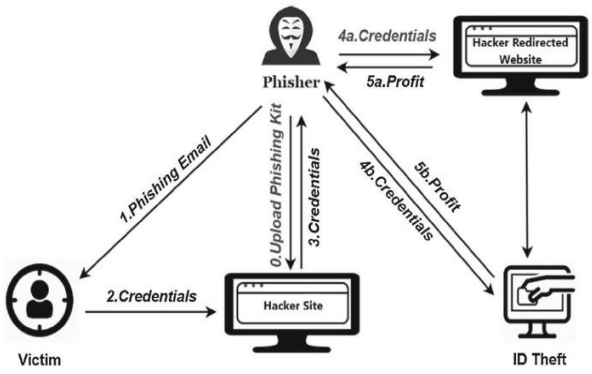

Figure 1 illustrates the phishing process via fake emails and URLs.

Phishing emails come very similar to emails from a trusted commercial website, which deceives the user into believing that the sender has the right to obtain the data required

RUJCST
ISSN-L 3007-5084
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

of him, such as data confirming the purchase and payment process [12].

IV. Objectives

- **Providing** a comprehensive overview of the most prominent recent research that limits phishing attacks and the techniques that have been used.
- Make a **comparison** of each technology used in each research and identify the weaknesses and strengths of each technology.
- Proposing **future** solutions and technologies to reduce phishing attacks.

V. Previous studies

In this part of the scientific paper, we will make a comparison about the most important techniques used to limit phishing attacks that attempt to obtain personal data. We will also focus more on phishing attacks via fake emails, URLs, and deceptive web pages.

In the misleading process in Google website as example, https://www.g00gle.com, instead of the original URL, that is, by putting (double zero) instead of (oo), The misleading people may also shorten the URLs using some websites, such as (Tiny Uniform Resource Locators (TinyURL)), as as in [13].

Authors in [14] highlights phishing attacks, as recently, phishing attacks have become one of the most prominent attacks faced by governments, organizations that provide services and Internet users that provide services, alike.

Detecting and protecting organizations from cyberattacks faces challenges, in most cases due to the unknown form of the attack or there is no real-time warning system. The difficulty of revealing the identity of these attacks or the extent of their impact (risk ratio) may greatly affect the infrastructure in organizations and national critical due to the absence of real-time detection or warning systems for attacks [15].

- **An intelligent identification and classification system for malicious uniform resource locators (URLs).**

A system for detecting and classifying URLs using machine learning [10], The researchers proposed a system to detect and

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

distinguish between benign and malicious URLs using a binary classifier. This system uses various combined learning methods to develop a high-performance detection and classification of malicious URLs. This system was divided into four main processing stages: preparation, control, evaluation, and publishing. It also classified URLs based on Its features are divided into five categories: benign, spam, phishing, malware, and distortion. This system also adopted the use of the collective approach method, the nearest neighbor approach. This model relies on machine learning by training on a group of real and fake addresses through which the model is fed with a set of addresses. One of the features of this model is that it relies on group learning, meaning that it detects and recognizes URL addresses and classifies them into five categories, as I mentioned previously. The model is very effective, especially if it is integrated by browser makers or programmers. The accuracy of the results showed in this the model indicates that it is 99% effective for detecting malicious URLs, 95% for detecting spam, 98% for detecting fraudulent web pages, and 97% for defacement.

- **Phishing Web Page Detection Methods: URL and HTML Features Detection.**
- **Compare URL AND HTML features with a set of conditions to be checked:**

The researchers indicated in [7] to a set of conditions and rules that must be met in any browser, it also relied on a set of advice for users to make the site more effective in terms of accuracy and the ability to detect pages faster, because its focus is on educating users, and to a very large extent, one of the most prominent basic and sensitive conditions that the researchers focused on. It is necessary to ensure that web pages and URLs are available and positively verified to allow the user to access the site.

- **Phishing Detection from URLs Using Deep Learning Approach**

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

In [16] Researchers have proposed convolutional neural network (CNN) technology, which represents a phishing detection system using deep learning techniques. The technology verifies addresses when a user logs in to a specific website. The IP address of the web page is matched with the IP address registered in the white list, in case there is a conflict. In the data recorded on the site, the system will warn the user. The technology works to detect phishing based on convolutional neural networks (CNN) and deep learning as a running model for education and training. What distinguishes this model is that it does not require any feature engineering, as CNN extracts features from URLs. Automatically through its hidden layers, the results showed that the model can identify URLs with an accuracy of up to 98%.

- **Email Anti-Phishing Detection Application**

Referring to the research [12] The researchers built a Tree algorithm approach to detect the source code of a phishing site linked to an email in order to better protect users - information from fake sites. Its primary purpose is to identify a phishing attack linked to the victim's email. Anti-phishing is used to detect, identify and block a phishing site. Or the email affected by a phishing site. This tool aims to prevent the phishing site from influencing the user's email when opening his emails, as it is characterized by combating phishing using the user's email and password. The second feature is to detect phishing sites attached to the user's email using the tree algorithm. The third feature detects the phishing site and generates a report for the user. The anti-phishing detection application can generate a report on the phishing site which is attached to the victim's email.

- **Multi-factor authentication (MFT).**

Adopting two-factor authentication, especially in financial and commercial transactions, as most commercial sites rely on authentication only on basic data such as user name and password, and this leads to a significant weakness in data security and ease of exploitation by fraudsters. It is necessary to use two-factor authentication (MFA), especially in commercial transactions and

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

transfers. Finance also strongly recommends this method, as when the customer enters the site and enters the name and password, MFA sends a response message for approval, such as text messages or other things, and thus the security of the information is improved [17].

Paper	Author	Year	Technology	performance	strength point
Page No 1. [10]	Q. Abu Al-Haija And others	2022	An intelligent identification and classification system for malicious uniform resource locators (URLs).	Performance: The overall performance rate in detecting all attacks reaches 97%. Weakness: The failure rate in this system reaches 3% of the total. In addition, it relies on only two layers in the scanning and learning process	1- It depends on self-learning. 2- Effective in detecting all types of phishing attacks with high accuracy.

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

Page No 2. [7]	Humam, Faris and others	2020	Phishing Web Page Detection Methods: URL and HTML Features Detection	Performance: Relatively effective. Weakness: based on a set of rules Specific that does not rely on machine learning It is also limited to several conditions.	Strengths: Relying heavily on user awareness. It also contains a set of the most important conditions that must be met by any site.
Page No 3. [16]	Shweta Singh And others	2020	Phishing Detection from URLs Using Deep Learning Approach	Performance: The system achieved 98% accuracy. Weaknesses: If a user visits a legitimate website for the first time, it will be considered suspicious by this system.	Strengths: It can detect anti-phishing by isolating the detected phishing site on the victim's email friends.

RUJCST
جامعة الرازي
Al-Razi University
ISSN-L 3007-5084
كلية الحاسوب وتقنية المعلومات
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

Page No. 4 [12]	R. A. Abbas Helmi And others.	2019	Email Anti-Phishing Detection Application	Performance: It showed high performance and efficiency in detecting forged mail messages. Weaknesses: It has the ability to detect emails only.	It has the ability to calculate the percentages of phishing emails stored in the user's email. It has the ability to link to the user's original emails and come up with an analysis of the phishing emails found in the user's email.

- **The most important suggestions and recommended solutions:**

First, we suggest integrating the models that were previously proposed by adding them to browsers, which leads to limiting and detecting different types by recognizing them directly by the browser.

- Educating users because no matter how many different models are built to detect phishing, it cannot be completely prevented.
- Use multi-factor authentication between the service provider and the user.
- Users should never believe or open spam messages.
- During online shopping, users must ensure that the web pages are legitimate by ensuring that the site is protected and contains the HTTPS protocol and not

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

HTTP. They should also ensure that the link contains three words and does not contain any other symbols such as (-, 0+).

VI. Conclusions and future solutions.

We suggest adding a unique primary field (personal or business ID field) as a more secure form of authentication, and adding it on every business site. So that each site gives each user this ID when creating an account on the site, and the user uses it every time he logs in to the site, where the first field in the login is to enter the private ID, and the second field is the username only, and the password field is not added in the login because it is added. In the process, the site may belong to the fraudster, so he takes it and goes to the real site as if he were the real user. Figure 2 shows how the normal method is done, while Figure 3 shows what the proposed process would be like. When the user enters his information, the legitimate site verifies the user's ID from the site's white list, and if it exists, it converts it. When purchasing and paying, the customer uses his password and payment card, as Table 1 shows the form of the user's ID during registration on the site. If the ID is not found in the list, the browser displays a warning message to the customer that the site is suspicious by applying some algorithms between the site and the browser. Table 1 shows an example of what the user's ID looks like during registration on the site.

Location	IDs
Amazon	USER5V7_KZA81
Ali Baba	USER0Kq_X33LY

RUJCST
ISSN-L 3007-5084
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

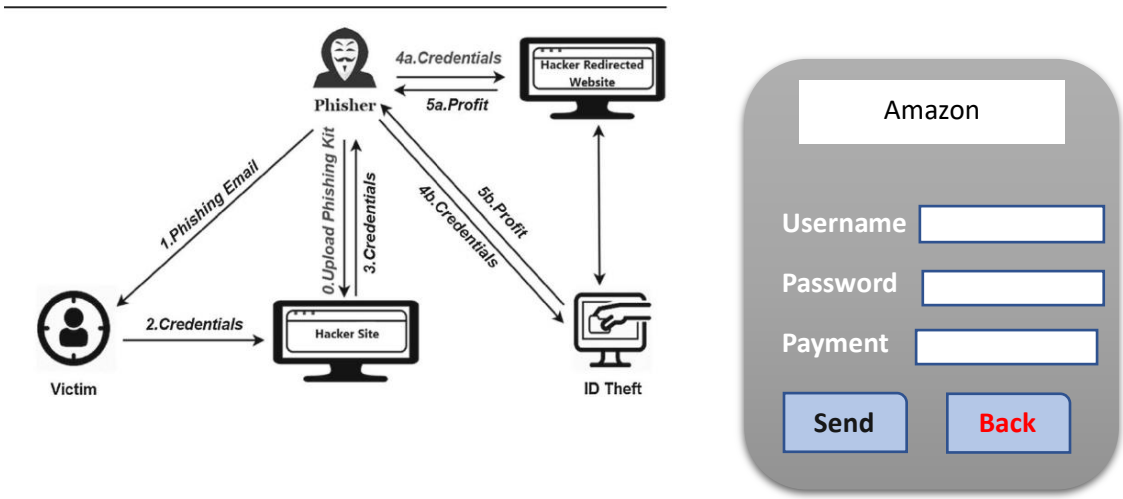

Figure 2 shows how the normal method is done.

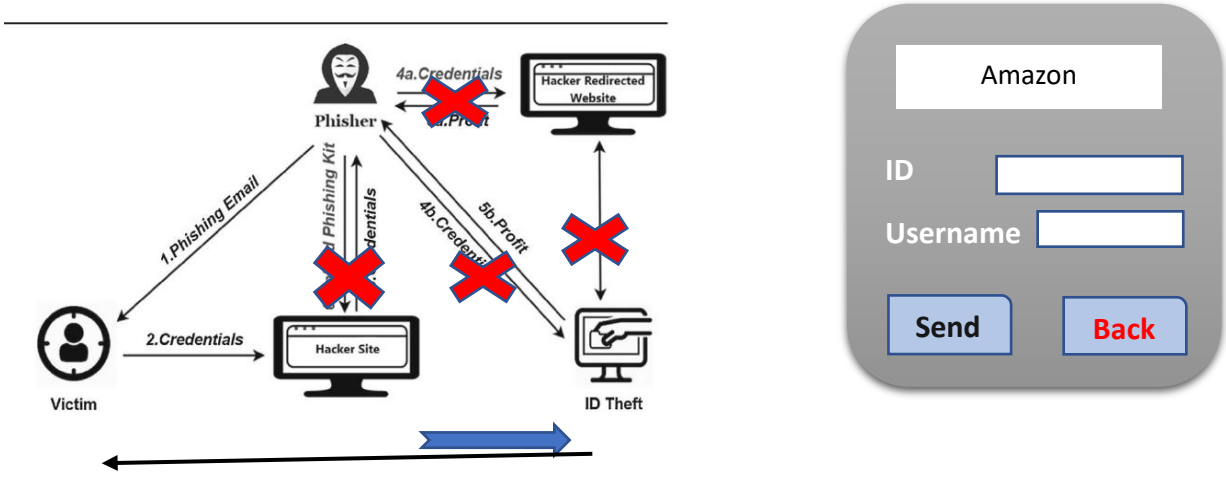

Figure 3 shows what the proposed process would look like

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

References

[1] N. Chou, Ledesma, R, Teraguchi , Y and Mitchell, J. C, "Client-Side Defence against Web-based Identity Theft," *11th annual netowrk and distributed system security symposium, San Diego, California, USA,* 2004.

[2] D. Lazarus, "Phishing expedition at heart of AT&T hacking," *San Francisco Chronicle,* 9.2016.

[3] R. Dhamija, Tygar J D and Hearst M, "Why Phishing works," *in CHI Conference on Human factors in Computing Systems, Montreal, Canada,* April 2006.

[4] G. Lawton, "E-mail authentication is here, but has it arrived yet," *Computer,* vol. 38, no. 11, pp. 17-9, 2005.

[5] J. Ragucci and S.A, "Societal Aspects of Phishing," *IEEE International Symposium on Technology and Society,* vol. ISTAS, pp. 1-5, 2006.

[6] A. Basit, M. Zafar, X. Liu, A. . R. Javed, Z. Jalil and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *pumed,* p. 16, 20 1 2021.

[7] H. , Faris and S. Yazid, "Phishing Web Page Detection Methods: URL and HTML Features Detection," *IEEE,* p. 10, 23 4 2020.

[8] Y. Wang and I. Duncan, "A Novel Method to Prevent Phishing by using OCR Technology," *IEEE,* p. 5, 1 12 2019.

[9] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection," *Pubmed,* p. 16, 1 1 2021.

[10] Q. Abu Al-Haija and M. Al-Fayoumi, "An intelligent identification and classification system for malicious uniform resource locators (URLs)," *Pubmed,* p. 17, 23 12 2022.

[11] V. K. Sharma, P. Mathur and D. K. Srivastava, "Secure Electronic Fund Transfer Model based on Two level Authentication," *IEEE,* p. 5, 18 1 218.

[12] R. A. Abbas Helmi, C. . S. Ren and A. Jamal, "Email Anti-Phishing Detection Application," *IEEE,* p. 8, 7 9 2019.

[13] S. A. e. al, "A survey of intelligent detection designs of html url phishing attacks," *IEEE Access,* 2023.

[14] B. A, Zafar M, Liu X, Javed AR, Jalil Z and Kifayat K, " A comprehensive survey of AI-enabled phishing attacks detection

RUJCST
جامعة الرازي
Al-Razi University
كلية الحاسوب وتقنية المعلومات
ISSN-L 3007-5084
مجلة جامعة الرازي لعلوم الحاسوب وتقنية المعلومات
Al-Razi University Journal of Computer Science and Technology
علمية محكمة تصدر عن كلية الحاسوب وتقنية المعلومات - جامعة الرازي

techniques," *Telecommun Syst,* no. 76, p. 139–54, 2021.

[15] G. K, Strategic cyber security—Kenneth Geers—Google books, Accessed 21 Feb 2022.

[16] S. Singh, M. Singh and R. Pandey, "Phishing Detection from URLs Using Deep Learning Approach," *IEEE,* p. 8, 22 7 2020.

[17] S. M. Toapanta Toapanta, H. A. Mera Caicedo, B. A. Naranjo Sanchez and L. E. Mafla Gallegos, "Analysis of Security Mechanisms to Mitigate Hacker Attacks to Improve e-Commerce Management in Ecuador," *IEEE,* p. 18, 11 7 2020.

[18] S. Roy and P. Venkateswaran, "Online Payment System using Steganography and," *IEEE,* p. 10, 17 1 2014.