International Journal of Information Security (2024) 23:819–848
https://doi.org/10.1007/s10207-023-00768-x

REGULAR CONTRIBUTION

A comprehensive survey of phishing: mediums, intended targets, attack and defence techniques and a novel taxonomy

Richa Goenka[1] · **Meenu Chawla**[1] · **Namita Tiwari**[1]

Published online: 19 October 2023

Abstract
The recent surge in phishing incidents in the post-COVID era poses a serious threat towards the social and economic well-being of users. The escalation in dependency upon the internet for meeting daily chores has made them vulnerable to falling prey to the ever-evolving menace of phishing. The objective of this article is: to explore different tactics and motivational factors behind phishing, identify the communication mediums through which phishing is circulated and perform a detailed review along with a comparison of the various surveys in this domain. Another objective is to determine the open research challenges in this genre and to identify the scope of research in the future. An extensive literature survey is performed, which includes articles from eminent online research databases. Barring a few initial articles related to phishing, the articles published in Science Citation/Scopus-indexed journals and survey/review articles published in the last ten years are considered. Highly cited works are given preference. The search query returned numerous articles, which were narrowed by title screening. Further screening of articles was performed by reading the abstract and eliminating the articles related to user-oriented phishing interventions. Eventually, 25 survey articles were shortlisted to be surveyed. This article is an effort to provide a novel taxonomy of phishing to academia that would assist in identifying the sections where phishing countermeasures are inadequate.

Keywords Phishing · Phishing techniques · Phishing circulation mediums · Intended targets · Phishing countermeasures

1 Background

"Dear valued customer, an unusual activity has been noticed in your bank account. To continue operating your account, please verify your details by clicking on the link below." If someone receives any such e-mail from their bank and clicks the given link without verifying its authenticity, then there is a high probability that they may fall prey to phishing. Phishing is a social engineering attack wherein the fraudsters manipulate the victim and exploit human error [1] with the sole purpose of getting access to restricted data, personal information, login credentials, or spreading malware. In most cases, it is performed by forwarding a URL or link through spoofed e-mails [2]. After clicking on the link, the victim is redirected to a fake website, which the fraudsters have created by replicating the authentic one. The details keyed in by the victim on the fake web page are captured by the attackers, following which the victim is redirected to the authentic website without raising any suspicion. The damage is irreversible by the time the victim realises that he or she has been phished.

The COVID-19 pandemic has significantly compelled individuals to be reliant on online services [3]. As per a report, after the pandemic, the total internet hits surged by between 50 and 70% [4]. With the spread of the pandemic and restrictions on human interactions being imposed in almost every region, organisations were left with no other choice but to take refuge in technology to continue with their operations. Moreover, this transformation occurred in almost every sector of society. From entertainment, education, e-travelling and shopping to working from home, meetings, and e-banking, there is not a single sphere of life that was untouched by the use of information systems or networks. Many organisations started believing this situation as the new normal and considered a permanent tran-

✉ Richa Goenka
richa132@gmail.com

[1] CSE Department, Maulana Azad National Institute of Technology, Bhopal, M.P., India

sition to a hybrid model of operation (working-from-home and working-from-office) [5, 6]. Members of the workforce found it difficult to come to terms with this shift in their work patterns.

Nonetheless, the sudden occurrence of this metamorphosis exposed the vulnerabilities of the system to cyber attackers. An unprecedented situation and struggle for survival forced organisations to involuntarily move towards online mode without enough planning, infrastructure, and training. Many businesses did not even have a cyber-security policy in place to guide users about the plan of action in case of a cyber-attack. This subservience was exploited by fraudsters. Stress, fear, and anxiety amongst users during the COVID pandemic contributed to their falling for various phishing attacks [8–12].

1.1 Phishing statistics

The fraudsters leveraged the widespread panic caused by the pandemic and endorsed different technical and psychological techniques to persuade the victims to click on the phishing link. According to the Anti-Phishing Working Group (APWG) report [13], beginning in March 2020, cybercriminals launched a variety of COVID-themed phishing and malware attacks against workers, healthcare facilities, and recently unemployed. In many countries, after COVID-19, government-aided financial assistance programs were started. The fraudsters made use of phishing to steal sensitive information from the beneficiaries and deceitfully applied for government benefits. As per another study, in the first quarter of 2020, COVID-based phishing e-mail attacks were up 600% [14]. The annual report of the Internet Crime Complaint Centre (IC3) of the Federal Bureau of Investigation (FBI), 2020, states that there has been an upsurge of 69% in the total number of cyber-crime cases in the USA, with losses exceeding $4.1 billion. Out of this, phishing scams accounted for over $54 million [15]. There has been an unprecedented increase in the number of detected phishing websites over the last 10 years, as shown in Fig. 1. October 2022 saw the highest number of monthly phishing attacks reported in AWPG history which almost doubled since early 2022. Figure 2 shows the monthly growth in phishing websites in the year 2022. Apart from some decline in the initial months of the first quarter of 2022, there has been a significant rise in the number of unique phishing websites. Figure 3 shows the industry domains most targeted by phishing attacks in the fourth quarter of 2022 [7]. Financial Institutions along with WebMail-based organisations and Social Media were the most targeted sectors by fraudsters. About 55% of the total phishing attacks in 2022 were observed in these organisations.

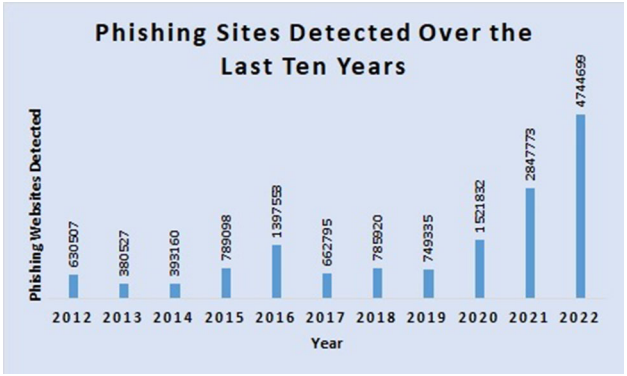

Fig. 1 Unique phishing websites detected in last 10 years [7]

Fig. 2 Unique phishing websites detected since January 2022 [7]

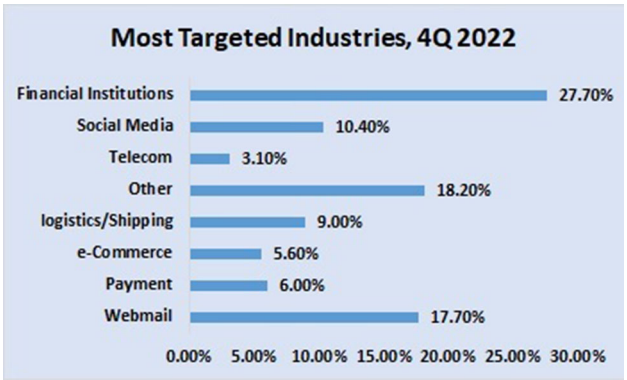

Fig. 3 Most targeted industries in Quarter 4,2022 [7]

1.2 Motivation

Scores of defence techniques against phishing have been proposed by researchers over the years. Still, phishing attacks are on the rise. The perpetrators involved are technically sound and outmanoeuvre the defence approaches being applied and devise new methods to deceive the users. The main reason for the disparity between anti-phishing and phishing attacks is the lack of sufficient knowledge about phishing strategies applied by criminals. Ever since its outset, there has been a quest amongst researchers to effectively summarise the vari-

ous dimensions of phishing attacks through surveys. There is an abundance of research work in this domain, but a considerable portion of that mainly focuses on anti-phishing. Not many authors have focused on the overall approach utilised by phishers to commit fraud. Very little importance is given to the mediums of phishing distribution and the category into which the said phishing attack falls. If there is precise knowledge about the classification of different types of phishing attacks, competent phishing detection techniques can be designed in a customised manner to combat them.

Through this survey, a novel taxonomy of phishing is presented, where phishing has been classified based on the mediums of circulation, intended targets, and the techniques used to perform phishing. The suggested taxonomy covers all the aspects of phishing attacks without being complex. We did not come across any existing literature, where an attempt is made to explain the different subcategories of phishing circulation mediums as thoroughly as in this work. To present the gravity of the situation, notorious case studies involving phishing are also discussed.

The statistics related to phishing as reported by prominent groups in this domain such as APWG, Kaspersky, Cisco, Verizon, etc. are mentioned throughout the course of this survey. Some previous surveys and reviews in this genre are studied and analysed from the perspective of their contribution towards phishing research.

1.3 Contributions

Following are the contributions of this survey, which we are sure will enable the researchers to move towards a better understanding of the rapidly advancing threat of phishing:

1. Illustrating the objectives behind a phishing attack and providing the researchers with an insight into the magnitude of the situation through case studies and statistical analysis.
2. Presenting a novel phishing profile that clearly identifies the intended targets of phishing, categorises the different mediums through which a phishing attack can be circulated, and explores the various phishing attack techniques that are currently employed by the phishers.
3. Critical evaluation of various phishing detection techniques along with their comparative analysis.
4. Identifying the open challenges in phishing countermeasures and suggesting future research directions.

The rest of the paper is organised as follows. Section 2 discusses surveys related to phishing, including a comparative analysis in the form of a table. Section 3 delves into the history of phishing, the phishing process, and the various goals of phishing attacks. Infamous phishing case studies have also been discussed. Section 4 presents a classification of phishing based on different heads, namely circulation mediums, intended targets, and techniques. Section 5 presents an analysis and discussion on various phishing attacks. Section 6 discusses various phishing detection approaches along with their pros and cons. Section 7 summarises the conclusions and future scope.

2 Related surveys

Diverse phishing-related literature is available throughout various libraries. Some of the earliest works in this domain were presented by [16, 17] who were among the first to enlighten the researchers on various aspects of phishing. Authors in [16] have discussed different conventional phishing attack techniques employed by the threat actors, along with a methodology for preventing them.

[17] have presented an extensive discussion on the social engineering factors and phishing attack vectors. However, phishing detection approaches were not discussed. Rather, they focused more on phishing prevention and suggested traditional methods to combat phishing. Even though these works are more than a decade old, they have proven to be of great help in understanding the basics of phishing.

The author in [18] has presented a high-level insight about phishing by describing the entire phishing structure, i.e. from the time the idea of the attack is conceived to the time when the illegally obtained benefits are received by the attacker. Different categories of brands being targeted are also discussed. The author has illustrated some variations in the phishing attacks that fraudsters employ to communicate phishing URLs. Some advanced phishing techniques and countermeasures, along with their merits are discussed as well.

In the survey presented in [19], the authors emphasise that different anti-phishing approaches need to be viewed with respect to the entire process of phishing. The user education-based phishing detection approach is evaluated against the software-based approach, and it is concluded that user education alone cannot guarantee a positive response towards phishing awareness. It needs to be complemented with a software-based solution. Various software-based phishing detection techniques have been reviewed against the metrics of detection accuracy and low false positives.

The survey in [20] focuses mainly on e-mail-based phishing and overlooks other phishing mediums. The authors have termed phishing as a type of spam that utilises two different approaches: *social engineering-based*, which depends on spurious e-mails to obtain victim's data, and *technical subterfuge-based*, which uses malware to exploit security gaps in the victim's system to perform frauds. A survey of feature sets for phishing e-mail detection is also presented, which classifies the features into three groups based upon

their method of extraction, i.e. (1) features extracted directly from e-mail like structural, link, element, spam filter, and word list (2) features based on some phishing keywords appearing together such as-click, URL/link, prize, account, etc. (3) text-based features. Anti-phishing approaches have been discussed and classified as per their relevance in the various stages of a phishing attack.

[21] have presented the life cycle of a phishing attack and focus mainly on web-based phishing attacks. The authors have grouped the phishing strategies adopted by phishers into three main categories based on different stages of the phishing attack: First, the attacker imitates and sends a fake message to the victims, instructing them to validate or update their credentials through a specific URL. These messages are carefully designed with logos and other visual details of the authentic sender. Second, when the victim clicks on the URL, a hoax web page opens that asks the victims for their details following which the victim is redirected to the original website through Man in the Middle (MITM) technique. Third, a variation of the MITM technique, that requires users to enter their details through a pop-up window. The authors have also studied and evaluated various phishing detection approaches.

Techniques of phishing detection and filtering have been discussed in [22], along with their advantages and disadvantages. The authors have performed a relative analysis of these different techniques. Some emerging phishing attack trends and the attacker's motivation behind them are also presented.

[23] categorises anti-phishing solutions as phishing prevention, user education, and phishing detection. Various phishing prevention techniques are mentioned, but to be successful, they are dependent on the user's ability to understand them. Also, to be implemented, they require modifications to the existing system and have proven to be complex and expensive. Hence, the authors have focused on phishing detection techniques. Detection schemes are classified on the basis of their approach along with their pros and cons, novelty factor, dataset, and accuracy. The authors have also made suggestions about the scope of improvement in different phishing detection schemes.

Along with conventional mediums, [24] have explored phishing attacks and phishing detection techniques in new channels such as social networking sites and mobile phones. A phishing taxonomy comprising various dimensions of phishing, such as phishing communication media, the devices being targeted, phishing execution methods, and anti-phishing measures has been presented. To depict the factual significance of phishing detection, a comparison of commercial anti-phishing tools and research anti-phishing tools has been presented. Furthermore, tools are analysed for their performance and ranked accordingly.

Apart from basic information about phishing, such as history, life cycle, types, and countermeasures, the authors in [25] have addressed open issues and challenges being faced in the current scenario. The menace of phishing in the emerging domain of IoT has been communicated. For a better comprehension of the issue, prevailing resolutions, and future outlook, the authors have also discussed different datasets and tools currently being used by academics.

The authors in [26] have presented a systematic review of software-based phishing detection techniques. Along with a taxonomy of phishing detection, evaluation datasets and evaluation metrics have been discussed. To facilitate zero-day attack discovery, a newfound feature called Network Round Trip Time has been studied. A timeline-based record of different phishing detection techniques proposed over the years has been presented. Phishing detection features based on URL, website content, and website visual similarity have been summarised. Research guidance related to dataset selection, feature selection, and the detection scheme to be applied is also provided.

In [27], a comprehensive review of old and current phishing attacks is presented. The medium of phishing, the vector being used to transmit, and the phishing techniques used to perform the attack are reviewed for each phishing attack being discussed. The survey focuses on phishing attack techniques with a detailed description of the technical subterfuge involved in each one of them. The authors have also presented a state-of-the-art forecast about how different phishing attack techniques can be combined in the future to launch attacks of higher sophistication.

In [28], critical scrutiny of different anti-phishing genres, i.e. legal, educational, computerised using human-designed mechanisms, and intelligent machine learning mechanisms is performed. Authors have also stressed the importance of a user education-based anti-phishing approach. Content-based phishing detection methods have been described in detail. A comparison of various machine learning-based phishing detection techniques has been illustrated on the basis of performance, merits, and demerits.

The survey [29] reexamines the available phishing research literature from the point of view of current security challenges namely zero-day attack detection, base-rate neglect, time taken for attack detection, and limited availability of diverse and good-quality (near to reality) datasets. In addition, the authors have categorised phishing detection techniques on the basis of different attack vectors. The features used, detection methods, dataset properties (availability, size and class ratio, diversity, etc.), and evaluation metrics in different phishing detection techniques have been enlisted. The need to include feature importance in research and the scarcity of diverse datasets has been highlighted.

In the study [30], a survey on machine learning (ML) based (Random Forest, SVM, K-star, Adaboost, etc.) and nature-inspired (NI) based (Particle Swarm Optimisation and Firefly Algorithm etc.) phishing detection techniques is presented. The survey focuses on phishing websites as well as e-mails.

Various drawbacks of existing solutions are discussed, such as insufficient dataset, use of third-party, small feature-set, and dependency on the database. The authors have suggested the development of deep learning-based and NI-based phishing detection algorithms to enhance the overall performance of the model.

In [31] along with a survey of past and current phishing attack techniques, an extensive review of conventional and modern phishing detection techniques is done. The authors have discussed prevailing challenges and trends in the domain of phishing.

[32] concentrates on studying the detection of UBEs (Unsolicited Bulk e-mails) that are spam and phishing e-mails through machine learning. Various UBE filtering approaches are broadly labelled as content-based and behaviour-based filters, case-based filters, heuristic filters, previous likeness-based, and adaptive filters. The working of various commercial UBE filters is summarised. A mechanism to process raw e-mail data based on forty distinguishing features is described. The readers are also enlightened about how to determine feature importance. Distinct feature extraction approaches are explained. e-mail classification using many machine learning algorithms is illustrated and evaluated on the basis of performance.

[33] performs an extensive systematic search pertaining to web phishing detection. The anti-phishing solutions are categorised on the basis of input dimensions (URL/website address-based approach, textual content of the web page-based/similarity-based approach, hybrid approach). Website address-based approaches are further categorised as heuristic, list, and learning-based approaches. Textual content-based approaches are further categorised as ML-based and rule-based approaches. Different solutions proposed for each category are mentioned in detail and compared depending on the evaluation metrics, performance, benefits, and drawbacks. The authors conclude by suggesting the use of hybrid techniques with deep learning methodologies for improved performance and efficiency.

The study in [34] is confined to investigating the attack techniques practised by a particular phishing attack group that targets public institutions in South Korea for the purpose of intelligence gathering. Common distinguishing features in phishing e-mails originating from this attack group are identified and analysed. Post the analysis, their purpose is determined, and suggestions regarding phishing countermeasures to be applied by mail service providers are given.

In [35], the characteristics which make some individuals more vulnerable to being exposed to phishing are recognised. The current situation of phishing attacks is identified and prevalent phishing detection techniques are reviewed. A new phishing anatomy is proposed, which includes stages of the attack and their types, mediums of phishing, susceptibilities, threats, targets, and techniques. Some counteractive measures are also suggested.

[36] thoroughly examines various phishing methods and anti-phishing techniques. The evolution of phishing, its life cycle, and attacker motivation are also covered. A detailed taxonomy of phishing attacks on desktops as well as mobile devices is presented. Various open research challenges or research gaps have been identified and discussed.

Characteristics of phishing attacks during the COVID-19 pandemic are studied in [37]. Scientific studies, along with government reports and other literature, that investigated phishing during the pandemic are reviewed, and a comparative analysis is presented. Noticeable phishing attacks that were detected during the initial months of the pandemic are listed along with a description of motive, target, attack vector, and date. Current challenges are highlighted, and future research avenues are identified, which include the need for large benchmark datasets, and the use of deep-learning-based methods to extract the features and perform attack detection. The authors also stress efforts to be made towards quantifying the impact of the attack.

Authors in [38] have provided a literature review on various Artificial Intelligence (AI)-based phishing detection techniques. A comparison chart that includes classification algorithms, feature selection methods, and accuracy of different artificial intelligence techniques, namely, machine learning, deep learning, hybrid learning, and scenario-based approach, proposed through the literature is illustrated. The current practices and challenges in this domain are discussed, and future research directions are proposed with the aim of developing adaptable and sturdy anti-phishing solutions.

A systematic literature review has been presented in [39], which studies the application of natural language processing (NLP) for phishing e-mail detection. The structure of an e-mail for extracting and selecting the features is explained. A brief summary of various machine learning algorithms used for phishing e-mail detection, some feature extraction techniques, tools used for evaluation, evaluation metrics, datasets used and their properties, and optimisation algorithms used across various studies have been summarised. Based on their research, the authors have concluded that more research needs to be performed on using deep learning as a phishing e-mail detection technique and attention needs to be given to phishing detection in languages other than English. Also, the authors found TF-IDF (Term frequency-Inverse Document Frequency) and word embedding to be the most frequently used NLP techniques for detecting phishing e-mails. However, the survey is limited to studying works only in the field of phishing e-mails, and little or no attention is given to other facets of phishing, such as phishing URLs, smishing, vishing, compromised domains, etc.

Another systematic review is done in [40] where phishing is divided into various types: website, web page, e-mail,

Table 1 Summary of existing phishing-related surveys

Survey	Main focus	Distribution mediums	Intended targets	Case studies	Attack techniques	Challenges	Future scope
[16]	Conventional phishing attacks	No	No	Yes	Basic	Yes	No
[17]	Phishing prevention	Yes	No	No	Basic	No	No
[18]	Phishing process, Conventional countermeasures	Yes	No	Yes	No	Yes	No
[19]	Detection, Prevention, Correction	No	No	No	No	Yes	Yes
[20]	Feature sets for phishing e-mail detection	e-mail	No	No	No	No	No
[21]	Web-based Phishing, Anti-phishing methods	No	No	Yes	No	No	No
[22]	Phishing Detection and filtering	No	No	No	No	Yes	No
[23]	Web Phishing detection approaches	No	No	No	No	Yes	Yes
[24]	Phishing attack and classification techniques	Yes	No	In Brief	Yes	Yes	Yes
[25]	Phishing attacks and detection	No	No	No	Yes	Yes	Yes
[26]	Phishing detection	No	No	No	No	Yes	Yes
[27]	Phishing attack techniques	Yes	No	No	Yes	No	Yes
[28]	Conventional and automated anti-phishing measures	No	No	No	No	No	No
[29]	Phishing detection approaches	No	No	No	No	Yes	Yes
[30]	ML and NI-based phishing detection	No	No	No	No	No	Yes
[31]	Phishing attack techniques, countermeasures in brief	Yes	No	Yes	Yes	Yes	No
[32]	Phishing and spam e-mail filtering through ML	e-mail	No	No	No	No	Yes
[33]	Web Phishing detection	No	No	No	No	Yes	Yes
[34]	Profiling the phishing attacks by one attack group	No	No	Yes	Yes	No	No
[35]	Phishing attack techniques, countermeasures in brief	Yes	No	Yes	Yes	Legal challenges	Yes
[36]	Phishing attacks, phishing detection	Yes	No	Yes	Yes	Yes	Yes
[37]	Characteristics of phishing attacks during COVID-19	No	No	Yes	No	Yes	Yes
[38]	AI-based Phishing detection	Yes	Devices	No	No	Yes	Yes
[39]	Phishing e-mail detection through NLP	No	No	No	No	Yes	Yes
[40]	Systematic phishing classification	Yes	No	No	No	Yes	Yes
This Survey	Phishing attack techniques, classification	Yes	Yes	Yes	Yes	Yes	Yes

SMS, tweet, financial data, and URL. The anti-phishing approaches proposed during the last decade are compared based on the type of phishing, classification algorithm used, type of dataset and performance evaluation method. Future research scope and insights have been provided, which include phishing research for non-English languages, expert validation of features, and standard threshold for performance evaluation. Table 1 summarises the above-discussed surveys and compares them with this work.

3 History, stages, modus-operandi, case studies

3.1 History of phishing

In the early nineties, when the internet had just made an appearance to be used by the general public, online security was a matter of concern only for government agencies. Private organisations were least responsible for the cybersecurity of their end users. America-On-Line (AOL) paid the price when its users reported the first incident of phishing in 1995 [16]. A small community of self-identified computer hackers, mostly teenagers, wrote a software program called AOHell [41]. It facilitated an automated method of stealing passwords and credit card details. In those days, AOL did not issue any warnings related to login and credit card scams to its users. Specifically, the 'New Member Lounge' chat rooms were targeted as they had users who were new to the use of the internet. Direct messages were sent to clueless users who were tricked into revealing their login credentials and became victims of the first-ever phishing incident. Even though the motivation behind the attack was to continue with uninterrupted access to the internet, it paved the way for other much more lethal phishing incidents that have occurred over the years.

3.2 Phishing case studies

1. **Phone Phishing attack on Twitter** [42, 43] In July 2020, some Twitter employees were subjected to a phishing attack via phone. The attackers, who professed to be Twitter employees, exploited human vulnerabilities and manipulated them to divulge their login credentials. Through these credentials, the adversaries were able to access Twitter administrator tools, which further equipped them to access the Twitter accounts of many celebrities, send fake tweets, and ask for Bitcoin contributions on their behalf. A huge fan following of the celebrities ensured a transfer of more than $100,000 in bitcoins to bogus accounts.
2. **Phishing attack on Google and Facebook** [44] The perpetrators sent forged e-mails with fraudulent invoices perceived to be originating from Quanta Computers in Taiwan to some employees of the two technology giants. Since Quanta Computers regularly carried out business with Google and Facebook, no suspicion was raised, and more than $100 million were transferred to the fake company's bank accounts between 2013 and 2015.
3. **Fake President Scam on FACC** [45] In 2016, an Austrian aerospace parts manufacturer company, FACC, lost around $61 million as a result of a phishing attack. The phisher masqueraded as the CEO of the company and sent a hoax e-mail to the finance department with instructions to transfer funds to the attacker-controlled bank account.
4. **Phishing attack on COVID vaccine supply chain** [46] In 2020, a phishing attack against global vaccine supply chain manufacturers was uncovered. The intention behind the attack was to access sensitive information related to the COVID vaccine cold chain distribution system. The attack was spread across multiple countries and targeted the employees of companies that were involved in the attempt to keep up the COVID vaccine supply chain, such as biomedical research organisations, medical equipment manufacturers, immunology experts and pharmaceutical firms. Logistics firms involved in the transportation of the vaccine were also targeted.

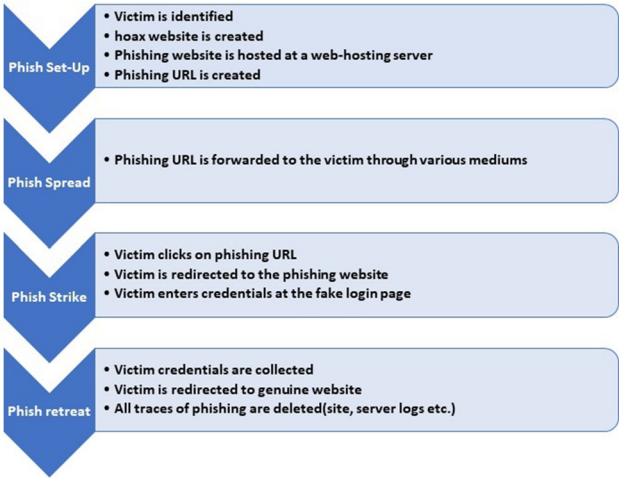

Fig. 4 Process of a phishing attack

3.3 The phishing process

Figure 4 shows the various stages of a phishing attack. A phisher takes the following course of action to steal information from the victims [21]:

1. **Phish Set-up** During this stage, a phisher identifies the victim or a group of victims, marks the information to be extracted, and sets up a phishing website to be used for deception. The phishing website is uploaded to a web-hosting server. The attacker can either acquire a domain and use it for malicious purposes or hack a legitimate domain and append the phishing web pages.
2. **Phish Spread** In this stage, the already identified victims are exposed to the already created phishing website through a phishing link. The different methods to transmit this link have been discussed in the next section.
3. **Phish Strike** Once the victims click on the phishing link, they are redirected to the fake website that has been created by the fraudsters. On the fake website, the

victims provide login credentials or other personal information that might be used by the scammers for malicious intentions. Malware can also be installed on the victim's computer. The fraud website has a similar look and feel as the original, including the same logo. The created link has almost the same orthography as the original brand name but with minute differences. This stage relies heavily on human error and a lack of online security awareness. It manipulates human psychology which tends to ignore certain security aspects when exposed to urgency or anxiety.

4. **Phish Retreat** After having fulfilled the purpose, the victims are redirected to the authentic website, and all the traces of phishing are attempted to be deleted (such as the fake website, server logs, etc.).

3.4 Phishing objectives

There might be multiple objectives behind a phishing attack. The pivotal ones are [47]:

1. **Financial gains** The attackers gain access to the online banking login credentials of the victims through the mimicked website and can perform monetary transactions. The majority of phishing attacks are motivated by a desire for financial gains.
2. **Defamation** Getting access to social media login details enables the phishers to send derogatory messages or upload obscene posts from the victim's profile with the intent of defamation.
3. **Impersonation** The attackers imitate the identity with a motive to execute malicious ventures. This can be done for financial benefits, criminal activities like committing fraud or to malign the reputation of an individual or an organisation. The stolen identity can also be used to perform further phishing attacks.
4. **Identity fraud** Stolen identities are in huge demand on the dark web [48]. Instead of utilising the victim's identity for financial gains directly, the phisher can sell them on the dark web, where they can be further utilised to perform unlawful activities and even acts of terrorism [49, 50]. The main reason that this nexus works is that there are no geographical limitations on the internet. Since the crime is spread across multiple locations, it becomes very difficult to track, and the scammers can continue with their endeavours for a longer duration.
5. **Espionage** Business rivals use phishing to spy on their counterparts to steal trade secrets. Business proprietary information includes details about products, pricing, corporate strategies, industrial research, and financial statements. This information is sensitive and, if revealed to a competing company, can be used by it to get ahead in acquiring a contract or to taint the public impression of an enterprise, leading to losses worth billions.
6. **Malware Installation** Another purpose of phishing is to install malicious software on the target computer. In most cases, an e-mail containing malware as an attachment is sent to the victim. Upon clicking, the malware installs itself on the victim's machine and performs desired tasks such as snooping, encrypting, corrupting the data, or even opening a backdoor to the system for the attackers for a much more hazardous attack later. The attack on the Ukrainian power grid [51] is a perfect example, when prior to the hijack of the SCADA system at the power grid, the employees of the power grid companies were sent phishing e-mails containing BlackEnergy3 malware. It enabled the attackers to gain access to the user credentials some of which were for VPN (Virtual Private Network) that the grid workers used to remotely log into the SCADA system. There are many variants of malware that pose a threat to the victim's computer. Some of them are:

   (a) Ransomware: Malware that denies the victims access to their data by encrypting it until a ransom is paid, mostly in the form of cryptocurrencies. Almost 70% of malware breaches between November 2020 and October 2021 had ransomware involved [52]. As per a survey [53] conducted across 31 countries, 66% of organisations were hit by ransomware in the year 2021. This amounts to an increase of 78% as compared to the previous year. Out of these organisations, 46% paid the ransom to get their data back. However, only 61% of the data could be restored after paying the ransom. Only 4% of the ransom-paying organisations got all their data back. There has also been a 4.8-fold increase in the average ransom payment. There is an increase in different variants of ransomware, which makes it easy for them to evade the anti-malware software [54].

   (b) Spyware: Malware that is delivered mainly through phishing e-mails with the purpose of snooping on the victim's data, tracking the websites being visited, and monitoring the online activity. The main motive to infiltrate spyware on the victim's computer is to gather information. Spyware can capture confidential data like passwords, account PINs, browsing habits, credit card details etc. and send it to the spyware authors, who can sell it or use it to carry out a more fatal cyber attack thereafter.

   (c) Viruses, worms and trojans: All these are malicious software that can cause damage to a computer, but some difference exists among the three. *Virus* is malware that attaches itself to another software program that needs to be executed to install the virus on the victim's computer. A virus when installed can corrupt

the data or hardware of the system. It has the capability to replicate itself and can also spread to another computer. However, a virus cannot spread without human intervention, such as executing or sharing an infected file through e-mails knowingly or unknowingly. *Worm* is malware that like a virus, spreads to other computers, but it does not attach itself to another program and also does not need human help to spread. It has the capability to replicate itself on the system and travel through networks. So, it can transmit thousands of copies of itself which spread further and create a disaster. The main objective of a worm is to eat up the system's resources, causing it to slow down, and allow a malicious user to control the system remotely. *Trojan Horse* is a malicious program that disguises itself as a genuine application. It does not replicate itself but can be equally catastrophic. The primary purpose of the Trojan Horse is to collect information.

(d) Adware: Adware is malware that automatically displays or downloads advertising content, such as pop-ups or banners once the user is online. It enters the system through software that a person downloads from the internet, usually freeware, and discreetly installs itself on the victim's computer. Adware harms the victim's computer by slowing it down and hijacking the browser.

(e) Keyloggers: A malware which covertly records the keystrokes on a keyboard with the purpose of getting unauthorised access to sensitive information related to the victim [55]. A keylogger can expose passwords, banking details, personal correspondence, and any other activity of the victim to the adversary who can use it to fulfil malicious aspirations.

4 Phishing classification

Figure 5 depicts the phishing classification suggested in this article.

4.1 Classification based on phishing circulation mediums

In order to trap the identified victim, the phisher needs to approach the victim through various mediums. The medium provides the fraudster with a mechanism to dispatch the phishing link, and then gather the sensitive information out of the details entered by the victim on the phoney web page. Following phishing circulation mediums have been recognised which are employed by phishers throughout the world:

4.1.1 Mobile phone phishing

An increase in the number of subscribers and the change in user requirements have led to a massive advancement in mobile technology in the last few years. Mobile phones have been replaced by smartphones, which have now become such an integral part of our daily routine that it is hard to imagine stepping out without carrying one. Netizens no longer need to worry about owning a computer or a laptop when they have access to a smartphone, which is compact, inexpensive, has a long battery life, and most importantly, performs similar functions. In the year 2016, there were 2.7 billion smartphone users worldwide. This figure rose to 6.5 billion in 2022 and is expected to reach about 7.7 billion by the year 2027 [56]. Mobile devices have evolved from a luxury to a utility and finally, a necessity. Apart from harbouring personal information like pictures, e-mails, and social media accounts, smartphones are also prevalently used to perform monetary transactions through e-commerce websites and for utility bill payments. Mobile devices are with users almost all of the time, and their increased reliance on them to perform their daily chores has made them an appealing option for phishers to carry out fraud using them as a medium. APWG report [57] states that there has been a 70% rise in mobile phone-based frauds in the second quarter of 2022 as compared to the first quarter. A mobile phone user is more susceptible to phishing attacks as compared to a desktop user [58]. The main reasons behind this are small screen size, the inability of the user to see background applications and lack of vigilance [59]. Mobile applications have simple user interfaces that can be easily fabricated [60]. In addition, mobile devices lack application identity indicators [61]. A high response rate of text messages as compared to e-mails is another reason for the growing interest of phishers in this domain. As per a study [62], the open rate of SMS is about 98%, 95% of which are read within the first 3 min. Consequently, mobile phishing can be carried out through the following means:

Phishing through SMS Also known as *smishing*, the perpetrators send a malicious link to the victims through an SMS message on their mobile devices. SMS phishing can be done through one of the following methods [63]:

- **Malware** The smishing message contains a link which on clicking installs malware on the victim's device. The malware can monitor the victim's online activity, capture the login credentials and other sensitive information and send the same to cyber-criminals. It can also disguise itself as a genuine app, tricking the users into revealing their confidential information.
- **Malicious website** The link/URL in the smishing message leads the victims to a malicious website that may be masquerading as a reputed website. The victims may

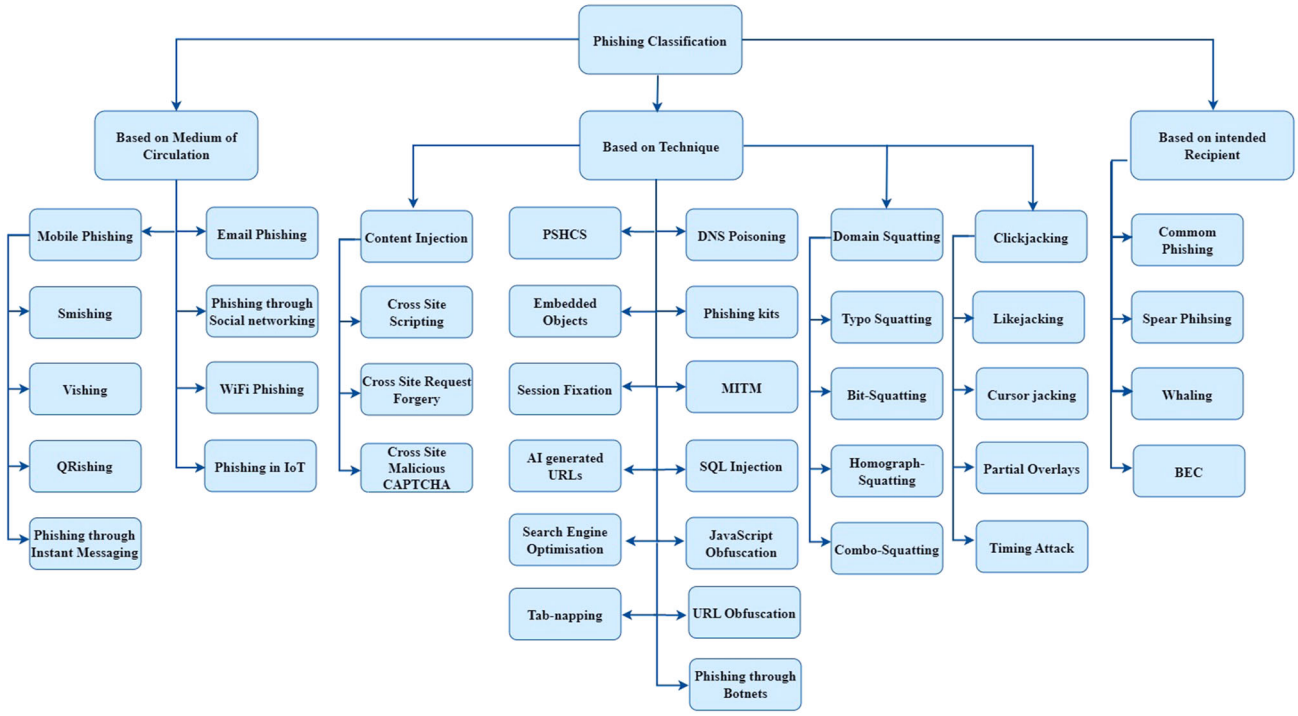

Fig. 5 Classification of phishing

end up filling in the login details or banking information at the phoney website and suffer from monetary losses.

- **Contact through phone/e-mail** Victims are asked to contact a given phone number or e-mail ID to claim complimentary gifts or vouchers. Upon being contacted, the attacker tricks the victims into divulging their credentials.
- **Self-replying SMS** Malicious links are sent to the victims through a self-replying SMS that asks them to agree or disagree with a subscription.

Phishing through calls A survey [64] reveals that people are more cautious about e-mails as compared to phone calls. When they receive a call on their phone, even from an unknown person, people tend to believe it is genuine without verifying its authenticity. Also known as *vishing* which means the use of voice to perform phishing. The victims receive phone calls purported to have originated from a trusted party say a bank or credit card company. The callers are trained to speak in a persuasive manner and try to create a sense of urgency or fear amongst the victims like the bank account being dysfunctional if the required details are not provided. In most cases, the attackers use VoIP (Voice over Internet Protocol) [65] or caller ID spoofing [66] to make the call. The frantic victims are convinced that they are left with no other option but to reveal their sensitive information like social media account details, bank login credentials, or credit card details. The disclosure of private and confidential attributes can lead to further victimisation like stalking or online harassment. The victims are also at risk of becoming suspects in crimes committed by an imposter.

Phishing through QR codes QR or *quick response* codes were first used in the Japanese auto industry in the '90 s [67] to abrogate the limitations of one-dimensional barcodes that could store only a small amount of data. Later, the QR codes carved their own path to being used in smartphones due to their ability to be machine-readable and to store large amounts of data.

QR codes became more prevalent during the pandemic as they proved to be an efficient means to ensure social distancing. Furthermore, they are easy to deploy, free of cost, fast, and convenient. They are damage resistant as compared to one-dimensional barcodes [68]. It is also not mandatory to know the domain name of a website. The URL of the website is encoded through the QR code and the users are required to scan the same through their smartphone camera. The QR code reading application installed on the phone does the rest. This sheer convenience has contributed to raising their popularity among the masses. QR codes are being used in a wide variety of applications like physical access control, ticketing and logistics, identification, and electronic payments. However, the enhanced use of QR codes has attracted criminal minds to consider them as a potential medium to access users' confidential information, spread malware, and most importantly for monetary benefits. Phishers can embed malicious URLs within the QR code [69] which upon scanning leads the victims to a mimicked website. They can also create an

entirely new QR code and stick it over an authentic QR code on a card or a flyer at the retail stores. Usually, the users can decide whether or not to open the link on the phone browser. However, there are certain applications that directly visit the web page without waiting for approval from the users. The use of 'URL shorteners' also limits the user's ability to assess a URL before visiting. Curiosity towards QR codes also leads the user to visit a web page without verifying its authenticity. As per a study [70], 85% of users who scanned a QR code chose to visit the URL in their phone browser even when the domain was unfamiliar. Since no conventional approach or authentication mechanism is employed to generate a QR code [71], they are always at risk of being used by fraudsters as a phishing medium until a comprehensive solution is achieved.

Phishing through Instant Messaging The youngsters of today, who are constantly active on their smartphones make use of instant messaging to get in touch with their acquaintances. Instant messaging applications like WhatsApp and Telegram offer features ranging from audio and video calling to hyperlinks, emojis, photos, videos, and file sharing. This medium of communication is much more popular than conventional SMS and has attracted phishers.

4.1.2 E-mail phishing

As per the Verizon Data breach Report [72], 96% of social attacks arrive through e-mails and almost 100% social attacks involve phishing. E-mails are also one of the dominant mediums to circulate malware as 46% of organisations receive malware through e-mails and 94% of malware are delivered through e-mails. This implies that e-mails are the most lucrative medium to propagate phishing. A report [73] suggests that malware delivered via e-mails have tripled in the fourth quarter of 2021. The e-mails designed for malicious purposes are generally spoofed [2], that is, they appear to have originated from a trusted sender such as a bank, educational institution, credit card company, or a business partner. The probability of a phishing attempt being successful is greatly enhanced when the phishing e-mail is received from someone known to the victim [74]. The contents of the e-mail give the impression to be genuine in terms of language and overall visual details but they instigate feelings of fear, urgency, or greed and impede the prudence of the victim. For example, issuing a warning about the credit card being blocked in case of non-compliance to the e-mail or social media account to be deactivated if the user fails to change his or her password. Some phishing e-mails also start with a congratulatory message about the victim winning a lottery and ask for account details to transfer the prize money. Prior to launching the phishing attack, the fraudsters perform social engineering and gather basic details about the victim. These details when mentioned in the e-mail, hinder the victim's power of reasoning and eliminate any suspicion. CISCO's 2021 Cyber-security threat trends report [75] suggests that at least one person clicked a phishing link in around 86% of organizations. As per Avanan Global Phish Cyber Attack Report 2021 [76], one out of every 99 e-mails is malicious and 54% of total phishing e-mails are sent with an objective to harvest credentials. Another key finding states that Business E-mail Compromise incorporates 20.7% of all phishing attacks and 2.2% of phishing attacks are for the motive of extortion. The report also suggests the use of Artificial Intelligence to mitigate phishing attacks.

4.1.3 Phishing through social media

Phishing through social media refers to phishing attacks that are carried out through social networking platforms like Facebook, Instagram, Twitter, etc. The busy personal and professional lives of users have created a geographical barrier between friends and loved ones. People find it difficult and sometimes cumbersome to personally go and meet each other. Social media has come to the rescue in this situation and has given netizens an opportunity to be a part of the lives of loved ones without being physically present. In 2022, out of 5 billion internet users throughout the world, 93% have an account on social media [77]. This wide outreach of social networking sites has enticed the phishers to use them as a potential medium to spread phishing through impersonation, credential theft and data gathering. The phishers can create a fake social media profile of the victim and send friend requests to their acquaintances. Furthermore, they can ask for money from the victim's friends and family on their behalf. Defamation can also be a reason to set up a fake profile. Fake profiles of business houses, brands, and celebrities also exist.

[78] states that 16% of all the accounts on Facebook are either fake or duplicates. According to a report by the Federal Trade Commission (FTC), in 2021, more than one out of every four persons lost money due to fraud that started on social media. As per a report [73] by PhishLabs, there has been a two-fold increase in phishing attacks with social media as a channel in 2021.

4.1.4 Phishing through Wi-Fi

The never-ending need for users to be online all the time has caused a rapid enhancement of wireless technologies. Users can access the internet even when they are not in the comfort of their homes or offices. They use Wi-Fi hotspot connections at various public places like airports, restaurants, shopping malls, etc. This has made users susceptible to becoming victims of Wi-Fi Phishing or *Evil-Twin Attack* [79]. The phishers create a fraud access point, also known as 'Evil Twin' with the same network name or Service Set Identifier (SSID) as the

authentic access point operating in that area. The login interface of Evil-Twin is reasonably forged to deceive the users and appears to be genuine. In most cases, Evil Twin promises free internet service, and once the user logs into the rogue Wi-Fi network, their sensitive information like passwords and bank details can be snooped by the phisher.

4.1.5 Phishing through Internet of Things (IoT)

The main purpose behind the development of IoT was to create a network of consumer-level devices such as security cameras, lighting systems, doorbells, refrigerators, televisions etc., that can interact over the network and be controlled through voice command or smartphones. However, their enhanced usage in various spheres of our lives has also enhanced the attack surface [80, 81] which is being exploited by the phishers. A network of entwined devices, if breached, can pose a serious threat to user privacy and security in the victim organisation. As discussed in [82], IoT systems are vulnerable to phishing attacks, as the web portal used by the consumer to configure the system is seldom accessed. Thus, the user is unfamiliar with the web portal and cannot discern a fake web page. Moreover, lack of standard technology or protocols to be used by the individual devices may lead to in-coordination and eventually, a vulnerable system [83]. As per a report [84], 98% of data over the IoT is unencrypted and more than half of the devices are exposed to cyber-attacks. Once compromised, a node in an IoT can be made to act as a botnet to affect other nodes or used to initiate a distributed denial of service attack.

4.2 Classification based on intended recipient

4.2.1 Common phishing

This is the most simplified version of phishing, which relies on the fact that some people are more focused on the task at hand and tend to ignore the subtle inconsistencies that exist in the received message. Phishers send the message with the phishing link/URL using any of the phishing distribution mediums to thousands of random victims and hope that some of them will respond. The messages are forged to convince the recipients to divulge their credentials. The probability of success is low.

4.2.2 Spear phishing

This version of phishing is targeted at specific individuals or organisations [85–87]. Prior to launching the attack, the offenders congregate social and personal information about the victims by monitoring their social media activity. After gathering sufficient information, a message is sent. The message seems to be from a trusted party, such as a friend, colleague, business associate, or an organisation such as a bank, credit card company, or admin of social media account. The language and impression of the message are forged and persuade the victims to disclose their sensitive information. A report [88] suggests that, out of all the known groups carrying out targeted cyber-attacks, two-thirds make use of spear phishing. According to research [89], the actual incidents of spear phishing may be much higher as its reporting is scarce. The authors have examined reasons behind the non-reporting of spear phishing by the users and concluded that self-efficacy, expected negative outcomes from reporting spear phishing e-mails, and cyber security self-monitoring are the factors that influence the likelihood of reporting spear phishing e-mails by the individuals.

4.2.3 Whaling

A variant of spear phishing in which the executives with high-level access to organisational resources and information are targeted. Being a targeted attack, it is more sophisticated as compared to a common phishing attack. The success rate is high and the phishers perform a great deal of preparation before launching such attacks. The whaling e-mails are crafted with much effort and include business terminology and tone. They also contain personal information about the targeted individual to avoid any suspicion. The phishing attack on Facebook and Google [44] executives mentioned in the previous section is an example of a whaling attack.

4.2.4 Business e-mail compromise

Business e-mail compromise (BEC) [90] is like whaling in the sense that both have company executives as victims. However, in a whaling attack, the business executive is the target but in BEC, the executive (mainly the CEO) or a trusted party is impersonated and the main target is the business itself. After posing as the CEO or an individual with high authority in the business, the phisher sends spoofed e-mails [2] to lower-level employees instructing them to transfer funds. The mail has a personalised or genuine appeal with business terminology leaving little scope of suspicion for the victim. The Fake president phishing attack on FACC [45] discussed in the phishing case studies is an example of BEC. As per APWG phishing activity trends report, in the second quarter of 2022 [57], 73% of BEC messages were sent from free webmail accounts 72% of which used Google webmail addresses. Gift cards, being used in about 40% of total BEC attacks, were the most popular means to avail benefits. In 26% of the cases, payroll diversion was attempted and wire transfer in 9.6% of the cases. 15.5% of the cases involved advanced fee fraud. BEC has also become more costly with a 14% rise in average wire transfer cost. Public Service Announcement made by Internet Crime Control Centre (IC3) [91] of the FBI men-

tions that from 2019 to 2021, there has been a 65% climb in global losses due to BEC. For almost the same duration, another report [92] mentions an increase in those BEC complaints that involve the use of virtual meeting platforms. As a result of the COVID-19 pandemic, most organisations conducted business through virtual meetings. The threat actors leveraged the opportunity and used the virtual meeting platforms to commit BEC scams. BEC attacks are difficult to detect as they rely on the social engineering skills of the phisher. Malicious URLs or software are not utilised and it is the ability of the fraudster to impersonate a high-ranking official which convinces the victim to transfer funds.

4.3 Classification based on technique

4.3.1 Phishing sites hosted on compromised domains (PSHCD)

To stage phishing, the fake web pages need to be hosted on a server. It is the domain of this server that a victim is being redirected to. However, hosting involves some cost that is to be borne by the phishers. The phishing URL might also be blacklisted or blocked by the search engine once it is reported. To overcome these obstacles, the phishers host their phishing site on a domain that is not owned by them but on the one which they have hijacked by exploiting vulnerabilities in the website content management system such as WordPress, Joomla or Drupal [93, 94]. In a study [95] as early as 2007, 76% of phishing web pages were found to be hosted on compromised domains. A recent report [96] mentions that only 24% of phishing websites are hosted on domains owned by the attacker. Rest are hosted either on hijacked domains or on free hosting services.

4.3.2 DNS poisoning

Domain Name Server (DNS) is used to convert web addresses into numeric IP addresses [97]. Whenever the users type in a web address in their internet browser, the record from the DNS cache is fetched and the user is directed to the corresponding IP Address. DNS Poisoning [98, 99] is an attack in which the record in the DNS cache is altered with an IP Address that serves malicious users. The internet traffic intended for genuine IP Addresses is redirected to the modified one. The attackers can host a phishing site on the fake IP Address and the victim can be duped. It is also known as *DNS cache Poisoning*.

4.3.3 Phishing through Botnets

A botnet is a network of systems that have been infected with malware, that can be remotely controlled and commanded by the cyber-criminals [100–102]. The malware is called a bot or a zombie and the controlling party of the botnet is termed as bot-herders. Instead of targeting specific individuals, the bot eyes those machines on the internet that are vulnerable. After infecting and adding another machine to the botnet, its defence is incapacitated such as impairing the anti-virus. The bot-herders can then communicate with the bot and issue instructions, receive vital information and specify the further course of action. The objective of the phishers behind using this technique is to perform automated tasks by employing the extensive processing capability of a large number of machines that have been cobbled together to create a botnet. The phishers can administer botnets to [103]: send phishing e-mails to millions of other users, act as proxy services, carry out distributed denial of service attacks, spy, spread malware, act as servers to host phishing sites, spamming.

4.3.4 Phishing through content injection

Content injection refers to the insinuation of malicious code, script, image, link, etc. into a legitimate site by exploiting vulnerabilities in its code. Apart from leading the victims to a phishing site, this attack can also induct malware on their computers. There are many variations through which an attacker can carry out a content injection attack.*Cross-site scripting* [104] or XSS is one such attack where the attacker injects malicious javascript code on the data entry field or in a URL on the graphical user interface of a legitimate website. XSS has the ability to bypass the same-origin policy [105], which restricts interaction between scripts and data originating from two different domains. The malicious script executes when the page is loaded on the web browser of the victim. Upon execution, the malicious script can access the victim's personal information stored on the browser and convey it to the attacker's server. The script can also present a counterfeit login form to the victim to capture his or her confidential details. The victim can also be redirected to a phishing site under the pretext of visiting a legitimate site. The PatchStack white paper [106] states that in 2021, out of total WordPress vulnerabilities, almost 50% belong to cross-site scripting which is about 36% more than the previous year.

Phishing through content injection can also be done by *Cross-site Request Forgery* [107, 108] or CSRF. To launch a CSRF attack, the phishers lure the victim to click on a malicious URL which can be dispatched to the victim through e-mail or social media. This URL is crafted to send an unauthorised request to the web application with which the victim already has an active session. The susceptibility of the target application to failing to distinguish between genuine requests by the user and unapproved requests by the user is exploited in CSRF. The CSRF attack can be used by the attacker to trick the victim and send an unfavourable request to the web

application server. Such requests can be to change passwords, transfer funds, add or delete records, etc. The report in [106] states that Cross-Site Request Forgery amounts to 11.18% of total WordPress vulnerabilities.

Another variation of content injection attack is *Cross-site malicious CAPTCHA* attack [109]. This attack also attempts to outmanoeuvre the same-origin-policy [105] as the XSS attack. The victim is tricked into revealing their sensitive personal information to a fraud website set up by the phishers. The sensitive information of the victim is displayed on the phoney website in an obscured manner like a CAPTCHA code and the victim is asked to complete the details. Other methods to lure the victim include a gaming challenge or typing test. The naïve victim fills in the personal details which are forwarded to the phishers.

4.3.5 Phishing through search engine optimisation

After fabricating a phishing website, the aim of the phisher is to induce the victims to click on the link of the website. Through the search engine optimization [110, 111], the phishers make sure that while searching for the particular goods or services on the search engine, their phishing site is displayed amongst the top results. To the victim, the phishing site appears to be indexed legitimately by the search engine. The phishers can also incorporate black hat search engine optimization [112] to improve the page rank of their phishing website resulting in better search engine indexing. This is achieved by including keywords of prominent occasions or trends in the designed malicious website. This technique boosts the probability of the victim clicking on the phishing link.

4.3.6 Phishing through domain-squatting

Domain-squatting [113] or *cyber-squatting* refers to the act of registering domain names that are much too similar to those of trusted legitimate websites. Apart from registration, the phishers also set up a clone website which is almost the same as the original. When a user makes a mistake while typing the domain name of a website, since the domain name with the same typographical error is already registered, he or she is directed to the phoney website which has the same visual appearance as the original website. The unwary victim enters the details in the login entry form at the phishing site and ends up disclosing their credentials. Apart from relying on the victims to make a typographical error, the phishers can also forward the phishing links to unsuspecting potential victims. There are many techniques that fraudsters employ to go through with domain-squatting.

Typo-squatting [114–116] is planned to manipulate the typographical mistakes in the target domain names typed by the victim. For example, use Citibak.com or Citibamk.com in place of Citibank.com. Typo-squatting can also be executed with well-known software packages. The users can unknowingly download malicious software on their systems.

Bit-squatting [117] is a variation of cyber-squatting which relies on bit-flip faults that occur while the user has made a DNS query. The bits can flip due to various reasons which include cosmic rays, malfunctioning hardware, and operating the device outside the permissible temperature range. As per the authors in [117], bit flips can also occur due to the absence of Error Correcting Code RAM in a majority of computers and smartphones. The Error Correcting Code RAM has the potential to identify and correct the bit flips. To leverage this fault, the phishers register domain names that have a letter differing by one bit from the corresponding letter in the target legitimate domain.

Sound-squatting [118] also known as homophone squatting is a domain-squatting technique that is based on homophones, i.e. words that are spelt differently but sound identical. For example, Citibank.com and Sitibank.com. The phishers can register multiple homophonic variations of a particular target domain and wait for a confused user to be redirected to their phoney website.

Homograph-squatting [119] technique exploits the visual similarity between different characters and registers bogus domain names to fool the users. An example can be googIe.com in place of google.com (use of upper case 'i' in place of lower case 'L'). This form of cyber-squatting has been further enhanced by the introduction of IDN (Internationalised Domain Name) homograph-squatting. In IDN homograph-squatting, the attacker replaces one or more characters of the target domain name with other indistinguishable characters from a different language. For example, the Greek letter omicron (ο) and Latin small o (o), Cyrillic 'а' and Latin 'a'.

Combo-squatting [120] suggests registering domains by leaving the original target domain intact but including additional keywords along with it (example-youtubelogin.com or facebooklive.com). Amongst all the domain-squatting mechanisms, combo-squatting attracts the most traffic [121]. After analysing more than 468 billion DNS entries over a period of six years, authors in [120] have concluded that even after 1000 days, 60% of combo-squatting domains continued to be alive. Table 2 shows various examples of domain-squatting.

The detection of domain-squatting can be performed by analysis of the target domain. Authors in [114] have proposed a model to predict different typo-squatted domains for a particular target domain. The authors in [122] have proposed the detection of bit-squatting by analysing the different permutations of bit-flips for all characters of the domain.

Table 2 Domain-squatting of the URL www.phishing.com

Domain	Squatting type
wwwphishing.com	Missing dot typo-squatting
www.hishing.com	Omission typo-squatting
www.phishiing.com	Insertion typo-squatting
www.hpishing.com	Permutation typo-squatting
www.fishing.com	Sound-squatting
www.phishingsite.com	Combo-squatting
www.Phishing.com	Homograph-squatting
www.phishinf.com	Bit-squatting

4.3.7 Phishing through URL obfuscation

The internet users of today are smart enough to characterise a genuine and a fake URL if there are visible differences between the two. In pursuance of a possible victim to click on a phishing URL without being suspicious, the cybercriminals practice URL obfuscation [17, 123]. The phishing URL is either hidden/shortened or imitated as the authentic URL. Several techniques to obscure a phishing URL have been identified: The attacker can register domain names similar to authentic popular websites and share the phishing URL with the victim mainly through e-mails. This technique is also known as *Bad Domain Names*. A popular way towards creating an obfuscated URL is by adding a subdomain. The phisher can register a domain (for example- mydomain.com) and share a URL with the victim by appending a subdomain of a popular website (for example- facebook.mydomain.com). Here, the victims who are unaware of the technicalities of a URL can consider this as genuine without noticing the actual domain. URL obfuscation can also be achieved by swapping the domain and subdomain. Changing the top-level domain or country code top-level domain can also result in a phishing URL. An unsuspecting user can visit the said URL and end up providing the login details to phishers.

URL shortening can also be used for the purpose of phishing. Third-party URL shorteners like tinyurl.com and smallurl.com provide a facility to shorten a URL. This helps to manage lengthy URLs which have various subdomains, multiple subfolders and query strings. Although these URL shortening services do not intend to, they are used by the phishers to fulfil their unlawful purposes. The fraudsters register a phoney website and use the third-party service to generate a shortened URL which they share with the victim. Since the actual URL is obscured, the victim is unaware as to whether the web page to be visited is genuine or forged. Table 3 lists some sample obfuscated URLs for the genuine URL www.new.legitimate.com.

Table 3 URL obfuscation for www.new.legitimate.com

Obfuscated URL	Obfuscation Type
www.legitimate.new.com	Swap domain and sub-domain
www.new.legitimate.co.in	Change of country code TLD
www.new.legitimate.n.com	Adding phished domain to URL
www.tiny.com	Use of URL Shortners

4.3.8 Phishing through JavaScript obfuscation

JavaScript is a scripting language used to design client-side web pages and is applied in most websites [124]. A fraudster can embed malicious script in the client side code which executes when the page is loaded in the victim's browser and redirects a victim to the phishing site or to install malware on the victim's computer. JavaScript can also be used to create a hoax address bar, padlock icon, and SSL certificate and make the URL of a phishing site appear legitimate. The user does not become suspicious in the absence of any apparent discrepancies. A survey [125] reports that 83% of the phishing sites use SSL certificates. However, with growing incidents of phishing, users employ anti-virus software to evade such attacks on client-side JavaScript code. But the attackers have come up with a mechanism of obfuscating the phishing code to dodge the anti-virus software [126, 127]. The obfuscated code is similar to machine-level code and is difficult to comprehend and analyse. Obfuscation also makes it difficult for the researchers to reverse-engineer the phishing code and understand its origin [103].

4.3.9 Phishing through SQL injection

The development of Internet technology has led to an increase in the number of web-based applications. An urgency among web developers to deliver the software by the deadline has become a cause of many security-related issues. As per a study [128], 75% of web applications and online businesses are susceptible to being attacked online and SQL Injection attack [129, 130] is one such attack. Through an SQL injection, the attacker sends malicious SQL commands to the database via the data entry form at the website's user interface. The attacker mainly targets those web applications that lack proper security measures such as user input validation, web application firewall, data sanitisation, etc. SQL injection attack, if successful, enables the attacker to get unauthorised access to information such as sensitive business attributes, personal details and financial information. The attacker can manipulate the back-end database and steal credentials for impersonation and privilege abuse. Administrative rights enabling the attacker to modify or delete entire tables can also be achieved through SQL injection attacks.

4.3.10 Phishing through man in the middle attack

As the name suggests, in Man-in-the-Middle (MITM) attack [8, 17, 27], the phisher places itself in between the user and the web application server and acts as a proxy to the server. By employing the MITM attack, the assailant can interject, eavesdrop and even alter the communication transpiring between the user and the server and get access to private and confidential information of the victim without raising suspicion. The phisher sets up two separate connections. One between itself and the victim, and the other between itself and the server. The victim perceives the attacker's server as the genuine server and discloses his or her credentials which the attacker can use to escalate the attack or store to be misused later. Also, the attacker's server communicates with the real web server masquerading as the actual user. Thus, both the victim as well as the server correspond with the attacker's proxy server. MITM attacks can be deployed in various communication channels [131] such as Bluetooth, Wi-Fi, and GSM. This attack is hard to detect as the two-way communication between the victim and the server happens seamlessly and there are no markers of anything going haywire. Even in cases of secured HTTPS communications, the attacker establishes its own SSL connections. To facilitate the exposure of the victim to itself, the attacker deploys various methods like DNS cache poisoning, URL obfuscation, transparent proxy, browser proxy configuration, etc. DNS cache poisoning and URL obfuscation have already been discussed in earlier parts of this section. The phisher reuses the source code of the genuine website to create its phishing version and sets up *transparent proxy cache* [17]. The victim's request for the genuine website is intercepted by the proxy, and the created phishing version is returned. This fraud version expropriates the communication between the victim and the genuine web server and can also store the victim's credentials and sensitive information. *Browser proxy configuration* [17] involves tweaking the proxy configuration at the victim's web browser and influencing the entire web traffic to pass through the attacker's server. In order to alter the proxy settings, the phisher initiates the attack in advance by deploying malware (maybe through an e-mail). After altering the proxy configurations in the browser, the attacker's server acts as a proxy between the victim and the authentic server destination.

4.3.11 Phishing through ClickJacking

Also known as *UI redressing Attack* [27, 132–134], ClickJacking attempts to expropriate the clicks done on the user interface of a web page, link of which the victim might have received through a spoofed e-mail [2]. The victim is lured to click on a web page element (such as "Click on this button to claim your reward") that is disguised as a genuine element. However, the clicking action leads the victim to perform unintended tasks like inadvertent download of malware, execution of malicious scripts, submission of credentials, making payments, etc. The attacker exploits the web designing functionality of iframes, which allow one web page to be displayed or overlapped within another parent web page. The target web page is invisible and is embedded on top of the legitimate web page, which the user recognises. The user has the perception that the web element being clicked is genuine, but in fact, he or she is clicking on a malicious invisible element affixed on top of it. There are many variants of ClickJacking attacks. *LikeJacking* [135] where the user is tempted to like a post on Facebook by transparently superimposing it over some other web page element (such as the 'Skip Ad' button). *CursorJacking* [133] where the attacker replaces the genuine cursor with a decoyed cursor. The victim gets disoriented about the actual position of the cursor and clicks on the unintended region of the page. Another type of attack is an attack on the pointer [136], where an invisible iframe is attached along with the pointer and moves throughout the screen with the pointer. Whenever the user clicks, regardless of the user's intentions, the invisible iframe is clicked. The attacker can confuse the victim through *Partial Overlays* [137] by hiding a part of the target web element. For example, on the online payment page, the attacker can obscure the receiver details and amount but leave the 'Make Payment' button as it is. The oblivious victim goes ahead with the transaction without knowing about the actual receiver of the amount. The attacker can also exploit the response delay exhibited by the users while clicking during any task [137] and launch a *Timing Attack*. The time it takes to react can be the time it takes to click while hovering over a display element or the time it takes between two individual clicks of a double click. The attacker can insert a web element (such as a 'Pay Now' button) over a decoy button just before the user clicks.

4.3.12 Phishing through embedded objects

Most of the phishing detection techniques [23, 138–142] rely on the source code and textual details of the suspicious website. They extract features accordingly and compare them with features of an authenticated site. If for a unique URL, the similarity is below a predefined threshold, the suspicious website is classified as a phishing website. To bypass these techniques, attackers replace the entire textual content of a web page with embedded objects such as images [103], flash, scripts, etc. Upon substituting the source code content with an embedded object (say, an image), the various phishing detection methods are not able to extract features, thus resulting in inaccurate classification.

4.3.13 Phishing through tab-napping

Also known as *tab-jacking* [143] i.e. hijacking of the browser tab. Users who have a habit of opening multiple tabs simultaneously while accessing the internet are the most vulnerable to becoming victims of this attack. The attacker shares a phishing link with the victim. Once, the victim clicks on the link, a phishing page resembling a genuine web page is opened. Nothing remarkable happens while the victim is on that phishing web page. However, once the victim navigates to another opened tab and the phishing site's tab becomes inactive, a malicious script embedded in the phishing site executes and loads a hoax login screen (such as for an e-mail or social media account) and modifies its favicon and title. As the user's focus again shifts to the phishing tab and he notices a login screen, he perceives that his earlier session has expired. Since the user had opened the same website earlier, even though on a separate tab, he or she remains unsuspicious about the phishing website and submits the login credentials, which are redirected to the attacker. The attacker exploits the presumption of the victim that once a tab is opened, its contents remain static. The victim is oblivious to the fact that even a pre-loaded tab can be led to open a phishing website by executing some malicious JavaScript code. [144] demonstrated the execution of this attack.

4.3.14 Phishing through session fixation

Also known as *preset session attack* [17], this attack focuses on the session identifiers that are used by the server to monitor the activity of the user throughout a session. After performing validation, when the user logs into a website and performs various endeavours, such as selecting items to buy, making payment, updating their profile, etc., a unique session identification ID (SID) is assigned to that particular session of the user. This SID keeps track of the user's activity when he or she navigates through various web pages within the website. The SID can be saved as a URL, cookie, or form field. In the session fixation attack, the phishers create a session ID before the victim logs into a web server, and lure the victim to start a session with it. Thus, there is no need to get the victim's session ID thereafter [145]. This may be done by sending an e-mail containing a URL with the created SID to the victim. The URL may be for a login form. As soon as the victim authenticates with his or her login details, the phisher can hijack the active session and unauthenticated transactions can be performed. Since the URL is that of a legitimate website, the victim does not become suspicious.

4.3.15 Phishing through phishing kits

Once a phishing site is reported, it can be efficiently blacklisted or blocked. It is a time-consuming process to create a phishing site from scratch. The use of a phishing kit [146] enables criminal minds to create a fresh hoax website easily just by following simple instructions without having advanced programming abilities. Rather than designing the phishing website themselves, the attackers deploy phishing kits which are readily available on the dark web [48]. Phishing kits are ready-made fake templates of famous websites that have a vast customer base. The fraudsters must execute the instructions provided along with the phishing template to carry out a phishing attack. Consequently, they need not possess advanced technical skills to be successful phishers. Most phishing kits also facilitate the hosting of the phishing site, mainly on compromised websites or on websites that provide free hosting services. Some sophisticated phishing kits may also contain means of transmitting the phishing website to the victim and scripts to capture victim credentials. In the year 2021, Kaspersky blocked 1.2 million phishing websites which were generated through 469 phishing kits [147]. Phishing-as-a-Service (Phaas) [31] is also available from a variety of online resources and can be purchased by anyone with money and malicious intent. Phaas is a business model through which an experienced cyber-criminal becomes a service provider for a novice phisher. Through Phaas, the phisher can develop, deploy and manage the money-related aspects of phishing sites without any hassles.

4.3.16 Phishing through AI-generated URLs

The software-based phishing detection techniques use Artificial intelligence (AI) to train the system for the detection of phishing websites. The phishers can also try to improve the standard of their attacks to bypass the anti-phishing approaches. In an attempt to outline the different approaches used by the threat actors to evade AI-based phishing detection techniques, the authors in [148] analysed more than a million phishing URLs and tried to understand various strategies that the phishers can utilise to create phishing URLs. The authors have simulated how deep neural networks may be used by attackers to improve their efficiency. Through Long-Short Term Memory Networks (LSTM), the authors have created an algorithm that generates synthetic URLs. It has been proved that these URLs have a much better likelihood of avoiding AI-based detection mechanisms.

5 Analysis and discussion on various phishing attacks

This section presents a detailed analysis of different categories of phishing attacks that have been identified in the previous section.

*Phishing Distribution Mediums:*In an attempt to widen their reach, the attackers are targeting mobile users through personal communication mediums at a much higher rate. A report [150], shows a 50% rise in the attacks on mobile devices in 2022 as compared to the previous year. Handheld devices are more vulnerable to phishing attacks as compared to a desktop, because of the architectural differences. There has also been a surge in the use of voice mail and text messages to carry out spear phishing and BEC attacks. The purpose behind this is to lend a sense of credibility to the sender. Attackers also target open Wi-Fi networks in public places to steal the user's credentials. The use of VPN (Virtual Private Networks) can provide an additional layer of safety against the same [151]. In 2022, attacks on another phishing distribution medium, IoT, rose by 65% [152]. The primary reason identified behind this surge is the lack of sufficient security mechanisms in IoT devices, the unregulated addition of devices with dubious supply chains which makes the network vulnerable to malware attacks, and uninterrupted access to the IoT devices through the internet which opens a back-door for the fraudsters.

Though the attackers are exploring a wide range of phishing distribution mediums, a majority (96%) of phishing attacks continue to be delivered through phishing e-mails [72]. The primary reason includes them being the most widely accepted, convenient and inexpensive means to share messages with millions of users at a mouse click. Moreover, the degree of anonymity that the attackers enjoy while sending a phishing e-mail is unparalleled. They can use spoofed or compromised sender addresses for the e-mails to appear genuine. As per [153], 21% of the users are unaware of the concept of e-mail spoofing. The attackers also launch targeted attacks that have the ability to bypass the e-mail filtering mechanism resulting in 18.8% of phishing e-mails entering the victims' inboxes. [154]

*Phishing Intended Recipients:*Amongst targeted attacks, spear phishing is the most common variant. Spear phishing e-mails are difficult to identify as they do not depict any signs of being fake. Also, most of them emerge from hijacked e-mail accounts. Due to the availability of various platforms to host fake web pages such as Microsoft Azure Custom domains, it has become arduous to single out fake ones. The presence of a Microsoft SSL (Secure Socket Layer) certificate further dampens any suspicion. However, more than 80% of the phishing sites are protected by SSL [125]. This implies that the mere presence of an SSL certificate can no longer be considered a sign of safe browsing.

The BEC attacks have shown an increase of 81% over the two halves of 2022 [155].The open rate of BEC attack e-mails was 28%, out of which, the reply rate was 15%. The major cause behind this substantial growth has been identified as the presence of victims' information on Linkedin, social networking sites, parent websites, etc. The threat actors can leverage these details and produce convincing e-mails with a higher likelihood of tricking the employees of the organisation.

*Phishing Techniques:*The phishing techniques discussed in the previous section employ a wide range of tactics to deceive the victims. Some of these techniques exploit weaknesses in the target organisation such as lack of sufficient security measures, firewalls, data validations, and coding vulnerabilities. Phishing through botnets, content injection, JavaScript obfuscation and SQL Injection attacks fall into this category. Initially, JavaScript obfuscation was used to prevent web scams by obscuring the actual code. But, the phishers are using the same strategy to evade detection. A report [156], uncovers that, out of 10,000 malicious JavaScript samples, at least 25% used obfuscation. SQL Injection attacks comprised 76% of all web application attacks in 2020 [157]. Cybercriminals are increasingly using botnets to launch a series of attacks on unaware victims. In 2021, botnet attacks saw a rise of 23% [158].

Phishing attacks through session fixation, ClickJacking and Tab-Napping work by relying on the technical skills of the attacker. In order to generate phoney clicks on hidden adverts, fraudsters have long depended on malware or automated scripts. However, in recent years, criminal organisations have begun to switch to methods that hijack actual user clicks. A study [159] collected data on 250,000 websites and found ClickJacking scripts on 613 popular websites. Although this is not an astounding figure, the fact that these 613 websites attracted a daily traffic of 43 million hits is alarming.

Phishing through domain squatting and URL obfuscation depends on the victim's ignorance or inability to distinguish between genuine and fake brand domain names. In both these attacks, popular brand names are either misspelt in the domain or used as it is in the subdomain or path with the intention to mislead the victim towards visiting the same. During the COVID pandemic, 55% phishing sites used brand names in the URL [160]. LinkedIn was the most impersonated brand in the first Quarter of 2022 [161].

A major percentage of phishing sites are hosted on compromised domains [95, 96]. The rationale behind the same is that PSHCD can bypass list-based detection measures as the legitimate domains on which they are hosted are not included in blacklists. Since these domains have been in existence for a long time, they can bypass those detection measures that take into account the domain registration time. Due to the reputation of legitimate domains, they are indexed in search engines and are also able to bypass search engine-based phishing detection approaches. It is very important for academia to devise an efficient technique to distinguish between maliciously owned and compromised domains. A different strategy needs to be followed once a phishing

Table 4 Analysis of various phishing attacks

Attack type	Attack sub-type	Analysis
Phishing Distribution Mediums	Phishing through E-mail	Contribute in 96% Phishing Attacks.Convenient, cheap, widely accepted, easy to deploy
	Phishing through Mobiles	Increasing at a rate of 50%. Reasons include architectural differences, simple user interface, lack of application identity indicators.
	Phishing through IoT	Unencrypted data, lack of standardisation leading to a rise of 65%
	Phishing through Social Media	Attacks doubled in 2021. Wide outreach and a trend to share personal details lead attackers to execute impersonation and credential theft.
	Phishing through WiFi	Urge within the users to always be online and promise of free internet leads the attacker to achieve success
Intended Targets	Spear Phishing, whaling	Targeted attacks, constitute 76% of phishing attacks. Main purpose is to gather credentials to be used for a bigger attack later on.
	BEC	Increase by 81%. Highly profitable. Reasons for success include the availability of victim's information online.
	Common Phishing	Less customised, lesser probability of success but fewer resources required
Phishing Techniques	Content Injection	Exploit System Vulnerabilities. Diverse attacks -malware upload, credential theft, malicious code/links insert, website defacement. Difficult to detect. Can be mitigated by proper code reviews and content security policies.
	SQL Injection	76% of all web-app attacks. Targets assailable applications. Capability of stealing sensitive data from the database. Can be prevented by ensuring proper input validations, web application firewalls, regular security tests and updates.
	Domain Squatting, URL Obfuscation	Exploit user ignorance. Target well-known brands (used in 55% of phishing sites). Researchers and domain registrars need to collaborate for effective prevention.
	Botnets	Distributed Attack Infrastructure. Spread phishing messages over a massive number of targets, operational for a longer duration.
	Session fixation, ClickJacking, Tab-Napping	Exploit the design functionalities of web browsers along with user distraction. Can be kept under check by user awareness and web browser security measures.
	PSHCD	Include majority(more than 70%) of phishing sites. Done to avoid being blacklisted. Once up, it remains undetected for a longer duration and bypasses URL-based and search engine-based detection measures.
	Phishing through embedded objects	The entire content of web page is replaced with images/flash etc. Can be caught by image-based detection approaches but they are computationally expensive
	DNS Poisoning	Targets DNS servers that are misconfigured. and are easy to compromise. Affects a large number of users. Operational till the time the DNS cache is refreshed. Also, rely on attacker's knowledge of DNS protocols.
	Phishing through Search Engine Optimisation	Performed to enhance the reputation of the phishing site by manipulating the search engine rankings. Can be prevented by refining the search engine algorithms and employing sufficient security measures.
	Phishing through Phishing Kits	Pre-assembled and customised collection of tools that make development and deployment of phishing sites easier. 3677 phishing kits uncovered in 2022 which is a 25% increase [149]. Security companies can develop a signature to identify and block these attacks.

domain is identified as compromised. The reason is that the owner of the compromised domain is also a victim [93]. If the compromised domain is blacklisted, its owner would suffer from monetary losses without being the culprit. Table 4 presents a summary of the critical analysis of various phishing attacks.

Phishing detection approaches mentioned here are discussed in detail in the following section.

6 Phishing countermeasures

This section presents a summary of different anti-phishing measures that are being applied to counter phishing. Throughout the literature, phishing countermeasures are segregated into user involvement-based countermeasures and software-based countermeasures. The main idea behind user education-based phishing countermeasures is that most phishers are able to fulfil their goal due to a lack of awareness among individuals. A large proportion of internet users

are unaware of basic security etiquette that must be followed while being online. So, it is of utmost importance for them to be provided with effective training and guidance about the response to be followed during diverse interactions on the internet (such as e-mails from banks or any other online service for routine maintenance and data updates).

User education is a very important and effective constituent of phishing countermeasures. But, its main drawback is its reliance on the user's skill and ability to understand the use of the system. Even the security experts are outwitted by the phishers, who are learning new skills to introduce new methods of deception. Moreover, users have to dedicate considerable man-hours towards learning the process.

6.1 Software-based phishing countermeasures

The researchers' preference for software-based phishing countermeasures stems from their ability to withstand a phishing attack with minimal user involvement. Based on the features being used, the software-based phishing approaches are categorised as list-based approaches, heuristics-based approaches and visual similarity-based approaches. Their further categorisation is discussed here.

6.1.1 List-based phishing detection approach

This approach is further broken down into blacklist and whitelist-based approaches. *Blacklist-based* phishing detection techniques [162–164]maintain a database of resources such as URLs, websites, images, DOM (Document Object Model), etc which are known to be reported as phishing sites. Whenever the user clicks on a URL or visits a web page, it is first verified with the black-list. If a match is found, the system warns the user about a possible phishing threat or even blocks the malicious web page from being loaded onto the user's browser. *Whitelist-based* phishing detection techniques [165–167] maintain a database of legitimate resources. The available resource of a suspicious website is matched with the white list and flagged as phishing if its similarity with an entry in the white list is above a predefined threshold but its domain mismatches. In the case of URLs, if the URL matches an entry in the whitelist, it is termed legal.

Pros and Cons The main advantage of list-based techniques is that they are simple and lightweight to implement in the client's browser. But their primary drawback lies with the fact that they are not able to detect *zero-day phishing attacks* i.e. those phishing attacks that are yet unknown to the users and the anti-phishing players in the industry. As per [168], most of the phishing websites (47%-83%) are updated after 12 h in the black-list. Moreover, the phishers make slight modifications to blacklisted URLs and are able to evade the phishing detection filters. Another drawback is that the list needs to be updated periodically to keep pace with the exponential growth of phishing websites.

Table 5 Some URL-based features

Feature type	Feature name
Count-based features	URL length
	Number of dots
	Number of subdomains
	Number of special characters (*,?,/,-)
	Number of digits
Binary features	Presence of IP address
	Presence of brand name
	Presence of suspicious words
	Presence special characters(*,?,/,-)
	Presence of https
	Presence of @ symbol
	Presence of URL shortener

6.1.2 Heuristic phishing detection approach

This category of phishing detection approach relies on the probable features/properties that are displayed by a known phishing website and trains a machine learning model or uses a rule-based approach to try and find these properties in a suspicious website. This approach is based on features extracted from URL, source code, and a third party. *URL-based heuristic phishing detection* techniques [138, 169–171] select features from the suspected URL to detect phishing. These features may be count-based or binary. Some of the URL-based features are listed in Table 5. *Source code-based heuristic phishing detection* techniques extract common features present in the content of the suspicious web page being loaded. These can be based on hyperlinks [139] or textual keywords [172, 173]. Table 6 lists some of the source code-based features. *Third-party-based heuristic phishing detection* techniques [142, 174, 175] rely on data provided by a service other than the software or the user such as search engine indexing, page rank, WHOIS information, domain age, etc. *Search Engine-based* techniques [176] extract keywords along with the title, meta description, copyright information, domain from website's source code and generate a query. This query is then fed to the search engine. The given website is classified as legitimate only if its domain is returned in the top search engine results of the query. Hybrid techniques [140, 141, 177–179] are also in prevalence which employ a combination of heuristic approaches (such as hyperlink with content-based or content-based with search engine-based).

Pros and Cons The primary advantage that heuristics-based approaches enjoy over list-based approaches is that

Table 6 Source code-based features

Feature type	Feature name
Hyperlink-based features	Number of null links
	No hyperlink present
	Ratio of internal and external hyperlinks
	Externally redirecting hyperlinks
Text-based features	TF-IDF of keywords present in the web page
	CSS features
	HTML tag-based features

they are capable of detecting zero-day attacks. Machine learning algorithms help to achieve high accuracy even though they increase the computational overhead and training cost. Search engine-based techniques have low complexity and work in real time but they fail when a newly registered genuine website is encountered resulting in high false positives. Keyword-based techniques are language-dependent and work only for the English language. Another limitation of heuristics-based approaches is that all phishing sites do not possess similar features. Once the fraudsters gain knowledge about the phishing detection scheme used, they can easily bypass the features and continue with their malicious designs. Also, they are not able to correctly classify phishing sites hosted on compromised domains (PSHCD).

6.1.3 Visual similarity-based phishing detection approach

To circumvent the heuristic phishing detection techniques, attackers replace the entire content of the web page with embedded objects such as images, Flash, JAVA Applets, etc. To contradict these phishing sites, *visual similarity-based* phishing detection techniques [180–183] are used, which are based on the assumption that the attacker tries to imitate the visual details of a targeted genuine website to deceive the victim. A database containing visual features (such as font details, images, logos, page layout, etc.) of known legitimate sites is maintained, and if the similarity score between the suspicious website and an entry in the pre-stored database is above a certain threshold for mismatching domains, the suspicious site is labelled as phishing.

Pros and Cons Visual similarity-based techniques can detect embedded objects in a web page that the heuristic techniques fail to detect. Moreover, these techniques use features that are common for the entire website. So, there is no need to extract different features for different web pages of a single website. However, they fail when non-pre-stored phishing sites are encountered. Insertion of empty HTML tags or deletion of unimportant tags also leads to their failure. These techniques suffer from large storage requirements and computational complexity. Furthermore, the attacker can evade these techniques by reducing the similarity in appearance.

Table 7 summarises various anti-phishing approaches on the basis of different properties exhibited by them.

6.2 Datasets

The researchers need access to a dataset of phishing and legitimate websites not only to test and train the proposed technique but also for performance evaluation. One of the benchmark datasets for malicious sites is *PhishTank*. Launched in 2006, PhishTank [184] is a community-based phishing verification system. A suspicious phishing site is

Table 7 Summary of various anti-phishing approaches

Approach	Zero-day detection	Database independence	Third party independence	Language independence	New legitimate site detection	Embedded object detection	PSHCD detection
Black-list-based	No	No	Yes	Yes	Yes	No	No
White-list-based	No	No	Yes	Yes	No	No	No
URL-based heuristic	Yes	Yes	Yes	Yes	Yes	No	No
Source code-based heuristic	Yes	Yes	Yes	No	Yes	No	No
Third party-based heuristic	Yes	Yes	No	Yes	No	No	No
Visual similarity-based	No	No	Yes	Yes	Yes	Yes	Yes

added to the dataset once it is verified after being reported. The dataset is updated periodically and has proved to be of great assistance to researchers in this genre. *Alexa Top Websites* was a resource for legitimate websites. Though it was discontinued in 2022, other players like *Ahrefs* [185], *Similarweb* [186] and *Majestic Million* [187] have filled in.

To tailor the data as per their requirements, researchers also create and publish their own datasets. [188], proposed by [169] is one such dataset. It incorporates 36,400 genuine URLs and 37,175 malicious URLs. Another dataset is *ISCX-URL2016* [189], prepared by researchers at the University of New Brunswick. A repository containing 1,14,250 phishing as well as legitimate URLs is created by consolidating the URLs from five different data sources. The *Mendeley* dataset [190], contains 58,000 legitimate and 30,647 phishing web instances having 111 features each. The computer emergency response team (CERT) of Japan released a dataset *JPCERT/CC* [191] of phishing URLs which were confirmed from January 2019 to June 2022. All the above-mentioned datasets are publicly available and are of cardinal value to the research community.

6.3 Analysis of various phishing counter measures

This section presents a comprehensive analysis of various phishing detection approaches that researchers have proposed over the years.

List-based approaches To tackle the matter of near duplicate blacklisted URLs, [164] proposed an approach which detects variants of blacklisted sites by generating all possible URLs from the blacklist. The generated URLs are checked for maliciousness through content similarity and DNS query. The technique fails when different URLs having the same phishing content are encountered. In [162], a third-party independent approach is proposed to detect replicas of existing blacklisted sites. The source code features of suspicious sites are compared with those of the blacklisted web pages and similarity is calculated using hamming distance. The approach does not give the desired results when the entire content of the web page is replaced with an image. [192] adopts a distinct viewpoint to identify malicious web pages and add them to blacklists. The new URLs are found by pursuing the phishing forms iteratively and tracking the redirections obtained from URLs that are blacklisted. The whitelist approach works by classifying the resources that are not in the list as malicious [165]. But this approach classifies newly visited web pages as phishing. To overcome this issue, in [166], the new web pages that a user visits are first checked for legitimacy via DNS details and hyperlink information, and if found genuine, updated in the whitelist. The approach is said to be able to detect zero-day attacks. A similar strategy is suggested in [167] but with a different method for comparing the domain name to one of the whitelist entries. However, both approaches suffer from third-party dependency.

Heuristics-based approaches In [170], 14 URL-based features are extracted to train Naive Bayes and SVM (Support Vector Machines) classifiers. An accuracy of 90% is achieved through SVM. [169] presents a method for phishing detection that combines seven different machine learning algorithms with word vectors, hybrid features, and NLP (Natural Language Processing) based features. The suggested method can be used in any language, is independent of third parties, operates in real-time, and can even identify newly launched phishing websites. With 73,575 URLs, the authors have produced a sizable phishing dataset. According to the trial findings, a Random Forest classifier combined with NLP-based features had the best accuracy of 97.98%. However, the method is less effective when used with phishing URLs that have short domains or no path. In [138], a phishing detection method with only 9 URL-based lexical features is proposed. The approach's major goal is to create a system that can be used in Android applications and IoT (Internet of Things) contexts and quickly identify malicious URLs. Although the method yields a 99.5% accuracy, using too few features can lower the accuracy in actual settings. In [171], 9 efficient features from the URL are extracted and used to train 7 different machine learning classifiers. The random forest model provided the best accuracy of 95.2%. In [194], 33 URL-based attributes are collected from more than 11,000 websites. These attributes are fed to various machine learning classifiers after preprocessing and the proposed ensemble classifier of LSD (Logistic Regression, SVM and Decision Tree) achieved the highest accuracy of 98.12%.
The title of the web page is appended to the domain name and given as a search query to the Google search engine in [174]. The website is considered legitimate if the domain name matches any of the domain names of the top search results. The user is warned of phishing if the domain is not on the search engine result page. A 95.95% accuracy is attained, although the method is language dependent and suffers from significant false positives when legitimate but less well-known freshly released websites are encountered. A client-side deployable, third-party independent approach is proposed in [139].To obtain a 98.4% accuracy rate, the suggested method extracts hyperlink-based features from the website source code and trains a logistic regression classifier. But if a phisher modifies the page source references (for instance, favicons, pictures, or javascript) or uses embedded objects, the approach can be circumvented. Integration of search engine-based and heuristic methods is presented in [176] to propose a language-independent and lightweight solution that achieves a true positive rate of 98.15%. [193] presents a methodology that combines a web content-based approach, heuristic features, and blacklist-based characteris-

tics. Extraction of comprehensive features from data acquired from four separate sources namely, phishing sites, suspicious sites, legitimate sites, and spoofed web improves detection accuracy.

[173] propose a novel technique for phishing detection using plain text and domain-specific word embedding from the HTML source code. To evaluate their model, they used several word embeddings by utilising ensemble and multimodal approaches. The proposed approach, however, depends solely on the website content, and might not work if the content is changed to images.

To detect PSHCD, [142] suggests the incorporation of similarity-based features in addition to a search engine-based approach. An accuracy of 98.61% is achieved but the technique fails when PSHCD which are indexed in search engines are encountered. Also, to minimise false positives, the similarity threshold is kept at 0. Another approach is proposed in [141], where PSHCD are determined by calculating the similarity score between the login and home page of the suspicious website using the Jaccard similarity coefficient. Other phishing websites are detected through hyperlinks and URL-based features with the TWSVM classifier. The accuracy achieved is 98.05%. The approach fails to detect phishing sites having the same login and home page

[59] trains a machine learning classifier with static and site popularity features extracted from the URL to present an efficient technique to detect mobile phishing. A detection accuracy of 93.85% is obtained through the Random Forest algorithm. [179] integrates the URL-based and text-based approaches to detect smishing SMS. The machine learning classifiers for both approaches are merged by a voting classifier to achieve an accuracy of 99.03%.

Visual Similarity-based Approaches: An improved approach of [162] is proposed in [182]. At the first level, similarity-based features (tag attributes, scripts, paths, filenames etc.) are used to establish similarity with a blacklisted site. To detect non-blacklisted or previously unseen phishing sites the authors have implemented a second-level heuristic filter. An ensemble machine learning model is trained with URL and source code-based features to obtain an accuracy of 98.72%. [181] develop a technique for detecting 'very similar' and 'locally similar' phishing websites. For 'very similar', the wHash (wavelet Hashing) process with the colour histogram has proved to be accurate and stable. For 'local similar', the SIFT (Scale-Invariant Feature Transform) approach is chosen. A cache is also included to shorten the detection time. In [183], visual renderings of target brand logos are extracted by HOG (Histogram of Oriented Gradients) in a scale-invariant manner. Further, an SVM classifier is used to reduce false positives. The technique was able to achieve 93.50% precision and 77.94% recall score. The approach is limited to already learnt logos.

Table 8 presents a tabular analysis of various phishing countermeasures.

7 Challenges, future scope and conclusion

The phishers use various distribution mediums to share malicious links with the victims. After a detailed analysis in Sect. 5, it is observed that along with e-mails, the use of other distribution mediums such as mobiles, IoT etc. is also skyrocketing. The adversaries are focusing on targeted attacks as they are more likely to succeed and provide higher returns. Furthermore, apart from relying on the victims' ignorance or unpreparedness, the phishers are creating opportunities for themselves by not only trying to find system vulnerabilities but also making themselves technically sound.

In this paper, a detailed anatomy of phishing is presented, which explores its multiple related genres namely, the motivations, case studies, mediums of circulation, intended recipients, attack techniques and countermeasure approaches. Distinct surveys in the domain of phishing are reviewed and compared with this survey. The paper emphasises the certitude that before beginning with the design of a phishing detection technique, understanding the different aspects of phishing, such as circulation mediums, targeted victims, intentions behind the attack and the attack technique being used is more important. The researchers can develop efficient solutions with high precision if they have knowledge about the different types of phishing attacks that they are dealing with.

This paper identifies three broad categories of phishing attacks, i.e. phishing on the basis of medium, phishing on the basis of intended targets and phishing on the basis of technique. All these categories are discussed in detail along with supporting reports from eminent players in the domain of cyber-security and phishing. Even though the main focus of this survey is on categorising phishing attacks and discussing attack techniques, we have also weighed upon phishing countermeasures and various phishing detection approaches that have been proposed by the researchers. The benefits and drawbacks of each of the phishing detection approaches are mentioned as well.

Some open research challenges in the current anti-phishing scenario have been identified that need to be addressed:

- The list-based approaches are easy to deploy, but they are dependent on a database, which needs to be updated frequently, and are unable to detect zero-day attacks. These approaches also fail when different versions of the same phishing site are encountered.
- The heuristics-based approach can be bypassed if the attacker gets to know about the detection algorithm and features being used.

Table 8 Analysis of proposed phishing detection techniques

Technique	Approach	Dataset	Evaluation metrics (%)	Remarks
[164]	Blacklist,URL-based Heuristic	Phishing-PhishTank, SpamScatter Legitimate- DMOZ, Yahoo	False Positives-3%, False Negatives-5%	Third-party dependent. Aims only for variants of blacklists
[162]	Blacklist, Source Code-based Heuristic	Phishing-PhishTank Legitimate-Alexa	84.36% of Phishing sites as replicas	Fails when phishing page content is replaced by a screenshot
[192]	Update blacklist by URL tracking	Phishtank (URLs posing as Paypal)	91.97% of Phishing URLs	Focus only on blacklist updation
[166]	Whitelist, hyperlink features	Phishing-PhishTank Legitimate-Alexa, Stuffgate, Wikipedia	Accuracy-89.38 True Positives-86.02 False Negatives-1.48	Third-party dependent, Small dataset, Fails when phishing web page has all hyperlinks for a common local page
[174]	Search engine-based Heuristic	Phishing-PhishTank Legitimate-Alexa	Accuracy-95.95 True Positive-99.5	Approach gives high False Positives for non-English web pages and newly launched legitimate sites
[170]	Heuristic, URL-based features to train ML Classifiers	Phishing-PhishTank Legitimate-DMOZ, Yahoo	Accuracy-91.28	Unable to detect PSHCD
[142]	Search engine-based Heuristic, similarity score comparison	Phishing-PhishTank Legitimate-Alexa	Accuracy-98.61 True positives-97.77	Third party dependent, Similarity threshold as 0 may lead to increase in False Negatives
[169]	Heuristic, NLP-based features from URL to train ML Classifiers	PhishTank and Yandex used to construct own dataset [188]	Accuracy-97.98	Fails for small domains and subdomains without any path
[139]	Heuristic, Hyperlink-based features to train ML classifier	Phishing-PhishTank Legitimate-Alexa, Stuffgate, Wikipedia	Accuracy-98.42 True Positives-98.39 Precision-98.8	The approach fails if the page source references (CSS, favicon, images etc) are altered
[176]	Heuristic with search engine based	Phishing-PhishTank, OpenPhish Legitimate- Alexa	True Positive-98.15	Unable to detect PSHCD
[182]	Blacklisted visual similarity, URL and source code-based features to train ensemble ML	Phishing-PhishTank Legitimate-Google Search	Accuracy- 98.63	Phishing websites with a low level of similarity with blacklisted sites can be difficult to detect, Genuine sites are also filtered twice
[183]	Visual Similarity-based with Machine Learning	Phishing-PhishTank, OpenPhish Legitimate-Alexa	Precision-93.5 Recall -77.94	Approach limited only for previously seen logos. Limited Data of snapshots for creating each brand detector
[141]	URL, hyperlink and similarity-based features to train ML classifier	Phishing-PhishTank Legitimate-Alexa	Accuracy-98.05 Recall-98.33	Fails when the same web page is used for login and as homepage
[167]	Whitelist, hyperlink features	Phishing-PhishTank Legitimate-Alexa	Accuracy-96.17 True Positives-95	Third-party dependent
[193]	Blacklist, heuristic and web content-based features with Machine Learning	Phishing-PhishTank Spoofed Web-Millersmiles Legitimate-Relbank	99.3	Some noisy features can cause overfitting
[138]	Heuristic, URl-based lexical features, Machine Learning	ISCXURL-2016 [189]	Accuracy-99.57	Very less features. The approach might not give the desired results in actual scenarios
[173]	Word Embedding with Machine Learning	URLs of Labeled dataset with phishing-5438, legitimate-5076 and HTML source code as text files	Accuracy-99.34	Approach is language dependent, Fails when entire content is replaced with an image

- Keyword-based solutions are language dependent.
- Heuristic approaches based on search engines suffer from latency and are unable to distinguish newly launched legitimate websites. Additionally, the phishing sites hosted on compromised domains (PSHCD) are not classified correctly as they are already indexed with the search engine and can bypass URL-based approaches.
- Heuristic techniques based on hyperlink features fail when the phishing web page has all the hyperlinks pointing to a common domain.
- In phishing sites where the entire content is replaced with an embedded object such as an image, the features cannot be extracted by detection approaches that are based on textual data such as HTML or DOM.
- Visual similarity-based approaches require high storage and computational costs. Moreover, if the fraudster creates a phishing website with a slight reduction in similarity, it results in high false negatives.

The bewildering rise in the number of phishing websites being reported demonstrates that, despite numerous researchers proposing a wide range of anti-phishing methodologies, the attackers always stay one step ahead and find a way to elude the phishing countermeasures. It is also worth mentioning that no single phishing detection approach is sufficient to detect all kinds of phishing attacks.
Based on our analysis of the literature related to phishing, we suggest the *scope of future research directions* in this domain:

- Layered or hybrid phishing detection techniques that are efficient as well as robust and incorporate the benefits of various existing phishing countermeasures should be developed.
- The approach should be such that the attacker is unable to evade the technique.
- The measures should be lightweight and database independent, with the capability of detecting PSHCD and embedded objects without latency.
- The need of the hour is that the researchers and developers can decide on a trade-off between accuracy and computation time depending on the organisational requirements and then settle on a plan of action for the phishing detection approach to be applied.

Author Contributions RG wrote the whole manuscript. MC and NT provided valuable inputs and reviewed the manuscript.

Funding No funding was received for conducting this study and the authors have no financial or proprietary interests in any material discussed in this article.

Data availability Since this work is a survey, no datasets were created or analysed. Hence, data sharing is not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Williams, E.J., Hinds, J., Joinson, A.N.: Exploring susceptibility to phishing in the workplace. Int. J. Hum. Comput. Stud. **120**, 1–13 (2018)
2. Maroofi, S., Korczyński, M., Hölzel, A., Duda, A.: Adoption of email anti-spoofing schemes: a large scale analysis. IEEE Trans. Netw. Serv. Manag. **18**(3), 3184–3196 (2021)
3. Pandey, N., Pal, A., et al.: Impact of digital surge during COVID-19 pandemic: a viewpoint on research and practice. Int. J. Inf. Manag. **55**, 102171 (2020)
4. Beech, F.M..: Covid-19 pushes up internet use 70% and streaming more than 12%, first figures reveal. https://www.forbes.com/sites/markbeech/2020/03/25/covid-19-pushes-up-internet-use-70-streaming-more-than-12-first-figures-reveal/?sh=1e813ced3104, (May 2020). Accessed June 2022
5. Akala, A.: More big employers are talking about permanent work-from-home positions. https://www.cnbc.com/2020/05/01/major-companies-talking-about-permanent-work-from-home-positions.html. Accessed June 2022
6. BBC News. Twitter allows staff to work from home "forever". https://www.bbc.com/news/technology-52628119 (2020). Accessed June 2022
7. APWG. Phishing activity trends report-4th quarter (2022). https://apwg.org/trendsreports/. Accessed July 2023
8. Abroshan, H., Devos, J., Poels, G., Laermans, E.: Covid-19 and phishing: effects of human emotions, behavior, and demographics on the success of phishing attempts during the pandemic. IEEE Access **9**, 121916–121929 (2021)
9. Lallie, H.S., Shepherd, L.A., Nurse, J.R.C., Erola, A., Epiphaniou, G., Maple, C., Bellekens, X.: Cyber security in the age of Covid-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Comput. Secur. **105**, 102248 (2021)
10. BNP Media T. Kelly: How hackers are using Covid-19 to find new phishing victims. https://www.securitymagazine.com/articles/92666-how-hackers-are-using-covid-19-to-find-new-phishing-victims. Accessed June 2022
11. Cision: Phishing in a pandemic: 1 in 4 Americans received a Covid-19 related phishing email. https://www.prnewswire.com/news-releases/phishing-in-a-pandemic-1-in-4-americans-received-a-covid-19-related-phishing-email-301134037.html (2021). Accessed June 2022
12. Security Boulevard: Phishing statistics: the 29 latest phishing stats to know in 2020. https://securityboulevard.com/2020/04/phishing-statistics-the-29-latest-phishing-stats-to-know-in-2020/. Accessed June 2022
13. APWG: Phishing activity trends report-1st quarter 2020. https://docs.apwg.org/reports/apwg_trends_report_q1_2020.pdf?_ga=2.30422460.2018635328.1665064249-1448730527.1654753557&_gl=1*a4rx10*_ga*MTQ0ODczMDUyNy4xNjU0NzUzNTU3*_ga_55RF0RHXSR*MTY2NTA2NDI0OC4xNS4xLjE2NjUwNjQ1MDYuMC4wLjA. Accessed April 2022
14. Stu Sjouwerman: Q1 2020 coronavirus-related phishing email attacks are up 600%. https://blog.knowbe4.com/q1-2020-coronavirus-related-phishing-email-attacks-are-up-600. Accessed 15 Jan 2022

15. FBI: Internet crime report. https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf. Accessed 16 Feb 2022
16. Jakobsson, M., Myers, S.: Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft. Wiley, New York (2006)
17. Ollmann, G.: The phishing guide understanding & preventing phishing attacks. NGS Software Insight Security Research (2004)
18. Ramzan, Z.: Phishing Attacks and Countermeasures, pp. 433–448. Springer, Berlin (2010)
19. Khonji, M., Iraqi, Y., Jones, A.: Phishing detection: a literature survey. IEEE Commun. Surv. Tutor. **15**(4), 2091–2121 (2013)
20. Almomani, A., Gupta, B.B., Atawneh, S., Meulenberg, A., Almomani, E.: A survey of phishing email filtering techniques. IEEE Commun. Surv. Tutor. **15**(4), 2070–2090 (2013)
21. Mohammad, R.M., Thabtah, F., McCluskey, L.: Tutorial and critical analysis of phishing websites methods. Comput. Sci. Rev. **17**, 1–24 (2015)
22. Tewari, A., Jain, A.K., Gupta, B.B.: Recent survey of various defense mechanisms against phishing attacks. J. Inf. Priv. Secur. **12**(1), 3–13 (2016)
23. Varshney, G., Misra, M., Atrey, P.K.: A survey and classification of web phishing detection schemes. Secur. Commun. Netw. **9**(18), 6266–6284 (2016)
24. Aleroud, A., Zhou, L.: Phishing environments, techniques, and countermeasures: A survey. Comput. Secur. **68**, 160–196 (2017)
25. Gupta, B.B., Tewari, A., Jain, A.K., Agrawal, D.P.: Fighting against phishing attacks: state of the art and future challenges. Neural Comput. Appl. **28**(12), 3629–3654 (2017)
26. Dou, Z., Khalil, I., Khreishah, A., Al-Fuqaha, A., Guizani, M.: Systematization of knowledge (SOK): a systematic review of software-based web phishing detection. IEEE Commun. Surv. Tutor. **19**(4), 2797–2819 (2017)
27. Chiew, K.L., Yong, K.S.C., Tan, C.L.: A survey of phishing attacks: their types, vectors and technical approaches. Expert Syst. Appl. **106**, 1–20 (2018)
28. Qabajeh, I., Thabtah, F., Chiclana, F.: A recent review of conventional vs. automated cybersecurity anti-phishing techniques. Comput. Sci. Rev. **29**, 44–55 (2018)
29. Das, A., Baki, S., El Aassal, A., Verma, R., Dunbar, A.: Sok: a comprehensive reexamination of phishing research from the security perspective. IEEE Commun. Surv. Tutor. **22**(1), 671–708 (2019)
30. Akinyelu, A.A.: Machine learning and nature inspired based phishing detection: a literature survey. Int. J. Artif. Intell. Tools **28**(05), 1930002 (2019)
31. Alabdan, R.: Phishing attacks survey: types, vectors, and technical approaches. Future Internet **12**(10), 168 (2020)
32. Gangavarapu, T., Jaidhar, C.D., Chanduka, B.: Applicability of machine learning in spam and phishing email filtering: review and approaches. Artif. Intell. Rev. **53**(7), 5019–5081 (2020)
33. Vijayalakshmi, M., Shalinie, S.M., Yang, M.H., Meenakshi, U.R.: Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions. IET Netw. **9**(5), 235–246 (2020)
34. Lee, J., Lee, Y., Lee, D., Kwon, H., Shin, D.: Classification of attack types and analysis of attack methods for profiling phishing mail attack groups. IEEE Access **9**, 80866–80872 (2021)
35. Alkhalil, Z., Hewage, C., Nawaf, L., Khan, I.: Phishing attacks: a recent comprehensive study and a new anatomy. Front. Comput. Sci. **3**, 563060 (2021)
36. Jain, A.K., Gupta, B.B.: A survey of phishing attack techniques, defence mechanisms and open research challenges. Enterp. Inf. Syst. **16**(4), 527–565 (2022)
37. Al-Qahtani, A.F., Cresci, S.: The COVID-19 scamdemic: a survey of phishing attacks and their countermeasures during COVID-19. IET Inf. Secur. **16**(5), 324–345 (2022)
38. Basit, A., Zafar, M., Liu, X., Javed, A.R., Jalil, Z., Kifayat, K.: A comprehensive survey of ai-enabled phishing attacks detection techniques. Telecommun. Syst. **76**(1), 139–154 (2021)
39. Salloum, S., Gaber, T., Vadera, S., Sharan, K.: A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access (2022)
40. Abdillah, R., Shukur, Z., Mohd, M., Murah, M.Z.: A systematic literature review on phishing classification techniques . IEEE Access (2022)
41. Rekouche, K.: Early phishing. arXiv preprint arXiv:1106.4692 (2011)
42. BBC News: Twitter hack: staff tricked by phone spear-phishing scam. https://www.bbc.com/news/technology-53607374. Accessed Jan 2022
43. Twitter: An update on our security incident. https://blog.twitter.com/en_us/topics/company/2020/an-update-on-our-security-incident (2020). Accessed Jan 2022
44. CNBC: How this scammer used phishing emails to steal over $100 million from google and facebook. https://www.cnbc.com/2019/03/27/phishing-email-scam-stole-100-million-from-facebook-and-google.html (2019). Accessed Jan 2022
45. Reuters: Austria's facc, hit by cyber fraud, fires CEO. https://www.reuters.com/article/us-facc-ceo-idUSKCN0YG0ZF (2016). Accessed Jan 2022
46. SecurityIntelligence: Ibm uncovers global phishing campaign targeting the covid-19 vaccine cold chain. https://securityintelligence.com/posts/ibm-uncovers-global-phishing-covid-19-vaccine-cold-chain/. Accessed Jan 2022
47. Weider, D.Yu., Nargundkar, S., Tiruthani, N.: A phishing vulnerability analysis of web based systems. In: 2008 IEEE Symposium on Computers and Communications, pp. 326–331. IEEE (2008)
48. Nazah, S., Huda, S., Abawajy, J., Hassan, M.M.: Evolution of dark web threat analysis and detection: a systematic approach. IEEE Access **8**, 171796–171819 (2020)
49. Bates, R.A.: Tracking lone wolf terrorists. J. Public Prof. Sociol. **8**(1), 6 (2016)
50. Weimann, G.: Going dark: terrorism on the dark web. Stud. Conf. Terror. **39**(3), 195–206 (2016)
51. E-ISAC and SANS: Analysis of the cyber attack on the Ukrainian power grid. https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2016/05/20081514/E-ISAC_SANS_Ukraine_DUC_5.pdf. Accessed Jan 2022
52. Verizon: Dbir:data breach investigations report. https://www.verizon.com/business/resources/Tcd0/reports/dbir/2022-data-breach-investigations-report-dbir.pdf (2022). Accessed Dec 2022
53. SOPHOS: The state of ransomware 2022. https://assets.sophos.com/X24WTUEQ/at/4zpw59pnkpxxnhfhgj9bxgj9/sophos-state-of-ransomware-2022-wp.pdf. Accessed Oct 2022
54. Hull, G., John, H., Arief, B.: Ransomware deployment methods and analysis: views from a predictive model and human responses. Crime Sci. **8**(1), 1–22 (2019)
55. Damopoulos, D., Kambourakis, G., Gritzalis, S.: From keyloggers to touchloggers: take the rough with the smooth. Comput. Secur. **32**, 102–114 (2013)
56. Statista: Number of smartphone subscriptions worldwide from 2016 to 2021, with forecasts from 2022 to 2027. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ (2022). Accessed June 2022
57. APWG: Phishing activity trends report-2nd quarter 2022. https://docs.apwg.org/reports/apwg_trends_report_q2_2022.pdf?_ga=2.45552807.263073049.1665252062-1448730527.1654753557&_gl=1*14k5jc7*_ga*MTQ0ODczMDUyNy4xNjU0NzUzNTU3*_ga_55RF0RHXSR*MTY2NTI1ODU1NS4xOS4xLjE2NjUyNTg1NTkuMC4wLjA. Accessed April 2022

58. Diksha Goel and Ankit Kumar Jain: Mobile phishing attacks and defence mechanisms: state of art and open research challenges. Comput. Secur. **73**, 519–544 (2018)
59. Jain, A.K., Debnath, N., Jain, A.K.: APuML: an efficient approach to detect mobile phishing webpages using machine learning. Wirel. Pers. Commun. **125**(4), 3227–3248 (2022)
60. Shahriar, H., Klintic, T., Clincy, V., et al.: Mobile phishing attacks and mitigation techniques. J. Inf. Secur. **6**(03), 206 (2015)
61. Felt, A.P., Wagner, D.: Phishing on mobile devices (2011)
62. Business2Community: SMS marketing: texting your way to success. https://www.business2community.com/digital-marketing/sms-marketing-texting-your-way-to-success-02388639. Accessed June 2022
63. Mishra, S., Soni, D.: Smishing detector: a security model to detect smishing through SMS content analysis and URL behavior analysis. Futur. Gener. Comput. Syst. **108**, 803–815 (2020)
64. Jakobsson, M.: The human factor in phishing. Privacy Security of Consumer Information (2007)
65. Singh, H.P., Singh, S., Singh, J., Khan, S.A.: VoIP: state of art for global connectivity—a critical review. J. Netw. Comput. Appl. **37**, 365–379 (2014)
66. Mustafa, H., Wenyuan, X., Sadeghi, A.-R., Schulz, S.: End-to-end detection of caller id spoofing attacks. IEEE Trans. Depend. Secure Comput. **15**(3), 423–436 (2016)
67. DENSO WAVE INCORPORATED. History of QR code. https://www.qrcode.com/en/history/. Accessed Dec 2021
68. Lin, P.-Y., Chen, Y.-H.: High payload secret hiding technology for QR codes. EURASIP J. Image Video Process. **2017**(1), 1–8 (2017)
69. Dabrowski, A., Krombholz, K., Ullrich, J., Weippl, E.R.: QR inception: barcode-in-barcode attacks. In: Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices, pp. 3–10 (2014)
70. Vidas, T., Owusu, E., Wang, S., Zeng, C., Cranor, L.F., Christin, N.: Qrishing: the susceptibility of smartphone users to QR code phishing attacks. In: International Conference on Financial Cryptography and Data Security, pp. 52–69. Springer (2013)
71. Focardi, R., Luccio, F.L., Wahsheh, H.A.M.: Security threats and solutions for two-dimensional barcodes: a comparative study. In: Computer and Network Security Essentials, pp. 207–219. Springer (2018)
72. Verizon: Dbir:data breach investigations report. https://www.verizon.com/business/resources/reports/2020-data-breach-investigations-report.pdf (2020). Accessed Dec 2021
73. PhishLabs: Social media attacks doubled in 2021 according to latest phishlabs report. https://www.phishlabs.com/news/social-media-attacks-doubled-in-2021-according-to-latest-phishlabs-report/ (2022). Accessed 3 Sept 2022
74. Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menczer, F.: Social phishing. Commun. ACM **50**(10), 94–100 (2007)
75. Cisco: Cybersecurity threat trends: phishing, crypto top the list. https://learn-umbrella.cisco.com/ebook-library/2021-cyber-security-threat-trends-phishing-crypto-top-the-list (2021). Acce ssed 3 Sept 2022
76. Avanan: 1h cyber attack report. https://www.avanan.com/hubfs/Content/Collateral/1H-Cyber-Attack-Report.pdf (2021). Accessed 3 Sept 2022
77. Statista: Number of internet and social media users worldwide as of July 2022. https://www.statista.com/statistics/617136/digital-population-worldwide/ (2022). Accessed 3 Sept 2022
78. Statista: 16% of all facebook accounts are fake or duplicates. https://www.statista.com/chart/20685/duplicate-and-false-facebook-accounts/ (2020). Accessed 20 Aug 2022
79. Song, Y., Yang, C., Gu, G.: Who is peeping at your passwords at starbucks? To catch an evil twin access point. In: 2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN), pp. 323–332. IEEE (2010)
80. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
81. Sharma, R., Mahapatra, R. P., Sharma, N.: The internet of things and its applications in cyber security. In: A Handbook of Internet of Things in Biomedical and Cyber Physical System, pp. 87–108 (2020)
82. Nirmal, K., Janet, B., Kumar, R.: Analyzing and eliminating phishing threats in IoT, network and other web applications using iterative intersection. Peer-to-Peer Netw. Appl. **14**, 2327–2339 (2021)
83. Tewari, A., Gupta, B.B.: Security, privacy and trust of different layers in internet-of-things (IoTs) framework. Future Gener. Comput. Syst. **108**, 909–920 (2020)
84. PaloAlto Networks. 2020 unit 42 IoT threat report. https://start.paloaltonetworks.com/unit-42-iot-threat-report. Accessed July 2023
85. Caputo, D.D., Pfleeger, S.L., Freeman, J.D., Johnson, M.E.: Going spear phishing: exploring embedded training and awareness. IEEE Secur. Priv. **12**(1), 28–38 (2013)
86. Parmar, B.: Protecting against spear-phishing. Comput. Fraud Secur. **2012**(1), 8–11 (2012)
87. Wang, J., Herath, T., Chen, R., Vishwanath, A., Rao, H.R.: Research article phishing susceptibility: an investigation into the processing of a targeted spear phishing email. IEEE Trans. Prof. Commun. **55**(4), 345–362 (2012)
88. Symantec: Istr:internet security threat report. https://docs.broadcom.com/doc/istr-24-2019-en (2019). Accessed 23 Apr 2022
89. Kwak, Y., Lee, S., Damiano, A., Vishwanath, A.: Why do users not report spear phishing emails? Telemat. Inf. **48**, 101343 (2020)
90. Al-Musib, N.S., Al-Serhani, F.M., Humayun, M., Jhanjhi, N.Z.: Business email compromise (BEC) attacks. Mater. Today Proc. (2021)
91. FBI. Public service announcement. https://www.ic3.gov/Media/Y2022/PSA220504 (2022). Accessed 23 Apr 2022
92. FBI. Public service announcement. https://www.ic3.gov/Media/Y2022/PSA220216 (2022). Accessed 23 Apr 2022
93. Le Page, S., Jourdan, G.-V.: Victim or attacker? A multi-dataset domain classification of phishing attacks. In: 2019 17th International Conference on Privacy, Security and Trust (PST), pp. 1–10. IEEE (2019)
94. Corona, I., Biggio, B., Contini, M., Piras, L., Corda, R., Mereu, M., Mureddu, G., Ariu, D., Roli, F.: Deltaphish: detecting phishing webpages in compromised websites. In: European Symposium on Research in Computer Security, pp. 370–388. Springer (2017)
95. Moore, T., Clayton, R.: Examining the impact of website take-down on phishing. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, pp. 1–13 (2007)
96. PhishLabs. Most phishing attacks use compromised domains and free hosting. https://www.phishlabs.com/blog/most-phishing-attacks-use-compromised-domains-and-free-hosting/ (2021). Accessed 14 Feb 2022
97. Pope, M.B., Warkentin, M., Mutchler, L.A., Luo, X.R.: The domain name system-past, present, and future. Commun. Assoc. Inf. Syst. **30**(1), 21 (2012)
98. Kim, H., Huh, J.H.: Detecting DNS-poisoning-based phishing attacks from their network performance characteristics. Electron. Lett. **47**(11), 656–658 (2011)
99. Perdisci, R., Antonakakis, M., Luo, X., Lee, W.: WSEC DNS: protecting recursive DNS resolvers from poisoning attacks. In: 2009 IEEE/IFIP International Conference on Dependable Systems & Networks, pp. 3–12. IEEE (2009)

100. Schiller, C.A., Binkley, J., Harley, D., Evron, G., Bradley, T., Willems, C., Cross, M.: Botnets overview. In: Botnets, pp. 29–75. Syngress (2007)
101. Vural, I., Venter, H.: Detecting mobile spam botnets using artificial immune systems. In: IFIP International Conference on Digital Forensics, pp. 183–192. Springer (2011)
102. Negash, N., Che, X.: An overview of modern botnets. Inf. Secur. J. Glob. Perspect. **24**(4–6), 127–132 (2015)
103. Milletary, J., CERT Coordination Center.: Technical trends in phishing attacks. Retrieved December 1(2007):3 (2005)
104. Gupta, S., Gupta, B.B.: Cross-site scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art. Int. J. Syst. Assur. Eng. Manag. **8**(1), 512–530 (2017)
105. Ruderman, J.: The same origin policy. http://www.mozilla.org/projects/security/components/same-origin.html (2001)
106. Patchstack. State of wordpress security in 2021. https://patchstack.com/wp-content/uploads/2022/03/Patchstack-%E2%80%93-State-Of-WordPress-Security-In-2021.pdf (2022). Accessed 15 May 2022
107. Nagar, N., Suman, U.: Prevention, detection, and recovery of CSRF attack in online banking system. In: Online banking security measures and data protection, pp. 172–188. IGI Global (2017)
108. Zhang, J., Hu, H., Huo, S.: A browser-based cross site request forgery detection model. J. Phys. Conf. Ser. **1738**, 012073 (2021)
109. Gelernter, N., Herzberg, A.: Tell me about yourself: the malicious captcha attack. In: Proceedings of the 25th International Conference on World Wide Web, pp. 999–1008 (2016)
110. Yalçın, N., Köse, U.: What is search engine optimization: Seo? Procedia Soc. Behav. Sci. **9**, 487–493 (2010)
111. Chaudhry, J.A., Chaudhry, S.A., Rittenhouse, R.G.: Phishing attacks and defenses. Int. J. Secur. Appl. **10**(1), 247–256 (2016)
112. Nagunwa, T.: Behind identity theft and fraud in cyberspace: the current landscape of phishing vectors. Int. J. Cyber-Secur. Digit. Forensics IJCSDF **3**(1), 72–83 (2014)
113. van der Toorn, O., Müller, M., Dickinson, S., Hesselman, C., Sperotto, A., van Rijswijk-Deij, R.: Addressing the challenges of modern DNS a comprehensive tutorial. Comput. Sci. Rev. **45**, 100469 (2022)
114. Wang, Y.-M., Beck, D., Wang, J., Verbowski, C., Daniels, B.: Strider typo-patrol: discovery and analysis of systematic typo-squatting. SRUTI **6**(31–36), 2–2 (2006)
115. Spaulding, J., Nyang, D., Mohaisen, A.: Understanding the effectiveness of typosquatting techniques. In: Proceedings of the fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies, pp. 1–8 (2017)
116. Moore, T., Edelman, B.: Measuring the perpetrators and funders of typosquatting. In: International Conference on Financial Cryptography and Data Security, pp. 175–191. Springer (2010)
117. Dinaburg, A.: Bitsquatting: Dns hijacking without exploitation (2011)
118. Nikiforakis, N., Balduzzi, M., Desmet, L., Piessens, F., Joosen, W.: Soundsquatting: uncovering the use of homophones in domain squatting. In: International Conference on Information Security, pp. 291–308. Springer (2014)
119. Holgers, T., Watson, D.E., Gribble, S.D.: Cutting through the confusion: a measurement study of homograph attacks. In: USENIX Annual Technical Conference, General Track, pp. 261–266 (2006)
120. Kintis, P., Miramirkhani, N., Lever, C., Chen, Y., Romero-Gómez, R., Pitropakis, N., Nikiforakis, N., Antonakakis, M.: Hiding in plain sight: a longitudinal study of combosquatting abuse. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 569–586 (2017)
121. Zeng, Y., Zang, T., Zhang, Y., Chen, X., Wang, Y.: A comprehensive measurement study of domain-squatting abuse. In: ICC 2019-2019 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2019)
122. Nikiforakis, N., Van Acker, S., Meert, W., Desmet, L., Piessens, F., Joosen, W.: Bitsquatting: exploiting bit-flips for fun, or profit? In: Proceedings of the 22nd international conference on World Wide Web, pp. 989–998 (2013)
123. Rader, M., Rahman, S.: Exploring historical and emerging phishing techniques and mitigating the associated security risks. arXiv preprint arXiv:1512.00082 (2015)
124. Skolka, P., Staicu, C.-A., Pradel, M.: Anything to hide? Studying minified and obfuscated code in the web. In: The World Wide Web Conference, pp. 1735–1746 (2019)
125. APWG. Phishing activity trends report-4th quarter 2020. https://docs.apwg.org/reports/apwg_trends_report_q4_2020.pdf?_ga=2.52213802.263073049.1665252062-1448730527.1654753557&_gl=1*1imdh26*_ga*MTQ0ODczMDUyNy4xNjU0NzUzNTU3*_ga_55RF0RHXSR*MTY2NTI1MjA2MS4xOC4wLjE2NjUyNTIzNTMuMC4wLjA (2021). Accessed April 2022
126. Sarker, S., Jueckstock, J., Kapravelos, A.: Hiding in plain site: detecting javascript obfuscation through concealed browser api usage. In: Proceedings of the ACM Internet Measurement Conference, pp. 648–661 (2020)
127. Romano, A., Lehmann, D., Pradel, M., Wang, W.: Wobfuscator: Obfuscating javascript malware via opportunistic translation to webassembly. In: Proceedings of the 2022 IEEE Symposium on Security and Privacy (S&P 2022), pp. 1101–1116 (2022)
128. Bagchi, K., Udo, G.: An analysis of the growth of computer and internet security breaches. Commun. Assoc. Inf. Syst. **12**(1), 46 (2003)
129. Loughran, D.T., Salih, M.K., Subburaj, V.H.: All about SQL injection attacks. J. Colloq. Inf. Syst. Secur. Educ. **6**, 24–24 (2018)
130. Patil, D.R., Patil, J.B.: Survey on malicious web pages detection techniques. Int. J. u-and e-Serv. Sci. Technol. **8**(5), 195–206 (2015)
131. Conti, M., Dragoni, N., Lesyk, V.: A survey of man in the middle attacks. IEEE Commun. Surv. Tutor. **18**(3), 2027–2051 (2016)
132. Sahani, R., Randhawa, S.: Clickjacking: Beware of clicking. Wirel. Pers. Commun. **121**(4), 2845–2855 (2021)
133. Shahriar, H., Devendran, V.K.: Classification of clickjacking attacks and detection techniques. Inf. Secur. J. A Glob. Perspect. **23**(4–6), 137–147 (2014)
134. Sinha, R., Uppal, D., Singh, D., Rathi, R.: Clickjacking: existing defenses and some novel approaches. In: 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014), pp 396–401. IEEE (2014)
135. Akhawe, D., He, W., Li, Z., Moazzezi, R., Song, D.: Clickjacking revisited: a perceptual view of {UI} security. In: 8th USENIX workshop on offensive technologies (WOOT 14) (2014)
136. Stone, P.: Next generation clickjacking. BlackHat Europe (2010)
137. Huang, L.-S., Moshchuk, A., Wang, H.J., Schecter, S., Jackson, C.: Clickjacking: attacks and defenses. In: 21st USENIX Security Symposium (USENIX Security 12), pp. 413–428 (2012)
138. Gupta, B.B., Yadav, K., Razzak, I., Psannis, K., Castiglione, A., Chang, X.: A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment. Comput. Commun. **175**, 47–57 (2021)
139. Jain, A.K., Gupta, B.B.: A machine learning based approach for phishing detection using hyperlinks information. J. Amb. Intell. Hum. Comput. **10**(5), 2015–2028 (2019)
140. Ramana, A.V., Rao, K.L., Rao, R.S.: Stop-phish: an intelligent phishing detection method using feature selection ensemble. Soc. Netw. Anal. Min. **11**(1), 1–9 (2021)
141. Rao, R.S., Pais, A.R., Anand, P.: A heuristic technique to detect phishing websites using TWSVM classifier. Neural Comput. Appl. **33**(11), 5733–5752 (2021)
142. Rao, R.S., Pais, A.R.: Jail-phish: an improved search engine based phishing detection system. Comput. Secur. **83**, 246–267 (2019)

143. Suri, R.K., Tomar, D.S., Sahu, D.R.: An approach to perceive tabnabbing attack. Int. J. Sci. Technol. Res. **1**(6), 90–94 (2012)
144. Raskin, A.: Tabnabbing: a new type of phishing attack. línea. http://www.azarask.in/blog/post/a-new-type-of-phishing-attack/. [Último acceso: 10 12 2013] (2010)
145. Kolsek, M.: Session fixation vulnerability in web-based applications. ACROS Security. http://www.acrossecurity.com/papers/sessionfixation.pdf (2002)
146. Kaspersky. Phishing-kit market: what's inside "off-the-shelf" phishing packages. https://securelist.com/phishing-kit-market-whats-inside-off-the-shelf-phishing-packages/106149/ (2022). Accessed 25 Aug 2022
147. Kaspersky. How scammers are creating thousands of fake pages using phishing kits. https://usa.kaspersky.com/about/press-releases/2022_quick-cheap-and-dangerous-how-scammers-are-creating-thousands-of-fake-pages-using-phishing-kits (2022). Accessed 25 Aug 2022
148. Bahnsen, A.C., Torroledo, I., Camacho, L.D., Villegas, S.: Deepphish: simulating malicious AI. In: 2018 APWG Symposium on Electronic Crime Research (eCrime), pp. 1–8 (2018)
149. Group-IB. https://www.group-ib.com/media-center/press-releases/phishing-kits-2022/ (2023). Accessed Sep 2023
150. CNBC. https://www.cnbc.com/2023/01/07/phishing-attacks-are-increasing-and-getting-more-sophisticated.html (2023). Accessed July 2023
151. Dupuis, M., Geiger, T., Slayton, M., Dewing, F.: The use and non-use of cybersecurity tools among consumers: do they want help? In: Proceedings of the 20th Annual SIG Conference on Information Technology Education, pp. 81–86 (2019)
152. SECTRIO. https://sectrio.com/iot-security-reports/2023-ot-iot-threat-landscape-report/ (2023). Accessed July 2023
153. Proofpoint. https://www.proofpoint.com/sites/default/files/threat-reports/pfpt-us-tr-state-of-the-phish-2023.pdf (2023). Accessed Sep 2023
154. Check Point. https://www.avanan.com/hubfs/2022-Defender-Report/WP_Avanan_Keeping_Your_Emails_Secure_Who_Does_It_Best.pdf (2022). Accessed Sep 2023
155. Abnormal Society. https://cdn2.assets-servd.host/gifted-zorilla/production/files/Read-Alert-Data-Shows-28-of-BEC-Attacks-Opened-by-Employees.pdf?dm=1675457683 (2023). Accessed Sep 2023
156. AKAMAI. https://www.akamai.com/blog/security/over-25-percent-of-malicious-javascript-is-being-obfuscated (2021). Accessed July 2023
157. AKAMAI. https://www.akamai.com/blog/security/web-application-and-api-protection-from-sql-injection-to-magecart (2020). Accessed July 2023
158. COMPARITECH. https://www.comparitech.com/blog/information-security/botnet-statistics/ (2022). Accessed July 2023
159. ZedNET. https://www.zdnet.com/article/clickjacking-scripts-found-on-613-popular-sites-academics-say// (2019). Accessed July 2023
160. F5 Labs. https://www.f5.com/content/dam/f5-labs-v2/article/articles/threats/22--2020-oct-dec/20201110_2020_phishing_report/F5Labs-2020-Phishing-and-Fraud-Report.pdf (2020). Accessed July 2023
161. CheckPoint. https://blog.checkpoint.com/security/social-networks-most-likely-to-be-imitated-by-criminal-groups-with-linkedin-now-accounting-for-half-of-all-phishing-attempts-worldwide/ (2022). Accessed July 2023
162. Rao, R.S., Pais, A.R.: An enhanced blacklist method to detect phishing websites. In: International Conference on Information Systems Security, pp. 323–333. Springer (2017)
163. Bell, S., Komisarczuk, P.: An analysis of phishing blacklists: Google safe browsing, openphish, and phishtank. In: Proceedings of the Australasian Computer Science Week Multiconference, pp. 1–11 (2020)
164. Prakash, P., Kumar, M., Kompella, R R., Gupta, M.: Phishnet: predictive blacklisting to detect phishing attacks. In: 2010 Proceedings IEEE INFOCOM, pp. 1–5. IEEE (2010)
165. Han, W., Cao, Y., Bertino, E., Yong, J.: Using automated individual white-list to protect web digital identities. Expert Syst. Appl. **39**(15), 11861–11869 (2012)
166. Jain, A.K., Gupta, B.B.: A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURASIP J. Inf. Secur. **2016**(1), 1–11 (2016)
167. Azeez, N.A., Misra, S., Margaret, I.A., Fernandez-Sanz, L., et al.: Adopting automated whitelist approach for detecting phishing attacks. Comput. Secur. **108**, 102328 (2021)
168. Sheng, S., Wardman, B., Warner, G., Hong, J., Zhang, C.: An empirical analysis of phishing blacklists. Lorrie Cranor (2009)
169. Sahingoz, O.K., Buber, E., Demir, O., Diri, B.: Machine learning based phishing detection from URLs. Expert Syst. Appl. **117**, 345–357 (2019)
170. Jain, A.K., Gupta, B.B.: Phish-safe: URL features-based phishing detection system using machine learning. In: Cyber Security, pp. 467–474. Springer (2018)
171. Ali, M.S., Jain, A.K.: Efficient feature selection approach for detection of phishing URL of Covid-19 era. In: International Conference on Cyber Security, Privacy and Networking, pp. 45–56. Springer (2021)
172. Jain, A.K., Parashar, S., Katare, P., Sharma, I.: Phishskape: a content based approach to escape phishing attacks. Procedia Comput. Sci. **171**, 1102–1109 (2020)
173. Rao, R.S., Umarekar, A., Pais, A.R.: Application of word embedding and machine learning in detecting phishing websites. Telecommun. Syst. 1–13 (2022)
174. Varshney, G., Misra, M., Atrey, P.K.: A phish detector using lightweight search features. Comput. Secur. **62**, 213–228 (2016)
175. Jain, A.K., Gupta, B.B.: Two-level authentication approach to protect from phishing attacks in real time. J. Amb. Intell. Hum. Comput. **9**(6), 1783–1796 (2018)
176. Gupta, B.B., Jain, A.K.: Phishing attack detection using a search engine and heuristics-based technique. J. Inf. Technol. Res. JITR **13**(2), 94–109 (2020)
177. Jain, A.K., Gupta, B.B.: Towards detection of phishing websites on client-side using machine learning based approach. Telecommun. Syst. **68**(4), 687–700 (2018)
178. Rao, R.S., Pais, A.R.: Detection of phishing websites using an efficient feature-based machine learning framework. Neural Comput. Appl. **31**(8), 3851–3873 (2019)
179. Jain, A.K., Gupta, B.B., Kaur, K., Bhutani, P., Alhalabi, W., Almomani, A.: A content and URL analysis-based efficient approach to detect smishing SMS in intelligent systems. Int. J. Intell. Syst. **37**(12), 11117–11141 (2022)
180. Mao, J., Tian, W., Li, P., Wei, T., Liang, Z.: Phishing-alarm: robust and efficient phishing detection via page component similarity. IEEE Access **5**, 17020–17030 (2017)
181. Chen, J.-L., Ma, Y.-W., Huang, K.-L.: Intelligent visual similarity-based phishing websites detection. Symmetry **12**(10), 1681 (2020)
182. Routhu Srinivasa Rao and Alwyn Roshan Pais: Two level filtering mechanism to detect phishing sites using lightweight visual similarity approach. J. Ambient. Intell. Humaniz. Comput. **11**(9), 3853–3872 (2020)
183. Ahmet Selman Bozkir and Murat Aydos: Logosense: a companion hog based logo detection scheme for phishing web page and e-mail brand recognition. Comput. Secur. **95**, 101855 (2020)
184. Phishtank. https://phishtank.org/. Accessed July 2023
185. Ahrefs. https://ahrefs.com/. Accessed July 2023
186. Similarweb. https://www.similarweb.com/. Accessed July 2023

187. Majestic million. https://majestic.com/reports/majestic-million. Accessed July 2023
188. https://github.com/ebubekirbbr/pdd/tree/master/input. Accessed July 2023
189. https://www.unb.ca/cic/datasets/url-2016.html. Accessed July 2023
190. Vrbančič, G.: Phishing websites dataset. Mendeley Data (2020)
191. Jpcert/cc. https://github.com/JPCERTCC/phishurl-list/. Accessed July 2023
192. Lee, L.-H., Lee, K.-C., Chen, H.-H., Tseng, Y.-H.: Poster: Proactive blacklist update for anti-phishing. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 1448–1450 (2014)
193. Barraclough, P.A., Fehringer, G., Woodward, J.: Intelligent cyber-phishing detection for online. Comput. Secur. **104**, 102123 (2021)
194. Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S.B., Joga, S.R.K.: Phishing detection system through hybrid machine learning based on URL. IEEE Access **11**, 36805–36822 (2023)