Telecommun Syst (2018) 67:247–267
https://doi.org/10.1007/s11235-017-0334-z

Defending against phishing attacks: taxonomy of methods, current issues and future directions

B. B. Gupta[1] · **Nalin A. G. Arachchilage**[2] · **Kostas E. Psannis**[3]

Published online: 23 May 2017

Abstract Internet technology is so pervasive today, for example, from online social networking to online banking, it has made people's lives more comfortable. Due the growth of Internet technology, security threats to systems and networks are relentlessly inventive. One such a serious threat is "phishing", in which, attackers attempt to steal the user's credentials using fake emails or websites or both. It is true that both industry and academia are working hard to develop solutions to combat against phishing threats. It is therefore very important that organisations to pay attention to end-user awareness in phishing threat prevention. Therefore, aim of our paper is twofold. First, we will discuss the history of phishing attacks and the attackers' motivation in details. Then, we will provide taxonomy of various types of phishing attacks. Second, we will provide taxonomy of various solutions proposed in literature to protect users from phishing based on the attacks identified in our taxonomy. Moreover, we have also discussed impact of phishing attacks in Internet of Things (IoTs). We conclude our paper discussing various issues and challenges that still exist in the literature, which are important to fight against with phishing threats.

Keywords Phishing · Security · Malware · Social engineering · Spam · Visual similarity · Data mining · Machine learning

✉ B. B. Gupta
gupta.brij@gmail.com

Nalin A. G. Arachchilage
nalin.asanka@adfa.edu.au

Kostas E. Psannis
kpsannis@uom.gr

[1] Department of Computer Engineering, National Institute of Technology Kurukshetra, Kurukshetra, India

[2] Australian Centre for Cyber Security (ACCS), Australian Defence Force Academy, The University of New South Wales, PO Box 7916, Canberra BC, ACT 2610, Australia

[3] Department of Applied Informatics, School of Information Sciences, University of Macedonia, 54006 Thessaloníki, Greece

1 Introduction

Over the past years, 'identity theft' has been one of the most threatening and lucrative crime of the all Internet user [1]. It refers to the stealing and use of a person's identity information (i.e. credit cards numbers, bank details or identity number, etc) in order to steal money or commit other crimes [2]. One such type of crime is phishing attack. Phishing is a social engineering crime, whose spread occurrences and technique advancements pose enormous challenges for researchers in both academia and the industry. Phishing aims to steal sensitive information such as username, password and online banking details from victims. Figure 1, shows the growth of Internet user which is exponential per year [3,4].

Figure 2 shows the size of the Internet users in the world by various geographic regions. This is the recent information according to Internet World Stats [5]. Similarly, Fig. 3 shows Internet Penetration Rates in the World by Geographic regions according to Internet World Stats [5].

According to [6], in the year 2013 phishing attacks were the most serious threats. According to Wombat Security's 2016 State of the Phish report, not only are more organizations falling victim to phishing attacks, the number and sophistication level of the attacks they are experiencing has gone up. Among all the organizations they studied two-thirds were reported to have gone through phishing attacks either personalized or targeted. PhishMe, an anti-phishing vendor

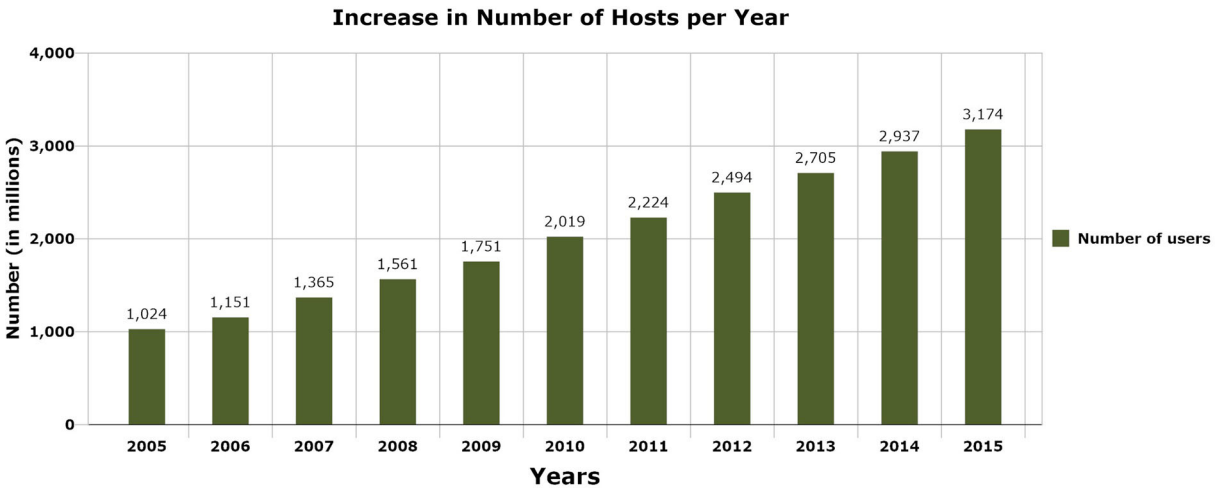
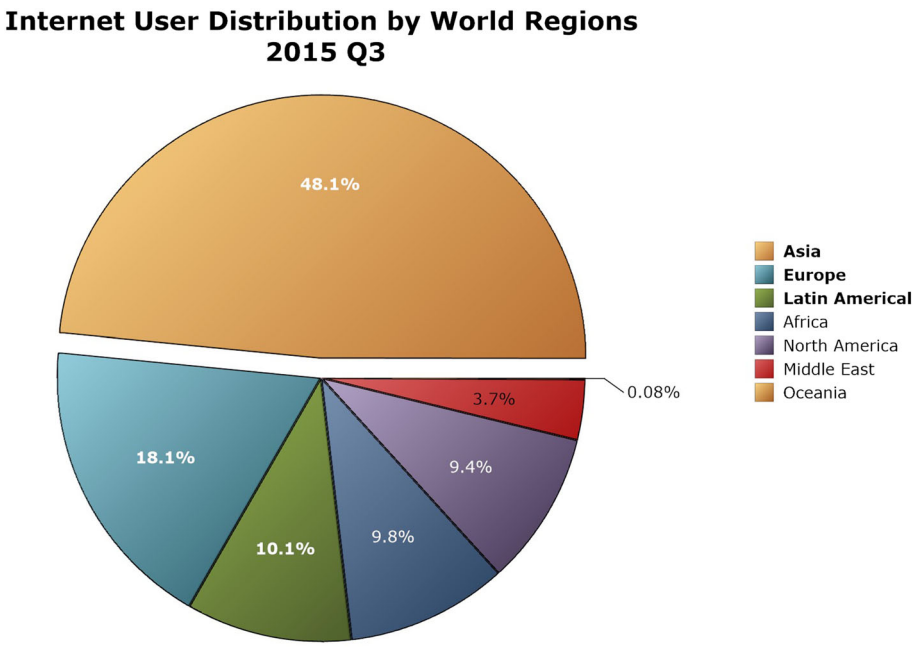


Fig. 1 Growth in number of internet users from 2000 to 2015 [3,4]

Fig. 2 Estimated internet users (in Millions) in the World by geographic regions [5]



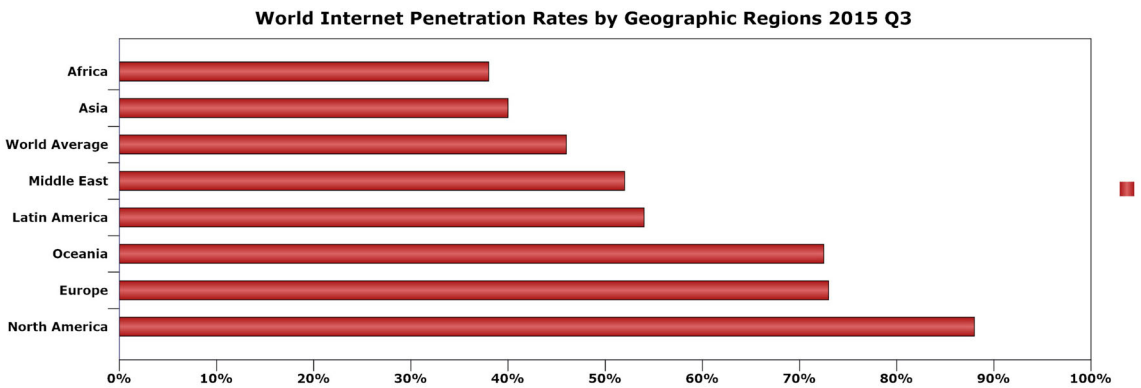


Fig. 3 Estimated world internet penetration rates in the world by geographic regions [5]

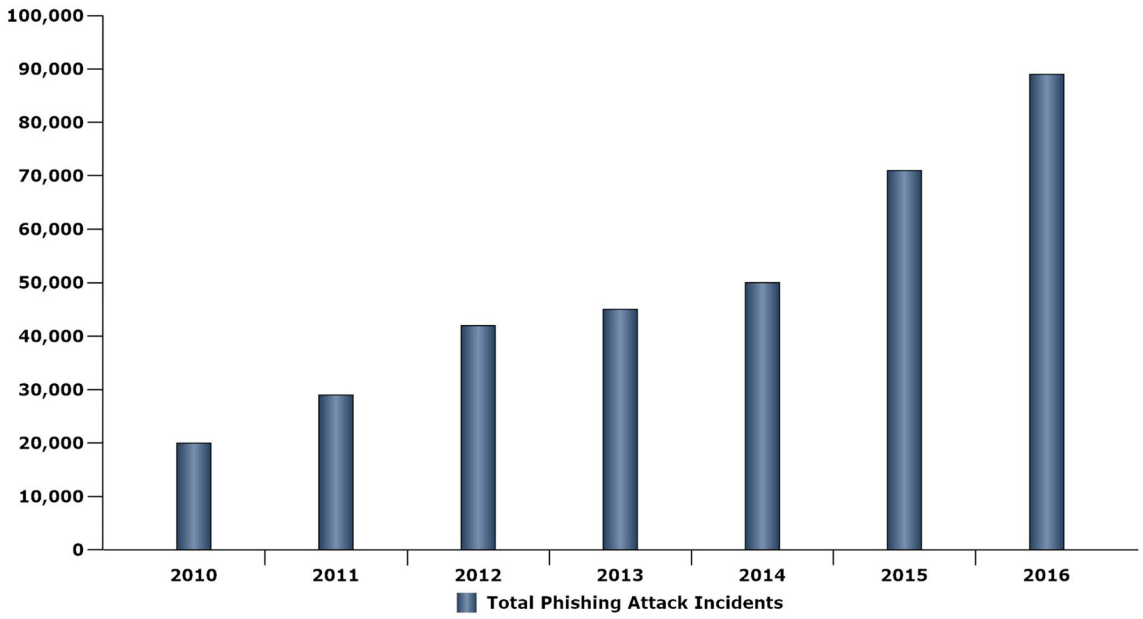


Fig. 4 Phishing attack incidents

reported in first quarter of 2016 that every 9 out of 10 phishing emails had a ransomware attached to it [7].

In the fourth quarter of 2013 Malcovery reported that Facebook, WhatsApp, UPS, Fargo and Companies House (UK) were the top five organizations targeted by the Phishers [8]. Sheng et al. showed that, women were more likely to be a victim of phishing than men. Similarly, people belonging to age group 18–25 years are likely to be a phishing victim [9–11]. Figure 4, shows the vigorous growth in phishing attacks over last few years (RSA monthly online fraud reports [12]).

The United States Computer Emergency Readiness Team (CERT) gathered security details from various agencies, which stated that there were 107,655 incident in 2011, 43,889 of which were on federal agencies [13]. In May of 2015, construction, engineering, transportation and telecommunication sectors were a target of Advanced Persistent Threat (APT3) phishing campaigns. FireEye identified it to be a zero day attack. The employees received phishing emails having malicious URLs, upon clicking them they redirected to compromised web server, and the target system downloaded an infected Adobe Flash Player SWF file and FLV file which made a backdoor [14].

Hillary Clinton presidential campaign chairman, John Podesta's Google email account was "hacked" in March 2016 prior to the US election [15]. To lure Podesta the hackers sent a phishing mail to his personal gmail account which made him give up his login credentials. In the phishing email, Podesta had been invited to click on a link (i.e. unified resource locator, so called "URL") warning him to change his password immediately. However, the URL did not link to a secure Google web page, instead directing the user blindly via bit.ly, which is a service used to shorten URLs. Podesta hack didn't require much technical skills as the hackers merely had to use basic social engineering skills to conduct the attack. Although, the simple techniques do not make this attack any less serious or illegal

Therefore, the aim of this paper is to look at the current phishing literature to determine seriousness of the problem. To give a brief overview of evolution of research in this field as well as current trends in phishing and its remedies to provides a view of the issues and challenges that are still prevailing in this area of research. In this paper, we describe the phishing attacks alongwith its history and attackers motivation to carry out these attacks. We will also provide taxonomy of various types of phishing attacks. We also present taxonomy of phishing detection techniques. Moreover, we have also discussed impact of phishing attacks in Internet of Things (IoTs).

Rest of the paper is organized as follows. In Sect. 2 we give an overview of phishing attack. Taxonomy of phishing attacks is then discussed in Sect. 3. Section 4 presents taxonomy of defence solutions. Phishing attacks in the Internet of things (IoTs) are discussed in Sect. 5. Current issues and challenges are discussed in Sect. 6. Finally, in Sect. 7 we conclude the paper and discuss the scope of future research in this field.

2 Phishing attack overview

2.1 Background and history

Security has been an issue in the field of computer technology since early 50's. During that time the computers had

techniques to make sure that an application is only using the memory that is given to it. Several encryption and access control techniques to protect passwords etc., were developed in 1960's. We have the concept of "Phone Phreaking" since the 1950's till 1980's, that is where the phrase "ph" in "Phishing" comes from replacing "f" in 'fishing'. In 1983, Thompson first described a security threat, which is called as "Trojan horse" [16,17].

In 1996, as described in [18], the term 'phishing' was used first time by hackers who stole America On-line by getting access to the passwords of AOL users. In addition, in 1997 first media publication warns customers of new threat called "phishing", also AOL cut down its direct access for Russian users due to increased level of fraud. In 2001, as described in [18], e-gold became the first victim among the financial institutions. Phishers started using spam messages to spread their network. In 2005, Bank of America lost 1.2 million usernames and SSNs of their customers. In 2006, phishers targeted VoIP first time. In 2007, according to Gartner study, about 1.5 millions of US citizen identities got stolen. In 2008, for advertising spams on the Facebook member's walls Wallace received $711 M. In 2011, Credit and Debit card details of more than 10M PlayStation Network and Sony Entertainments users are stolen and damaged approximately $1 to $2 billion making it the costliest cyber-hack ever. According Microsoft Safer security Index's report in 2014 the losses due to phishing were about $5 billions [19]. eCrime Trends Reports of the year 2012 shows that phishing attacks are increasing by 12% per year. Various solutions have been given by the researchers varying from content filtering to message authentication still the attackers are able to conduct these scams successfully [2,20–23]. Of course, it is easy to exploit humans rather than breaking into the system straightway.

2.2 Phishing statistics

In the fourth quarter of 2104 .com (41%) was the most commonly used domain to carry out phishing scams followed by .net (7%), .org (5%) and .br (3%) as shown in Fig. 5. Figure 6 gives an overview of the most targeted organizations according to APWG's report for the fourth quarter of 2014.

2.3 Phishing statistics and incidents of 2015 [24–28]

- Google Safe Browsing mentioned in a report that between years 2014 and 2015, the number of malicious web pages fell down from 18,454 to 14,977 whereas the number of phishing pages rose up from about 24,864 to 33,571.
- According to study conducted in 2016, employees take about 1 minute and 22 seconds to click on a spear phishing email.

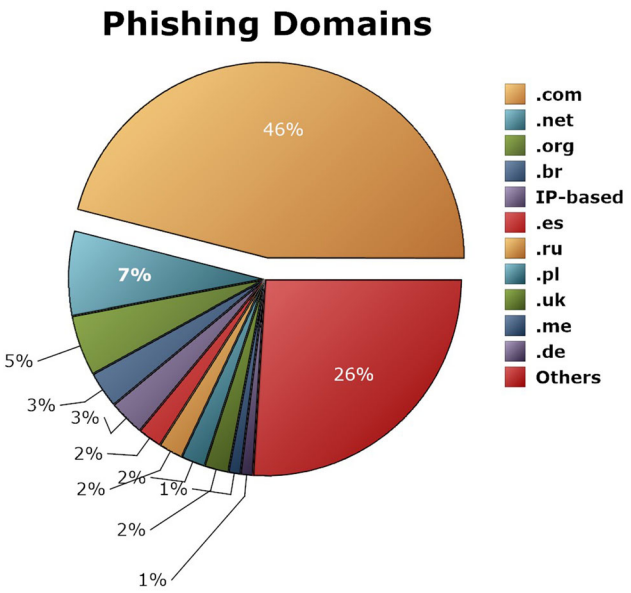


Fig. 5 Statistics of phishing websites based on domain (E-crime Report 2013 Q4)

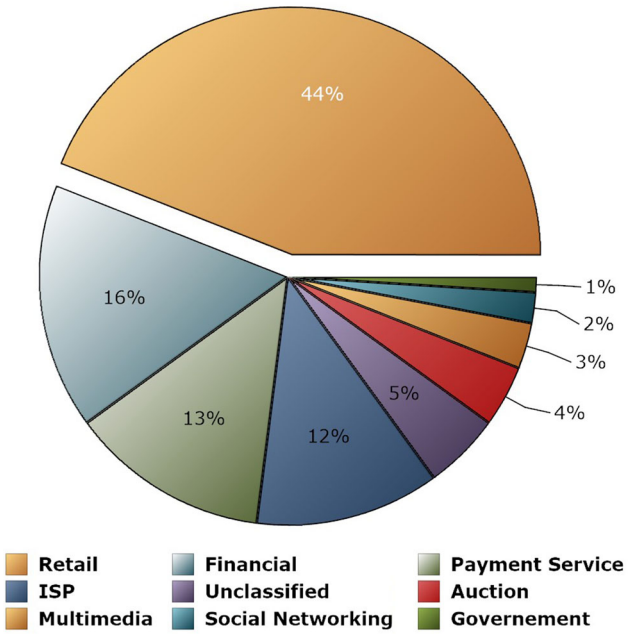


Fig. 6 Statistics showing most targeted industries in 2014

- In 2016, it was reported that about 70% of the IoT devices are vulnerable to cyber attacks.
- Only 6% of the organization claim to have a incident response system and law enforcement.
- In January 2015, three new android families were found, and about 80% of the malware attacks are a part of phishing scams.
- In 2016, it was shown that one out of every 230 emails have a virus, also if the emails were carrying any technical information they were most likely to be clicked open.

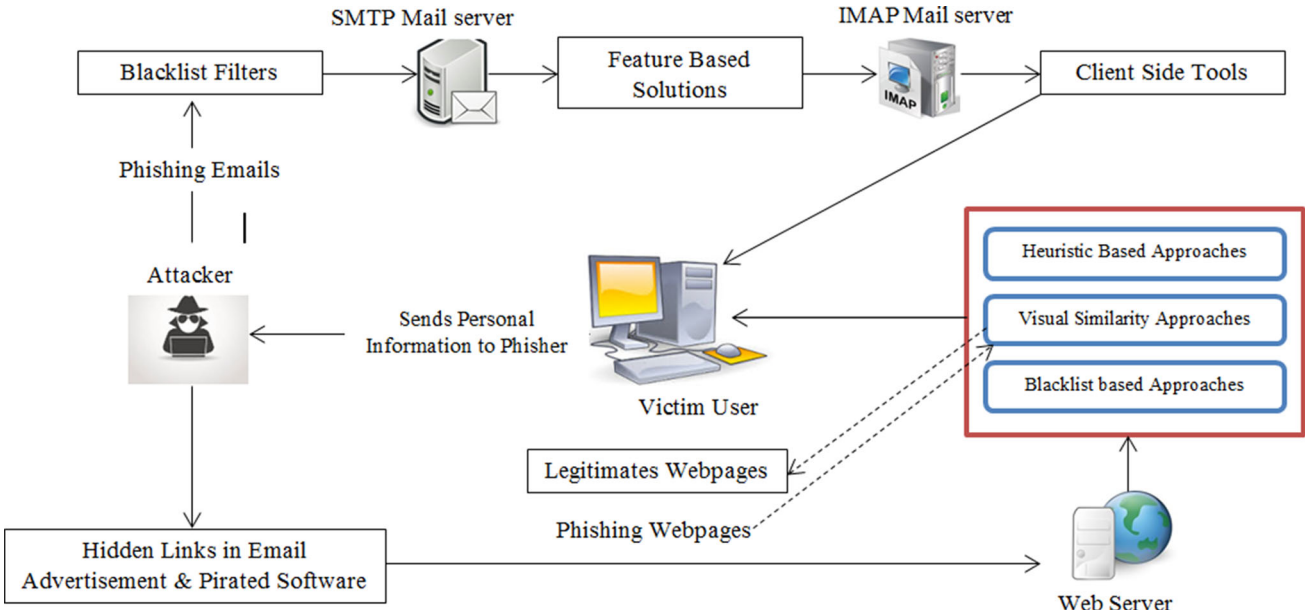


Fig. 7 Lifecycle of phishing attacks based on phishing emails and phishing websites

2.4 Phishing lifecycle

There are various phases to the phishing cycle. Still, we can visualize three main phases comprising the phishing cycle which the phishers iterate [1,16,29–34]. In first phase, the phisher explores organizations and selects a target and then, creates a phishing website and send numerous spam emails among the various users in Internet community. Second phase starts with reading of these emails. Whenever the user "bites" on the phish i.e. click on the link, third phase starts and user is redirected to the phishing site.

This section briefly discusses the phishing cycle as shown in Fig. 7. Every first emails goes the DNS based blacklist over the mailing system. If the domain of sender is found in blacklist, the email is blocked before reaching the SMTP mail server. Based on structural properties of emails, various solutions filters email before it reaches to the user's inbox. There are also various solutions available to check emails based on features of any email on client side. In case of phishing webpages, the links are embedded in emails sent to the user or any other advertisement. There are various solutions available on the client side as Internet is vast enough to control it. Some blacklist-based applications block the website if domain falls under blacklist. Unlike the blacklist solution for emails that block emails before they reach the SMTP mail server, it blocks the website when browser of client side request for the URL mentioned in the list. Some more solutions like heuristic feature and visual similarities block the webpage only when the browser request for any phishing webpage [35].

2.5 Motivation

Phishers always take the benefit of human factors that generally ignore the critical warning messages. Some of the driving factors behind these crimes are as below [32]:

- Banking credential theft—theft of credit card details, CVV number and online credentials for websites like PayPal and eBay etc.
- Identity theft- attackers steals sensitive identity details of a person such ID number, address, credit card details etc.
- Trade secrets and confidential document theft—targets specific companies to steal sensitive information, spear phishing are most commonly used.
- Notoriety—new hackers or criminals sometimes carry out scam to gain popularity and recognition among peers.
- Wring security loopholes: to discover errors and vulnerabilities in a system or new technology for future goals.
- Attack propagation—to hide their location.

3 Taxonomy of phishing attacks

The classification of phishing attacks can be done on the basis of mechanisms using which attackers are able to obtain personal details about the victim. Either a phisher uses a way in which he/she frauds to victim or he/she uses any malicious code to access victim personal information. A phisher may fraud to any innocent user either by using spoofed emails or by using fake websites. A basic classification of phishing attacks is shown in Fig. 8.

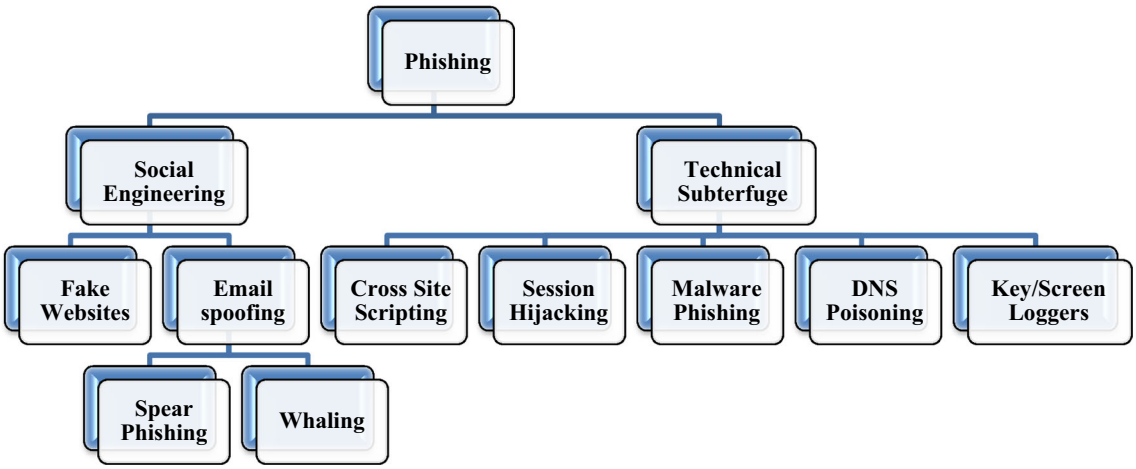


Fig. 8 Classification of phishing attacks based on how a phisher fraud to victims

3.1 Social engineering

Social engineering refers to the skills of luring a person to fall a victim to a malicious scam in order to accomplish certain goals which may harm the victims. During the years 2014, according to a global phishing survey, Apple was most targeted brand by the phishers. Fraudulent emails were sent by the attackers which asked the victims to update their iCloud accounts using a link that directed user to a malicious webpage where they were asked to update their credentials [36,37]. As defined in [38]:

"Social engineering schemes use spoofed e-mails purporting to be from legitimate businesses and agencies, designed to lead consumers to counterfeit websites that trick recipients into divulging financial data such as usernames and passwords." The phishing based on social engineering is further classified based on [39–44]:

I. *Spoofed emails*

a) We also called them phishing emails. These emails are used to convince its victim, so that he/she may send his/her personal information. Email phishing can be done in any of the following ways: (i) concatenating some string to start or end of a legitimate domain to generate a fake link; (ii) the actual links are not the same as the links visible to the user; (iii) use bugs to redirect a link to a malicious website; (iv) exchange certain characters of legitimate URL with similar characters that are different to detect; (v) use of Javascripts etc, to hide the address bars. Spear phishing- A new term "spear phishing" has also come into picture, where the target is a specific person or organization. Spear-phishing is also being used against any group of people in an organization working at any position, "whaling" specifically targets the high-rank employees of an organization [45]. A spear phishing attack targeting to a specific user may leverage information such as his/her user name and email address to craft an email that is personalized to the user. This spear phishing technique will certainly improve the success rate of the attack and techniques that can be leveraged by an attacker to find contextual information [46]. In the year 2009, major organizations such as Google, Yahoo, Adobe and Symantec became victims of spear phishing and malware attacks by a group Operation Aurora attacks.

b) Whaling- It is also called CEO fraud; here hackers use social engineering to phish users to give away their bank credentials employee data etc. These attacks are very difficult to trace because they don't deploy any kind of malware or fraudulent webpages.

II. *Fake websites*

These appear to be visually similar to a legitimate website, aim to phish the user into giving up some sensitive information; these are also called phishing websites. Generally, the links of the fake websites are embedded with phishing emails, advertisements or within crack of licenced software. The links are also very similar to the URI pattern of the legitimate webpage.

3.2 Technical subterfuge

Technical subterfuge is another popular way to fraud in which a phisher send some malicious code either attached with emails, or with websites, or with some self-executable code (generally crack of any software). As defined in [38]:

"Technical subterfuge schemes plant crime-ware onto PCs to steal credentials directly, often using systems to intercept consumers online account user names and passwords and to corrupt local navigational infrastructures to misdirect consumers to counterfeit websites." The phishing based on technical subterfuge is further classified based on [39–44]:

a) Cross-site scripting—XSS is a vulnerability due to weak security techniques in web applications. XSS happened when dynamic web page displayed input without properly validate. This allows attacker to install malicious JavaScript into generated page that is viewed on victim side. Following this, it is simple for the attacker to hijack the login credentials and misuse them. XSS can influence any site that allows the user to enter data.
b) Session hijacking—Session hijacking is a common and serious thread in WLAN. This is also known as cookies hijacking. Here, the session key is stolen via DoS attacks in order to steal the identity and access the resources illegally. The attacker forces the mobile station to break the connection with a particular access point.
c) Malware phishing—In malware based phishing, malware is used to store credentials in victim computer and send it to the owner i.e. the phisher.
d) DNS poisoning—In DNS poisoning attacks the phisher has a fake DNS server and somehow tempts the client to communicate with it, and once the victim connects they are directed to malicious webpages or might install malware into their systems.
e) Key/screen loggers- These are very difficult to trace and screen loggers the virtual keyboards are no longer useful. They take screenshots and mouse movements and send it to the attacker at a remote location.

4 Taxonomy of defense mechanishms

In this section, we present a classification of various phishing detection and protection techniques based on [39–44]. One of the classifications is used to classify various available solutions based on email filtering schemes [44] and other classification is used to classify various available solutions based on detecting phishing websites [41–43]. In email filtering classification, there are some schemes like network based protection that is based on blacklist schemes, heuristic schemes in which phishing emails either detect on server side or on client side, are based on some features which is introduced by phisher to redirect victim to phishing websites or other features used to fool victim. There are both server side (CANTINA, PILFER etc.) and client side solutions (blacklisting and whitelisting) available in email based filtering method, but during our literature survey, we only found client side solutions in websites detection schemes.

4.1 User education

Why do people fall for phishing? Dhamija et al. [47] conducted a laboratory-based study showing twenty-two participants to twenty websites, asking them to differentiate phishing website from legitimate ones. Authors revealed that those participants made mistakes on the test 40% of the time. Furthermore, authors noted that 23% of their participants missed out all.

Cues in the web browser address bar and status bar as well as all security indicators. Nevertheless, to date, a considerable amount of literature work has been discussed that "humans' incapability to interact with the systems" is one of major reasons why people still fall for phishing attacks [10,11,37, 47–57]

It is vital to state that users' perception of such phishing threats may encourage users to prevent from potential vulnerabilities. Downs et al [53] have employed a role-playing study aiming at understanding why people fall for phishing emails and what cues they look for to prevent such attacks. The results revealed that while people can protect themselves from known risks, they also have difficulties of understanding their known to unfamiliar risks. Wu et al. [54] stated that several people do not know how phishing works or how complex such attacks are. We can say that this can be attributed to people's lack of perception of phishing threat. In some cases, participants did not notice warning signals and in other cases they assumed warnings were not valid though they noticed them.

Previous research has been shown that both academic and government organizations have made a significant effort to deliver end-user education to enable public understanding of the importance of cyber security, especially in anti-phishing context [55]. The APWG (anti-phishing work group) which is a non-profit organization covers number of areas: (1) what can be a phishing threat? (2) how can it be severe? (3) what is the usefulness of having a safeguarding measure? (4) where and how to report a suspected phishing email or website? and (5) anti-phishing education to thwart phishing attacks. The US Computer Emergency Readiness Team also offers people free advice on its website about common cyber security breaches for computer users who have a limited computer literacy.

In the past efforts have been made in order to resolve the phishing threat either by prevention or detection of emails, URIs and web pages, still in the field of user awareness only little efforts are being made [55]. Therefore, we need to aim further research over anti-phishing education to save users from phishing attacks.

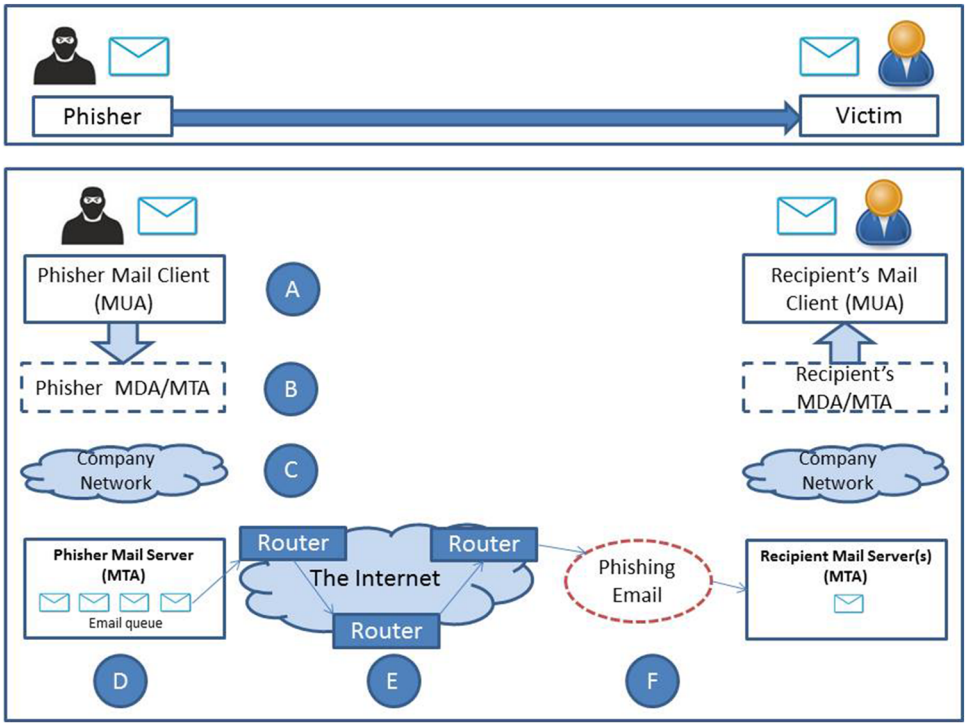


Fig. 9 Phishing email message transportation

4.2 Protection from phishing emails

Phishing email message transportation is shown in Fig. 9. The framework to detect phishing emails from a set of emails in real time situations is present between Message Transfer Agent (MTA) and Mail User Agent (MUA) to stop phishing email before reaching the victim's inbox.

MTA (Message Transfer Agent) Acts as mail carrier and storage location.
MUA (Mail User Agent) Application that retrieves emails such as "MS Outlook".
MDA (Message Delivery Agents) It is the mailbox, it saves messages until the user sees them.
Phisher Person or group with malicious intentions targeting a victim.
Victim People who might get lured by a phishing email and become a target.

An overview of email data parts is shown in Fig. 10 [58].
Phishing email Features:
Figure 10 shows the most significant features of email which are obtained by C and D data parts of the message. A generally used approach in extracting features found in A and B is by the use of blacklists [58]. The groups of most effective features of an email are discussed in Table 1.

Table 1 shows four groups of features: External Features (group 1), body-based features (group 2),

URL based features (group 3) and features header (group 4). Phishing emails are the traditional and one of the common ways for phishing frauds. The Fig. 11 shows the process of phishing email being transferred to a computer network [44].

Based on life cycle of phishing email, as discussed in [59] phishing emails are classifying into the following categories.

4.2.1 Network level protection

Implementation of network level protection restricts a set domains or IP addresses to enter in a network. The network level protection is also known as 'blacklist filters', because it based on the mechanism in which some particular range of IP addresses or domain name listed as blacklist and not allowing any communication from the list. Some example of this filter schemes as given below:

i. Anti-spam filters
The anti-spam techniques [44] can determine the origin of emails and decrease the attacks to great extent. Emails are sent in bulk to mark potential victims; these emails contain fake sender details and a false route information etc. These emails appear to be coming from a legitimate resource thus attackers are able to hide themselves.
ii. DNS-based blacklist
The blacklist approaches are reactive in nature requiring attention from Internet Service Providers to continuously update the list by monitoring the network traffic.

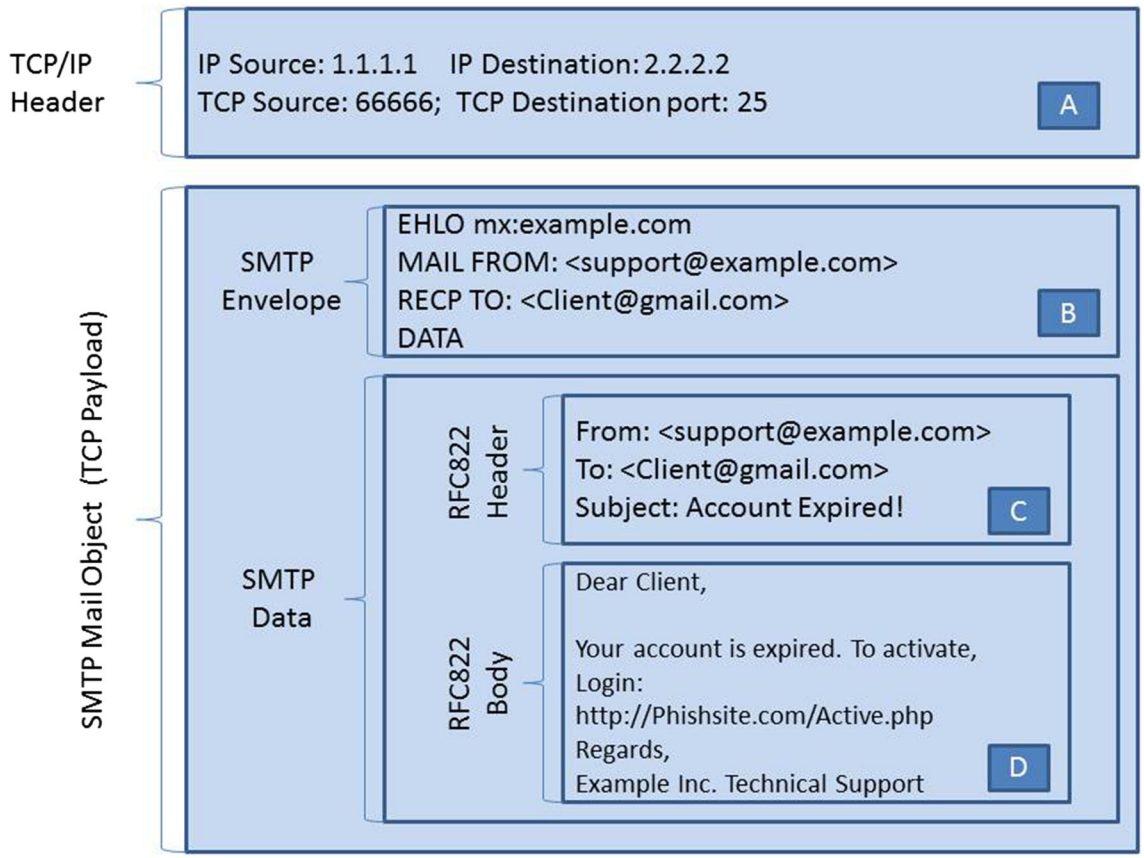


Fig. 10 An overview of email data parts [58]

The DNS-based blacklists [60] make use of DNS protocol. However, a server optimized to handle large DNS resource records is required or else other service handling DNSBLs may face several limitations in terms of performance and speed.

4.2.2 Authentication

User and server authentication approaches check whether the attacker is not pretending to be a valid sender of an email or a resource request, it increases the security at both server and user level. At user level authentication is ensured by use of passwords, but it is evident in past that passwords can be cracked by the phishers [61]. Authentication at domain level [62] is ensured by the service provider. Microsoft Sender ID [61] and Yahoo based Domain Key [62] provides some examples of domain level authentications. Email level authentication [63] is also used to authenticate email based on domain name and hash of password as digital signature.

4.2.3 Feature based email classification

A very common strategy behind using phishing emails is to embed a link clicking on which leds the user to a fake webpage. The phishing email used email structure with embedded URI to ask the user to disclose confidential data. However, these phishing emails are extracted by using some features, which can easily detect by using previous knowledge as phisher repeat some pattern to fool their victims by disturbing the email feature.

i. **Link features**
A hyperlink structure is as : <a href="URI">"Visual Text"<\a>, where URI is the actual address of link of "Visual Text." URI contents are not displayed in web browser, but "Visual Text." Phisher uses this fact in 'bait' e-mails to redirect victim to a phishing websites. LinkGuard algorithm [64], examines the true and fake link for all the discrepancy. Use of IP address directly also signal that the website might be phishing, but it is not definite, if the destination information is missing actual DNS is examined, in case of encoded links, decoding is done followed by recursive execution of LinkGaurd.

ii. **Structural features**
Support Vector Machine (SVM) [65] which is deployed as server site to classify emails before they reach the client is the most commonly used classification mechanism for phishing emails. However, the experimentation

Table 1 Most effective features of an Email

Group features	No	Features	Abbreviation of features
External features	1	Spam features (included 50 sub-features)	Spamfeatures
Body-based features	2	HTML e-mail	body_html
	3	Body of Multi part	body_multipart
	4	Verify your account phrase	body_Verifyphrase
	5	"OnClick" JavaScript event	body_JSonclick
	6	Code of JavaScript to change the status bar	body_JSchangebar
	7	Code of Java script	body_javascript
	8	Code of Java script to open popup windows	body_JSpopup
	9	Forms in email body	body_forms
	10	Ratio of the number of words to the number of Characters	body richness
URL-based features	11	html-links	url_htmllink
	12	Number of dots in a link	url_nodots
	13	Non matching between target and text of urls	url_ TarDiflink
	14	URL IP address	url_ip
	15	Image links	url_imagelink
	16	URL bag of word links	url_bagword
	17	URL has two domains	url_twodomain
	18	Non-standard port in the URL	url_nonstport
	19	URL containing hexadecimal characters or @ symbol	url_hexorat
Header based features	20	Subject replay word	sub_replay
	21	Difference between the sender domain from the domain of the embedded links	Diffsenlindom
	22	Subject (bank, verify, debit)	sub_words
	23	Sender e-mail address uses different replay address	Sendiffreplyto
	24	Total number of words in the subject line	subj noWords
	25	Total number of characters in the email's subject	subj noCharacters
Sender based features	26	Total number of words in the send field	send noWords
	27	Total number of characters	send noCharacters
	28	Difference between the sender's domain and the reply-to domain	send diffSenderReplyTo
	29	Sender's domain is different from the email's modal domain	send nonModalSenderDomain

was done only on a small data, i.e. they used only 25 features to distinguish these email which were selected using simulated annealing.

iii. **Word list features**
The main approach of the classification is based on the machine-learning algorithm.

This model has various shortcomings, it requires a large number of features, it has high complexity in terms of time and memory, it is also not able to detect zero-day attacks.

i. k-Nearest Neighbour (k-NN)—Gansterer proposed a k-NN classification techniques [66] that filters phishing emails based on k-nearest training data and the training data is selected using a predefined similarity function.
ii. Naive bays classifiers—it uses Bayes theorem to perform probabilistic classification, it is mostly used for text classification and keyword filtering. The features used for Naïve Bayes classifier are statistically independent to maintain the accuracy.

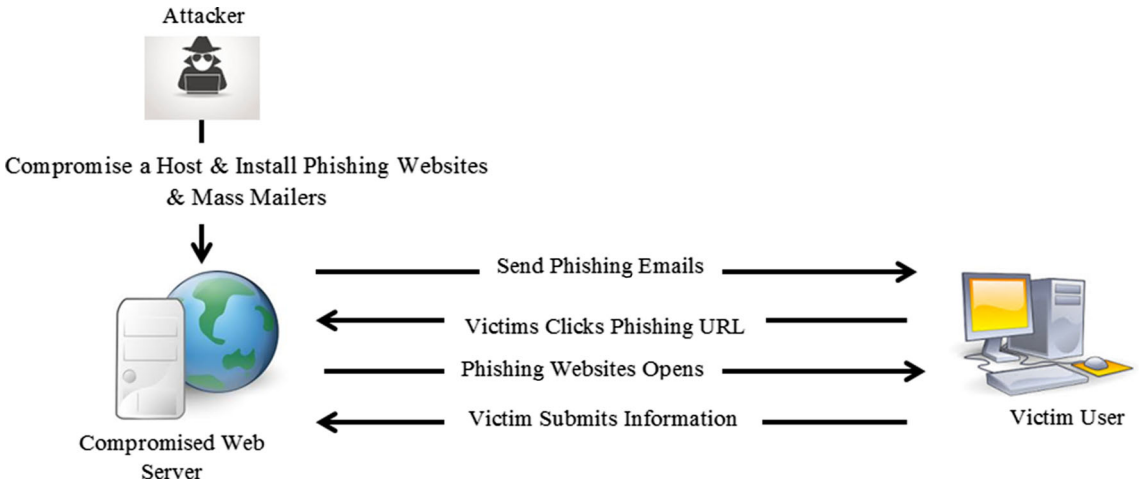


Fig. 11 Various phases in phishing email attack

4.2.4 Comparison of existing solutions

Comparisons of various existing solutions for phishing website detection is shown in Table 2.

4.3 Protection from phishing websites

A fake website seems to be similar with any legitimate websites in look and design. The URL pattern of any fake page also seems to be similar in first glance. Phishers try their best in both look and feel as well as in URL pattern so that more and more victim attract toward their fake page without knowing that they are under phishing attacks. As whole Internet, cannot be control from server side, so solutions available to detect these fake pages are only for the client side. Based on these characteristics, we classified many solutions for phishing website [41–43], under the following categories:

4.3.1 Blacklist and whitelist

Blacklists consists of a list phishing URLs and IP addresses detected in the past, which are updated in certain intervals, whereas, whitelist is collection legitimate addresses and URLs. They do not provide security against zero-day attacks, as new address or site cannot be detected by these blacklists. Whitelists generally used to reduce 'false positive' rates. However, blacklist has lower FP rates then heuristic.

i. Google safe browsing API
   Google safe browsing API [76] allows the user application to verify if a given URL is blacklisted or not. Although the protocol is still experimental, various browsers including Google Chrome and Mozilla Firefox use it. Google provides two blacklists: 'goog-phish-shavar' (phishing) and 'goog-malware-shavar' (malware).
   It allows the client side applications to examine if a URI is blacklisted from a list which is updated by Google regularly. Although the protocol is still experimental, various browsers use it.
ii. PhishNet: Predictive Blacklisting
   PhishNet [77] solves the problem of exact matching (if a URL is slightly changed version of blacklisted one then it remains undetected). It produces all the possible variants of a given URI using five heuristics:
   i. Replace top level domains (TLD)
   ii. Directory structure similarity
   iii. IP address equivalence
   iv. Query string substitution
   v. Brand name equivalence
iii. Automated individual white-list
   Automated Individual White-List (AIWL) [78] contains a list of whitelisted websites where users have provided any type of sensitive data referred to as trusted login user interfaces (LUIs) features.
   Ye et al., gave an AIWL based approach where Naïve Bayes classfier uses a feature vector based on successful or failed login attempts, then construct a model to calculate probability of any future login attempts on the basis of predefined thresholds.

4.3.2 Heuristic solutions

Heuristics refer to set of rules based on previous results and experiences, to solve a problem or learning purposes. The solutions based on heuristics are not necessarily optimal, but

Table 2 Comparisons of various existing solutions for phishing emails

Reference	Solution	Utility	Approach	Limitation/remarks
DNSBL information [60]	DNS-based blacklist	In anti-spam filters	Blacklist a range of IP address and domain names	Zero-day phishing
Lyon et al. [61]	Sender ID	In Microsoft Sender ID	Domain level authentication is used by sending Sender ID	Both side must use the same technology
Delany et al. [62]	Domain Key	In Yahoo Domain Key	Domain level authentication is used by sending Domain Key	Both side must use the same technology
Adida et al. [63]	Email authentication	Gmail, Hotmail, Yahoo	Authenticate by password hashing with domain name	Most user do not use email authentication
Chen et al. [64]	LinkGuard algorithm	In Windows it check all mailing application	Find the similarities between domain of actual and visual links	Check only for emails and higher false positive rate
Chandrasekaran et al.[65]	SVM based structural properties	Check emails before it reach to inbox	Implemented between MTA and MUA using SVM classifier	Use a very small set of features (25 only)
Gansterer et al. [66]	k-Nearest Neighbour	Rank emails in Ham, Spam and phishing	Detect emails based on similarities in k-sample phishing emails	High false positive rate then the spam filters
Fette et al. [67]	PILFERS	Uses 10 features including Spam Assasin output.	Random forest and Support vector machine (SVMs) for classification	Large amount of emails are misclassified
Bazarganigilani [4]	Based ontology concept and a set of heuristics		The working is divided in 5 steps and uses Information Gain (IG), Naïve Bayes classifier	Low accuracy of classification
Bergholz et al.[39]	Uses Dynamic Markov Models	Study statistical filtering of the phishing Emails	Dynamic markov chains are used to train the classifier on feature set	High time and storage complexity
[35,68–73]	User Education based Approach	Assess the effectiveness of training materials, oniline tests, embedded training approaches etc	Use of short training materials will enable users to read, immediate training after person becomes a phishing victim	Participants were more educated than the average Intenet users
MP-Shield [57]	Blacklist and Data mining based approach	Detects phishing activities in android based devices	Uses google balcklist API to extract attributes from network traffic and then performs classification	There is no method for the updation of the model to increase it's knowledge
Park et al [67]	Based ontology concept	The syntactic similarity for sentences, and the subject and object of verb comparison	To determine the hidden intention of email from the computer perspective so that machines could more accurately detect phishing emails	High complexity
Tayal et al [74]	Data mining based approach	Particle Swarm Optimization trained Classification Association Rule Mining	A new rule pruning scheme in order to reduce the number of rules and increasing the generalization aspect of the classifier	High time and storage complexity
PhishWHO [75]	Difference between the target and actual identities of a webpage	Phishing webpage detection via identity keywords extraction and target domain name finder	Exploit URL patterns based on the proposed N-gram model to extract identity keywords	Cannot address visual cloning

the results are near optimal and ease the decision making process.

Heuristics based phishing detection techniques are found to be effective in case of zero-day phishing attacks. These approaches are based on the data collected from real phishing scams, but they have high false positives but gives better results than blacklist based approaches. However, Browsers such as, Mozilla Firefox, Internet Explorer etc., use heuristic based solution for phishing detection.

i. SpoofGuard
SpoofGuard [79] is browser plug for Internet Explorer in developed by Stanford university. It uses a set of heuristics to detect anomalies in the webpage content. It detects phishing scams based on HTTP. It defines a certain threshold value and if the results of the heuristics cross the threshold level a warning is given to the user. SpoofGuard checks if the URL is similar to a whitelisted one. Then it detects for the presence of a hidden attribute in the URL. If the URL in the text attribute is different from the actual one then the site malicious. It has traffic indicator system, which calculates the threat level by navigating the site. The passwords fields if present also increase the level of threat (though they are mostly harmless) as they might copy a login form.

ii. Collaborative intrusion detection
In CIDS [80], untrusted data is exchanged between different intrusion detection systems (IDSs). Each CIDS analyzes DNS cache to extract high Resource records zones and low Time-to-live values. This list is sent to a global CIDS. The systems then examine the listed addresses and find the infected ones. The origin of the infected data can be detected by monitoring incoming and outgoing connections of suspicious IP address. But this approach is not implemented due to complexity regarding examining the fast flux attacks.

iii. PhishGuard: a browser plug-in
Phishguard [81] is a heuristic based approach, which performs phishing detection based on HTTP authentications. PhishGuard starts it test when web page requires user credentials. It sends the same user ID but a different password to the pages, if the response is HTTP 200 then the page is phishing. In case of HTTP 401 response, either it is wrong password error or the website is blindly signaling failed authentication. The site is regarded as phishing if it already has hash value, else it is legitimate and the user is requested to re-enter the password.

iv. CANTINA: a content-based approach
CANTINA [82] is a toolbar that examines webpage content for phishing detection by calculating term frequency-inverse document frequency (TF-IDF) for each term on the webpage, the top n terms with highest values are then used to represent that document. If the webpage is one of these entries then it is considered to be legitimate else not. The following set of heuristics is used to lower the false positives:

(i) A domain more than 12 months old is likely to be legitimate.
(ii) Presence of—or @ in the linkor URL indicates page is phishing.
(iii) Presence of more than five '.' Indicates page is phishing.
(iv) Embedded HTML forms indicate page is phishing.

4.3.3 Visual similarities

It is one of the important properties, which is maintained by almost every phishers, so that the victim cannot easily distinguish between a fake page and a trusted target page. If any fake page is not similar in visual appearance then there is very less chance to make fool any victim easily. Based on these properties, researcher also proposed their solutions as given below:

i. Visual similarity based phishing detection
Visual similarity based phishing detection (VSBPD) [83] monitors whether a user is giving away any kind of sensitive data to a suspicious webpage. It keeps a check on the forms filled by the user, it looks for the similarities of text and images embedded on the page. It also stores user credentials and where they are to be sent. If the website is not on the trusted list, the processes is interrupted and a warning is generated. The warning is generally raised when there is similarity between two pages, in case both of them require same information, however it is less likely for any of these websites to be fake if they not similar in appearance. This approach is inspired by anti-phish (plug-in) and DOMAntiPhish (browser extension).

ii. BaitAlarm
According to [84], the increase in similarity between phishing and legitimate page increase the chance of user falling for that phishing scam. They use CSS technology to maintain the consistency of the page.
The BaitAlarm has of three components: The Pre-Processor, the Layout Monitor, and Network Library. Pre-Processor extract the layout of any new loaded page, and when user enter credentials in this page, browser hold this page and the layout information is sent to Layout Monitor. On receiving the information the Layout Monitor's comparison-unit obtains CSS from page and checks Network Library for the victim pages comparison-unit. Then, a similarity score is calculated between targets and malicious page, and if score is less than the current threshold, the page is legitimate, otherwise it is a phishing page.

4.3.4 Miscellaneous solution

There are various solutions present, which cannot classify among any of the above categories. These solutions can categories in miscellaneous solutions, as these solutions are not recent. However, these solutions have the historical impact as they came with evolution in the Internet technologies. These solutions are described below:

i. TrustBar
TrustBar [85] monitors the top portion of browser window that has logos and graphical icons. It must be present for every window in the browser so that attacks in which a fake site hide browser indicators or exchange them with other indicators are detected and prevented.
ii. Dynamic security skin
It is an extension for Mozilla browser. It requires secure remote password (SRP) protocol to authenticate Webpages. This extension has a trusted window for entering username and password; it deploys images to create a reliable path between the window and user to prevent any fake webpage and text field entries [86].

4.3.5 Comparison of existing solutions

Comparisons of various existing solutions for phishing website detection is shown in Table 3.

5 Phishing and Internet of Things

The IoT is now changing our ways of living and becoming a part of our daily lives. Devices are made to be smart and connected to the Internet (Fig. 12 shows the architecture of IoT network). IoT is everywhere from our homes, schools vehicles to our bodies [91]. However, along with it comes a new vector of threats which are far wider in the IoT scenario and with its growth the threats will grow too. In a highly connected environment the security threats can target organization, governments or common people and can result privacy exploitation or data thefts and so on [92,93]. SANS institute reports stated that in first half of 2015 phishing attacks resulted into 37% intrusions in IT organizations [94]. The attackers use emails to lure the victims into falling for these attacks. These emails might have links to some malicious webpages; in general, these are available only for few hours, thus reducing the significance of blacklists and other heuristic approaches.

In January 2015, proofpoint reported a cyber-security an attack on IoT devices where spam emails were sent in bulk thrice a day and 25% of the devices were televisions and refrigerators and routers [95]. In these cases, we cannot blame the user for ignorance as a phishing attack can be successful only if all the eight layers of security have been compromised, thus we need to make sure that these attacks get blocked at the initial levels. In the IoT the security is so weak that the hacker are able to use the software in the thingbots for relaying malicious emails without even sending a virus or Trojan, such devices can be easily used for DDos attacks without the user knowing it as it will not at all effect the devices functionality. And the only way to make these devices infection free is to take them offline regularly and update their software [96].

The IoT devices also are need are needed to be brought offline from time to time and updated as it the only way to disconnect it if it has become a part of botnet etc [97]. In the year 2013, 20 billion devices were connected to Internet and this number will increase to 32 billion by the year 2020. Smart thing are the future and everyone is appreciating it but these devices are also making the job of attackers easy. Proofpoint showed in survey that during two weeks more than 100,000 gadgets were compromised to send over 750,000 malicious emails [98].

Today an object mainly communicates with another object who is in the same application system, but there's no doubt that the technical future is connecting every application system and with the growth of the Internet of Things the communication between different systems will become more and more frequent for the collaboration. Figure 13 shows the contribution of IoT to various fields. Only if we can solve the interoperability problem we can have the Internet of Things with better connectivity. Some researchers [91] have come up with a solution that is addition of a coordination layer into the Internet of Things' architecture design. The coordination layer responds to process the structure of packages from different application systems. Then reassemble them to a unified structure which can be identified and processed by every application system. Of course, if the standards of the Internet of Things are completed then the systems which based on the standards will have no problem in interoperability.

To secure the IoT devices from such attacks we have designed a simple algorithm for detection of such email traffic to secure the household devices. Sometimes even legitimate emails are classified as spams. This gives rise to a need of filtering phishing emails so that the user is warned not to open those mails. Most of the approaches use encryption [56]. Thus, the attacks on these devices are highly distributed in nature which results in malicious emails successfully reaching inboxes [97,99].

6 Current issues and challenges

In the literature survey, various solutions have been proposed in order to prevent and detect phishing attacks but there is no one perfect solution that can fight phishing attacks all at

Table 3 Comparison of existing phishing website solutions

Reference	Solution	Utility	Approach	Limitation/remark
Google Developer [76]	Google safe browsing API	Chrome, Firefox, etc.	Provide a blacklist and when any hit occur, browser block the page	Not able to detect Zero-day phishing and when IP change
Prakash et al. [77]	PhishNet	Predictive blacklist	Remove the exact match limitation of blacklist by various techniques	Not able to detect Zero-day phishing
Han et al. [78]	AIWL	Tool to maintain individual whitelist	User maintained their own individual whitelist & list of features of these legitimate webpages	AIWL warn whenever any information is sent to any other page then in the list
Chou et al. [79]	SpoofGuard	Internet Explorer plug-in	It detect spoofed pages based on URIs with the help of certain rules	If URIs are not as defined in rules, it cannot detect it
Wu et al. [80]	CIDS	Intrusion detection system	Exchange data among IDS globally	Not implemented yet
Joshi et al. [81]	PhishGuard	A browser plug-in	A phishing website don't respond correctly while asking of credentials	Credentials theft if phishing website reply unauthorised
Zhang et al. [82]	CANTINA	Internet Explorer toolbar	Search top TF-IDF in search engine and find current URI in top list	Higher false positive rate when TF of any other term is high
Medvet et al. [83]	Visual similarity based phishing detection	Approach to detect phishing	Find similarities based on text pieces, image embedded and overall visual appearance	Not distinguish if text pieces are replace with image of same appearance
Mao et al. [84]	BaitAlarm	Google Chrome	Compare CSS of two websites, where first is victim and second is phished	Selection of victim site is manually, which is not feasible in practical
Herzberg et al. [85]	TrustBar	Secure user interface add-on to browsers	Identify SSL certificate and shown in browser	Implemented in browser itself and not used separately
Dhamija et al. [86]	Dynamic Security Skin	Extension for Mozilla browser	Generate unique 'skin' for each user and each transaction	Depend on human, whether or not he/she understand the 'skin'
Chen et al. [87]	Normalized Compression Distance	Approach to detect phishing Website attacks	Used the concept of Gestalt theory and super signals to treat webpages as indivisible	Not robust when alter to web pages e.g. image colour, relocation of objects and text content

Table 3 continued

Reference	Solution	Utility	Approach	Limitation/remark
Gastellier-Prevost et al. [88]	Phishark	Anti-phishing toolbar	Define 20 heuristics for detecting phishing webpages	Not robust when alter to web pages and need to add more heuristics for detecting phishing webpages
Moghimi et. al [89]	Rule-based phishing detection method	PhishDetector extension	Two feature sets to determine the webpage identity, a rule-based method by extracting the hidden knowledge from our model	Not entirely reliable in detecting phishing attacks
Solanki et al. [90]	Heuristic based approach	Approach to detect phishing Website attacks	Extract the features then apply this features to machine learning techniques it will identify website are phished or legitimate	Heuristic evaluation on an interface is minimal due to organizational constraints

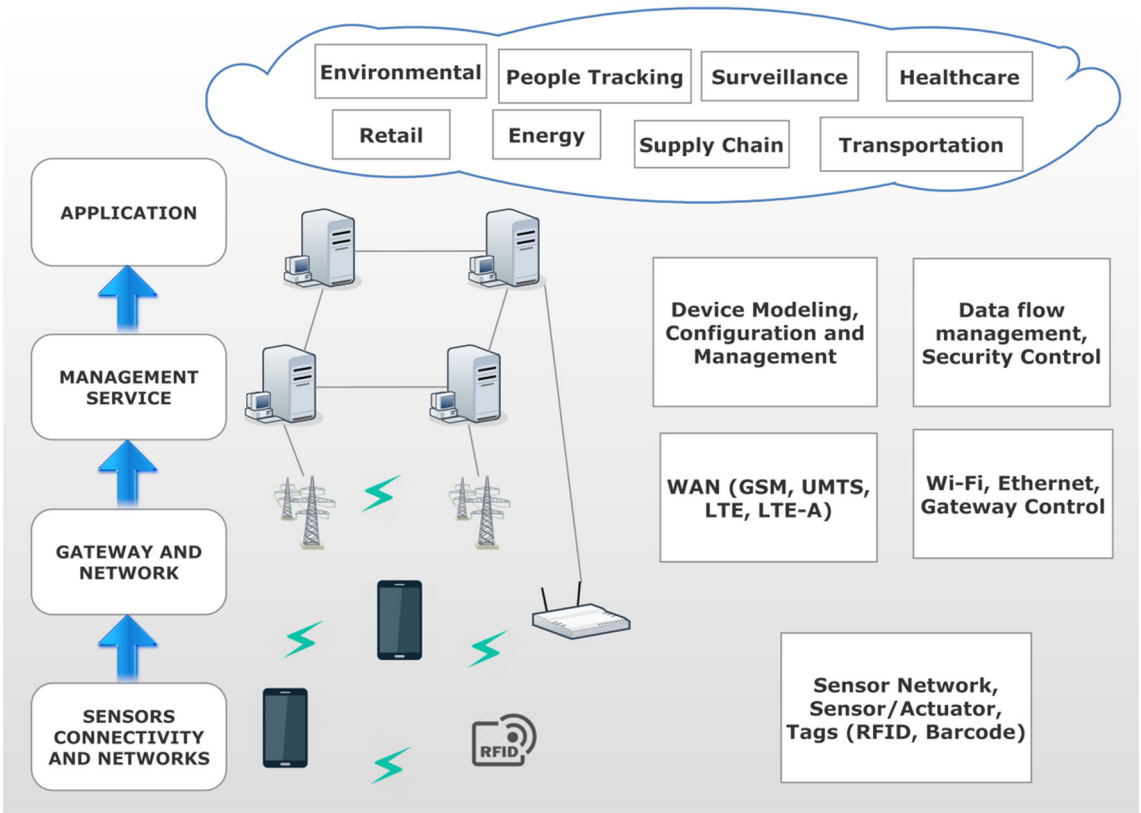


Fig. 12 IoT architecture

once. We have previously mentioned that phishing attacks can be carried out either by malware or social engineering where the attackers either use fake web pages or emails [21,22,100,101]. Thus, there are many solutions are available to detect these emails and websites. To control phishing emails successfully, there were various solutions proposed as

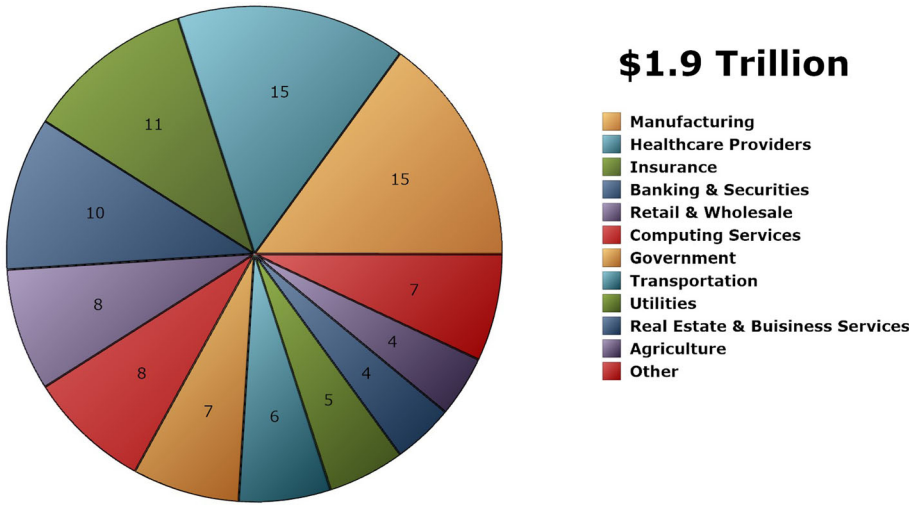


Fig. 13 Profits that IoT will contribute by 2020

discussed in [44,60–66]. Similarly, to control phishing websites attacks, there were also various solutions proposed in [41–43,76–88]. Phishing email is not the only way to fraud, but phishers also uses fake websites and phishing emails have also the link of fake websites. Phishing websites found more harmful as compare to phishing email, as many of phishing emails are filter before receiving by user inbox and filtered mail found under the spam mail. Any educated user can easily neglect these filtered spam mail if filtered correctly.

DNS-based blacklist (DNSBL) [60] deploy the DNS protocol to control phishing emails. The blacklists are required to be updated periodically, which requires an interactive behavior and the attackers take advantage of it if they have access to a legitimate PC or by changing IP addresses. In addition, it is unable to prevent from 'zero-day' phishing, i.e. initial victims cannot protect from phishing. Spam filtering techniques used at server side are also not very effective in case of phishing emails, as they perform classification on the appearance of certain words or phrases, if phisher changes in its statistics of phishing attacks, i.e. if there is solution present in server-side, which based on bag-of-words, then phisher have to avoid only these words and then rest is easy as 'piece-of-cake'.

User and server authentication approaches check whether the attacker is not pretending to be a valid sender of an email or a resource request, it increases the security at both server and user level. At user level authentication is ensured by use of passwords, but it is evident in past that passwords can be cracked by the phishers [61]. Authentication at domain level [62] is ensured by the service provider. Email level authentication [63], is also used to authenticate email based on domain name and hash of password as digital signature.

Spoofed hyperlinks in the phishing mails are very common feature. Support vector machine (SVM) [65] which is deployed as server site to classify emails before they reach the client is the most commonly used classification mechanism for phishing emails. However, the experimental results in this work were not sufficient for large data, i.e. they used only 25 features to distinguish these email.

Google safe browsing API [76] asks the client side applications to check if a URI is blacklisted from a list which is updated by Google. Although the protocol is still experimental, various browsers use it. PhishNet [77] addresses the exact match limitation found in blacklists. As life of these, phishing attacks are very less, a large amount of data is consumed to store these blacklisted URIs and domain, which have no use in near future. In addition, the complexity of comparing every URI with blacklist data is very high.

After blacklist scheme, some heuristic scheme proposed to detect phishing. Unlike blacklist detection technique, heuristic techniques can identify 'zero-day' attacks. But these schemes have large number of 'false positive' then blacklist schemes. Due to more advanced and complex scams it is difficult to design heuristics without false positives.

Visual based similarity, on the other hand is appropriate scheme to detect any website as phishing. Visual similarity based phishing detection (VSBPD) [83] gives a warning to user whenever he tries gives his credentials to an untrusted website. It checks the visual appearance of a page including images and font etc., and also remembers the details user is giving to a page and where it is to be sent.

BaitAlarm [84] is comparatively more efficient as VSBPD compare the text and their style in two websites, but if text content is replaced with picture then this scheme cannot able to compare these pages. BaitAlarm on the other hand compare the CSS of two websites, though the phishing site give same look yet they have very subtle differences with respect to the content, this approach uses visual similarities for phishing website detection. However, both scheme not describe that, how we choose a legitimate site from which, we compare any suspicious site. Storing information that describes image can be expensive as image take more space then data of

any page used in heuristic schemes. In addition, large 'false positive' found as in heuristic schemes.

In the recent years, phishing attacks have become one of the most serious threats faced by the Internet users, organizations and service providers. Several approaches have been proposed in the literature for the detection and filtering of phishing attacks, however Internet community is still looking for a complete solution to secure the Internet from these attacks. These are virtually no ways to detect any breaches in these devices but to take them offline and update their software manually [102].

Several security techniques are there for the IoT environment, but there are no such approaches for the detection of spam and phishing emails. Most of the approaches use encryption. Thus, the attacks on these devices are highly distributed in nature, which results in malicious emails successfully reaching inboxes.

7 Conclusion and scope for future work

It has been a couple of decades since the phishing problem arose, but it is still used to steal personal information, online credentials, and credit card details. There are various solutions available, but whenever any solution proposed to overcome these attacks, phishers came with the vulnerabilities of the solution to make the attach successful. From these attacks, we focus on the social engineering attacks, as it creates negative effect on online commerce. Phishers always used communication media for the fraudulent activities using spoofed emails and fake websites. It creates bad impression on e-commerce, which is very much necessary in this new era of Internet. Our survey helps new researchers to understand the history, current trends of attacks and failure of various available solutions.

We classified social engineering phishing based on spoofed email attacks and fake websites. We have also classified various solutions either in the spoofed email filtering or in fake page detection. We further classify these solutions as per some common properties share between them. These classifications are based on blacklist, network, heuristics, some feature and various other properties. After the classification, we also described various issues and challenges in current solutions to understand the idea for future study to help the humanity by defending against phishing attacks.

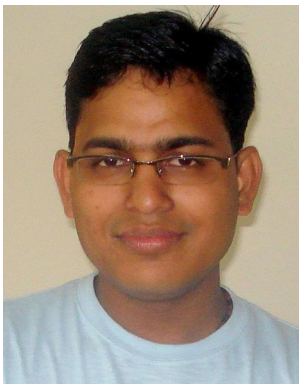
References

1. Ramanathan, V., & Wechsler, H. (2012). phishGILLNET—phishing detection methodology using probabilistic latent semantic analysis, AdaBoost, and co-training. *EURASIP Journal on Information Security, a Springer Open Journal, 1*, 1–22.
2. Arachchilage, N. A. G., Love, S., & Beznosov, K. (2016). Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior*, *60*, 185–197.
3. The Statistics Portal. *Global number of Internet Users 2005–2015*. http://www.statista.com/statistics/273018/number-of-internet-users-worldwide/.
4. Gupta, B. B., Joshi, R. C., & Misra, M. (2009). Defending against distributed denial of service attacks: Issues and challenges. *Information Security Journal: A Global Perspective*, *18*(5), 224–247.
5. Internet World Stats. *Internet User Statistics—The Big Picture: World Internet Users and Population Stats*. http://www.internetworldstats.com/stats.htm.
6. Goggi, C. (2013). The 13 worst security threats of 2013," December, 2013. http://www.gfi.com/blog/the-13-worst-security-threats-of-2013/.
7. Christina, G. *The 13 worst security threats of 2013*. http://www.gfi.com/blog/the-13-worst-security-threats-of-2013/.
8. Ragan, S. (2013). *Senior executives blamed for a majority of undisclosed security incidents*. http://www.networkworld.com/article/2171678/data-center/senior-executives-blamed-for-a-majority-of-undisclosed-security-incidents.html.
9. Sheng, S., Holbrook, M., & Kumaraguru, P. (2010). *Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions*, CHI, pp. 373–382, Atlanta, GA.
10. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F. & Downs, J., (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *28th international conference on human factors in computing systems*, 10–15 April, 2010, Atlanta, GA.
11. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J. & Nunge, E. (2007). Anti-phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on usable privacy and security*, Pittsburgh, PA, July 2007.
12. EMC. *RSA Online Fraud Resource Center*. http://www.emc.com/onlinefraud#!resources.
13. McCaney, K. (2012). *To hackers, government users are phish in a barrel*. http://gcn.com/articles/2012/03/19/phishing-goverment-cyber-attacks-us-cert.aspx.
14. FireEye. *Operation Clandestine Wolf—Adobe Flash Zero-Day in APT3 Phishing Campaign*. https://www.fireeye.com/blog/threat-research/2015/06/operation-clandestine-wolf-adobe-flash-zero-day.html.
15. Krieg, G., & Kopan, T. (2016). *CNN News, Is this the email that hacked John Podesta's account?*http://edition.cnn.com/2016/10/28/politics/phishing-email-hack-john-podesta-hillary-clinton-wikileaks/. Accessed 19 November 2016.
16. Ollmann, G. (2004). *The Phishing guide—understanding & preventing phishing attacks*. IBM Internet Security Systems.
17. Breen, C., & Dahlbom, C. A. (1960). Signaling systems for control of telephone switching. *Bell System Technical Journal*, *39*(6), 1381–1444.
18. The Trembling Uterus Blog. http://tremblinguterus.blogspot.pt/.
19. 20% Indians are victims of Online phishing attacks: Microsoft. IANS. news.biharprabha.com. Retrieved 11 February 2014.
20. IID. *eCrime trends report*. http://internetidentity.com/resources/.
21. Arachchilage, N. A. G., & Love, S. (2014). Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, *38*, 304–312.
22. Arachchilage, N. A. G., & Love, S. (2013). A game design framework for avoiding phishing attacks. *Computers in Human Behavior*, *29*(3), 706–714.
23. Arachchilage, N. A. G. (2015). User-centred security: A game design to thwart phishing attacks. In *International Conference: Redefining the R&D Needs for Australian Cyber Security on November 16, 2015*, University of New South Wales at the

Australian Defence Force Academy, Canberra. arXiv preprint arXiv:1511.03459.
24. The statistics portal. *Phishing: Distribution of attacks by country 2015*. http://www.statista.com/statistics/266362/phishing-attacks-country/.
25. ThreatSim. *State of Phish 2015*. http://threatsim.com/wp-content/uploads/2015/04/ThreatSim-Phish2015_FINAL.pdf:stateofart.
26. Wombat Security Technologies. *The latest in phishing*, June 2015. http://info.wombatsecurity.com/blog/latest-in-phishing-june-2015.
27. Wombat Security Technologies. *The latest in phishing*, September 2015. http://www.marketwired.com/press-release/the-latest-in-phishing-september-2015-2058330.htm.
28. Symanctec. *Symantec intelligence report*, January 2015. https://www4.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-report-volume-20-2015-social_v2.pdf. Last accessed on March 11 2016.
29. Downs, J. S. et al. (2007). Behavioural response to phishing risk. In *Proceedings of the A. C. M. conference on anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 37–44). Pittsburgh, PA.
30. Chen, J., & Guo, C. (2006) Online detection and prevention of phishing attacks. In *Proceedings of the fifth Mexican international conference in computer science, IEEE conference*, pp. 1–7.
31. GARTNER. *Gartner survey shows phishing attacks escalated in 2007; More than $3 billion lost to these attacks*, December 17, 2007. http://www.gartner.com/it/page.jsp?id=565125.
32. Yu, W. D., Nargundkar, S., & Tiruthani, N. (2008). A phishing vulnerability analysis of web based systems. In *Proceedings of the 13th IEEE symposium on computers and communications (ISCC 2008), IEEE*. Marrakech, pp. 326–331.
33. Bergholz, A., Paaß, G., Reichartz, F., Strobel, S., & Chang, J. H. (2008). Improved phishing detection using model based features. In *Proceedings on conference on email and anti-spam (CEAS)*. Mountain View, CA.
34. Toolan, F., & Carthy, J. (2009). Phishing detection using classifier ensembles. In *IEEE conference eCrime researchers summit*, (pp. 1–9). Tacoma, WA.
35. Tally, G., Thomas, R., & Vleck, T. V. (2004). *Anti-phishing: Best practices for Institutions and Consumers, Mcafee research technical report*, September.
36. Security Response. *Symnatec*. http://www.symantec.com/connect/blogs/apple-ids-targeted-kelihos-botnet-phishing-campaign. Last accessed on December 2014.
37. Li, J., Li, J., Chen, X., Jia, C., & Lou, W. (2015). Identity-based encryption with outsourced revocation in cloud computing. *IEEE Transactions on Computers.*, *64*(2), 425–437.
38. APWG. (2016). *Phishing activity trends report*. http://www.antiphishing.org/resources/apwg-reports/.
39. Almomani, B., Gupta, B., Wan, T., et al. (2013). Phishing dynamic evolving neural fuzzy framework for online detection "Zero-day" phishing email. *Indian Journal of Science and Technology*, *6*(1), 3960–3964.
40. Srivastava, B., Gupta, B., Tyagi, A., Shamn, A., & Mishra, A. Recent survey on DDoS attacks and defence mechanisms. In *Advances in parallel distributed computing, communications in computer and information science*, Vol. 203, pp. 570–580.
41. Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. *IEEE Communications Surveys & Tutorials*, *15*(4), 2091–2121.
42. Aburrous, M. et al. (2008). *Intelligent phishing website detection system using fuzzy techniques, IEEE conference*, Damascus, Syria, pp. 1–6.
43. Aburrous, M. et al. (2010). Predicting phishing websites using classification mining techniques with experimental case studies. In *IEEE conference on seventh international conference on information technology* (pp. 176–181). Las Vegas, NV.
44. Almomani, A., Gupta, B. B., Atawneh, S., Meulenberg, A., & Almomani, E. (2013). A survey of phishing email filtering techniques. *IEEE Communications Surveys & Tutorials*, *15*(4), 2070–2090.
45. Hong, J. (2012). The state of phishing attacks. *Communications of the ACM*, *55*(1), 74–81.
46. Chuenchujit, T. (2016). *A taxonomy of phishing research*. University of Illinois at Urbana-Champaign, Doctoral dissertation.
47. Dhamija, R., Tygar, J. D. & Hearst, M., (2006). Why phishing works. In *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '06*, pp. 581–590, Montréal, Québec, April 22–27, 2006. New York, NY: ACM Press. doi:10.1145/1124772.1124861. Accessed 25 November 2016.
48. Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E., (2007). Protecting people from phishing: The design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on human factors in computing systems*, San Jose, CA, April–May 2007.
49. Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., & Cranor, L. F., et al. (2007). *Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer*, APWG eCrime Researchers Summit, 4–5 October 2007, Pittsburgh, PA.
50. Marforio, C., Masti, R. J., Soriente, C., Kostiainen, K. & Capkun, S., (2016). Hardened setup of personalized security indicators to counter phishing attacks in mobile banking. In *Proceedings of the 6th workshop on security and privacy in smartphones and mobile devices* (pp. 83–92). New York: ACM.
51. Vishwanath, A. (2016). Mobile device affordance: Explicating how smartphones influence the outcome of phishing attacks. *Computers in Human Behavior*, *63*, 198–207.
52. Zhao, M., An, B. & Kiekintveld, C. (2016). Optimizing personalized email filtering thresholds to mitigate sequential spear phishing attacks. In *Proceedings of the 30th AAAI conference on artificial intelligence (AAAI)*.
53. Downs, J. S., Holbrook, M. & Cranor, L. F. (2007). Behavioural response to phishing risk. In *Proceedings of the anti-phishing working groups—2nd annual eCrime researchers summit*, pp. 37–44, October 2007, Pittsburgh, PA. doi:10.1145/1299015.1299019. Accessed 25 November 2016.
54. Wu, M., Miller, R. & Garfinkel, S., (2005). Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on human factors in computing systems*, Montreal, Quebec, 22–27 April 2006.
55. Kirlappos, I., & Sasse, M. A. (2012). Security education against phishing: A modest proposal for a major rethink. *IEEE Security and Privacy Magazine*, *10*(2), 24–32.
56. Gupta, B., Agrawal, D. P., & Yamaguchi, S. (eds.) (2016). *Handbook of research on modern cryptographic solutions for computer and cyber security*. IGI Global.
57. Bottazzi, G. et al. (2015). MP-shield: A framework for phishing detection in mobile devices. In *Proceedings of the 3rd IEEE international workshop on cybercrimes and emerging web environments,* Liverpool, October.
58. Khonji, M., et al. (2012). Enhancing phishing e-mail classifiers: A lexical URL analysis approach. *International Journal for Information Security Research*, *2*(1/2), 236–245.
59. Khonji, M. et al. (2011). A study of feature subset evaluators and feature subset searching methods for phishing classification. In *Proceedings of the 8th annual collaboration, electronic messaging, anti-abuse and spam conference, ACM conference*, Perth.
60. DNSBL Information. Spam Database Lookup. http://www.dnsbl.info.
61. Lyon, J., & Wong, M. (2006). *Sender ID: Authenticating E-mail*. RFC 4406, April.

62. Delany, M. (2007). *Domain-based email authentication using public keys advertised in the DNS (Domain Keys)*. RFC 4870, May.
63. Adida, B., Hohenberger, S., & Rivest, R. L. (2005). Fighting phishing attacks: A light-weight trust architecture for detecting spoofed emails. In *USENIX steps to reducing unwanted traffic on the internet workshop (SRUTI)*.
64. Chen, J., & Guo, C. (2006). Online detection and prevention of phishing attacks. In *Communications and networking in China, ChinaCom '06*, Beijing, pp. 1–7.
65. Chandrasekaran, M., Narayanan, K., & Upadhyaya, S. (2006). Phishing email detection based on structural properties. In *NYS cyber security conference*, pp. 1–7.
66. Gansterer, W. N., & Polz, D. (2009). E-mail classification for phishing defence. In *Proceedings of the 31th ECIR research on advances in information retrieval*. Springer conference, Toulouse, pp. 449–460.
67. Park, G., & Taylor, J. M. (2015). *Using syntactic features for phishing detection*, May.
68. Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A. & Pham, T. (2009). School of phish: A real-world evaluation of anti-phishing training. In *Proceedings of the 5th symposium on usable privacy and security*, 15–17 July 2009, Mountain View, CA. doi:10.1145/1572532.1572536. Accessed 25 October 2011.
69. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F. & Hong, J. (2008). *Lessons from a real world evaluation of anti-phishing training*. eCrime Researchers Summit, 15–16 October, pp. 1–12.
70. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J., (2007). *Teaching Johnny not to fall for phish*. Technical reports, Cranegie Mellon University, http://www.cylab.cmu.edu/files/cmucylab07003.pdf. Accessed 12 June 2011.
71. Arachchilage, N. A. G., (2016). *Serious games for cyber security education*. Lambert Academic Publishing, pp. 1–244, ISBN-13: 978-3-659-85318-0. [arXiv preprint arXiv: 1610.09511]. Accessed 15 November 2016.
72. Schuetz, S., Lowry, P. B., & Thatcher, J. (2016). Defending against spear-phishing: Motivating users through fear appeal manipulations. In *20th Pacific Asia conference on information systems (PACIS 2016)* (pp.1–12). Chiayi, Taiwan.
73. Arachchilage, N. A. G. (2012). *Security awareness of computer users: A game based learning approach*. Ph.D. dissertation, Brunel University, School of Information Systems, Computing and Mathematics. http://bura.brunel.ac.uk/handle/2438/7620. Accessed 19 November 2016.
74. Tayal, K., & Rav, V. (2016). Particle swarm optimization trained class association rule mining: Application to phishing detection. In *Proceedings of the international conference on informatics and analytics ICIA-16*, Article No. 13, Pondicherry, August 25–26.
75. Tan, C. L., Chiew, K. L., Wong, K., & Sze, S. N. (2016) *PhishWHO: Phishing webpage detection via identity keywords extraction and target domain name finder*. Decision Support Systems Vol. 88, August, pp. 18–27.
76. Safe Browsing API—Google Developer. https://developers.google.com/safe-browsing/.
77. Prakash, P., Kumar, M., Kompella, R. R., & Gupta, M. (2010). PhishNet: Predictive blacklisting to detect phishing attacks. In *Proceedings of the INFOCOM-2010 IEEE*, San Diego, pp. 1–5.
78. Han, W., Cao, Y., Bertino, E., & Yong, J. (2012). Using automated individual whitelist to protect web digital identities. *Expert Systems with Applications*, *39*(15), 11861–11869.
79. Chou, N., Ledesma, R., Teraguchi, Y., Boneh, D., & Mitchell, J. C. (2004). Client-side defence against web-based identity theft. In *NDSS*. The Internet Society.
80. Wu, Y. -S., Foo, B., Mei, Y., & Bagchi, S. (2003). Collaborative intrusion detection system (CIDS): A framework for accurate and efficient IDS. In *Proceedings of the computer security applications conference, 2003*, pp. 234–244.
81. Joshi, Y. et al. (2008). PhishGuard: A browser plugin for protection from phishing. In *2nd international conference on internet multimedia services architecture and applications, 2008*. IMSAA 2008. IEEE.
82. Zhang, Y., Hong, J. I., & Cranor, L. F. (2007). Cantina: A content-based approach to detecting phishing web sites. In *Proceedings of the 16th international conference on World Wide Web*. New York: ACM.
83. Medvet, E., Kirda, E., & Kruegel, C. (2008). Visual-similarity-based phishing detection. In *Proceedings of the 4th international conference on Security and privacy in communication networks, SecureComm '08, Article no 2*, pp. 1–11.
84. Mao, J., Li, P., Li, K., Tao, W., & Zhenkai, L. (2013). BaitAlarm: Detecting phishing sites using similarity in fundamental visual features. In *5th international conference on intelligent networking and collaborative system intelligent networking and collaborative systems (IN-CoS)*, Xi'an, 2013, pp. 790–795.
85. Amir, H., & Gbara, A. (2004). *Trustbar: Protecting (even naive) web users from spoofing and phishing attacks*. Computer Science Department Bar Ilan University, July, pp. 1–28.
86. Dhamija, R., & Tygar, J. D. (2005). The battle against phishing—dynamic security skins. In *Proceedings of the 2005 symposium on usable privacy and security, SOUPS '05*, pp. 77 – 88.
87. Teh-Chung, C., Scott, D., & James, M. (2010). Detecting visually similar web pages: Application to phishing detection. *ACM Transactions on Internet Technology (TOIT)*, *10*(2), 5.
88. Gastellier-Prevost, S., Granadillo, G. G., & Laurent, M. (2011). Decisive heuristics to differentiate legitimate from phishing sites. In *IEEE conference on network and information systems security (SAR-SSI)*, 2011.
89. Moghimi, M., & Varjani, A. Y. (2016). New rule-based phishing detection method. *Expert Systems with Application*, *53*, 231–242.
90. Solanki, J., & Vaishnav, R. G. (2015). Website phishing detection using heuristic based approach. In *Proceedings of the third international conference on advances in computing, electronics and electrical technology*.
91. Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, *54*, 2787–2805.
92. Bertlucci, J. (2016). *Internet of thingbots: The new security worry*. http://www.informationweek.com/big-data/big-data-analytics/internet-of-thingbots-the-new-security-worry/d/d-id/1234973.
93. Gorman, M. *The internet of things isn't safe: Thousands of smart gadgets hacked to send spam and phishing emails*. http://www.engadget.com/2014/01/17/internet-of-things-hacked-malicious-email-phishing/.
94. Proofpint. *Proofpoint uncovers internet of things (IoT) cyber-attack*. http://investors.proofpoint.com/releasedetail.cfm?releaseid=819799.
95. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660.
96. Roman, R., Najera, P., & Lopez, J. (2011). Securing the internet of things. *Computer*, *44*(9), 51–58.
97. Tang, D. (2009). *Event detection in sensor networks*. School of Engineering and Applied Sciences, The George Washington University.
98. Koroneous, G. L. (2016). *Enterprise Tech Spotlight: IoT Tipping Point, Phishing Scams, Retail Breaches*. http://news.verizonenterprise.com/2015/08/iot-retail-breaches-phishing-security/.
99. Arachchilage, N. A. G., & Cole, M. (2011). Design a mobile game for home computer users to prevent from "phishing attacks".

In *IEEE International Conference on Information Society (i-Society)*, 2011, pp. 485–489.
100. Gupta, B. B., et al. (2016). Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*. doi:10.1007/s00521-016-2275-y.
101. Jain, A. K., & Gupta, B. B. (2016). A novel approach to protect against phishing attacks at client side using auto-updated white-list. *EURASIP Journal on Information Security*. doi:10.1186/s13635-016-0034-3.
102. Mayer-Schönberger, V. (2011). Failing to forget the "Drunken Pirate". In: *Delete: The virtue of forgetting in the digital age (new in paper)*, 1st edn. (pp. 3–15). Princeton: Princeton University Press.



B. B. Gupta received PhD degree from Indian Institute of Technology Roorkee, India in the area of Information and Cyber Security. In 2009, he was selected for Canadian Commonwealth Scholarship awarded by Government of Canada. He published more than 100 research papers (including 02 books and 14 book chapters) in International Journals and Conferences of high repute including IEEE, Elsevier, ACM, Springer, Wiley, Taylor & Francis, Inderscience, etc. He has visited several countries, i.e. Canada, Japan, Malaysia, China, Hong-Kong, etc to present his research work. His biography was selected and published in the 30th Edition of Marquis Who's Who in the World, 2012. He also received Young Faculty research fellowship award from Ministry of Electronics and Information Technology, government of India in 2017. He is also working as principal investigator of various R&D projects. He is serving as associate editor of IEEE Access, Associate editor of IJICS, Inderscience and Executive editor of IJITCA, Inderscience, respectively. He is also serving as reviewer for Journals of IEEE, Springer, Wiley, Taylor & Francis, etc. He is also serving as guest editor of various reputed Journals. He is senior member of IEEE. He was also visiting researcher with Yamaguchi University, Japan in January 2015. At present, he is working as Assistant Professor in the Department of Computer Engineering, National Institute of Technology Kurukshetra India. His research interest includes Information security, Cyber Security, Cloud Computing, Web security, Intrusion detection and Phishing.

Nalin A. G. Arachchilage currently works as a Lecturer in Cyber Security in the Australian Centre for Cyber Security (ACCS) at the University of New South Wales (UNSW Canberra at the Australian Defence Force Academy). He holds a PhD in Usable Security entitled "'Security Awareness of Computer Users: A Game Based Learning Approach" from Brunel University London, UK where he developed a game design framework to protect computer users against "phishing attacks". Prior to undertaking his current position at UNSW Canberra, he worked as Research Fellow in Usable Security and Privacy in the Laboratory of Education and Research in Software Security Engineering (LERSSE) at the University of British Columbia (UBC), Canada. Before moving to Vancouver, he was a Postdoctoral Researcher in Systems Security Engineering in the Cyber Security Centre, Department of Computer Science at Oxford University.

Kostas E. Psannis was born in Thessaloniki, Greece. He received a degree in Physics from Aristotle University of Thessaloniki (Greece), and the Ph.D. degree from the Department of Electronic and Computer Engineering of Brunel University (UK). From 2001 to 2002 he was awarded the British Chevening scholarship sponsored by the Foreign & Commonwealth Office (FCO), British Government. He was awarded, in the year 2006, a research grant by IISF (Grant No. 2006.1.3.916), Japan. Since 2004 he has been a (Visiting) Assistant Professor in the Department of Applied Informatics, University of Macedonia, Greece, where currently he is Assistant Professor (& Departmental LLP/Erasmus-Exchange Students Coordinator and Higher Education Mentor) in the Department of Applied Informatics, School of Information Sciences. He is also joint Researcher in the Department of Scientific and Engineering Simulation, Graduate School of Engineering, Nagoya Institute of Technology, Japan. He has extensive research, development, and consulting experience in the area of telecommunications technologies. Since 1999 he has participated in several R&D funded projects in the area of ICT (EU and JAPAN). He was invited to speak on the EU-Japan Coordinated Call Preparatory meeting, Green & Content Centric Networking (CCN), organized by European Commission (EC) and National Institute of Information and Communications Technology (NICT)/Ministry of Internal Affairs and Communications (MIC), Japan (in the context of the upcoming ICT Work Programme 2013) and International Telecommunication Union (ITU) SG13 meeting on DAN/CCN, July 2012, amongst other invited speakers. He has several publications in international Conferences, books chapters and peer reviewed journals. His professional interests are: Multimodal Data Communications Systems, Haptic Communication between Humans and Robots, Cloud Transmission/Streaming/Synchronization, Future Media-Internet of Things, Experiments on International Connections (E-ICONS) over TEIN3 (Pan-Asian), Science Information Network (SINET, Japan), GRNET (Greece)-Okeanos Cloud, and GEANT (European Union) dedicated high capacity connectivity. He is Guest Editor for the Special Issue on Architectures and Algorithms of High Efficiency Video Coding (HEVC) Standard for Real-Time Video Applications (2014), Journal of Real Time Image Processing, for the Special Issue on Emerging Multimedia Technology for Smart Surveillance System with IoT Environment (2016), The Journal of Supercomputing, for the Special Issue on Emerging Multimedia Technology for Multimedia-centric Internet of Things (mm-IoT) (2016), Multimedia Tools and Applications, for the Special Issue on Recent Advances in Mobile Cloud Computing (2017), Wireless Communications and Mobile Computing, for the Special Issue on Security and Privacy of Multimedia Big Data in the Critical Infrastructure (2017), Multimedia Tools and Applications. He is currently GOLD member committee of IEEE Broadcast Technology Society (BTS) and a member of the IEEE Industrial Electronics Society (IES), TC on Cloud and Wireless Systems for Industrial Applications. From 2017 he is serving as an ASSOCIATE EDITOR for IEEE ACCESS. He is also a member of the European Commission (EC) EURAXESS Links JAPAN and member of the EU-JAPAN Centre for Industrial Cooperation.