A literature review on classification of phishing attacks

Article *in* International Journal of Advanced Technology and Engineering Exploration · April 2022

DOI: 10.19101/IJATEE.2021.875031

2 authors:

Chanti Surya prakasam
Pondicherry University
9 PUBLICATIONS **41** CITATIONS
SEE PROFILE

T. Chithralekha
Pondicherry University
35 PUBLICATIONS **267** CITATIONS
SEE PROFILE

International Journal of Advanced Technology and Engineering Exploration, Vol 9(89)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.875031

Review Article

A literature review on classification of phishing attacks

S. Chanti[1*] and T. Chithralekha[2]
Research Scholar, Department of Banking Technology, Pondicherry University, Puducherry[1]
Professor, Department of Computer Science, Pondicherry University, Puducherry[2]

Received: 15-October-2021; Revised: 25-April-2022; Accepted: 27-April-2022

Abstract
Phishing is a type of security threat that loots users' personal credentials such as online banking, credit card numbers, card verification value (CVV) numbers, automated teller machine (ATM) pins. Phishing scams are done by sending spoofed emails, instant messaging that carry hyperlinks that redirect the users to fake/spoofed sites, and steal their sensitive information. Phishers mainly concentrate on internet users who perform E-banking. Since these E-transactions are inevitable in today's digital world, many anti-phishing tools are developed to secure the user from phishing attacks. This paper proposes a new definition of phishing based on the intention of phishing and a complete classification of phishing attacks starting the email phishing to the very recent ransomware. This literature provides the classification of phishing attacks and the different possible ways the attacker targets the victims. A statistical analysis on phishing attacks is performed using the data collected from anti-phishing working group (APWG) technical reports to find: (i) top three countries hosting phishing, (ii) top three most affected countries hosting phishing, (iii) top three least affected countries, (iv) top three industry sectors affected by phishing, (v) top three malware used for phishing, and (vi) hypertext transfer protocol secure (HTTPS) enabled phishing uniform resource locator (URL). This study is helpful in understanding the different ways of performing phishing attacks.

Keywords
Phishing, Pharming, Vishing, Ransomware, DNS level phishing, Credential stealing, Social engineering phishing, Malware based phishing, User information control, Domain hijacking, DNS spoofing.

1.Introduction
Phishing is a fraudulent activity through which the phisher/attacker tries to lure internet users into stealing their personal information/credentials to gain financially [1]. Phishing is comparable to fishing, which gives a different meaning; in phishing, the attacker uses bait (sending an email with an embedded hyperlink that redirects to a fraudulent site) to capture the internet users' credentials. Earlier, the hackers were known as Phreaks (a Phreak is someone who breaks into the telephone networks illegally to make free long-distance phone calls or to tap phone lines) and are closely related to each other. The reason for replacing *"f"* with *"ph"* is to link the phishing scams with phreaks [2, 3]. For the past two decades, phishing has become the most dangerous attack, and plenty of attack is taking place every day [4–6]. The first phishing scam was recorded on 2nd Jan 1996 on American online (AOL) that provides internet services [2].

*Author for correspondence

The phisher randomly generates credit card numbers, and by using those credit card numbers, they created AOL accounts.

Later, by applying the AOL instant messengers or email system, they send the email to the customers asking them to verify their account details by clicking on the embedded hyperlink provided in the email. If the user clicks on the hyperlink and enters the credentials, this information is automatically transferred to the attacker. Therefore, the attackers make use of those credentials for fraudulent activities.

Every time the attacker comes with a different technique to fool the internet user to steal their personal credentials (for example, bank account details, E-banking account details, social media accounts, email accounts, etc.). In anti-phishing working group (APWG) 2nd quarter 2021[7], for the month of June 2021 alone, the number of unique phishing attacks reached 2,22,127. *Figure 1* shows

the unique phishing sites developed by the phishers during 2020-2021. Pharming is a refined form of phishing i.e., "phishing without lure" [8, 9]. In a phishing attack, the attacker sends a spoofed email with an embedded hyperlink to the individual users. Hence the users are redirected to fraudulent sites that deceptively steal user credentials. However, in pharming, the attacker tries to redirect the users to a fraudulent site that looks very similar to the original site by poisoning the domain name system (DNS).

Thus, pharming is a DNS-based attack where the attacker gains unauthorized access to the DNS and modifies the host file records through which all the users who access the information from that DNS are redirected to the fraudulent website. It is challenging to detect and is more dangerous as it impacts many users who are victimized by DNS poisoning. More recently, phishing is also found to be done by usurping control of access to user information by means of malware (ransomware) and blackmailing the users to pay some ransom [4, 10]. In 2015, internet crime report (ICR) received 2,453 complaints that were identified as ransomware with a loss of $1.6 million [11]. According to the APWG report (1st quarter 2019) [12], 1,80,768 unique phishing websites were detected. Among that, 36% of phishing scams are on software-as-a-service (SaaS) and webmail services. To secure users from phishing attacks, various researchers and organizations have developed a lot of anti-phishing tools. These anti-phishing tools mostly work at the user end, and only a few works on the server-side. In our previous work, a complete classification of anti-phishing solutions is presented [13]. The main objective of this paper is to provide a complete classification of phishing attacks.

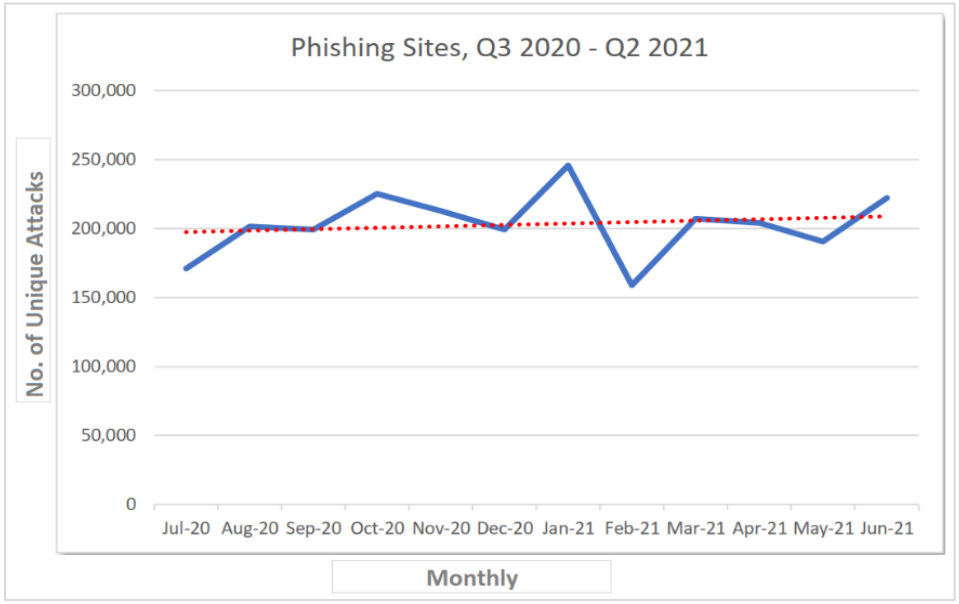

Figure 1 Unique phishing sites detected during 2020-2021 in an APWG survey report [7]

This paper explains phishing and pharming attacks, proposed a new definition of phishing, and provides a complete classification of phishing as follows. Section-2 provides the selection process of relevant literature. Section-3 is a brief introduction of phishing. In section-4, the motivation of phishing classification is presented. Section-5 provides the classification of phishing attacks based on the intention of phishing. Section-6 provides the phishing statistics based on APWG technical reports. Section-7 discusses the major challenges faced in detecting phishing attacks. Section-8 is about the discussion of the paper. Section-9 is the conclusion of the work.

2.Research methodology

2.1Research questions

The primary goal of this paper is to provide a comprehensive classification of phishing attacks. To accomplish this, we formulate the following study questions:

RQ1. There are so many definitions of phishing given in various sources. What would be the concise definition of phishing, which encompasses the semantics of most of the definitions?

RQ2. What could be the possible classification of all the phishing attacks starting from the oldest to the most recent phishing attacks?

RQ3. What are the current statistics on phishing attacks with respect to the impact on different countries, most prevalent kind of phishing attack, most targeted industry sectors, etc., around the globe?
RQ4. What are the major challenges to be addressed in phishing attacks?.

The answers to all the research questions are provided in section-8.

2.2Selection of relevant literature

The primary focus of this paper is on the complete classification of phishing attacks. A preliminary search was done to find the papers related to phishing attacks from various digital sources like Springer, ScienceDirect, IEEE Xplore, Emeralds, and others. The keywords like phishing attacks, types of phishing attacks, social media phishing, malware-based phishing, and social media phishing are used to find the relevant papers. The first level filtering is done based on the title and abstract of the papers. Later, the final set of papers was identified by studying the complete articles. The screening process found that the number of papers under this study's scope was very minimal as shown in *Figure 2.* Year-wise publication of of journal articles, books, reports, thesis, and conference proceedings are depicted in *Figure 3.*

To cover most of the phishing attacks in our classification, we try to gather the relevant literature from various sources like books, thesis, technical reports, websites, and blogs that are yet to be provided in the research articles. The websites and blogs referred to in this work are from very authentic sources and provide legitimate information. The website sources are used for multiple purposes to identify the new/recent phishing attacks, real-time phishing scams reported globally, technical reports, and datasets to analyze the impact of phishing attacks and their growth. *Figure 4* shows the various sources from there the relevant literature was collected. Reliable and trusted website references are cited in our paper for gathering the recent phishing attacks, definition of phishing and ransomware attacks. As we proposed a new definition of phishing, it is required to compare various other definitions of phishing to justify the proposed definition of phishing. *Table 1* contains the final set of references that have been considered for review.

Table 1 Selected paper for the review

Reference type	Author	Reference type	Author	Reference type	Author
Conference	[1]	Website	[34]	Journal	[67]
Website	[2]	Website	[35]	Book	[68]
Report	[3]	Website	[36]	Conference	[69]
Website	[4]	Report	[37]	Journal	[70]
Conference	[5]	Website	[38]	Conference	[71]
Report	[6]	Journal	[39]	Thesis	[72]
Report	[7]	Report	[40]	Conference	[73]
Journal	[8]	Conference	[41]	Conference	[74]
Report	[9]	Report	[42]	Journal	[75]
Report	[10]	Conference	[43]	Conference	[76]
Report	[11]	Website	[44]	Journal	[77]
Report	[12]	Website	[45]	Conference	[78]
Journal	[13]	Journal	[46]	Journal	[79]
Conference	[14]	Conference	[47]	Conference	[80]
Conference	[15]	Journal	[48]	Journal	[81]
Website	[16]	Journal	[49]	Conference	[82]
Conference	[17]	Book Chapter	[50]	Journal	[83]
Website	[18]	Journal	[51]	Book	[84]
Journal	[19]	Conference	[52]	Conference	[85]
Journal	[20]	Conference	[53]	Thesis	[86]
Journal	[21]	Conference	[54]	Conference	[87]
Journal	[22]	Journal	[55]	Report	[88]
Journal	[23]	Thesis	[56]	Journal	[89]
Journal	[24]	Website	[57]	Journal	[90]
Journal	[25]	Journal	[58]	Website	[91]
Conference	[26]	Journal	[59]	Dataset	[92]
Journal	[27]	Conference	[60]	Dataset	[93]
Journal	[28]	Report	[61]	Dataset	[94]

Reference type	Author	Reference type	Author	Reference type	Author
Journal	[29]	Journal	[62]	Journal	[95]
Website	[30]	Conference	[63]	Journal	[96]
Website	[31]	Journal	[64]	Website	[97]
Journal	[32]	Conference	[65]	Website	[98]
Website	[33]	Conference	[66]	Journal	[99]
xxxx	xxxx	xxxx	xxxx	Journal	[100]
xxxx	xxxx	xxxx	xxxx	Report	[101]

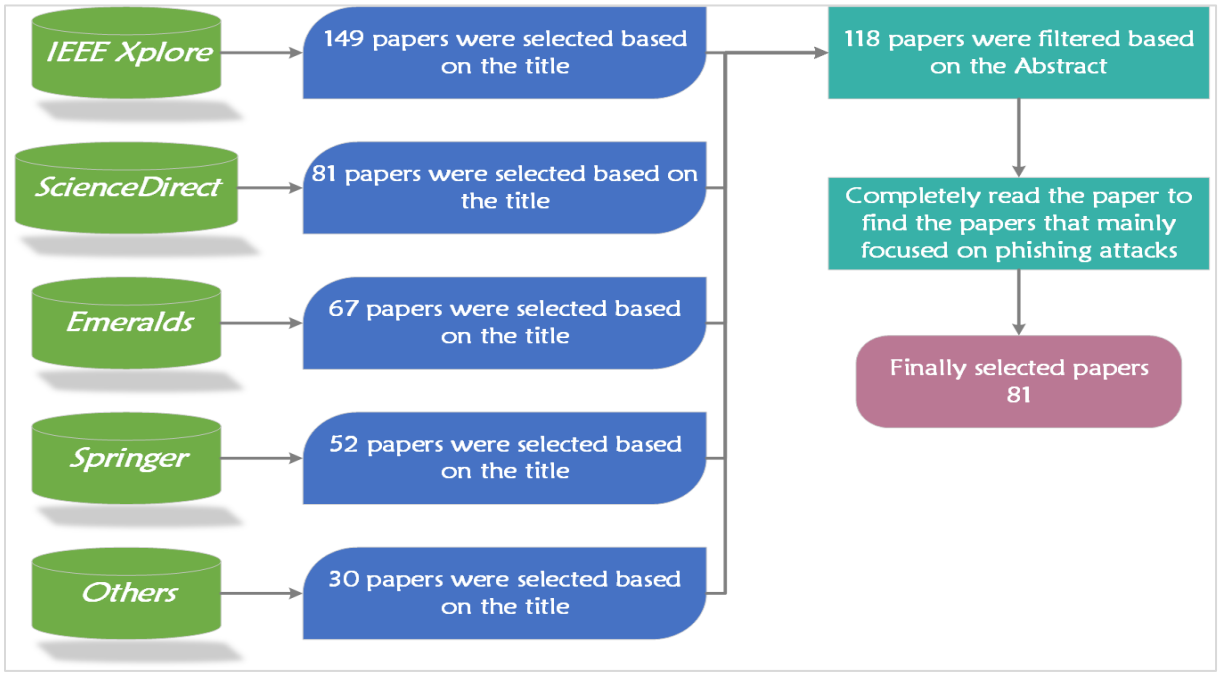

Figure 2 Prisma style representation of journal articles, books, reports, thesis, and conference proceedings

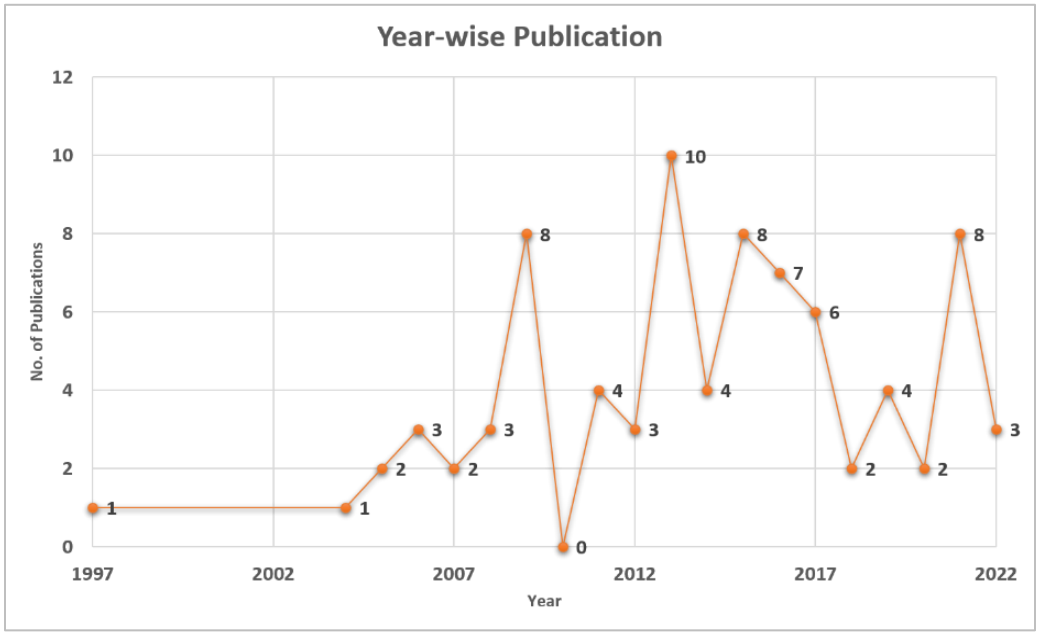

Figure 3 Year-wise publication of journal articles, books, reports, thesis, and conference proceedings

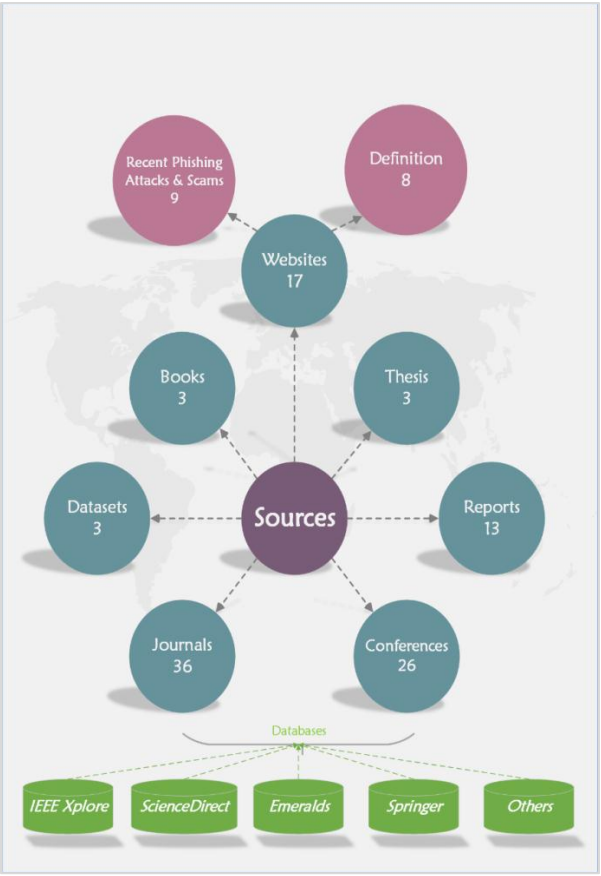

Figure 4 Different sources to obtain relevant literature

3.Phishing

Phishing is a form of online identity theft [1, 14, 15] which is exploited by the phisher for obtaining monetary gains. The users' online identity comprises username, password (example for online banking), credit/debit card number, card's pin, card verification value (CVV) number, and the validity of the credit/debit card. The compromised online identity details will help the phisher to perpetuate an online financial theft by spoofing the captured user's identity. The phisher steals the identity of the users either by calling over the phone and asking for online banking identity details (pretending to be a bank employee) or by sending a spoofed email with a malicious link. The link can either redirect the user to a malicious site or a malware attachment that downloads and installs on the user's machine and acts as a Trojan by divulging users' credentials to the phisher when the user performs online transactions. While this phishing attacks target users individually through emails, mass phishing could also be done by DNS poisoning. Here the poisoned DNS entry can redirect all users whose uniform resource locator (URL) resolution is performed through the poisoned DNS to unintended malicious sites. Thereby, users could be hijacked to the malicious, but seemingly legitimate site(s) (spoofed or fake online banking sites) or to unintended websites where the user's identity details are captured deceptively. A complete classification of phishing attacks is given in the upcoming section.

The real-time phishing attack incidents reported by the cybersecurity team for the past few years have been observed [8], [16–18]. From these attacks, it is observed that more than $280 million are lured from the victims.

4.Motivation for the proposed classification

The motivation for the proposed classification derives by considering the existing available classifications [19–22]. It is observed that most of the existing phishing classifications focus discretely on either phishing attack techniques used by phisher or prevention methods or the new phishing attack techniques and their prevention methods [23–29]. However, a comprehensive classification comprising of all the phishing attacks is not available in the literature. Also, most of the existing classifications ignore the impact of phishing on various sectors and the loss/damage caused due to phishing. In this paper, a new definition of phishing and complete classification of phishing attacks with a real-time incident of every attack mentioned in the classification is presented. It also provides a statistical analysis of phishing attacks. A comparison of existing classifications with the proposed classification is presented in *Table 2*.

5.Proposed classification of phishing based on the intention of phishing

The phishing classification is performed based on the intention of phishing. In order to achieve the same, the various definitions of phishing and ransomware are considered. From these definitions, a consolidated definition of phishing is obtained. This consolidated definition is the basis for the classification of phishing attacks. The existing definitions of phishing and ransomware are given below:

"*Phishing is a type of computer attack that communicates socially engineered messages to humans via electronic communication channels in*

order to persuade them to perform certain actions for the attacker's benefit "[28].

"*The activity of tricking people by getting them to give their identity, bank account numbers, etc. over the Internet or by e-mail, and then using these to steal money from them* " [30].

"*Phishing is a fraudulent attempt, usually made through email, to steal your personal information. The best way to protect you from phishing is to learn how to recognize a phish. Phishing emails usually appear to come from a well-known organization and ask for your personal information - such as credit card number, social security number, account number or password. Often times phishing attempts appear to come from sites, services and companies with which you do not even have an account*" [31].

"*Phishing is an attempt to trick someone into giving information over the internet or by email that would allow someone else to take money from them, for example by taking money out of their bank account* " [4].

"Phishing can be referred to as an automated identity theft, which takes the advantage of human nature and the Internet to trick millions of people and take a large amount of money" [32].

From the above definitions, it is obvious that the phisher intends to steal the identity of the users and spoof the same for financial gain.

More recently, instead of stealing and spoofing user identities for financial gain, phishers have found an alternate way of achieving financial gain by usurping control of access to user information by means of a malware (ransomware) and blackmailing users for a ransom. Many phishing scams are now ransomware [33].

Thus, it is also interesting to understand ransomware by the following definitions:
"*Ransomware is a type of software that is designed to block access to a computer system until a sum of money is paid*" [34].

"*Ransomware is a software designed by criminals to prevent computer users from getting access to their own computer system or files unless they pay money*"[35].

Table 2 Comparison with the existing classification of phishing attacks

	[48]	[54]	[50]	[49]	[51]	**Proposed classification**
Phishing Definition			*			*
Phishing attack(s) used	*	*		*	*	*
Comprehensive classification of phishing attacks						*
Realtime incidents of phishing attacks	*	*				*
Statistical analysis of phishing attacks						*

"Ransomware is a type of malware that prevents or limits users from accessing their system, either by locking the system's screen or by locking the users' files unless a ransom is paid. More modern ransomware families, collectively categorized as crypto-ransomware, encrypt certain file types on infected systems and forces users to pay the ransom through certain online payment methods to get a decrypt key" [36].

"Ransomware is similar to other malware in that it installs itself on a computer and runs in the background without the user's knowledge. But unlike malware that hides and steals valuable information, ransomware doesn't hide. As soon as ransomware has locked a user's machine and/or encrypted files, it notifies the user of its presence to make the ransom demand" [37].

"Ransomware is a form of malware that encrypts files on an infected device and holds them hostage until the user pays a ransom to the malware operators" [38].

Ransomware is relevant in the context of phishing because it is affected through phishing i.e., the attacker sends the spoofed email with malicious hyperlink/attachments. By clicking on the link or attachment, ransomware installs in the user system and encrypts the very important files and displays an alert with a timer asking to pay some ransom to decrypt the data.
From the above description, we evolve a comprehensive definition of phishing as a *fraudulent activity in which the attacker tries to gain illegal financial gain either by:*
–stealing and spoofing user identity/credentials or

–usurping control of access to user information.

This definition also helps to perform a classification of phishing based on the intention of phishing. Phishing is found to be carried out either for credential-stealing or user information control. Credential-stealing is found to be done either through social engineering or through malware distribution [32]. User information control is done by locking the screen of the victim's system or by encrypting the entire hard disk by means of malware (Ransomware) [35, 36]. A further detailed phishing classification has also been carried out in this work and described in the following sections. This classification is depicted in *Figure 5*.

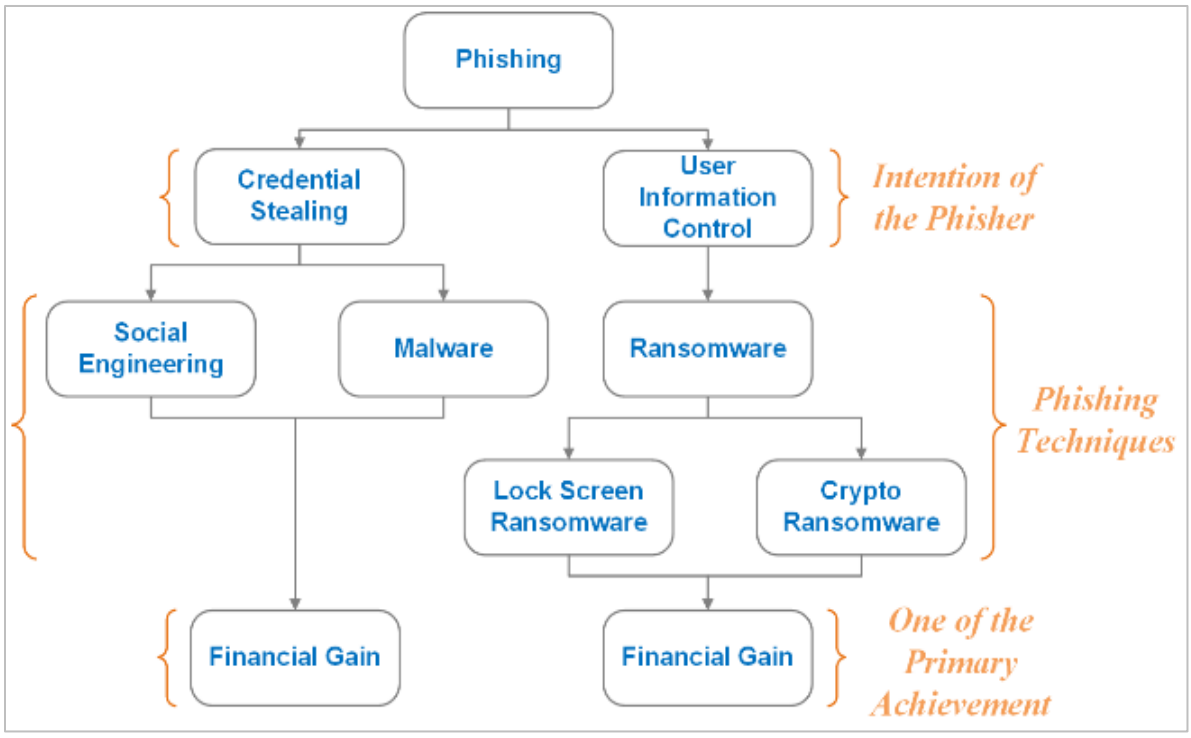

Figure 5 Phishing classification

5.1Classification of phishing based on user credential stealing

A phisher loots the user credentials for fraudulent activities. Credentials are the user's authentication information such as username, password, automated teller machine (ATM) card number, validity, pin, CVV number, one time password (OTP) etc. These credentials are used to access the bank accounts, transfer funds, online purchasing, and share files (through email or social media). It motivates the attacker to lure the users and steal their credentials to gain financially. The attacker steals these credentials either by social engineering or by installing malicious software on the victim's system. In social engineering, the user personally clicks on the hyperlink within the email sent to him (the attacker sends the email with a hyperlink) and gets redirected to a fraudulent website or ends up in a fake website where he enters his credentials by trusting the website i.e.

- A fraudulent website lures users' money for the services provided by them, but they do not provide any services (for example, www.rio2016.com/en/tourist-information/north-america).
- A spoofed website has a look and feel of an original site (for example, http://www.sbi.cgi-co.com) where a user is made to compromise his credentials by making him believe that he is in the original site.

To attract the user to a spoofed website the phisher uses a seemingly authentic URL and to attract the user to a fraudulent website a fake URL is used. A seemingly authentic URL looks exactly the same as the original website's URL, and a Fake URL is a random URL where the domain name of the actual

site and the URL do not match. In malware- based phishing, the attacker will not redirect the user to visit a fraudulent site. Instead, some malware (maybe an image, audio, video, document etc.) gets downloaded and installed in the user system and works in the background. It helps the attacker to control the system remotely and thereby enables him to steal users' credentials. The user credential-stealing can happen either through *i)* social engineering or *ii)* through injecting malware into the victim's system by the attacker. A further classification of phishing is done based on the above two aspects of phishing.

5.1.1Classification of social engineering-based phishing attacks

Social engineering is a psychological manipulation of people into divulging confidential information [39, 40]. The main purpose of social engineering is information gathering, accessing the user system maliciously. The attacker always targets the weakest part of the security system, and humans are the most vulnerable aspect of the security system. Social engineering comes in many forms, and the very common way to fool the internet customer is by phone calling or through email. The attacker sends an email with a hyperlink that looks seemingly authentic (www.sbi.cgi-co.com) but redirects the users to a spoofed site to steal their information. Sometimes malware is also installed to gain unauthorized access to the user system [39, 41, 42]. Phishing through social engineering can be further classified into three categories as shown in *Figure 6* and given below:

- Message based
- Website based
- Search engine based

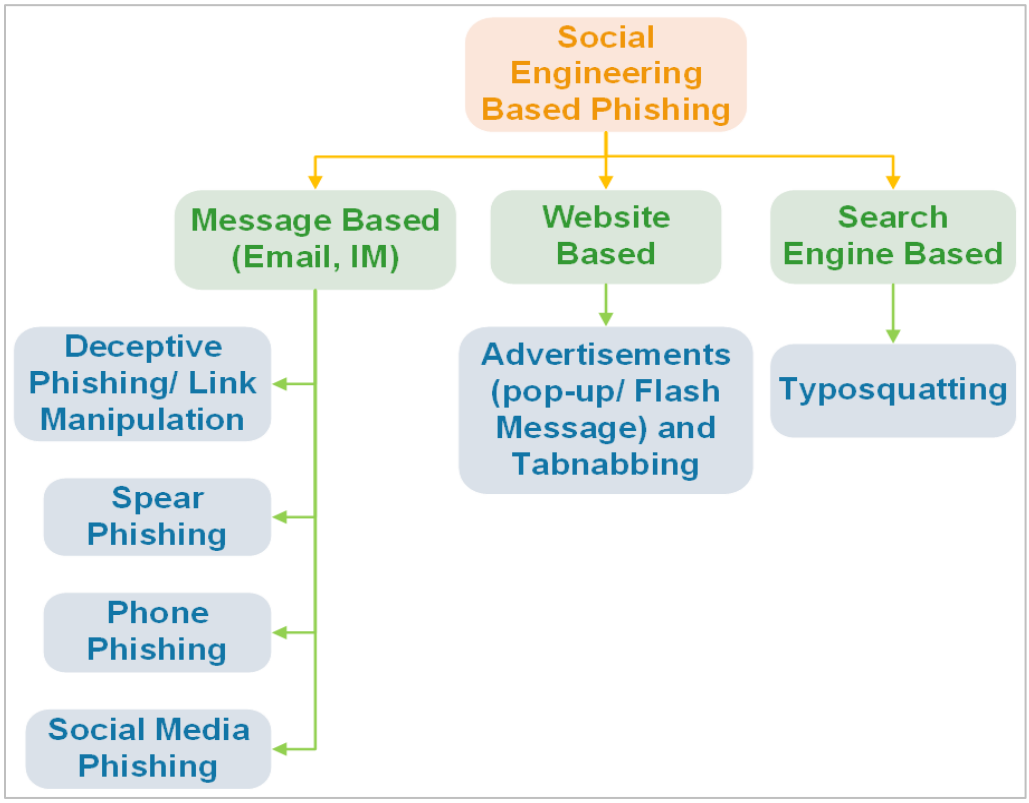

Figure 6 Classification of social engineering-based phishing attacks

5.1.1.1Message based (email and instant messaging)

Message based phishing can be classified into four types which are defined below in detail:

Deceptive phishing/ link manipulation

In Deceptive phishing, the phisher sends thousands of emails to the internet users with a hyperlink that redirects the users to a suspicious website or a seemingly authentic website to steal the users' credentials. The phishing email appears to be from a trusted entity (for example, banking website, e-commerce websites, social media etc.) with an emerging alert that invokes the user to click on the

link [43, 44]. Link manipulation is a subdomain attack, where the attacker creates the legitimate-looking domain in the subdomain part and sends the URL as a hyperlink to the victims through email. When a victim clicks on the suspicious hyperlink, they get redirected to a seemingly authentic website, and the user has to authenticate himself by entering the username and password. Once the user authenticates himself, the credentials are sent to the attacker [2, 43, 44].

Spear phishing

In Spear phishing, a suspicious email is received from a known person (for example, the organization's team manager and colleague) for malicious purpose. To perform this attack, the phisher first gains unauthorized access to any computer (through malware) in that particular organization and sends an email to other employees with the compromised email account (the email account is hacked and is controlled by the attacker) to fool the employees. The email may contain a hyperlink that redirects to a suspicious site or contains downloadable files (PDF, images, videos, audio files etc.). By clicking on the link or attachment, the malware gets downloaded on the victim's machine that runs in the background to gain unauthorized access [45, 46].

Phone phishing

Vishing is the other name of phone phishing, which uses social engineering techniques to gain unauthorized access to the user information through the telephone. The attacker uses the facilities like voice over internet protocol (VoIP), caller ID spoofing, & automated system (IVR) to call/message the victim to ask their personal credentials by claiming themselves as trusted entities [47]. Phone phishing through short message service (SMS) will ask the victim to visit a suspicious site by clicking on an URL to verify their account details or ask the victim to call the number given in the text message. A typical vishing maybe like "someone used your credit card information to perform some fraud transaction, so please call to this number and secure your credit card information", thereby instigating fear of loss of money and psychologically forces the user to call the number given in SMS [2, 47].

Social media phishing

Online scams are increasing drastically, and social networking sites (SNS) make the task much easier for phishers. Most of the users knowingly or unknowingly, sharing their personal information on SNS. Social engineering-based attacks can be performed on SNS to steal their credentials. Social media phishing is when attackers use social media sites like Facebook, Twitter, and Instagram instead of email to obtain your personal information or click on malicious links [48]. Social engineering on social media can be done in the following ways:

- through suspicious URL or attachments
- through bot account
- through compromised account
- through spam account

5.1.1.2Website based

Advertisements and tabnabbing are the primary sources for website-based phishing. Phishing attacks through advertisements are explained below.

Advertisements (through a pop-up or flash message) & tabnabbing

Nowadays, the internet is the best platform for business organizations, e-commerce sites to advertise their products. Most of the companies advertise their new products and services online. The phishers also advertise their websites in popular search engines, which is a unique way to redirect internet customers to fraudulent sites. Instead of sending email to a large number of customers, it is an easy and better way to advertise the phishing websites in any search engine through advertising networks (a company that post adds which connects the hosting site). When the users search for any keywords, the search engine displays the search results, including the advertisements, as a pop-up. These advertisements are displayed at the top of the search results, and they can also be displayed in a particular webpage as a pop-up. Most internet users click on the topmost displayed content without verifying the URL. It redirects the users to a seemingly authentic or fraudulent site that asks for user authentication/user credentials [49, 50]. Tabnabbing is a phishing technique used to steal user credentials when a victim visits a malicious site that looks like a legitimate site and opens multiple tabs simultaneously. If the malicious tab is inactive or idle, the malicious script inside the website will automatically execute and open a fake login page to steal users' credentials. When the victim opens the malicious tab, he does not pay much attention to the URL, which is modified and enters their login details. The attacker will take advantage of the victim's trust on the site to steal their credentials [51, 52].

5.1.1.3Search engine based

In Search engine-based phishing, the attacker registers for the domain very similar to the legitimate site and these links will be available during the search results to redirect the users. Typosquatting is a search engine-based phishing attack and is explained as follows:

Typosquatting

Typosquatting is a type of cybersquatting used for a phishing scam, where the attacker registers multiple domains that are almost similar to the legitimate domain to fool the internet users. The URLs with these domain names are then indexed in the search engine.

Most of the users misspell the keyword while searching in any search engine so that the URL of the phisher page is also displayed in the search results. If users click on that suspicious URL, they will be redirected to a seemingly authentic website to steal user credentials. For example, www.annozon.com, www.amazn.com etc., are the seemingly authentic URL's for www.amazon.com [53–55].

Table 3 provides a summary of social engineering-based phishing attacks in terms of a carrier, sender/caller, phishing source, type of URL embedded in email/website, destination, credential stealing, and phishing type.

- **Carrier:** The carrier of the phishing content may be an URL embedded in an email to the user, or it can be a website, or it can be a phone call made by the phisher.
- **Sender/Caller:** The sender is a phisher who sends thousands of emails to internet customers to lure their personal credentials. He may be a known or unknown person.
- **Phishing source:** In the case of email phishing, the sender of an email is a known person to the user or an unknown person. The Phishing source can be an embedded URL in an email or an URL within an advertisement or pop-up displayed on a website to redirect the traffic. Sometimes the phishing source can be a phone conversation by the phisher to the customer.
- **Type of URL embedded in email/website:** The URL embedded in an email can be a fake URL (for example, www.rio2016.com/en/tourist-information/north-america seems to be an authentic website to sell the Rio Olympics ticket but it is a fraudulent site that only collects the money from users and provides fake tickets) or a seemingly authentic URL (for example, http://www.sbi.cgi-co.com). *Figure 7* shows the classification of phishing URLs.

Table 3 Different ways of performing social engineering-based phishing

S. No.	Carrier	Sender	Phishing Source	Type of URL Embedded in email/ Website	Destination	Credential-Stealing by exploiting User Ignorance/ through Malware	Name of the Attack
1	Email	Known/ Unknown	URL embedded in email	Fake	Fraudulent site	User ignorance	Deceptive phishing
2	Email	Known/ Unknown	URL embedded in email	Seemingly authentic	Spoofed site	User ignorance	Deceptive phishing
3	Email	Known	URL embedded in email	Seemingly Authentic/ Fake	Spoofed site/ Fraudulent site	User ignorance	Spear phishing
4	Email	Known/ Unknown	URL embedded in email	Fake	NA	Through malware	-
5	Website	NA	Advertisements & pop-ups with URL's embedded in websites	Seemingly Authentic/ Fake	Spoofed site/ Fraudulent site	User ignorance	-
6	Website	NA	Advertisements & pop-ups with URL's embedded in websites	Seemingly authentic	Spoofed content	User ignorance	Tabnabbing
7	Phone call	Known/ unknown	Phone conversation by phisher	NA	NA	NA	Vishing/ Smishing
8	Social Media Sites	Known/ Unknown	URL/ Advertisements	Seemingly Authentic/ Fake	Spoofed site/ Fraudulent site	User ignorance	Social Media Phishing

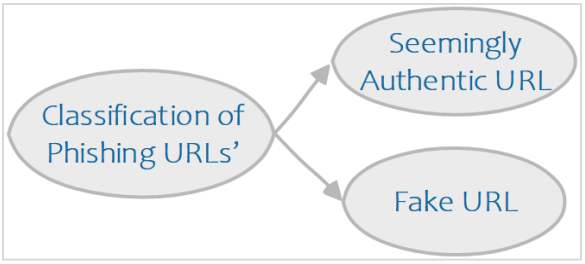

Figure 7 Classification of phishing URLs

–**Destination:** The destination of these phishing URL's can be a
–Fraudulent site
–Spoofed site

Fake URL leads to fraudulent site for example, www.rio2016.com/en/tourist-information/north-america lures the user to pay for services and does not provide any services. *Seemingly authentic URL leads to Spoofed site* (for example, http://www.sbi.cgi-co.com) that looks and feels like the original SBI bank site and fools the internet users [56, 57]. *Figure 8* shows the classification of Destination reached through phishing URLs.

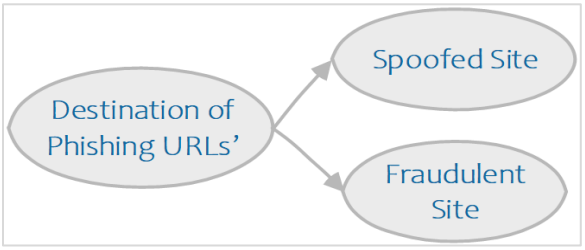

Figure 8 Classification of destination reached through phishing URLs

– **Credential stealing:** Credential stealing can happen through phone calls, by user ignorance or through malware.
– **Phishing type (Social engineering):** Depending on the carrier, sender, destination and credential stealing, different types of social engineering-based phishing can be identified viz. deceptive phishing, spear phishing, phone phishing and tabnabbing. From *Table 3*, two things can be extracted:

- Social engineering-based phishing is mostly because of user ignorance and
- Fake URL leads to a fraudulent site, and seemingly authentic URL leads to a spoofed site.

5.1.2Classification of malware-based phishing

Malware Based Phishing is a scam that runs malicious software on user's computers. Malware can be installed in victims' system through an email attachment or a downloadable file from the website or by exploiting the vulnerabilities in the system (whose system updates are not updated regularly). Most of the Small and Medium Scale Enterprises (SMEs) are affected by malware.

The attacker tries to find the loophole in the system (client/server) to install the malware to steal the personal credentials of the internet users. If the DNS server is compromised, then the traffic of a particular site will be redirected to a fraudulent server controlled by the phisher. To redirect the traffic, the attacker replaces the internet protocol (IP) address of the legitimate server with the attacker's IP, and the URL still looks exactly the same as the legitimate URL [2, 43, 44, 58]. *Figure 9* shows the classification of phishing attacks based on malware at three levels:

– Host based phishing attacks
– DNS based phishing attacks
– Server based phishing attacks

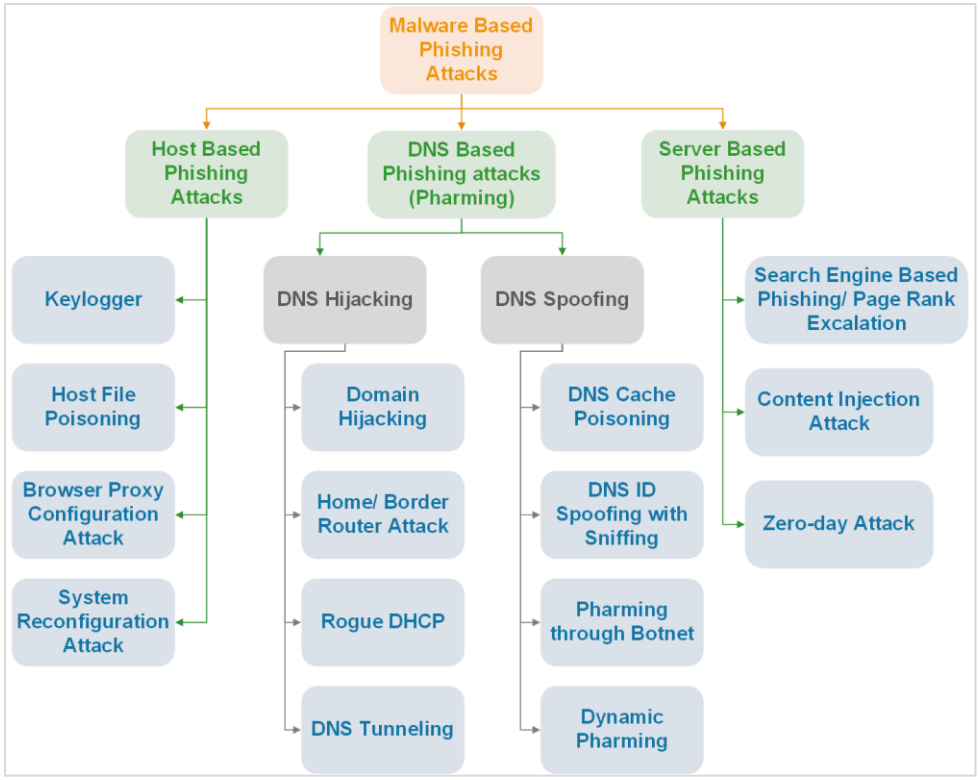

Figure 9 Classification of malware-based phishing attacks

5.1.2.1Host based phishing attacks

Phishing at the host level can be installing keyloggers, poisoning the host file record, system reconfiguration attacks etc. All these attacks can be done by installing malware in the user system. Botnets and Trojans are the most popular among the malware based host level phishing [59–61]. Host level phishing attacks are further explained below in details.

Keylogger

Keylogger is a type of spyware used to read every keystroke on the keyboard. If the keylogger is installed in any system, then the keylogger software can record any information (i.e., instant messaging, emails, log in details) which are entered through the keyboard. All these information is stored in a log file with the help of the keylogger and transferred to the phisher. Some keylogger programs can even record the email addresses used and the URLs of the website visited by the user. Keyloggers are of two types, namely; software based keyloggers and hardware based keyloggers. Software-based keyloggers should be installed in the victim system in order to monitor and steal their personal credentials like username, password, credit card number, pin, validity etc. However, in Hardware-based keyloggers, a device is plugged in between the central processing unit (CPU) and keyboard and implemented via bias level firmware. It works as soon as the system is turned on and records all the keystrokes and transmits them to the Phisher [32, 62, 63].

Host file poisoning

Host file poisoning is a client-side attack where the phisher modifies the host file records with the help of a Trojan such that the user can easily be redirected to a fraudulent website where his credentials could be stolen (a host file stores the domain name and its corresponding IP Address and every time when the user searches for some keyword in the search engine, the system first refers the host files to get the IP address. If the IP address is not available, then it asks the DNS server for the IP address of that particular domain and stores those results temporarily for later use) [2, 32, 44, 58, 59].

Browser proxy configuration attack

Browser proxy configuration attack will take the preference over the victim's web browser proxy configuration to redirect the entire traffic to a fraudulent proxy server controlled by the phisher to steal the user's personal credentials [64, 65]. DNS Rebinding is one such kind of attack that enables a remote attacker to breach the network firewall of a victim (by executing the malicious JavaScript on the victim's browser) to connect directly to the

computers on its private home network using their browser [66, 67]. Some other possibilities are, the intruder can steal the personal credentials of the victim or send phishing emails to the machines within the network.

System reconfiguration attacks

System reconfiguration attacks change the victim's personal computer (PC) setting for malicious purpose. Suppose there are 10 URLs in our bookmarks or the favourite list; the phisher modifies those bookmarks by replacing them with the URL that redirects the traffic to the website under the control of the phisher. To do this, the attacker needs unauthorized access to the victim's system to modify the records. For example, abcbank.com can be replaced with ababank.com or abcbanc.com. If the user visits the website through bookmarks, the suspicious website (developed by the phisher) that is very similar to the legitimate one is displayed to steal the user credentials [44, 68].

5.1.2.2DNS based phishing attacks

DNS level attacks can be done either by hijacking the DNS servers or by spoofing the DNS entries. Different methods for DNS hijacking and spoofing are described in detail below.

DNS hijacking

In DNS Hijacking, the attacker overrides a computer's transmission control protocol/internet protocol (TCP/IP) setting through malware. To do this, the attacker needs unauthorized access to the user system. When the user clicks on any fraudulent links, downloads files from fraudulent sites, the malware is installed in the user system, and it provides unauthorized access to the attacker. Once the attacker gains access, he changes the DNS settings and redirects all user requests to a Rogue server [9, 69–71]. DNS hijacking can be done by hijacking the domain by compromising the home/border router, rogue dynamic host configuration protocol (DHCP), changing the browser configuration, and tunnelling. All these attacks are explained below:

- **Domain hijacking** is a process by which Internet domain names are stolen from its legitimate owners [69]. To hijack a domain name, the domain registrar name and administrative email address need to be known. This information can be obtained from WHOIS data. To hijack a domain name, access to the domain control panel is required. The email address associated with the domain is the key to gain access to the domain control panel. To do this, the attacker has to hack the email address. Once done, he can log in to the domain control panel, and by clicking on forgot password, he can reset the password. The new password is sent to the email associated with that domain, which is already under the control of the attacker. Later the webserver of that domain is changed to the attacker's webserver to redirect the traffic [54, 71].
- **Home router or border router:** The phisher attacks home routers to redirect the traffic to the fraudulent site. Suppose a client load a website from the attacker's server through the home router, the page is rendered, and the attacker can identify the client's internal IP address either by guessing or using an applet program that identifies the IP address. After identifying the IP address, the attacker uses JavaScript to interpret the client run page to discover the router and modifies the router's settings. DNS poisoning can also be done to compromise, the home router and redirect the users to a malicious site even though they specify the legitimate site's URL (IP address of the legitimate site's web server changes to the attacker's web server IP) [72, 73].
- **Rouge DHCP:** In this attack, the attacker installs a fake (DHCP) server into the client's network to redirect the client's DNS request to the attacker's website. For example, there is a fake DHCP server and a legitimate DHCP server. The first DHCP client broadcasts a DHCP discover the packet, and the server broadcast a DHCP offer packet. Later, the client sends the DHCP request to the server from whom it is getting the DHCP offer packet. If the attacker sends the DHCP offer packet first, the client requests the Fake DHCP server, and this fake DHCP server will assign the IP address (contains the IP, subnet mask, default gateway) and the server IP. Once the client receives the IP address and DNS server IP from the rogue DHCP server, then all the client request will be redirected to the DNS server, which is under the attacker's control [55, 74, 75].
- **DNS tunneling** is the ability to encode the data to other programs in DNS queries and responses. Mostly DNS tunnelling is used for malware-based data exfiltration from Business networks. Data exfiltration is a malicious activity through which the attacker copies, transfers, or retrieves the data from a computer or server [76]. To perform tunneling, the client system should be compromised by sending a suspicious email with a malware attachment that redirects the user to a fraudulent site controlled by the attacker [77–79].

DNS spoofing

DNS spoofing involves modifying or inserting false entries into the DNS server by the phisher to whom there is no authority for modification [42]. This helps the phisher redirect the internet user to fraudulent or spoofed sites under his control to steal their personal credentials [78, 80]. Cache poisoning, DNS ID spoofing with sniffing, man-in-the-middle attack etc., are the different ways to spoof the DNS records, and each of these attacks is illustrated below in detail:

– **DNS cache poisoning** is a type of attack that corrupts the cache database of DNS name server. The DNS name server maintains the Domain name and their IP address that helps the users to connect to a particular website. The phisher sends the forged response to the DNS name server, and this corrupted information provided by the attacker is cached in the original DNS name server. Now the entire traffic on that original site, is redirected to the attacker site for malicious activities[17].

– **DNS ID spoofing with sniffing:** The user datagram protocol (UDP) protocols are used for querying the DNS, which does not use any handshaking process. For every UDP request, a unique ID is generated for DNS lookup responses. To successfully perform the attack, the phisher has to compromise the user systems' network traffic to capture the user DNS request (by sniffing the network traffic) and send the spoofed response to the user before the DNS server responds to that user. Now, the user visits the spoofed site designed and controlled by the attacker and enters the personal credentials by believing it to be a legitimate site [9].

– **Phishing through botnets:** Botnets can also be used for performing phishing scams. Phishing through botnet can be done in two ways as follows:

- *Botnet name server hijacking* helps the phisher host fake websites with several copies, and each has a different IP address. If any of this host/IP is closed by internet service provider (ISP), it points to the alternate location with a different IP address. This can be done by changing the ISP's DNS server as phisher's authoritative domain server [9].
- *The fast flux* is a DNS based technique that uses botnets that help the attackers to hide their phishing and malware content in websites whose network is continuously changing, and the compromised hosts are used as proxies to perform this attack [81]. Single flux and double flux are the two main classifications of fast flux networks [78]. The technique behind this attack is to keep a domain name with several IP addresses. Randomly the IP addresses are changed among the given list of IPs and make the attack more complicated to detect. Each record in DNS has a time to live (TTL) for mapping and set some time limit. The TTL value for DNS is 86400 seconds. The attacker uses a Round Robin IP address and less TTL for a DNS resource record (RR). If the TTL limit reaches, it changes to another infected computer in the same domain. In addition, they use a load-distribution scheme which removes the unresponsive node from the flux and maintains the availability of the content [82, 83].

– **Dynamic pharming:** Dynamic pharming is an advanced phishing scam that compromises the authentication schemes of the victim's browser [72]. Let us assume that the attacker controls the results of DNS queries for www.phishy.com (this domain is chosen only for explaining this attack, not for any other). Initially, the attacker gives the IP address of the server under his control say 1.3.4.9 in the DNS for www.phishy.com instead of 1.2.4.8 which the actual IP of that domain. The pharmer also says that requesters must not cache this result by setting the TTL=0 in the DNS record. If the user visits www.phishy.com/login.html for authentication, the browser asks for secure sockets layer (SSL) verification. The attacker submits a false certificate or no certificate. Immediately the browser will alert the user that the details of the certificate are not trusted, and it is your interest whether to accept those certificates or not. The attack works only when the user accepts the certificate. Once the user accepts the attacker's certificate, the user's browser will establish an SSL connection to the attacker IP 1.3.4.9 and its request for the login.html page. In return, the attacker sends a Trojan login.html document. This Trojan document helps in monitoring the user interaction with the legitimate domain www.phishy.com. Now the attacker manages the browser and loads the legitimate domain www.phishy.com/login.html with its actual IP (i.e., 1.2.4.8) in <*iframe*> to the user. If the user authenticates himself to the legitimate server www.phishy.com/login.html in the <*iframe*>, the malicious JavaScript in the outer document will take control and monitor the user's interaction in the <*iframe*> with the legitimate server. Here the outer document and <*iframe*> contain the same domain www.phishy.com and protocol. Now the same origin policy (SOP) will allow the malicious JavaScript running in the outer document to access the content in the <*iframe*>. The Trojan hijacks user session control for stealing personal

credentials, forge transactions sniff the password [72, 74].

5.1.2.3Server based phishing attacks

There are three types of phishing attacks that can be performed on the server side.

Search engine phishing/ page rank escalation

Pharming can also be done by escalating the page ranks in the search engine. The main purpose of page rank escalation is to display the fraudulent/spoofed site in the search results of any search engine (for example, Google). This can be done when the page is indexed by a search engine. When internet users click on that link, they will be redirected to a fraudulent/ spoofed site to steal their personal credentials. For example, the phisher develops a fraudulent/spoofed site with discounts, free offers, job alerts to attract the internet user to visit their websites [50, 84].

Content injection

In a content injection attack, the phisher inserts malicious content into a legitimate website to deceive the user into giving up their personal details. Sometimes the malicious content may be a downloadable file from the infected site. This later installs the malware in the user system and transfer the user's personal credentials to the phisher remotely [2, 44, 74, 85].

Zero-day attack

Zero-Day Attack refers to a loophole in the software such as browser software or operating system [86, 87]. The phisher then exploits this security hole before the vendor becomes aware of it and this exploit is called a zero-day attack. Zero-day attacks can include infiltrating malware, spyware or allowing unwanted access to user information. The term "zero-day" refers to the unknown nature of the hole to those outside of the hackers, specifically the developers.

5.2Classification of phishing attacks based on user information control

In user information control, the attacker controls the user information by locking the login screen or encrypting the entire hard disk (can encrypt some critical files or not allowing the user to access web browser) with the help of malware. To gain unauthorized access to the victim's computer, the attacker sends an email with an attachment that contains malicious code, and it is installed in the user system when they download the attachment or simply by clicking on the suspicious URL. Later, the attacker starts encrypting the data with a symmetric key algorithm and displays a flash message saying that the system is encrypted and to decrypt the data, some ransom needs to be paid. Here, the attacker does not require the user credentials; instead, the attacker blackmails the user for ransom/financial gain. User Information Control through ransomware can be classified into two categories as shown in *Figure 10* and explained below.

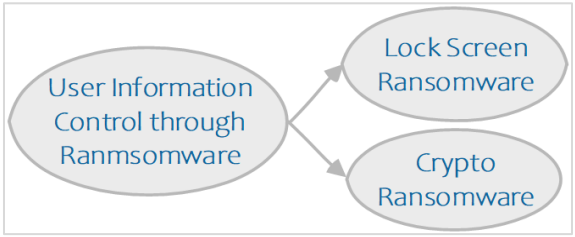

Figure 10 Classification of phishing attacks based on user information control

5.2.1Ransomware

Ransomware is a type of malware that blocks users from accessing their details from a PC. This can be done by installing the malicious code in the user system when the users click on any suspicious link. The ransomware is of two types:

- Lock screen ransomware and
- Crypto ransomware

In both the attacks, the user is not allowed to access their PC's. In Lock screen ransomware, the phisher displays an alert screen and warns the user that his system is locked and demands a ransom to unlock the screen. In Crypto ransomware, the phisher encrypts the file in the PC so that the user cannot open those files without the key. Any files can be encrypted by using ransomware. Only the attacker can decrypt the file with his/her private key. There is no guarantee that the user will get the key after paying the ransom in the form of bitcoin (BTC)). Cryptolocker, locky, alphlocky, key BTC, Tescrypt, crowti, Fakebsod etc., are the different types of ransomwares available [35, 88, 89]. For example, WannaCry ransomware is a recent attack that compromised more than 230000 systems around 150 Countries [90]. The details about WannaCry ransomware attacks are given below:

5.2.1.1WannaCry ransomware

WannaCry or WannaCrypt is a ransomware attack that infects the Windows operating systems. WannaCry ransomware is very dangerous, and it encrypts the entire hard disk of a computer and stops the user from accessing the system and demands some ransom in the form of Bitcoins decrypt the data. This ransomware can spread either through phishing email attachments or through hacking tools. The National Security Agency (NSA) found a vulnerability in implementing the server message block (SMBv2) protocol on the Windows operating system, which helps execute a malicious code remotely. For offensive purpose, NSA developed two

hacking Tools (Eternal Blue and Double Pulsar) which was stolen by the hacker team "Shadow Brokers" with bad intentions. On May 12, 2017, the shadow brokers' team hacked more than 230000 systems around 150 countries in a single day with the help of those hacking tools and demanded a ransom of $300 for decryption [88]. *Figure 11* shows how the screen looks when the system is infected with WannaCry ransomware. It is one of the biggest ransomware attacks that affect most developing and developed countries. Later, Microsoft released a patch on May 14, 2017, to address the vulnerability in their operating system. According to Kaspersky lab reports [90], Russia, Ukraine, India, and Taiwan are the four most affected countries, and 98% of affected computers were using the Windows7 operating system. Initially, the infection is likely to spread through a vulnerable SMBv2 protocol rather than email phishing.

Figure 11 The screen of ransomware affected system [88]

Summary: From *Table 3*, it is observed that phishing attacks are increasing drastically. According to Google, more than 18 million phishing emails and malware attempts are made each day [91]. Recent phishing scam from 2008 to 2020 is considered by covering all the types of phishing attacks listed in the classification of phishing attacks. More than $280 million are lured from the victims.

6.Phishing statistics

There are few organizations, international consortium that publish reports on phishing attacks (For example, APWG, PhishMe, OpenPhish and PhishTank provides technical reports on phishing) [33], [92–94]. These reports provide a clear idea of current phishing scams, the loss incurred etc., in detail. Time series analysis is chosen for phishing data analysis. The method of least square is the most common and widely used method to analyze time-series data.

"A least-squares method is a form of mathematical regression analysis used to determine the line of best fit for a set of data points providing a visual demonstration of the relationship between the data points being studied" [95]. The linear or ordinary method is the commonly applied method that aims to create a straight line as given Equation 1, that minimizes the sum of squares of the errors that are generated between the actual and predicted values of Y. After fitting a straight line by using the method of least squares, we can tell whether the data follows an increasing trend or decreasing trend along with the pattern followed by the data over the years.

$$Y = a + bx \quad (1)$$

Where,
Y = Dependent variable
x = Independent variable
a = Intercept
b = slope o f a line

"R squared or co-efficient of determination (R^2) in Equation 2 below is a statistical measure of fit that indicates how much variation of the dependent variable(Y) is explained by the independent variable (X) in a regression model" [64, 96]. The R^2 value is obtained by taking the data points of dependent and independent variables and finding the line of best fit from the regression model. The R^2 value is normally

represented in percentage (ranges from 0% to 100%). The more the R^2 value as shown in Equation 2, the more the model fits the data. If the model fits the data correctly, then it is possible for better forecasting/prediction and understanding the behaviour/pattern of the data.

$$R^2 = \frac{Variance\ explained\ by\ the\ model}{Total\ Variance} \quad (2)$$

For the purpose of data analysis, the trend reports of APWG, Phishing data from OpenPhish and PhishTank [92−94] have been used as data sources. The primary data from these sources have been analyzed to obtain the secondary data using which the following six graphs are drawn. Two more graphs on ransomware attack are also included from the existing works to show how critical the ransomware attacks are [97].

- top three countries hosting phishing
- top three countries most affected by phishing malware
- top three countries least affected by phishing malware
- top three most targeted industry sector in phishing
- top three malware used for phishing
- hypertext transfer protocol secure (HTTPS) enabled phishing URLs
- top five ransomware affected countries with five different ransomware families
- top twenty countries affected by WannaCry ransomware.

The above said graphs are shown in *Figures 12, 13, 14, 16, 18, 20, 22, 23,* and *24* respectively.

6.1Top three countries hosting phishing attacks

Experimental setup

To perform the phishing statistics, APWG (non-profit organization) trend reports for every quarter from the year of 2012 to 2019 are considered. The APWG trend reports provide the statistics related to phishing attacks reported globally. The primary intention of this organization is to focus on *"unifying the global response to cybercrime through data exchange, research and promoting public awareness".* And publish these reports quarterly for the internet users to know the impact of phishing attacks on different sectors.

Analysis

From the reports, we extracted the data required for identifying the top three countries hosting phishing scams. Trend analysis is performed on the collected data for forecasting and analyzing the trend of phishing attacks hosted by those three countries individually over the years. The step-by-step process of analysis is as follows:

1. The quarterly reports from the year 2012 to 2019 are considered.
2. Each of those trend reports provides the top 10 countries hosting more phishing contents. The quarter-wise data for all the years from 2012 to 2019 are collected.
3. To find the Top three countries hosting phishing, we have two criteria to fulfil. a) should contain a high score in hosting phishing, b) should appear in the maximum number of quarters. Now we combine the four quarters in every year to get years averages. Only countries that meet the above criteria is considered. There are few countries with a high phishing hosting score, but have appeared in few quarters only. For example, countries like Belize, Irelands and Brazil have a high score in hosting phishing, but they appeared in a minimum number of quarters, and hence they are not considered.
4. Now, we have to sort the countries in ascending order of the average computed above for the analysis so that the countries with the highest rates in hosting phishing will appear on the top.
5. The top three countries from every year are filtered and listed separately.
6. Finally, we got more than 20 countries as the top three in all the years. Again, these countries are sorted in ascending order to get the overall top three countries hosting phishing.
7. A graph is plotted with data points collected from the top 3 countries hosting phishing.
8. The method of least squares is applied to draw a trend line for forecasting. The outcome will be either an increasing trend or decreasing trend depending on the data.
9. The R^2 value is generated to know whether the model fits the data or not. The forecasting will be more accurate if the R^2 value is more.

Result

From the analysis, we found that the United States, Germany, and United Kingdom (UK) are the top three countries hosting phishing scams and have appeared in the maximum number of quarters as shown in *Figure 12*. Among these countries, the USA stood at the top every year with an average of 49% of phishing scams followed by Germany (2.40%) and UK (1.66%). *Figure 13* shows the trend analysis and pattern observed in the top countries hosting phishing scams.

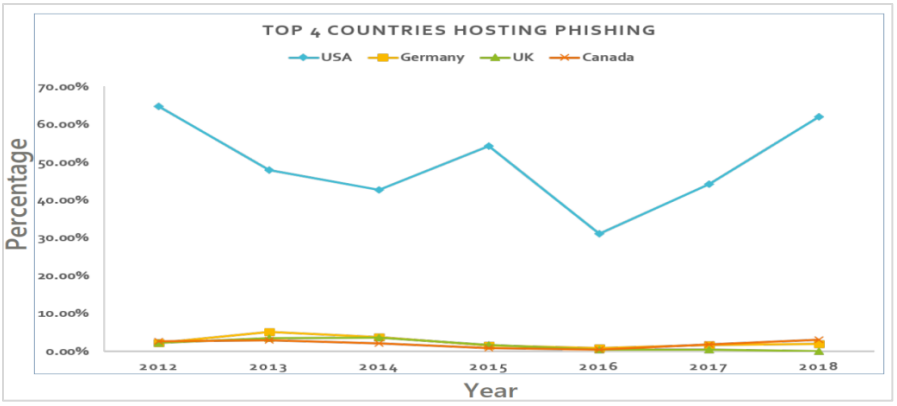

Figure 12 Top 4 countries hosting phishing attacks

Inference

Even though the countries like USA, Germany, and the UK are in the top three in hosting phishing, the trend analysis performed on individual countries tells us that they are following a decreasing trend. This can be observed in *Figure 13*. In the USA, the R^2 value is 3%, which means the model does not fit the data correctly, and with a lower R^2 value, the prediction for the upcoming years will be of 3% accuracy. For Germany, the prediction will be 30% accurate because the R^2 value is 30%. In the case of Germany, there is a pattern that every year, the percentage value is decreasing significantly. Therefore, it follows a decreasing trend with R^2 value as 69%. The experiment setup and analysis part of the upcoming graphs are the same as explained in this section. So, the result and inference part of the data analysis alone are discussed in the upcoming sections.

Figure 13 Trend analysis on countries hosting phishing

6.2Top three countries most affected by phishing malware

Result

From the graph shown in *Figure 14*, it is clear that China, Turkey and Taiwan are the top three most affected countries by phishing malware, among which China is at the top in every year (2012-2016). Trend analysis of individual countries is shown in *Figure 15.*

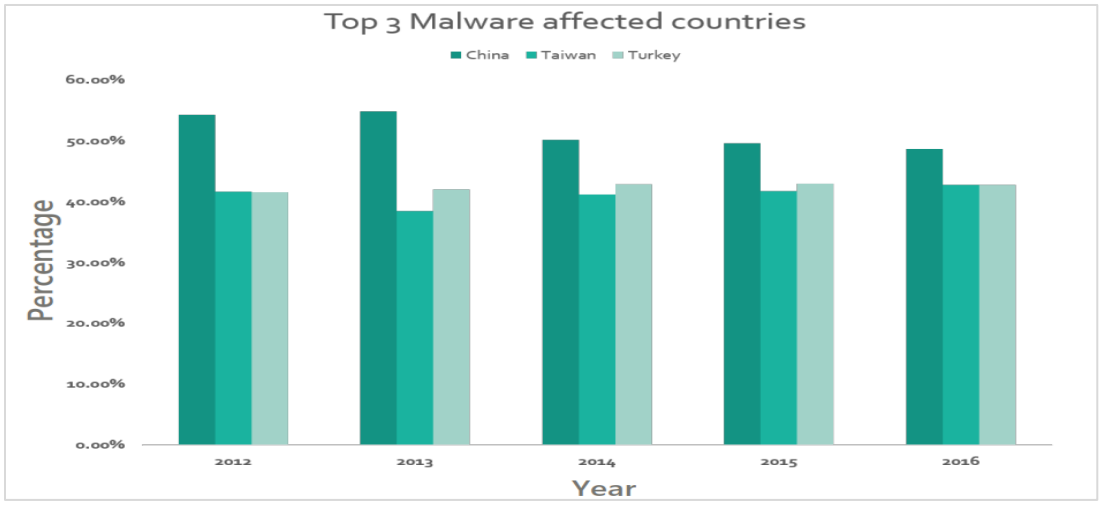

Figure 14 Top 3 countries affected by phishing

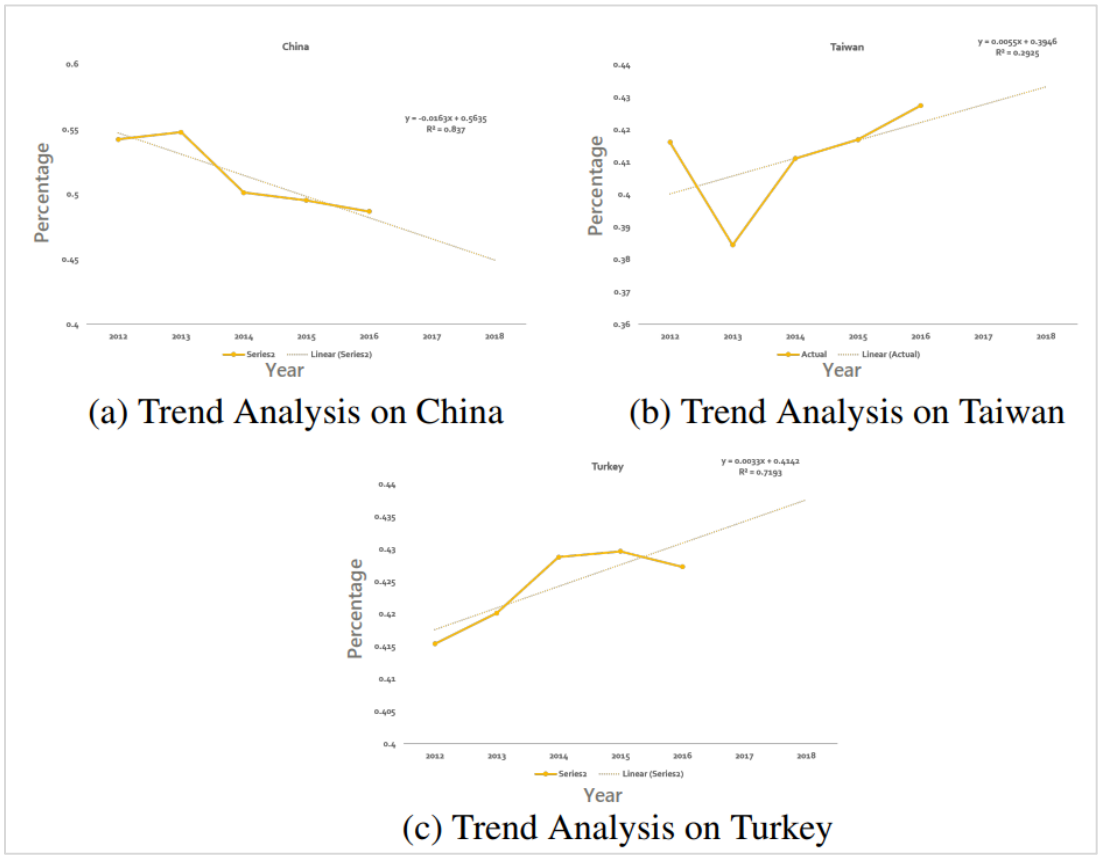

Figure 15 Trend analysis on the top 3 countries that are affected through phishing

Inference

From *Figure 14* it is so obvious that China is a top phishing attack affected country with an average of 51.46%, followed by Turkey with 42% and Taiwan with 41.11%. Although China is the top country affected by phishing, as shown in *Figure15(a)*, the R^2 value (83%) is high, and it follows a decreasing trend. That means the percentage of phishing is decreasing year by year. If the phishing affected rate is decreasing, it might be due to better security measures. In the case of Turkey and Taiwan, they follow an increasing trend as shown in *Figure15(b)* and *15(c)*. Compared to China, Turkey and Taiwan are less affected, but they are following an increasing trend over the years.

6.3Top three countries least affected by phishing malware

Result

The top three least affected countries through phishing malware are Japan, Norway and the Netherlands, which can be seen in *Figure16*. *Figure17* shows the trend analysis on individual countries, and they all follow an increasing trend.

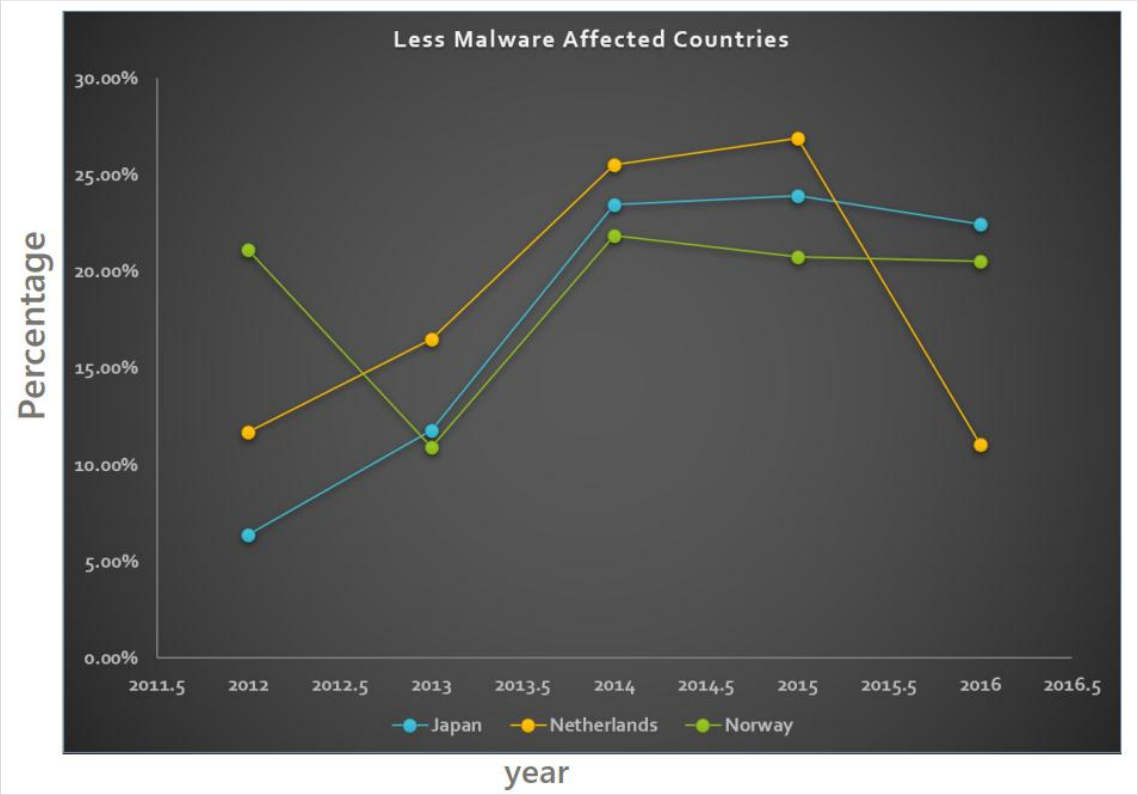

Figure 16 Least phishing malware affected countries

Inference

Japan is the least affected country with an overall average of 17.59%, followed by the Netherlands and Norway with 18.34% and 20.55% respectively. *Figure17 (a)* shows the trend analysis of Japan, which follows an increasing trend with an R^2 value of 76%.

Even though Japan is the least affected county, the percentage of phishing affected rate is in- creasing year by year. In the case of Norway and the Netherlands, they follow an increasing trend with R^2 value as 3% and 8%, respectively. Due to the low R^2 value, the prediction may not be accurate, and the phishing affected rate varies a lot in both countries every year.

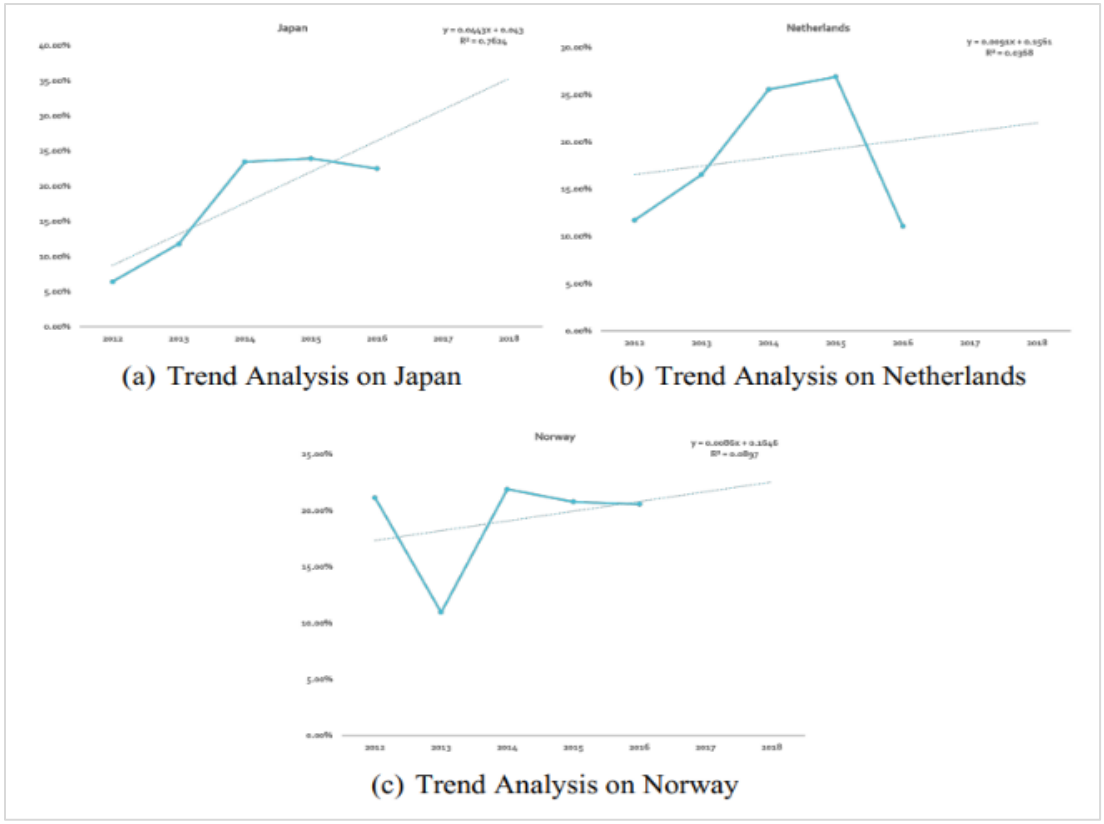

Figure 17 Trend analysis on the less phishing malware affected countries

6.4Top three most targeted industry sector in phishing

Result

Payment system, Financial, and Retail/Service sectors are the top three most affected industry sectors because of phishing. *Figure18* shows all the three sectors' infection rate year wise. *Figure19* shows the trend analysis of individual sectors along with predicted values for the future.

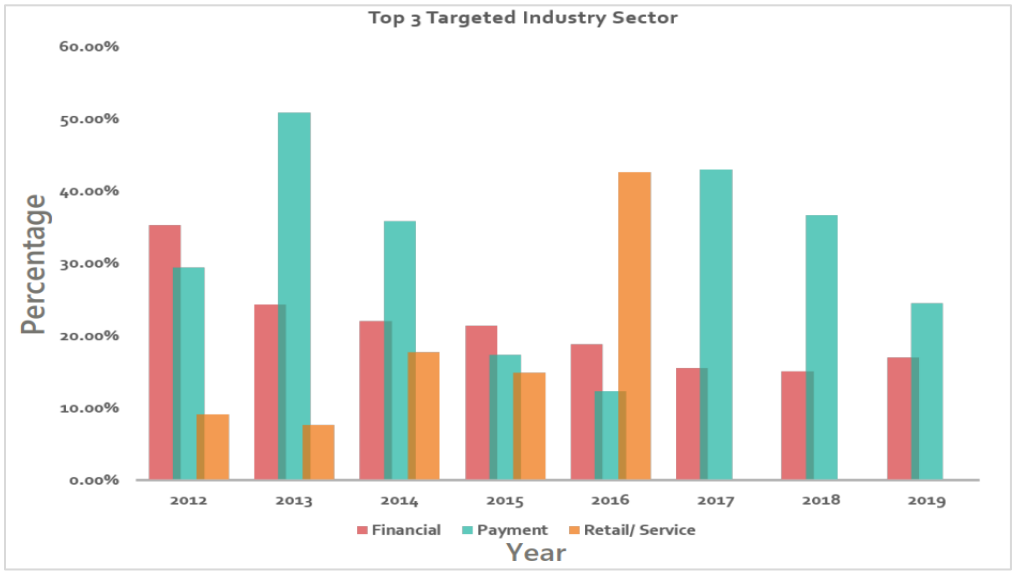

Figure 18 Top 3 most targeted industry sector by the phisher

Inference

From *Figure 18*, it is found that the payment system, financial, and retail/service sectors are the most targeted industry sectors. The payment system is the most targeted industry sector compared to the financial and retail/service sectors. However, when we observe year-wise, the payment sector is affected a lot between 2013-2019, the financial sector was highly affected in 2012, and the retail/service sector was highly affected in 2016. When these countries are observed individually as shown in *Figure 19*, the financial sector follows a decreasing trend with a high R^2 value (i.e.,75%), payment sector follows an increasing trend with a low R^2 value (i.e., 5%), and retail/service sector is observed to have an increasing trend with high R^2 value(i.e.,76%). In the financial and retail/service sectors, the R^2 value is greater than 50%, which gives better forecasting and the model fits the data. But, in the case of the payment sector, the R^2 value is less than 50% (i.e., 5.27%) which means the model does not fit the data.

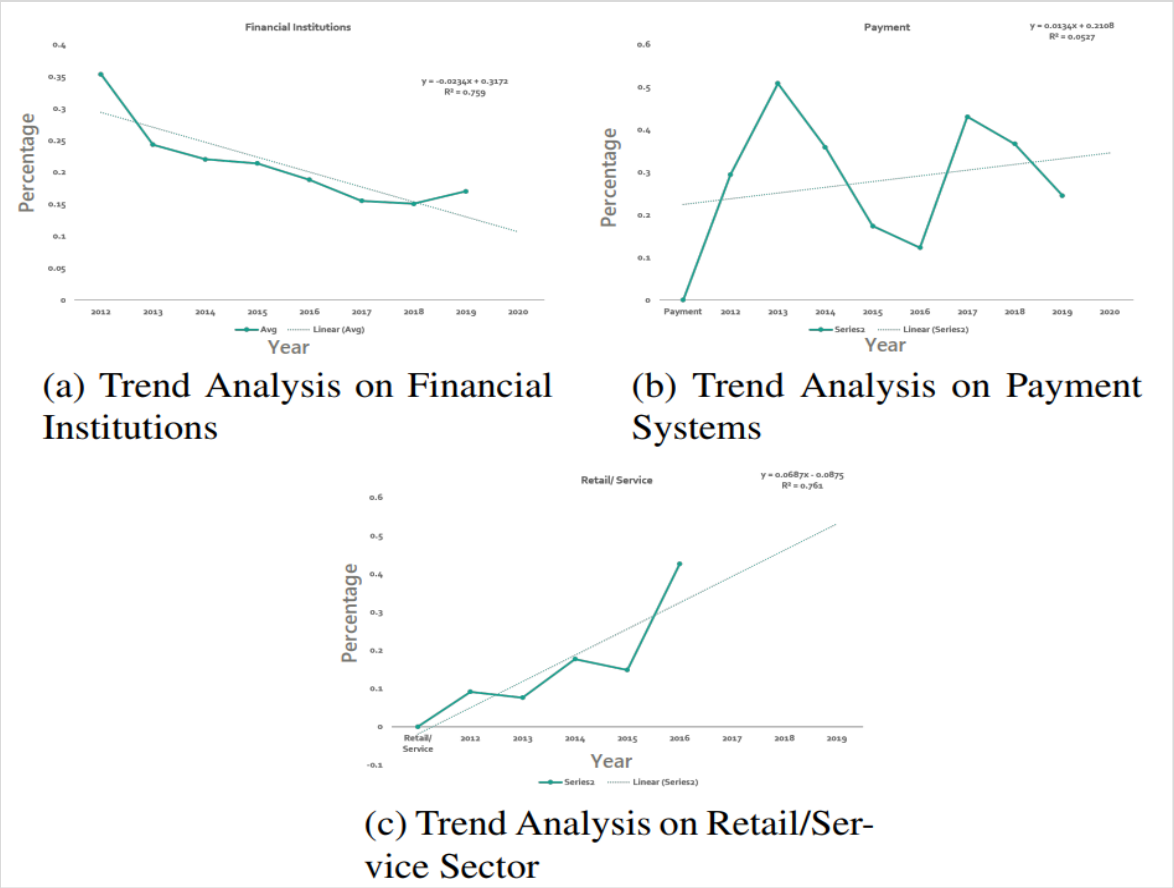

Figure 19 Trend analysis on most targeted industry sector

6.5Top three malware used for phishing

Result

Trojans, virus and worms have been identified as the top three malware types used for phishing as shown in *Figure 20*. Trend analysis of individual malware shows a decreasing trend.

Inference

Trojans, virus, and worms are the top three most commonly used malware for phishing from the analysis. *Figure 20* shows the impact of malware over the years. The Trojan is at the top among this malware almost every year, followed by a virus, and worms. When this malware is observed individually by plotting a trend line, an interesting thing is observed: all the malware follows a decreasing trend. *Figure 21* show the pattern and type of trend followed by each malware. For Trojans, Virus, and worms, the R^2 values are 95%, 92%, and 86%, respectively, which gives an accurate prediction.

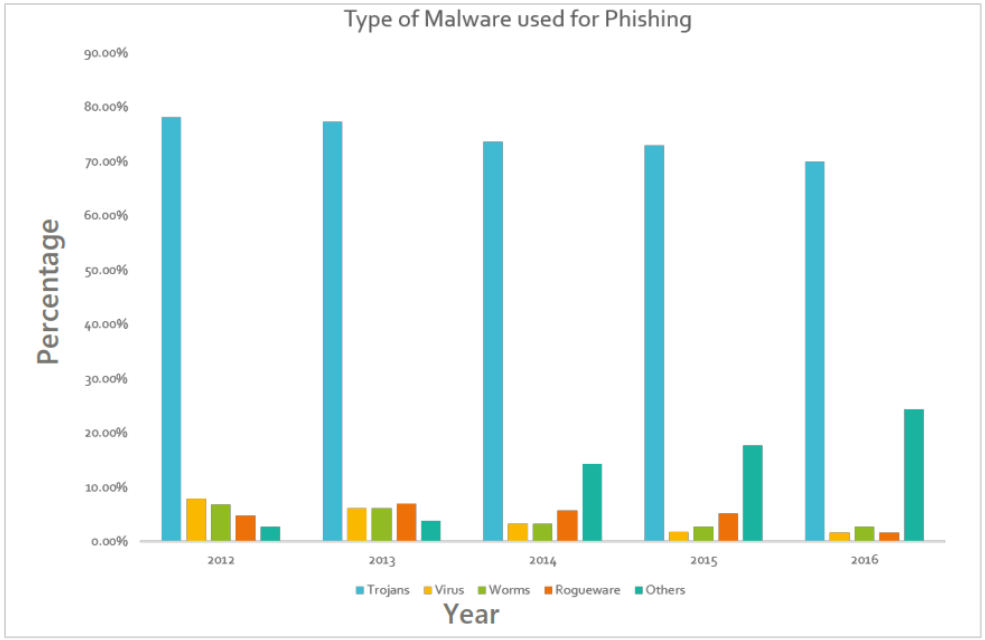

Figure 20 Type of the malware used for phishing

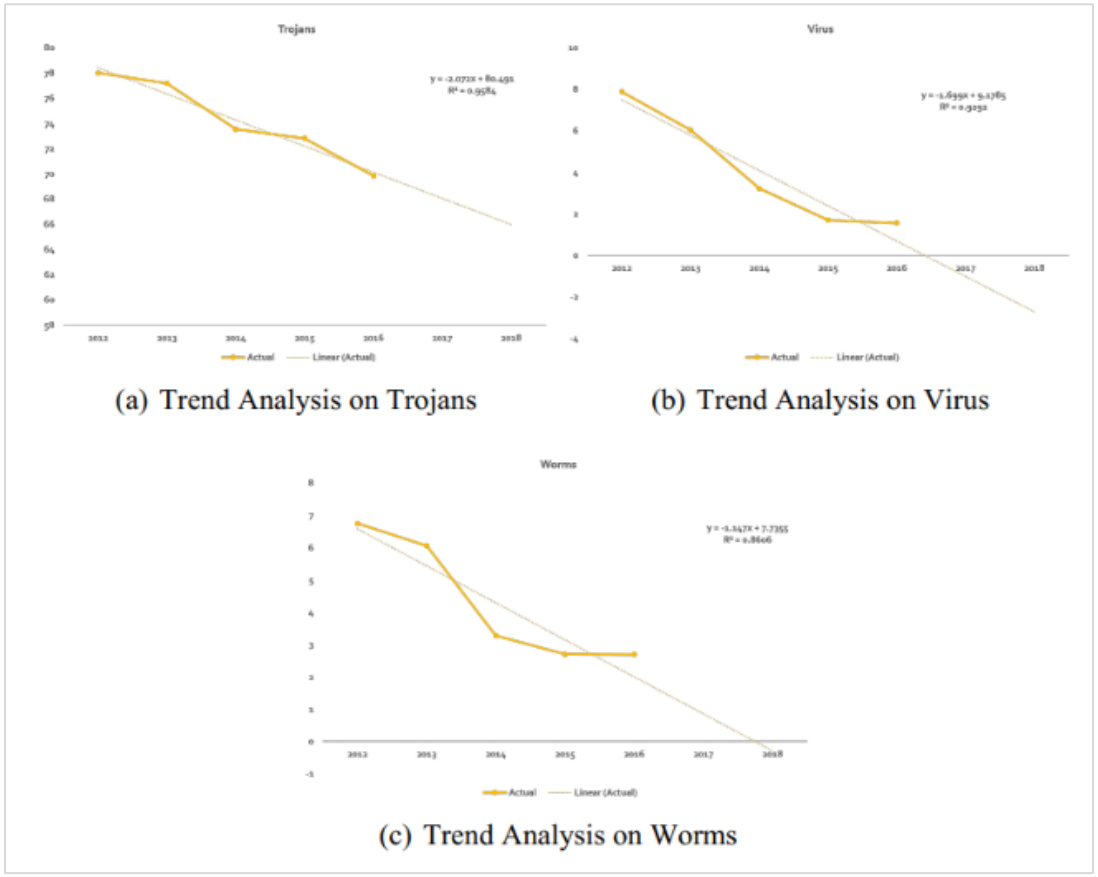

Figure 21 Trend analysis of top three malware

6.6HTTPS enabled phishing URLs

Result

From the *Figure 22* it clear that phishing URLs with HTTPS protection is increasing.

Inference

Nowadays, most phishing URLs contain SSL certificates that fool the victims into trusting the URLs as protected ones. The percentage of phishing URLs with SSL certificates is increasing gradually. In January-2019, the dataset downloaded from PhishTank [94] contains 15000 phishing URLs among that 4200 URLs (i.e., 28%) are HTTPS-enabled and shown visually in *Figure 22(a).* In February-2020, the URLs with HTTPS protection reached 43% as shown in *Figure 22(b).* The Phishing URLs are collected from OpenPhish in the month of February-2020 for 15 days and out of it, 45% of URLs are HTTPS protection enabled [93]. The results are shown in *Figure 22(c).* So, the HTTPS enabled URLs also have to be checked for security very exhaustively to verify if they belong to phishing URLs.

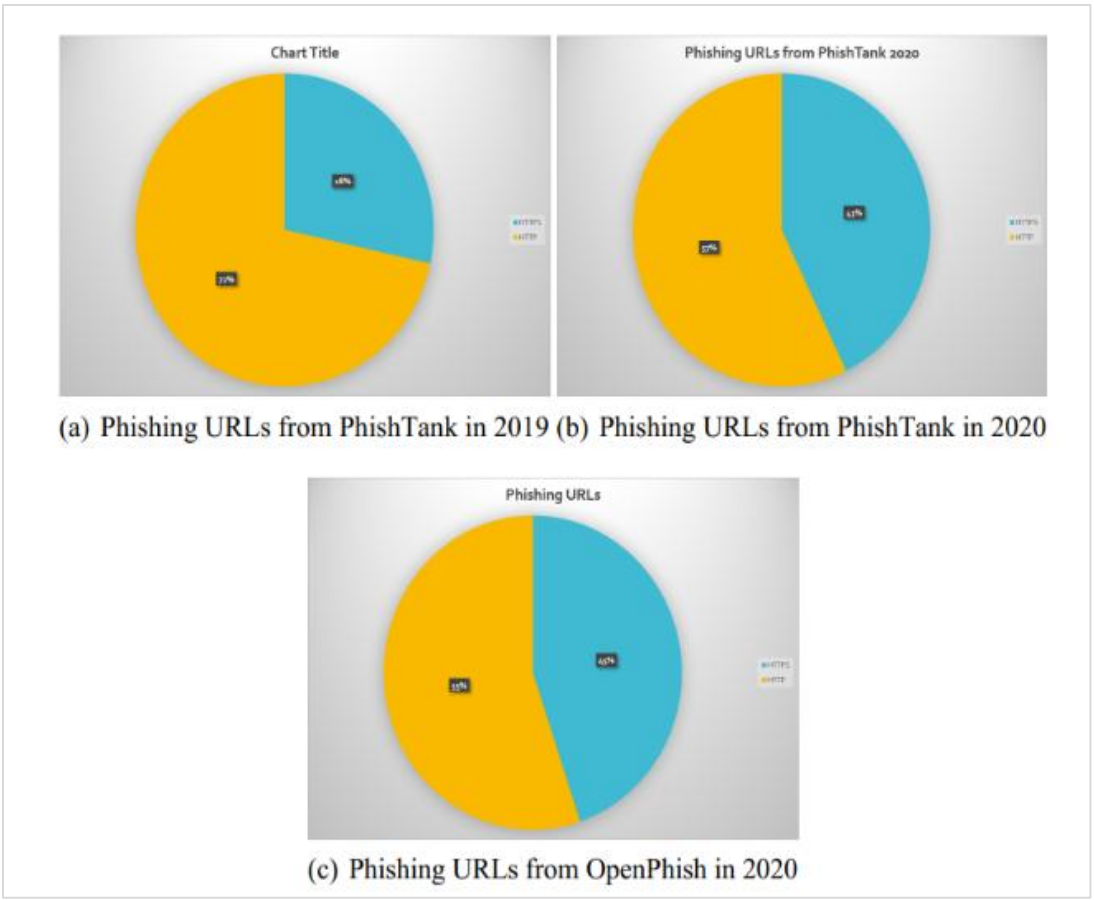

Figure 22 Comparing HTTPS enabled phishing URLs with other phishing URLs

6.7Top five ransomware affected countries with five different ransomware families

Ransomware becomes more dangerous and cause lots of damage to the Internet users. *Figure 23* shows top five different types of ransomware attacks in the top five counties during the year 2016 [97].

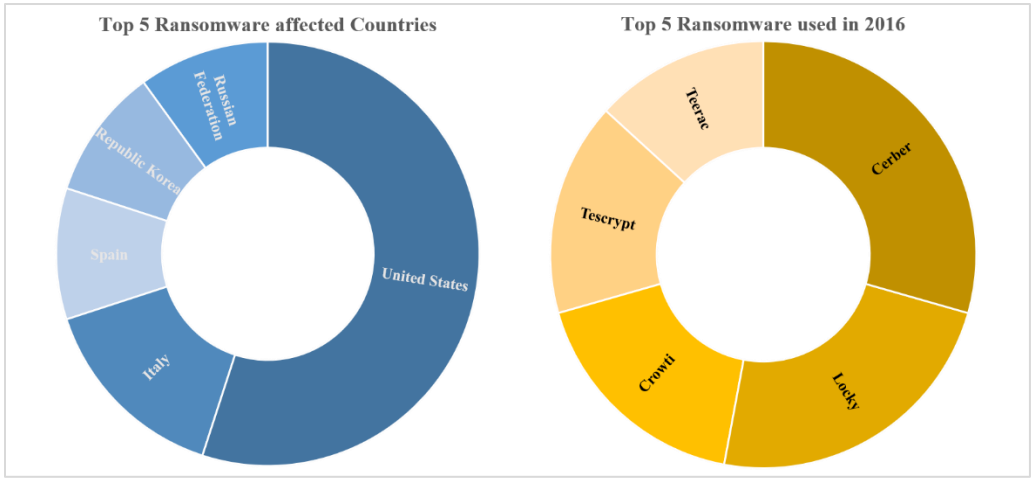

Figure 23 Top 5 countries affected with 5 different ransomware attacks

6.8Top twenty countries affected by WannaCry ransomware

According to [98], Russia is the most affected country with more than 70 percent of systems affected with WannaCry ransomware. The Railway system and the Ministry of Internal Affairs of the Russian Federation are mainly affected in Russia. In India, the Andhra Pradesh Police system, Government of Gujarat, Government of Kerala, Government of Maharashtra, Government of West Bengal are affected. There are many other counties like China, Spain, Italy, Brazil etc., which are affected due to WannaCry ransomware. *Figure 24* shows the top 20 counties affected by WannaCry Ransomware.

A complete list of abbreviations is shown in *Appendix I*.

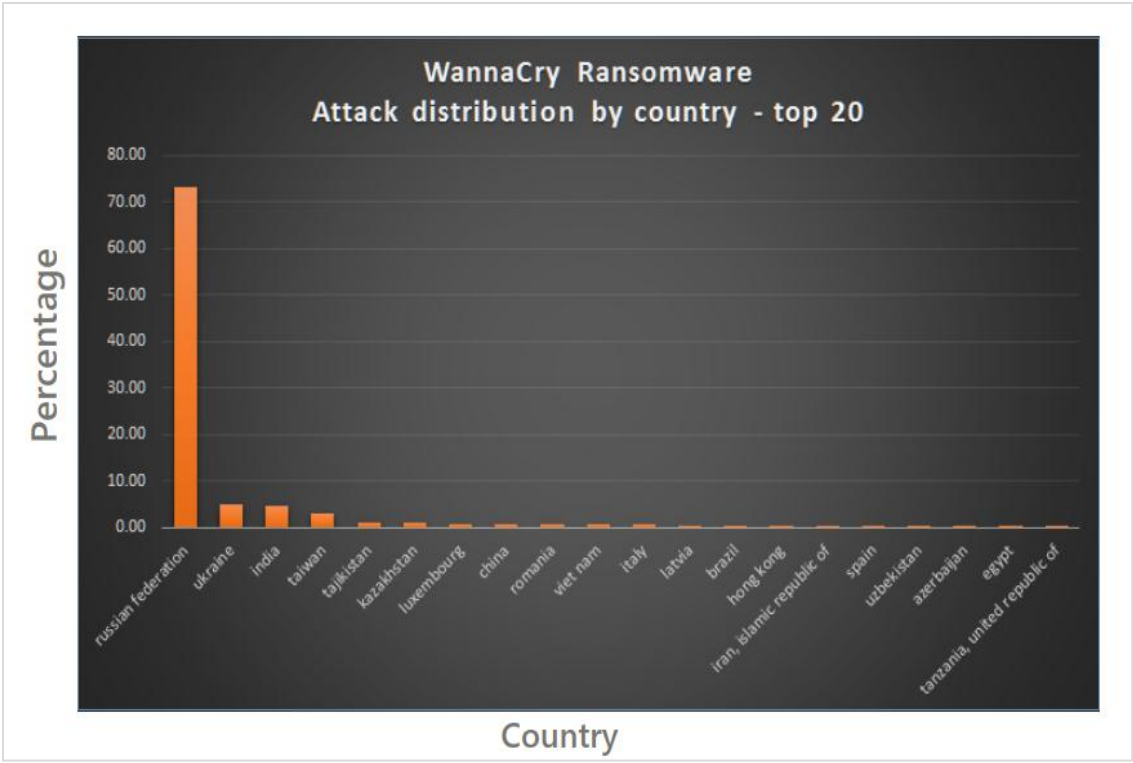

Figure 24 Top 20 countries affected by WannaCry ransomware [98]

7.Major challenges to be addressed

These challenges have been identified by means of our complete survey and study of the phishing attacks, which have been explained in this paper:

- Phishing scams on SNS like Facebook, Instagram, Twitter, etc., are increasing drastically.
- In SNS, the attackers spread fake news, malicious URLs & attachments that gain unauthorized access to the victim's computer. In Twitter, the attacker creates malicious bot accounts and spread phishing. These bot accounts are automated codes that perform particular tasks without human interaction.
- Our study found that detecting phishing attacks on SNS is much complicated compared to other sectors. At first, we have to see whether the URL posted is phishing or not. Knowing the URL is phishing or legitimate is not enough; we have to see the origin of the post (i.e., whether a human or bot poses it).
- More focus is required to identify the bot accounts on SNS that spread phishing. Even though there are plenty of works available in detecting bot accounts, very few works are available to detect bot accounts from a phishing perspective. More focus is needed to develop a better technique to identify bot accounts on SNS that spread phishing.
- Since the majority of the internet users in the world use mobile phones to access the SNS, it is essential to protect the privacy of the users.
- The attacker Develops cloned domains that look exactly same as some popular companies/organizations and fools the users to visit their spoofed links to gather the personal credentials for malicious intentions.
- Machine learning-based phishing detection approaches are the commonly used approach with better performance. The main challenge in machine learning-based techniques is to increase the model's performance by minimizing the number of features. More focus is needed in filtering the minimal set of features that improves the performance of the machine learning models.
- Malware-based phishing attacks are challenging to detect as it installs and works in the background without the users' knowledge. Even though the machine learning-based techniques perform better, there is still a scope to improve the performance by minimizing the number of features.
- The majority of the anti-phishing techniques fail in identifying phishing attacks because of the constant improvement in the attacking strategy followed by the attacker.
- Since the life span of the Phishing URLs is very short, it is challenging to have a standard dataset to analyse the behaviour and pattern followed by the attacker for executing the phishing attack.

8.Discussion

This paper discusses a comprehensive classification of phishing attacks. The phishing attacks are grouped into two types: social engineering-based and malware-based, and each type is explained in detail. The following are the responses to the research questions raised in this paper:

RQ1. There are so many definitions of phishing given in various sources. What would be the concise definition of phishing, which encompasses the semantics of most of the definitions?

There are different sources for the definition of phishing, and for the time being, new phishing attacks are identified, so the definition of phishing needs to be updated. A consolidated definition is proposed to cover all the possible means of performing a phishing attack encompassing the existing definitions. The proposed definition of phishing is: phishing a fraudulent activity in which the attacker tries to gain illegal financial gain either by:

– stealing and spoofing user identity/credentials or
– usurping control of access to user information.

The proposed definition of phishing also helps in the complete classification of phishing attacks based on the intention of phishing, as explained in section 5.

RQ2. What could be the possible classification of all the phishing attacks starting from the oldest to the most recent phishing attacks?

The complete classification of all phishing attacks from oldest to the most recent attack is wholly based on the intention of phishing. The early phishing attack started with a phishing email that looked authentic and fooled internet users by stealing their personal credentials. Over the period, the attacker has improved their attacking strategies like spreading fake news, phishing through phone calls, SMS phishing, malware-based phishing, and so on to the very recent ransomware attack. The proposed classification of phishing cover most of these phishing attacks into their respective groups, as shown in Section 5:

– Social engineering-based phishing
– Malware based phishing

– Phishing through ransomware

RQ3. What are the current statistics on phishing attacks with respect to the impact on different countries, most prevalent kind of phishing attack, most targeted industry sectors, etc., around the globe? From our study, the impact of phishing attacks on internet users is explained clearly in section 6. The points identified from the phishing statistics are:

– China is the most affected country by phishing, with 51.46% of overall phishing attacks reported.
– America stands at the top position in hosting new phishing attacks
– Payment, financial, and retail/service sectors are the most targeted industry sectors by the phisher.
– Out of 100%, 45% of phishing URLs are HTTPS protected, which tells us that identifying phishing URLs from legitimate URLs is very difficult.
– Phishing through ransomware becomes much more dangerous to internet users. Because the attacker uses symmetric key encryption to encrypt the victims' system, since the attacker uses the same key for encryption and decryption, only the attacker can decrypt the data. The attacker demands a ransom to provide the decryption key.

RQ4. What are the major challenges to be addressed in phishing attacks?

Based on our complete classification of phishing attacks survey, we identified the major challenges that must be addressed in the future, and these challenges are explained in section 7.

Current trends in phishing

As the technology advances, the attacking strategy of the phisher also advances. As a result, the attacker deploys new techniques and methods to lure the internet users. Based on the recent studies on phishing, it is found that the attackers adopt new trends to perform phishing [99–101]:

– Targeting the companies with ransomware attacks has been increased. According to APWG trend report, manufacturing industries are the most affected industry sector by ransomware [101].
– The attackers not only steal the personal credentials of the users, but also the documents related to their identity (like voter card, driving license etc.).
– Targeting the top brands by developing the spoofed site with seemingly authentic URLs to redirect the victims.

9.Conclusion and future work

In this paper, we have discussed about phishing, pharming, a new definition of phishing and presented a complete classification of phishing based on the intention of phishing. This paper provides the complete classification of phishing attacks and provides the different possible ways of performing phishing through different means (i.e., through email, advertisements, Instant messaging, phone calls, social media sites, malware, website, DNS etc.).

Some real-time phishing attacks corresponding to every type of attack described in the classification are also included. The statistical analysis of phishing attacks is done with the attack information extracted from APWG survey reports and OpenPhish phishing feeds. The analysis found that China is the most affected country by phishing, whereas America is top in hosting phishing. Payment sectors are the most targeted industry sector by phishing, and Trojans are the most popularly used malware to perform phishing. It is also found that more than 45% of phishing URLs are HTTPS protected. The results are illustrated graphically along with Trend analysis.

In future studies, we will focus more on the phishing attacks on social networking sites and mobile phones. Most of the social media users use their mobile phone to access Twitter, Facebook, Instagram etc.

Limitations

In this paper, we focused only on phishing attacks and the complete classification of phishing attacks, some real-time phishing scams reported globally, and analysis of these attacks. We did not cover the anti-phishing approaches that help detect phishing attacks, different types of approaches, and the most commonly used technique for phishing detection.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

S. Chanti: Developed the complete classification of phishing attacks and performed the statistical analysis. **T. Chithralekha:** Helped in improving the work by providing critical feedback to shape the manuscript.

References

[1] Kirda E, Kruegel C. Protecting users against phishing attacks with antiphish. In annual international computer software and applications conference 2005 (pp. 517-24). IEEE.

[2] http://www.phishing.org/history-of-phishing. Accessed 19 February 2018.

[3] Mei Y. Anti-phishing system: detecting phishing e-mail. School of Mathematics and Systems Engineering. 2008.

[4] https://dictionary.cambridge.org/dictionary/english/phishing. Accessed 8 March 2022.

[5] Yadav S, Bohra B. A review on recent phishing attacks in internet. In international conference on green computing and internet of things 2015 (pp. 1312-5). IEEE.

[6] IRONSCALES. How modern email phishing attacks have organization on the hook. 2017.

[7] APWG. APWG phishing trends report 2nd quarter 2021. 2021.

[8] Alfayoumi IS, Barhoom TS. Client â [euro]" Side pharming attacks detection using authoritative domain name servers. International Journal of Computer Applications. 2015; 113(10):26-31.

[9] Ollmann G. The vishing guide. IBM Global Technology Services. 2007:1-16.

[10] PhishMe. Q1 2016 malware review. 2016; 1–15.

[11] https://www.ic3.gov/Media/PDF/AnnualReport/2015_IC3Report.pdf. Accessed 8 March 2022.

[12] Anti-phishing working group. APWG Phishing activity trends report, 2nd quarter 2012.

[13] Chanti S, Chithralekha T. Classification of anti-phishing solutions. SN Computer Science. 2020; 1(1):1-8.

[14] James D, Philip M. A novel anti phishing framework based on visual cryptography. In international conference on power, signals, controls and computation 2012 (pp. 1-5). IEEE.

[15] Krishnakumar L, Varughese NM. High speed classification of vulnerabilities in cloud computing using collaborative network security management. In international conference on advanced computing and communication systems 2013 (pp. 1-6). IEEE.

[16] https://www.bbc.com/news/world-us-canada-41116177. Accessed 30 June 2020.

[17] Musashi Y, Kumagai M, Kubota S, Sugitani K. Detection of Kaminsky DNS cache poisoning attack. In international conference on intelligent networks and intelligent systems 2011 (pp. 121-4). IEEE.

[18] https://www.cisa.gov/uscert/ncas/alerts/TA18-201A. Accessed 5 April 2022.

[19] Arshad A, Rehman AU, Javaid S, Ali TM, Sheikh JA, Azeem M. A systematic literature review on phishing and anti-phishing techniques. arXiv preprint arXiv:2104.01255. 2021.

[20] 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. 2021; 9:80866-72.

[21] Sadiq A, Anwar M, Butt RA, Masud F, Shahzad MK, Naseem S, Younas M. A review of phishing attacks and countermeasures for internet of things-based smart business applications in industry 4.0. Human Behavior and Emerging Technologies. 2021; 3(5):854-64.

[22] Jain AK, Gupta BB. A survey of phishing attack techniques, defence mechanisms and open research challenges. Enterprise Information Systems. 2021:1-39.

[23] Aleroud A, Zhou L. Phishing environments, techniques, and countermeasures: a survey. Computers & Security. 2017; 68:160-96.

[24] Almomani A, Gupta BB, Atawneh S, Meulenberg A, Almomani E. A survey of phishing email filtering techniques. IEEE Communications Surveys & Tutorials. 2013; 15(4):2070-90.

[25] Chiew KL, Yong KS, Tan CL. A survey of phishing attacks: their types, vectors and technical approaches. Expert Systems with Applications. 2018; 106:1-20.

[26] Gupta S, Singhal A, Kapoor A. A literature survey on social engineering attacks: Phishing attack. In international conference on computing, communication and automation 2016(pp. 537-40). IEEE.

[27] Jampen D, Gür G, Sutter T, Tellenbach B. Don't click: towards an effective anti-phishing training. A comparative literature review. Human-centric Computing and Information Sciences. 2020; 10(1):1-41.

[28] Khonji M, Iraqi Y, Jones A. Phishing detection: a literature survey. IEEE Communications Surveys & Tutorials. 2013; 15(4):2091-121.

[29] Lastdrager EE. Achieving a consensual definition of phishing based on a systematic review of the literature. Crime Science. 2014; 3(1):1-10.

[30] https://www.oxfordlearnersdictionaries.com/definition/american_english/phishing. Accessed 5 April 2020.

[31] https://www.phishtank.com/what_is_phishing.php. Accessed 19 February 2020.

[32] Gupta BB, Tewari A, Jain AK, Agrawal DP. Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications. 2017; 28(12):3629-54.

[33] https://cofense.com/phishing-ransomware-threats-soared-q1-2016/. Accessed 19 February 2020.

[34] https://www.oxfordlearnersdictionaries.com/definition/english/ransomware. Accessed 20 March 2020.

[35] https://dictionary.cambridge.org/dictionary/english/ransomware. Accessed 20 March 2022.

[36] https://www.trendmicro.com/vinfo/us/security/definition/ransomware. Accessed 19 February 2018.

[37] Murphy R. How does ransomware work. Retrieved from. 2017.

[38] https://digitalguardian.com/blog/what-is-ransomware-and-how-to-protect-against-attacks. Accessed 19 February 2018.

[39] Mouton F, Leenen L, Malan MM, Venter HS. Towards an ontological model defining the social engineering domain. In IFIP international conference

on human choice and computers 2014 (pp. 266-79). Springer, Berlin, Heidelberg.
[40] Culpepper AM. Effectiveness of using red-teams to identify maritime security vulnerabilities to terrorist attack. Naval Postgraduate School Monterey Ca; 2004.
[41] Bhakta R, Harris IG. Semantic analysis of dialogs to detect social engineering attacks. In proceedings of the international conference on semantic computing 2015 (pp. 424-7). IEEE.
[42] Emigh A. The crimeware landscape: Malware, phishing, identity theft and beyond. Journal of Digital Forensic Practice. 2006; 1(3):245-60.
[43] Huang H, Zhong S, Tan J. Browser-side countermeasures for deceptive phishing attack. In fifth international conference on information assurance and security 2009 (pp. 352-5). IEEE.
[44] https://www.phishing.org/phishing-techniques. Accessed 21 March 2021.
[45] https://usa.kaspersky.com/resource-center/definitions/spear-phishing. Accessed 19 February 2018.
[46] Caputo DD, Pfleeger SL, Freeman JD, Johnson ME. Going spear phishing: Exploring embedded training and awareness. IEEE Security & Privacy. 2013; 12(1):28-38.
[47] Castiglione A, Prisco RD, Santis AD. Do you trust your phone? In international conference on electronic commerce and web technologies 2009 (pp. 50-61). Springer, Berlin, Heidelberg.
[48] Silic M, Back A. The dark side of social networking sites: understanding phishing risks. Computers in Human Behavior. 2016; 60:35-43.
[49] Abad C. The economy of phishing: a survey of the operations of the phishing market. 2005.
[50] Ganesan S. Detection of phishing websites using classification algorithms. In cyber security and digital forensics 2022 (pp. 129-41). Springer, Singapore.
[51] Suri RK, Tomar DS, Sahu DR. An approach to perceive tabnabbing attack. International Journal of Scientific & Technology Research. 2012; 1:90-4.
[52] Singh A, Tripathy S. TabSol: an efficient framework to defend Tabnabbing. In international conference on information technology 2014 (pp. 173-8). IEEE.
[53] Li X, Geng G, Yan Z, Chen Y, Lee X. Phishing detection based on newly registered domains. In international conference on big data 2016 (pp. 3685-92). IEEE.
[54] Chen G, Johnson MF, Marupally PR, Singireddy NK, Yin X, Paruchuri V. Combating typo-squatting for safer browsing. In international conference on advanced information networking and applications workshops 2009 (pp. 31-6). IEEE.
[55] Patel J, Panchal SD. A survey on pharming attack detection and prevention methodology. IOSR Journal of Computer Engineering. 2013; 9(1):66-72.
[56] Emilin SC. Detecting and preventing phishing websites DPPWS. Anna University.2014.
[57] https://securelist.com/the-rio-olympics-scammers-already-competing/74754/. Accessed 19 December 2019.
[58] Mishra M, Jain A. Anti-phishing techniques: a review. International Journal of Engineering Research and Applications.2012; 2(2):350-5.
[59] Zhenfang ZH. Study on computer Trojan horse virus and its prevention. International Journal of Engineering and Applied Sciences. 2015; 2(8):257840.
[60] Li C, Jiang W, Zou X. Botnet: Survey and case study. In fourth international conference on innovative computing, information and control 2009 (pp. 1184-7). IEEE.
[61] Micro T. Botnet threats and solutions: phishing.2006.
[62] Damopoulos D, Kambourakis G, Gritzalis S. From keyloggers to touchloggers: take the rough with the smooth. Computers & security. 2013; 32:102-14.
[63] Divya R, Muthukumarasamy S. An impervious QR-based visual authentication protocols to prevent black-bag cryptanalysis. In 9th international conference on intelligent systems and control 2015 (pp. 1-6). IEEE.
[64] Yaokumah W. Predicting and explaining cyber ethics with ethical theories. International Journal of Cyber Warfare and Terrorism. 2020; 10(2):46-63.
[65] Gastellier-Prevost S, Laurent M. Defeating pharming attacks at the client-side. In 5th international conference on network and system security 2011 (pp. 33-40). IEEE.
[66] Gastellier-Prevost S, Granadillo GG, Laurent M. Decisive heuristics to differentiate legitimate from phishing sites. In conference on network and information systems security 2011 (pp. 1-9). IEEE.
[67] Jackson C, Barth A, Bortz A, Shao W, Boneh D. Protecting browsers from DNS rebinding attacks. ACM Transactions on the Web. 2009; 3(1):1-26.
[68] Sarbazi-Azad H, Zomaya AY. Large scale network-centric distributed systems. John Wiley & Sons; 2013.
[69] Kim YG, Cho S, Lee JS, Lee MS, Kim IH, Kim SH. Method for evaluating the security risk of a website against phishing attacks. In international conference on intelligence and security informatics 2008 (pp. 21-31). Springer, Berlin, Heidelberg.
[70] Kaur D, Kaur P. Empirical analysis of web attacks. Procedia Computer Science. 2016; 78:298-306.
[71] Houser R, Hao S, Li Z, Liu D, Cotton C, Wang H. A comprehensive measurement-based investigation of DNS hijacking. In international symposium on reliable distributed systems 2021 (pp. 210-21). IEEE.
[72] Karlof CK. Human factors in web authentication. University of California, Berkeley; 2009.
[73] Stamm S, Ramzan Z, Jakobsson M. Drive-by pharming. In international conference on information and communications security 2007 (pp. 495-506). Springer, Berlin, Heidelberg.
[74] Gastellier-Prevost S, Granadillo GG, Laurent M. A dual approach to detect pharming attacks at the client-side. In IFIP international conference on new technologies, mobility and security 2011 (pp. 1-5). IEEE.

[75] Purkait S. DHCP-enabled LAN prone to phishing attacks. IUP Journal of Information Technology. 2013; 9(1):24-40.
[76] Steadman J, Scott-Hayward S. Dnsxd: detecting data exfiltration over DNS. In conference on network function virtualization and software defined networks (NFV-SDN) 2018 (pp. 1-6). IEEE.
[77] Farnham G, Atlasis A. Detecting DNS tunneling. SANS Institute InfoSec Reading Room. 2013; 9:1-32.
[78] Maksutov AA, Cherepanov IA, Alekseev MS. Detection and prevention of DNS spoofing attacks. In Siberian symposium on data science and engineering 2017 (pp. 84-7). IEEE.
[79] Jaworski S. Using splunk to detect DNS tunneling. SANS Institute InfoSec Reading Room. 2016.
[80] Steinhoff U, Wiesmaier A, Araújo R. The state of the art in DNS spoofing. In proceeding of international conferences applied cryptography and network security (ACNS) 2006.
[81] McGrath DK, Kalafut A, Gupta M. Phishing infrastructure fluxes all the way. IEEE Security & Privacy. 2009; 7(5):21-8.
[82] Holz T, Gorecki C, Rieck K, Freiling FC. Measuring and detecting fast-flux service networks. InNdss 2008.
[83] Zhou S. A survey on fast-flux attacks. Information Security Journal: A Global Perspective. 2015; 24(4-6):79-97.
[84] Gupta M. Pharming attack designs. In encyclopedia of information ethics and security 2007 (pp. 520-6). IGI Global.
[85] Kathrine GJ, Praise PM, Rose AA, Kalaivani EC. Variants of phishing attacks and their detection techniques. In international conference on trends in electronics and informatics 2019 (pp. 255-9). IEEE.
[86] Blasi M. Techniques for detecting zero day phishing websites. Iowa State University; 2009.
[87] Bu SJ, Cho SB. Integrating deep learning with first-order logic programmed constraints for zero-day phishing attack detection. In international conference on acoustics, speech and signal processing 2021 (pp. 2685-9). IEEE.
[88] Ronald F. Clayton. E Y Technical Intelligence Analysis - WannaCry Attack. 2017.
[89] Tandon A, Nayyar A. A comprehensive survey on ransomware attack: a growing havoc cyberthreat. Data Management, Analytics and Innovation. 2019:403-20.
[90] Mohurle S, Patil M. A brief study of wannacry threat: Ransomware attack 2017. International Journal of Advanced Research in Computer Science. 2017; 8(5):1938-40.
[91] https://blog.google/threat-analysis-group/findings-covid-19-and-online-security-threats/. Accessed 20 March 2022.
[92] https://apwg.org/trendsreports/. Accessed 20 March 2022.
[93] https://openphish.com/. Accessed 20 March 2022.
[94] http://www.phishtank.com/index.php. Accessed 20 March 2022.
[95] Cameron AC, Windmeijer FA. An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of Econometrics. 1997; 77(2):329-42.
[96] Akossou AY, Palm R. Impact of data structure on the estimators R-square and adjusted R-square in linear regression. International Journal of Mathematics Computation. 2013; 20(3):84-93.
[97] https://www.microsoft.com/security/blog/2017/02/14/ransomware-2016-threat-landscape-review/. Accessed 29 August 2021.
[98] https://securelist.com/wannacry-ransomware-used-in-widespread-attacks-all-over-the-world/78351/. Accessed 19 February 2018.
[99] Das S, Nippert-Eng C, Camp LJ. Evaluating user susceptibility to phishing attacks. Information & Computer Security. 2022.
[100] Abuadbba A, Wang S, Almashor M, Ahmed ME, Gaire R, Camtepe S, et al. Towards web phishing detection limitations and mitigation. arXiv preprint arXiv:2204.00985. 2022.
[101] APWG. APWG phishing trends report 4th quarter 2021.

S. Chanti received his M.Sc. Computer Science from Pondicherry University in 2015. He is currently pursuing his Ph.D. in the Department of Banking Technology, Pondicherry University, Pondicherry, India. Interested areas are Information Security, Machine Learning, and Blockchain Technology. Present three papers in various national and international conferences. Published two papers and one book chapter in international journals.
Email: schanti14@gmail.com

T. Chithralekha received her MTech Computer Science Engineering from Pondicherry University. She also received her Doctorate in Computer Science and Engineering from the same University. She is currently working as a Professor in the Department of Computer Science, Pondicherry University, Puducherry, India. She has published a number of papers in international conferences and journals and her research interests include information security for Banking and Financial sectors, Machine learning and Multi-agent Systems.
Email: tchithralekha@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	AOL	American Online
2	APWG	Anti-Phishing Working Group
3	ATM	Automated Teller Machine
4	CPU	Central Processing Unit
5	CVV	Card Verification Value
6	DHCP	Dynamic Host Configuration Protocol
7	DNS	Domain Name System
8	HTTPS	Hypertext Transfer Protocol Secure

9	ICR	Internet Crime Report
10	IP	Internet Protocol
11	ISP	Internet Service Provider
12	IVR	Interactive Voice Response
13	NSA	National Security Agency
14	OTP	One Time Password
15	PC	Personal Computer
16	RR	Resource Record
17	SaaS	Software-as-a-Service
18	SMBv2	Server Message Block Protocol
19	SMEs	Small and Medium Scale Enterprises
20	SMS	Short Message Service
21	SNS	Social Networking Sites
22	SOP	Same Origin Policy
23	SSL	Secure Sockets Layer
24	TCP	Transmission Control Protocol
25	TTL	Time to Live
26	UDP	User Datagram Protocol
27	URL	Uniform Resource Locator
28	VoIP	Voice Over Internet Protocol