STATE OF RESEARCH ON PHISHING AND RECENT TRENDS OF ATTACKS

By

ANIKET BHADANE *

SUNIL B. MANE **

* M.Tech Scholar, Department of Computer Engineering, Government College of Engineering Pune (COEP), Shivajinagar, Pune, Maharashtra, India.
** Associate Professor, Department of Computer Engineering and Information Technology, Government College of Engineering Pune (COEP), Shivajinagar, Pune, Maharashtra, India.

Date Received: 06/03/2018 Date Revised: 08/03/2018 Date Accepted: 25/06/2018

ABSTRACT

Phishing attacks cause companies and individuals huge economic as well as intangible damages. Phishing attacks employ a litany of attack vectors. To deal with such attacks, counter work needs to be done in several areas. In this paper, the authors have presented a survey of literature on phishing detection and the current trends in phishing. The authors have also mentioned that phishing detection can be classified into three main categories namely, disallowing attacks to reach the users, user training, and more useful user interfaces. The goal of this paper is to cover all important aspects involved in phishing detection as compared to existing surveys on phishing detection that have focused on individual aspects. There has been a continuous increase in phishing attacks, with a sharp rise in Spear phishing and attacks over Social Media.

Keywords: Phishing, Social Engineering, Phishing Detection, Usable Security, Human Computer Interaction.

INTRODUCTION

Normally, phishing messages and websites impersonate as trusted entities. These often use social engineering techniques to deceive targets into doing actions which are of the attacker's benefit (Social engineering (security) page Version ID: 800193757. 2017). Business E-mail Compromise (BEC) has reported a loss of 3 Billion USD (Internet Crime Complaint Center (IC3), 2016) and the numbers of attacks are continually increasing (Phishing Activity Trends Report). More than 80% of the organizations have faced phishing attacks (Crowe, 2016). Phishing attacks take advantage of recent events such as Equifax hack to hook users into traps (Equifax or Equiphish? - Krebs on Security, 2017; Kennedy, 2017). An attack impersonating Google Docs affected almost a million users (Robertson, 2017), and the DNC impersonation hack led to leak of political data (Vaas, 2016).

The past work on phishing detection can be categorized into three categories: understanding why people fall for phishing automated software techniques to detect phishing, training people to not fall for phishing attacks, and better user interfaces to help people make better decisions when confronted with an attack (Zhang et al., 2007). Previous studies focused on these individual aspects of phishing, with limited number of studies such as presenting organized study of all aspects involved in phishing (Hong, 2012). There are two main categories considered, in the study on phishing: one which tries to understand human psychology involved in phishing, and the other which develops techniques to mitigate phishing.

Hong (2012) have published an article mentioning different aspects involved in phishing. An ordered study of the current state of phishing, considering all aspects involved has been provided for readers to get acquainted with all the aspects in phishing so that they can further study about specific aspects in details. The details of the user psychology involved in phishing and the taxonomy of Anti-phishing techniques and the different directions of research that need to be taken for complete mitigation of phishing have also been provided. This survey also gives details of recent studies on specific trends in phishing that are seen nowadays. The related topics such as spam are not covered in detail, but such points are touched.

In Section 1, a Background on details of Phishing is given. In Section 2 Anti-Phishing techniques as per the taxonomy have been proposed. The learnings of the study with the different lines of research in phishing has been provided, and concluded in last section.

1. Background

Spam messages are "unsolicited" messages sent for commercial benefits, often in bulk quantities. Spam messages, in general, need not to be pretended as someone else. For example, a health and medicine company advertising its products. Whereas, Phishing messages are subset of spam messages, which "pretend to be someone else", such as, a phisher pretends to be from PayPal to get sensitive information from user. Gmail provides the option to divide the inbox into Primary, Social, Promotions, Update, and Forums Tabs. The emails in later four categories are also sent in bulk quantity. But these are the ones that users have "signed/subscribed for" and can sometimes be personalized by the user. For example, a user has signed for receiving weekly newsletter from some site. The site will be sending bulk emails to all its subscribers. Such emails will go into user's Promotions tab since user has subscribed for it. Outlook also provides the option to divide the inbox into Focused and Other Tabs, with the Other tab having mails such as those in Gmail's later four category tabs. Figure 1 shows diagrammatically, the classes of messages in the vicinity of phishing messages. Gmail adds Phishing emails to the Spam folder itself. On viewing mails in the spam folder, it shows a warning message in red or yellow banner describing the type of message. Also, Gmail maintains sender reputation as a feature to decide whether a message should be treated as spam or not, among other spam related features. A spam/ phishing mail sent from a reputed sender landed in the receiver's inbox, whereas the same mail sent from a less reputed sender landed in Spam folder, has been observed. Self-reputation can be seen in Gmail Postmaster (Postmaster Tools – Google. 2017). Mail service providers implement anti-phishing measures on server side. Also, various anti-phishing browser extensions are available on the client side which can be installed by individual users.

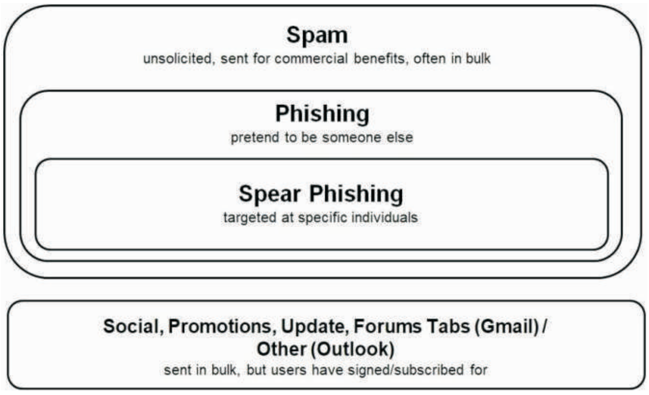

Spam
unsolicited, sent for commercial benefits, often in bulk
Phishing
pretend to be someone else
Spear Phishing
targeted at specific individuals
Social, Promotions, Update, Forums Tabs (Gmail) / Other (Outlook)
sent in bulk, but users have signed/subscribed for

Figure 1. Classes of Messages in the Vicinity of Phishing Messages

Phishing is a semantic attack. It exploits human vulnerabilities by targeting the way humans interpret content displayed on the system. This is commonly done using social engineering.

A phishing attack or a phishing taxonomy using phishing messages has three major stages: First stage corresponds to Attack Lure where the victim(s) receive a socially engineered phishing message. Second and Third stages correspond to Attack Exploit, where victim(s) perform action as suggested in the message, and attacker exploits the action performed by the victim(s) for his benefit.

Phishing messages persuade users to click on an URL in the message to enter sensitive information on resulting page, or replying to the message with sensitive information, or performing money transfer, etc. When using phishing websites to steal users' credentials and sensitive information tied to the target site, phishers may use free web hosting services, or register a new domain, and may also use compromised machines to host their files. Many phishing attacks made no attempt of disguising URL as target site, and were successful. Very few phishers registered domain names that were confusingly similar to the brands, which shows that phishers do not need deceptive URLs to fool users (Phishing Activity Trends Report). Today, most of the phishing sites are created using Toolkits, which makes it very easy for attackers to create websites visually similar to their targets. The phishing website's URL may be customized to create innumerable URLs by adding random parameters, with all these URLs pointing to the same phishing site.

Phishers may use vulnerabilities in the Domain Name System (DNS) to divert internet traffic to their phishing websites. In case of DNS cache poisoning, the DNS returns IP of the phishing website instead of the correct IP of the domain name. In another technique called domain shadowing, phishers compromise a legitimate domain name's DNS to set up new subdomains. These new subdomains can then be used to point to the phishing content.

Social media platforms such as Twitter have their own phishing and malware detection mechanisms but are frequently by-passed with the use of URL shorteners and multiple redirections (Aggarwal et al., 2012). Automated spear phishing tools have also been developed to generate phishing tweets. One such tool uses machine learning techniques to generate tweets based on existing spear phishing data, and the topics extracted from timeline posts of the target and of those they retweet or follow (Fox-Brewster, 2016).

Phishing messages also employ Technical Subterfuge. These schemes plant crimeware onto PCs. An attacker can send an email to an employee masquerading as IT department of the company, asking the employee to install a security patch which is actually a malware.

If the victim performs action as desired by the attacker, the attacker then exploits this action performed by the victim. For example, if the attack involves stealing of sensitive information or credentials, phishers may monetize the information themselves or sell the information on underground network marketplaces. Or if the attack involves installation of a malware, the attacker fetches/monitors sensitive data.

A particular type of phishing which constituted 90% of all phishing attacks is Spear Phishing. Spear Phishing is one of the biggest threats to corporate companies today with 95% of all these attacks are being on corporates (Weinberg, 2013). In contrast to normal phishing attacks, which target general public, spear phishing attacks are targeted at specific individuals or employees of an organization. Attackers gather specific information about their targets through their social network or acquaintance etc., and use this information to create customized phishing messages for the particular target group.

A form of spear phishing attack, called Whaling, is directed specifically at senior level executives in businesses and other high-profile targets. At toy making company Mattel, a high level financial executive received an email requesting money transfer of $3 Million, impersonating as from the newly assigned CEO (Cimpanu, 2016).

The calculation of cost of damages caused by phishing attacks differ extensively, and are largely dependent on the assumptions made by the organizations when calculating the damages and different departments affected within the organization. Phishing attacks not only cause direct damage, but also indirect damages such as damage of reputation, leakage of source code and intellectual property, etc (Khonji et al., 2013).

Phishing Honeypots (network decoys) can be used by researchers and organizations to bait attackers. They are purposely kept vulnerable, so that attackers can be lured to use such resources. These honeypots are isolated and monitored. They can be used to capture activities of phishers, which can then be used for research purposes to get better understanding of attack flow and trends.

Many phishing websites are found to redirect automated scripts or bots, such as web crawlers, to legitimate domain, but redirect browsers to phishing domain. This is done using robot.txt file, which is used by websites to communicate with web crawlers (Aggarwal et al., 2012).

Other recent trends of phishing observed are:

- Social media phishing attacks increased by 500% in Q4 of 2016. Huge increase was seen in the use of fraud accounts masquerading as customer support of popular brands. This tactic is relatively new and is called angler-phishing, where attackers register and use fake Twitter accounts masquerading as customer support of some brand (Muncaster, 2017).
- Specific brands were attacked more than regular on specific occasions, such as holiday season (Phishing

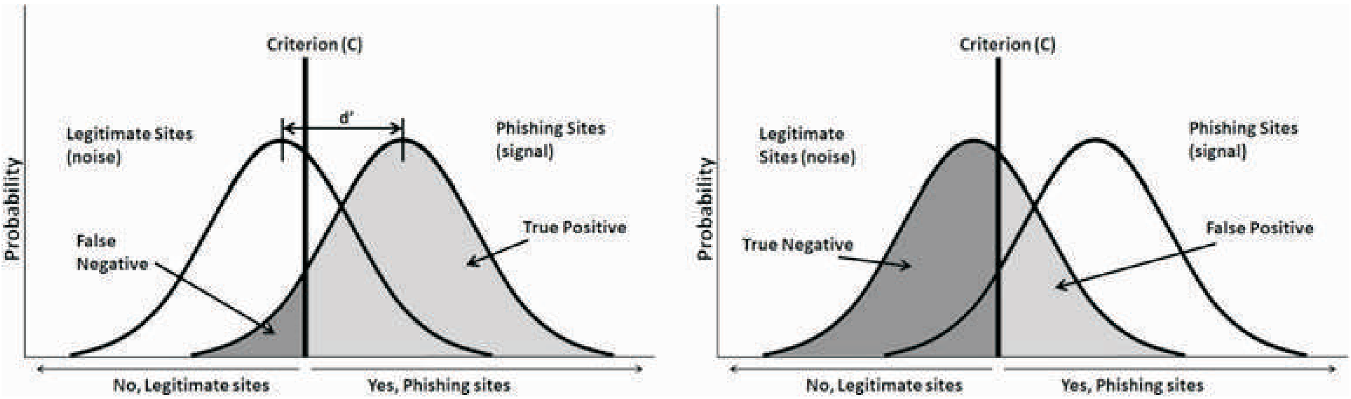


Figure 2. Use of Signal Detection Theory (SDT) in Phishing

Activity Trends Report).

- The top ten targets faced over 75% of all the phishing attacks in 2016 (Phishing Activity Trends Report).
- Attackers also were found to be using Internet Protocol (IP) filters on Phishing sites, to disallow people from other countries to visit the site and even people from the target company (Phishing Activity Trends Report).
- It is important to block an attack in the initial few hours of the attack, as a large percentage of users read the phishing message till the blacklists are updated. As found by (Jagatic et al., 2007), most of the users fell victim to a phishing attack in the first few hours of the attack itself.

Signal Detection Theory (SDT) has been used by various studies to measure user vulnerability to phishing attacks (Sheng et al., 2007; Canfield et al., 2016; Kumaraguru et al., 2010; Nicholson et al., 2017). SDT is used to quantify or measure the ability of users to distinguish between signal (phishing) and noise (legitimate). It has two factors involved: Sensitivity (d') and Criterion (C). Sensitivity measures users' ability to differentiate between signal and noise. It is the gap between the means of the two distributions. The further apart the distributions, the greater the sensitivity or d'. Criterion is defined as the user bias or tendency when making a decision. It is measured by how far their decision threshold (criterion line) is from the intersection of the two distributions. Regions C<0 and C>0 indicate more cautious/alert users and less cautious/alert users respectively. Figure 2 shows how the criterion line divides the graph to depict false positives, true positives, false negatives, and true negatives. A User Training approach should typically aim to i) increase users' cautiousness/alertness by shifting the criterion line to the right; ii) increase users' sensitivity, by increasing the separation between the two distributions, so people would be able to better differentiate between phishing and legitimate; or iii) a combination of i) and ii) (Sheng et al., 2007).

2. Anti-Phishing Techniques

The tree classification/taxonomy is presented in Figure 3, from the article by Hong (2012).

Solutions to mitigate phishing can be briefly classified into three.

- Make it Invisible
- Train Users
- Better Interfaces.

2.1 Make It Invisible

The solutions in this category focus on preventing phishing attacks from reaching the end user.

2.1.1 Filtering Messages

Email remains to be the most used carrier for phishing attacks. But there is considerable increase in phishing attacks carried over social media (Phishing Activity Trends Report). But phishing messages are not limited to emails, and are also sent over Social Media, Online Multiplayer Games, SMS, VoIP, etc. The phishing messages used over different media have different properties and challenges to be detected. There has been extensive research on spam detection, but phishing messages have different

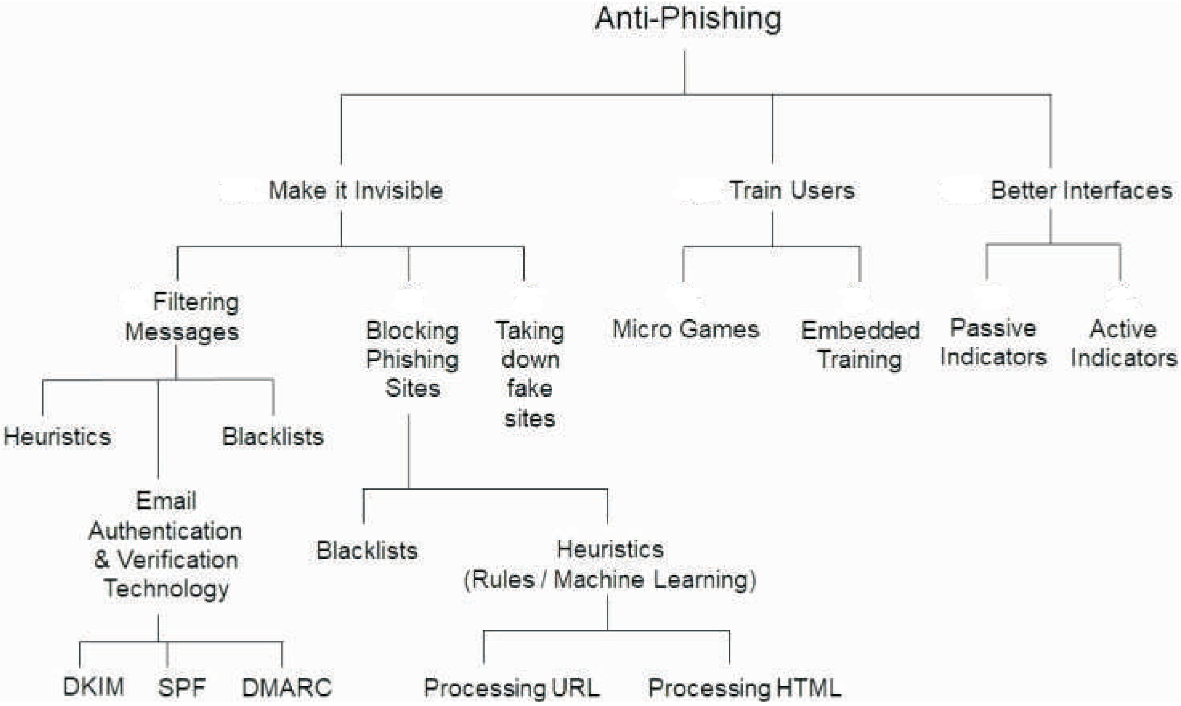


Figure 3. Classification/Taxonomy of Anti-Phishing Techniques

characteristics as explained earlier, and there exists comparatively lesser research on detecting various types of phishing messages.

The recipient of the phishing message falls prey to the attack when he performs action intended in the phishing message, such as clicking on URL in the message and entering sensitive information on resulting page, or replying to the message with sensitive information, or performing money transfer, etc. Phishing messages also use sentiments such as curiosity, fear, urgency, etc. in order to make the user perform the action as intended by the attacker. The ways to detect and prevent phishing messages as present in literature are described in this paper.

2.1.1.1 Heuristics

Heuristics, to detect phishing emails at server side, mainly uses machine learning techniques.

Fette et al. (2007) presented a method, PILFER, which is a machine learning based method for classification of phishing emails. They used Random Forest as classifier, with 10 decision trees. The classifier is trained and tested using 10-fold cross validation, and used only 10 features for email classification for more intuitive and faster decision process.

Almomani et al. (2013) provided a comprehensive survey of techniques used for detection of phishing emails. The general categorization of features of phishing emails is shown in Table 1. Large number of features can be used for classification of phishing emails. For example, Beghohlz et al. (2010) as mentioned in (Almomani et al., 2013) used 81 features. Detecting phishing attacks at network level is based on domain and IP address blacklisting, such as Domain Name System-based Blackhole List (DNSBL) and Snort. They also mention a survey of techniques to detect DNS-poisoning and mention k-nearest neighbor to be the best performer. Server side filters, and classifiers use machine learning algorithms. They also provide category wise comparison of the algorithms. There does not exist a standard classifier for detection of phishing email detection. Most classifiers used are based on supervised learning and most of the work is done offline. The authors provided a detailed comparison of these techniques. The techniques which they compared are shown in Figure 4. Khonji et al. (2013)

FeatureType	Description
Structural Features	Features based on structure of email; MIME Structure
Link features	Number of links, Number of domains, Features mentioned in Tables 3 and 4
Element features	Presence of Hypertext Markup Language (HTML), JavaScript, etc.
Spam filter features	Most of the researchers use the SpamAssassin tool whichuses very large number of features
Semantic Keywords features	Keywords and Groups of keywords commonly used by phishers. Ex: verify; confirm; urgent; "click" and "account";"market", "plan", and "prices".
Dynamic Markov Chain Features	Features of email text based on bag-of-words, makes classes of messages. 47 phishing classification features mentioned in series of papers.

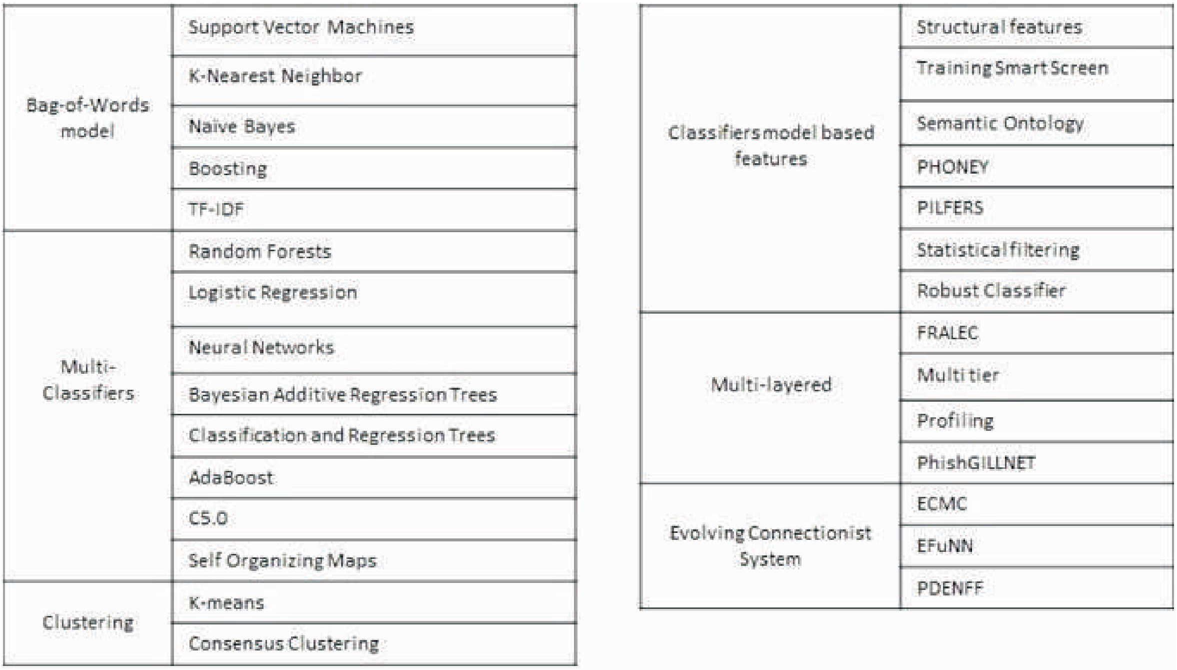
Table 1. General Categorization of Features of Phishing Emails

mentioned that phishing classifiers having good performance do not rely on Natural Language Processing (NLP) techniques, as these techniques have shown low detection rates. According to them, among techniques available in public literature, Berghloz's model-based email classifier (Bergholz et al., 2010) has the highest accuracy in detection of phishing emails, followed by Toolan and Carthy (2009).

A type of phishing which has considerably risen in recent times is Spear Phishing. As Spear Phishing was described earlier, it is important to mention the work done on detection of emails used to carry out this type of attack.

Khonji et al. (2011) proposed Anti-Spear phishing Content-based Authorship Identification (ASCAI). This framework builds a white-list of trusted e-mail senders based on the stylometric features of their email bodies, which are then used to detect identity impersonation attempts for daily email use, by comparing the stylometric features of incoming email bodies with those trusted senders.

Stringhini and Thonnard (2015) proposed Identity Mailer which aims to identify whether the claiming sender has really written the email or not. The system works at sender's side and uses email header information and stylometric features. By comparing a user's email with the emails of those written by others, it extracts features characterizing the user's behavior such as email writing habits, email composition habits, and recipient interaction habits. Similar to Identity Mailer, Duman et al. (2016) proposed Email Profiler which aims to identify whether the email



Bag-of-Words model	Support Vector Machines
	K-Nearest Neighbor
	Naive Bayes
	Boosting
	TF-IDF
Multi-Classifiers	Random Forests
	Logistic Regression
	Neural Networks
	Bayesian Additive Regression Trees
	Classification and Regression Trees
	AdaBoost
	C5.0
	Self Organizing Maps
Clustering	K-means
	Consensus Clustering

Classifiers model based features	Structural features
	Training Smart Screen
	Semantic Ontology
	PHONEY
	PILFERS
	Statistical filtering
	Robust Classifier
Multi-layered	FRALEC
	Multi tier
	Profiling
	PhishGILLNET
Evolving Connectionist System	ECMC
	EFuNN
	PDENFF

Figure 4. Server Side Phishing Email Filters using Machine Learning Algorithms (Almomani et al., 2013)

originates from the claimed sender, and detects Spear Phishing emails using features extracted from email metadata and stylometric information. They use a Support Vector Machine classifier. As opposite to Identity Mailer, Email Profiler is implemented at recipient's side and builds a behavioral profile for each email sender.

Dewan et al. (2014) made use of features of user's profile on an online social network to detect spear phishing emails. They used dataset from Symantec's enterprise email scanning service of 14 companies. They tested four machine learning models trained on these Social features extracted from LinkedIn profiles of employees, alongwith the stylometric features of emails (27 features in total; 9 social,18 stylometric). Their results showed that the social features that they extracted from LinkedIn did not help in determining spear phishing emails.

Very recently, Ho et al. (2017) developed a new approach for detecting credential spear phishing attacks in enterprise settings called Directed Anomaly Scoring (DAS), which is an unsupervised, non-parametric technique. The authors focused on emails using social engineering techniques to trick user to click a link in the email and enter credentials on resultant phishing website. They do not focus on malware attachment involved spear phishing attacks which, as they describe, can be dealt with having strong security infrastructure which is frequently updated, in which case the attacker will need zero day malwares. The authors mentioned two stages of spear phishing attacks: Lure and Exploit Payload. Further, they mentioned four different ways of impersonation, namely, Address Spoofer, Name Spoofer, Previously Unseen Attacker, and Lateral Attacker. Since organizations can deploy Domain Keys Identified Mail (DKIM)/Domain-based Message Authentication, Reporting and Conformance (DMARC) to mitigate address spoofing, they focus on detecting later three types of impersonation. Their security goals are to produce extremely low false positives, like 10, per day and their detector should detect real spear phishing attacks (true positives). They use 4-year dataset of over 370 million emails from a large organization. Their work draws on the Simple Mail Transfer Protocol (SMTP), Network Intrusion Detection System (NIDS) and Lightweight Directory Access Protocol (LDAP) logs from the organization, whose email contents were anonymized for privacy reasons by the organization before they received this dataset. These logs were used for looking at logins from new IPs, total logins per employee, inactivity periods, and others. They mentioned two major challenges in detecting spear phishing attacks, which are Senders with Limited Prior History and Churn in Header Values. They described two classes of features: sender reputation features in Lure stage, and domain reputation features in Exploit stage. They achieved false positive rate of less than 0.005% and detected 17 out of 19 spear phish attacks. They also mentioned the limitations of Standard Detection Techniques, and mentioned their technique to be better, since it leverages specific domain knowledge. They mentioned that machine learning techniques to detect spear phishing attacks have false positive rate of more than 1%, which is too high for their setting.

There has been extensive work done on spam detection in social networks, but not much on phishing detection in social networks.

Aggarwal et al. (2012) proposed PhishAri, a Chrome extension, to detect phishing Tweets on Twitter in real time. The features that they used are URL based, WHOIS based, Content based, and Sender based. They found that Random Forest classifier was the best performing classifier (in comparison with Naive Bayes and Decision Trees) with an accuracy of 92.52%. They mentioned that phishers use trending tags to a large extent in phishing tweets. At hour zero, they detected more phishing URLs than PhishTank's and Google Safe Browsing's blacklists, and Twitter's own defense system.

Over a set of multiple studies Dewan et al. (2016); Dewan and Kumaraguru, 2015; 2017 studied malicious posts and pages on Facebook. They do not work explicitly on phishing posts, but work on class of malicious posts which also contain phishing posts. They used comprehensive list of features based on facebook profile, post's text, metadata, and link-based features to detect malicious content on Facebook at hour zero. They use labeled dataset to train and evaluate multiple supervised learning models. They experimented with Naive Bayes, Logistic Regression, Decision Trees, Random Forest, AdaBoost, Support Vector Machines and Neural Networks Classifiers. For detection of malicious posts, they found Random Forest classifier to be the best performing. For detection of pages posting malicious content, they found Artificial Neural Networks trained on a fixed sized bag-of-words performed best achieving an accuracy of 84.13%.

Summarizing, heuristics using mainly machine learning techniques are implemented at server side to detect phishing messages. There does not exist a standard classifier for detection of phishing messages and no heuristic technique proposed till now is foolproof. Most classifiers are based on supervised learning and most of the work is done offline. Sender reputation plays a big role in detection of social engineering messages. Although heuristics are able to detect some of the attacks at zero-hour, they also have side-effects of false positives. False positives are a major liability when deploying aggressive technical solutions in corporate settings. A business related legitimate message being flagged as phishing can have serious implications on the business.

2.1.1.2 Email Authentication and Verification Technology

Sender Policy Framework (SPF), DKIM (Domain Keys Identified Mail) and DMARC (Domain-based Message Authentication, Reporting and Conformance) email authentication techniques have been discussed in this section. All these three technologies provide the framework for businesses to combat spoofed emails.

Email authentication techniques protect against email spoofing and operate at server level. They are used to combat phishing techniques using forged sender addresses that appear to originate from legitimate organizations.

SPF tells the receivers which mail servers are used to send mail from sender's domain. SPF, using TXT entry in Domain Name System (DNS), specifies which servers are allowed to send email on behalf of a domain.

DKIM authenticates the source and its contents. In DKIM, sender of the message has to sign the message, both headers and content, using his private key and put the signature into a “DKIM-Signature” field. The receiver (at the Message Transfer Agent (MTA) level) checks validity of the signature using Public key obtained by querying DNS servers. The public keys are published in DNS TXT records.

DMARC builds upon SPF and DKIM. DMARC defines a scalable mechanism by which a mail sender can express using DNS records (DMARC records), domain level policies governing how messages claiming to come from his or her domain should be handled when they do not fully align with DKIM and SPF lookup results.

If a receiver performs DMARC record, it lookups on a message claiming to come from a sender's domain, and if it does not align with SPF, DKIM, or both, the sender's DMARC record can tell receiver how it wants the receiver to handle the messages that are unaligned with SPF & DKIM. The sender's DMARC record can specify whether it wants the receiver to accept, quarantine, or reject unaligned messages, and also specify what percentage of unaligned messages sender wants receiver to reject or quarantine based on sender's policy preferences. Additionally, it provides a reporting mechanism of actions performed under those policies. For example, Gmail uses DMARC to reject all unauthenticated mails from eBay and PayPal.

Strict SPF policies can prove to be troublesome for employees wanting to send emails from their private email address, when they're unable to use their corporate address. SPF verifies immediate sender of an email, hence it does not work well with forwarded messages, since only the forwarded message will be verified and not the original message. SPF can also cause inconvenience for organizations that to add and move mail servers frequently, since SPF requires explicitly mentioning these servers to send mail on behalf of the domain (Cardinal, 2012).

DKIM does not help in cases where attackers buy their own domain names and set up DKIM on them. Their sending address would now seem to be legitimate. Attackers can use SPF and DKIM to make their message look legitimate, and can also compromise resources for carrying out attacks. DKIM has also been subjected to replay attacks (Cardinal, 2012; Blog, 2014).

While DMARC is an important factor in a multi-pronged spear phishing defense strategy, it does not offer total protection. DMARC does not counter more indirect attacks spoofing a company's brand using the Display Name and deliberately misspelled "cousin" domains. In addition, if a DMARC "reject" policy is moved too soon, legitimate email that is not properly authenticating might get blocked along with the suspicious mail. Then, the company will be at the risk of blocking important internal business emails like payslips and internal broadcasts as well as communication from suppliers like invoices. Before moving to "reject", full visibility into the email authentication ecosystem is needed.

Every domain must have an SPF, DKIM set up for these techniques to be truly effective. Out of the top million mail servers listed on Alexa, only 47% have setup SPF, and only 1% have DMARC; and out of those who have setup SPF policies, 29% have too many number of addresses specified (Durumeric et al., 2015). 2017's DMARC adoption report by phishing prevention specialist (Agari dmarc adoption report Open Season for Phishers) mentions that two-thirds of the Fortune 500 are yet to implement DMARC. Even among those who are aware of DMARC, 25 percent of the Fortune 500, 26 per cent of the Financial Time Stock Exchange (FTSE) 100, and 23 per cent of the ASX 100 are using it for anything more than monitoring. "Quarantine" or "reject" only appeared in eight per cent of Fortune 500 companies, 7 percent of FTSE 100 companies, and four per cent of ASX 100 companies. At RSA Conference 2011, experts suggested that the best way to fight phishing was collaborating with webmail providers (Cobb, 2011).

So if these authentication techniques have global adoption and correct implementation, they can be prevented against phishing attacks using email address spoofing, but not against phishing attacks which do not use email address spoofing, and as discussed earlier, many phishing attacks don't need email address spoofing to be successful.

2.1.1.3 Blacklists

DNS-based Blackhole Lists (DNSBLs) (also known as a 'Blacklists' or 'Blocklists') can be used by mail servers to get suggestion on an email's origin. Some DNSBLs, such as Spamhaus DBL, maintain list of domains used for sending phishing mails and claim almost zero false positives, so that it is low-risk to be used by production mail systems to reject emails that are flagged by it. As mentioned by Spamhaus DBL (The Spamhaus Project Frequently Asked Questions (FAQ) 2017), domains found in spam messages, domain reputation, Internet Service Provider (ISP) reputation, including domain's A, NS, MX and website DNS records are some of the factors involved in inclusion of domains in DBL.

Wikipedia has a list and comparison of DNS blacklists (Comparison of DNS blacklists 2017). Policies used by DNSBL depends mainly on its Goals, Nomination, and Listing Lifetime. One can not find the details of detection of attacks from domains used for the first time for phishing, by DNSBLs. But as with all blacklists, it is suspected that DNSBLs too have low detection rate of attacks from new domains.

There have been criticisms against DNSBLs such as blocking legitimate mails from shared mail servers, treating IP address from ISPs as dynamic and hence not suitable for sending mails directly, etc. (DNSBL 2017). DNSBLs can be used for not accepting emails from domains used for sending phishing emails. But attackers can use reputed or new domains for carrying out phishing attacks, and users do fall for these attacks in the first few hours before the domain is blacklisted. After presenting various techniques to filter phishing messages, the techniques for detecting and blocking phishing websites are presented below.

2.1.2 Blocking Phishing Sites

There exists a significant body of academic work focusing on phishing sites detection.

2.1.2.1 Heuristics

A new website can be classified as phishing or not by analyzing its Uniform Resource Locator (URL) and/or HTML content. From existing literatures, the features used for the analysis of phishing sites' URLs are presented in Tables 2 and 3 and the features used for the analysis of phishing sites' HTML are given in Table 4. These are not exhaustive lists, but they mention different types of features to give generalized idea about the features used. These features are used to define rules or train machine learning classifiers to detect phishing.

Ma et al. (2009) categorized URL features as Lexical-based and Host-based. Lexical features are the textual properties of the URL. Host-based features include where

Feature Type	Description
IP address properties	Whether IP present in a blacklist; Which autonomous systems or prefixes do the IP addresses of DNS records belong.
WHOIS properties	Registration, expiration, update dates; Registrar, Registrant; Whether WHOIS entry is locked.
Domain name properties	Time-to-Live (TTL) of DNS records of hostname; Whether host's PTR record exists; Whether one of host's IP addresses is resolved by PTR record.
Geographic properties	Continent/Country/City of IP address; Connection speed.
PageRank, popularity/reputation of the URL	A reputed page has lesser chances of being phishing. E.g. Using Google PageRank to get popularity of URL.

Table 2. Host-based Features used for the Analysis of Phishing URLs (Ma et al., 2009)

Feature Type	Description	Example
Top Level Domain (TLD) variant URL in blacklist	Checking whether different TLDs of the URL exist in blacklist	www.legitsite.com www.legitsite.fr www.legitsite.in etc.
IP Address Equivalence	URL with similar directory structure to a phishing URL pointing to same IP address	www.login-site.com/online.paypal.com and www.site-login.com/online.paypal.com both pointing to same IP address
Query String Substitution	Changing various parameters in URL	www.login-site.pr/online/paypal.php?abc and www.login-site.pr/online/paypal.php?xyz
Nonmatching URLs	Link where text is URL but href attribute is differentthan the URL	<a href="www.paypallogin.com">https://www.paypal.com</a>
More number of domains	Vulnerabilities in a site's code can be used to redirect users from legitimate site to phishing site, by adding phishing URL in the legit URL.	www.goodsite.com/info.php?url=http://www.goodsite-login.com hastwo domains and would redirect the user to goodsite-login domain which would be phishing.
More number of subdomains	Phishers make use of subdomains to make the URL look like legitimate	https://in.paypal.login-us.com/verification.asp?d=1
Presence of certain tokens/keywords	Keywords commonly used by phishers to fool users	E.g. login, banking, -, @, etc.
Shortened URL	Use of URL shorteners toobfuscate URL	bit.ly/cikl0z Does not reveal real URL
URL in Browsing History	URLs that user has visited before have lesser chance ofbeing phishing	—
Domain Age	Domains existing from longer time have lesser chances of being phishing	E.g. more than 30 days

Table 3. Lexical Features used for Analysis of Phishing URLs

Feature Type	Description
TF-IDF with search on Google	Finding terms with highest tf-idf in page, searching them on google, and checking whether current URL is among first k results.
Comparison of DOM Trees	Feature Vector from HTML tags, Hash of individual HTML tags, etc.
Hyperlink relationships in page	E.g. Relation between phishing and target page, page having links pointing to legit site, images, etc.
Snapshot/Image Comparison	Visual Similarity Comparison
"URL Login User Interface (LUI) - DOM IP" mapping	Storing this feature of a page as a vector in whitelist and checking whether any of whitelisted vectors exist on other visiting pages (Cao et al., 2008).
Presence of Form	Whittakes et al. (2010) found most phishing pages to have password fields.
Login form's brand not in domain part of URL	Detecting login form belonging to a brand, but placed on site not belonging to the brand.
Login or banking trademarks of legit site	Finding images on a page depicting a brand, but placed on site not belonging to the brand.

Table 4. Types of Features used for Analysis of Phishing Websites' HTML Contents

the site is hosted, who owns the site, and how is it managed. They mention extensive list of features used for the analysis of URLs. They use Statistical methods for classification of phishing pages and report prediction accuracy of 95-99%.

PhishNet exploits some patterns used by attackers in making new phishing URLs (Prakash et al., 2010). PhishNet makes many variations of a URL to check whether these variations are present in blacklist. It uses five heuristics for the variations (Types 1-5 in Table 3), and also performs tests of DNS, Transmission Control Protocol (TCP), Hyper Text Transfer Protocol (HTTP), Content Similarity to filter nonexistent and innocent children URLs.

Spoof Guard was a browser plugin compatible with Internet Explorer proposed by Chou et al. (Chou et al., 2004). It calculates a score to evaluate the possibility of the page being phishing, by analyzing URL (obfuscation, non-standard port numbers, user browsing history, http(s) based phishing attempts, analyzing domain), analyzing logos in page, analyzing links in page, and whether page contains password field.

CANTINA was proposed by Zhang et al. (2007). To detect phishing sites, it uses various lexical-based, host-based, and content-based features, along with TF-IDF (Term Frequency-Inverse Document Frequency) to get important terms in the page and observing whether the current page is present in the top few entries of Google search result.

More studies such as by Garera et al. (2007); Ludl et al. (2007) and Ma et al. (2009) use various features extracted from URL and HTML of site to find whether its phishing or not.

Rosiello et al. (2007) proposed a method based on the layout similarity between a phishing site and its target. Their analysis is based on the comparison of the Document Object Model (DOM) trees. Chen et al. (2010) compared the visual similarity of the phishing site and its target, applying Gestalt theory on the image screenshot of the sites. Afroz and Greenstadt (2011) proposed PhishZoo which uses the profiles of the targeted sites' appearances to detect phishing sites, using computer vision techniques. Jain and Gupta (2017) provided analysis of visual similarity based approaches for phishing websites' detection.

Wenyin et al. (2012) proposed a method using embedded links: a Web graph of links is built and the so-called "parasitic coefficient" is used to measure the relationship between the phishing sites and their targets. Ramesh et al. (2014) also looked at the hyperlink relationships between phishing sites and their targets.

Whittaker et al. (2010) described the technique used by Google to identify phishing websites and create their phishing blacklist (Google Safe Browsing Wikipedia, 2017). They used 3 main strategies: URL analysis, Hosting information, Contents analysis. The URL and hosting features extracted were: IP address for its hostname; whether the page has many host components; characteristic strings such as login, sign in; URL metadata such as PageRank; domain reputation score from Gmail's anti-spam system; IPs, nameservers, nameservers' IPs, city, region, country. The Page Content features extracted were: Page HTML, iframes, image, and javascript content; Extent to which pages link to other domains in terms of both HTML hyperlinks and images; Highest TF-IDF terms; whether page has password field. They used a proprietary implementation of the online gradient descent logistic regression learning algorithm, with five cross-validation models, and a candidate model. The properties of their classifiers are: Prioritizes precision over recall; Tolerates noisy data; Have low latency. User submitted phishing pages and reported error are manually examined. They achieved false positive rate of below 0.1%. They found that no phishing pages have high PageRank scores, and high proportion of phishing pages have a password field.

Cui et al. (2017) monitored a total of 19,066 phishing attacks over a period of ten months and found that over 90% of these attacks were actually replicas or variations of other attacks in the database. They recorded numerous attacks that stayed active throughout their observation period. They proposed a technique to find similar phishing pages, to find repetition of same attacks. They defined a feature vector based on the DOM of the web pages obtained from the attacks (counting occurrences of tags and making vector of the counts in arbitrary pre-defined order of tags), and then used Hierarchical Clustering to cluster together vectors that are close to one another. A typical in-depth defense strategy against phishing consists of a first line of fast altering, which will flag some sites as potential phishing sites, followed by a second line of much slower, in-depth analysis to confirm that the flagged sites are indeed attacks. The authors proposed their strategy to be implemented between the two steps as pre-processing step to reduce the amount of links sent to second line of defense.

Their clustering results show that a small group of attackers could be behind a large part of the current attacks, and taking down such groups could potentially have a large impact on the phishing attacks observed today. They showed their false positive rate to be 0.08%.

2.1.2.2 Use of Whitelist with Heuristics

A Whitelist of sites is a user-maintained or community/ organization-maintained list of legitimate websites.

Cao et al. (2008) presented Automated Individual White-List (AIWL) which is a user personalized whitelist of vectors of Login User Interfaces (LUIs) of sites. These LUIs are of the sites where the user has submitted his/her credentials and trusted the site. A typical LUI contains URL, page features, and DNS-IP mapping. When the user visits a site, its LUI is calculated and if LUI is not present in the whitelist and user tries to submit credentials to this site, AIWL warns user of a potential attack. Also, Naïve Bayes classifier is used as whitelist maintainer, to identify a successful login attempt.

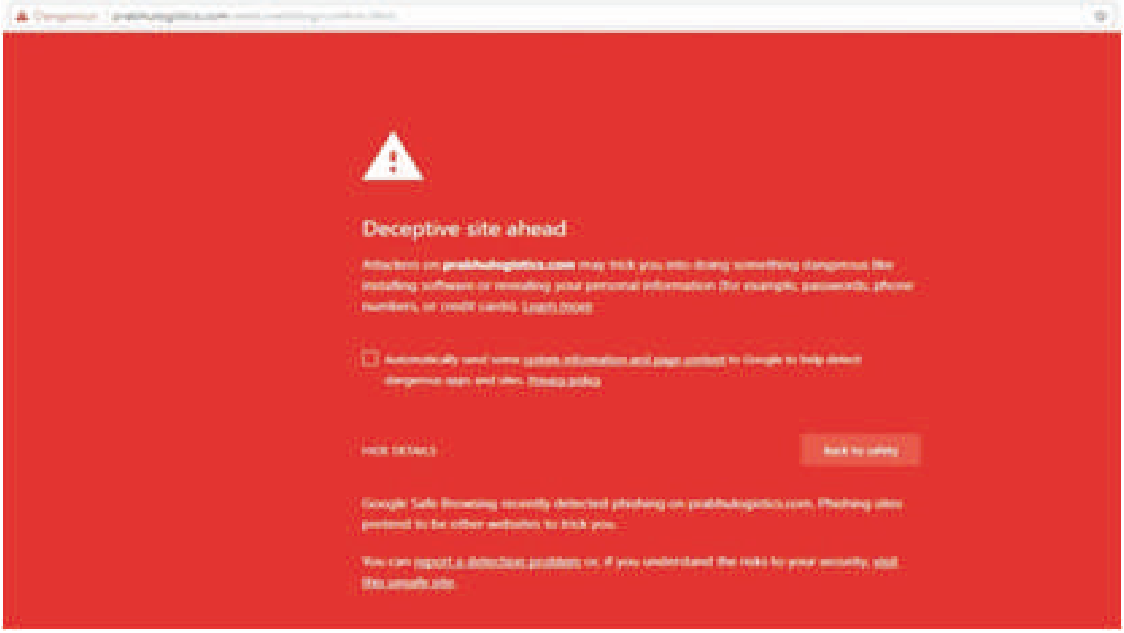
Ardi and Heidemann (2016) proposed AuntieTuna, which also uses user personalized whitelist. Their chrome extension finds cryptographic hash of chunks of user trusted page's rendered Document Object Model (DOM). It then adds URL of the page to the whitelist and stores the corresponding hashes on local storage of client. When user navigates to an unknown page, the tool finds cryptographic hash of chunks of that page and compares them with that of user trusted pages. If the number of matches of a user trusted page is greater than threshold, the tool warns the user of a possible phishing page.

Among visual comparison approaches using whitelists, Chen et al. (2009) maintained a whitelist of sites (ecommerce, banking, etc.) which are likely to be targeted. When a user visits a site, the browser takes a snapshot, which is then visually compared with sites in whitelist, using Harris-Laplace algorithm, Euclidean distance, and k-means clustering algorithm. Hara et al. (2009) used imgSeek (open source image similarity tools) to match suspected website's snapshot with websites in the database. Zhang et al. (2011) used Naive Bayesian classifier for comparing visual characteristics of suspected and legitimate websites.

2.1.2.3 Blacklists

A phishing websites' blacklist is a list of websites which have been marked as phishing by the particular blacklist service. Three such popular blacklist services are discussed here: Google Safe Browsing, PhishTank, and Anti Phishing Working Group (APWG). Google Safe Browsing is a blacklist service provided by Google for browsers and client applications to consult whether a site poses phishing or malware threat, using their public Safe Browsing API v4. Here, the technique used by Google to identify phishing websites are discussed in the previous section of Heuristics. Browsers such as Google Chrome, Safari, Firefox, Opera, and Vivaldi are known to use Google Safe Browsing blacklist (GoogleSafeBrowsing Wikipedia. 2017). Chrome browser, for example, periodically downloads the latest copy of this blacklist and stores it in the user's local system. When a user visits a site which is present in this list, Chrome contacts Google with partial hash of the URL to confirm whether the URL is actually dangerous (Google Chrome Privacy Whitepaper, 2017). If the page is confirmed to be phishing, Chrome shows a warning page as shown in Figure 5.

PhishTank is a community where users can submit phishing websites, which are then verified using community voting, and added to their blacklist of phishing websites. A user submitted site needs more than one to vote from the community to be verified as a phish, and the number of votes is dependent on the history of those who are voting. PhishTank provides developers with their free APIs to access the blacklist. The blacklist is used by organizations and services like Yahoo! Mail, McAfee, Kaspersky, APWG,



Dangerous | prabhulogistics.com [illegible]

Deceptive site ahead

Attackers on **prabhulogistics.com** may trick you into doing something dangerous like installing software or revealing your personal information (for example, passwords, phone numbers, or credit cards). Learn more

□ Automatically send some system information and page content to Google to help detect dangerous apps and sites. Privacy policy

Google Safe Browsing recently detected phishing on prabhulogistics.com. Phishing sites pretend to be other websites to trick you.

You can report a detection problem or, if you understand the risks to your security, visit this unsafe site.

Figure 5. Warning Page shown by Chrome Browser when Visiting a Site present in Google Safe Browsing Blacklist

Opera and many more (PhishTank> Friends of PhishTank. 2017).

The Anti-Phishing Working Group (APWG) is an international group which prepares and distributes phishing reports to its paying members (PhishTank > Frequently Asked Questions (FAQ) 2017). But APWG does provide trends, white papers, and articles on phishing on their website to general public. APWG has more than 3200+ members from more than 1700 companies and agencies worldwide.

A survey by Sheng et al. (2009) found zero false positive rates by blacklists on a set of 15,345 URLs. But blacklists were ineffective when protecting users against attacks at hour zero, as most of them caught less than 20%. Also, blacklists were updated at different speeds and extent, with 47% to 83% phishing websites detected by blacklists after 12 hours.

2.1.3 Taking Down Sites

Once a phishing page is detected, it must be reported to the respective ISP/hosting company in order to take down such content. Several organizations try to report such sites to ISPs. The IP address of the phishing site can be used to identify ISP/hosting company, and the owner of the domain can be found using WHOIS lookup. If the phishing page is placed on a legitimate server by compromising it, the legitimate website's owners/admins are contacted. ISPs are informed of such phishing pages. Phishing reports are normally handled by Abuse departments at ISPs. The time required to take down a site depends on the provider (Sachs, 2013).

In an approach called Fast flux, attackers try to hide a phishing website behind a large number of compromised proxies. IP addresses are changed at a very high rate by changing DNS resource records. Such attacks are more difficult to detect. Fast flux is of two types: single-flux which is a simple type of fast flux, and double-flux which is a more sophisticated type. On an average, its takes 196 hours to take down fast flux based attacks as compared to 62 hours for taking down a normal phishing attack (Hong, 2012).

Cui et al. (2017) observed that even though a phishing website is taken down, the attackers launch the website from different domains, IPs and URLs, without much modifying the website. They also mentioned that certain attacker groups are responsible for large number of attacks, and they use same resources for carrying out different attacks. Identifying these groups can stop a large number of attacks.

Taking down phishing sites and then replacing the phishing page with an educational message for people who land on the page is an innovative way of educating users. This idea is developed by APWG and Carnegie Mellon University (Welcome to APWG & CMU's Phishing Education Landing Page, 2017). This can help in educating people when they visit such phishing pages when they fall victim to such attacks. Although it is suspected that blacklists would have been updated by the time phishing sites are taken down, and browsers would show users warning to user (Figure 5) before accessing such sites. So users landing on such page would be on their own discretion ignoring the browser warning. The educational message would be even more effective in this case to educate the users about their incorrect decision.

There exists no software mechanism till date which is able to filter all types of phishing messages. Some phishing messages do reach the end users. So it is important for users to make better decisions and not fall for these attacks. This can be done by providing efficient User Training and Better Interfaces.

2.2 Train Users

User Training can be used for increasing user awareness about phishing attacks and teaching them how to detect such attacks so that they can take better informed decisions in such circumstances. User Education and Training is an important part of online security, but it has not received enough attention from the phishing community.

In a recent study by Deloitte, more than 70% organizations mentioned lack of security awareness of employees to be their major vulnerability. More than 4 out of 10 organizations don't provide security education to their education (Phishing Scamsat All-TimeHigh, Employee Training NotKeepingPace | Wombat Security. 2017). A survey by PWC showed that organization providing security education to employees are half as likely to fall for such attacks (Phishing threatens today’s economy ny times final.pdf. 2017).

Training increases the cost to the company and consumes time. But an employee falling for a phish can cause considerable amount of damage to the company. Some studies have shown training programs to be helpful (Kumaraguru et al., 2009; Sheng et al., 2010; Kumaraguru et al., 2007; Alnajim and Munro, 2009), while some disagree or say that training has mixed benefits (Caputo et al., 2014; Gorling, 2006; Gaffney, 2011).

Security is largely considered as a secondary goal, and educating users about things which are not related to their primary tasks may hit their cognitive limits (Gorling, 2006).

As evaluated by Sheng et al. (2010), after training people with the best available training programs, users were able to detect more phishing attacks, but still could not detect 29% of the attacks. Most companies have annual policy based training that's required for compliance, which have shown to be ineffective to change employees' behavior. Caputo et al. (2014) surveyed employees' behavior with phishing emails in a large organization. They found that the anti-phishing training did not help in changing employees' behavior when dealing with spear phishing emails. Users are seen not to be utilizing in their daily behavior, the knowledge they received during the training.

Users are found to not retain the knowledge (Oliveira et al., 2017) learned during the training and tend to forget after short period of time. Hence, there are studies advocating that training needs to be continuous. Organizations, banks, etc. send periodic security notices via emails or SMS's to their employees or clients about phishing threats. But such periodic notices have been found to be ineffective in changing users' behaviors (Kumaraguru et al., 2007). Studies by Kumaraguru et al. (2009); Kumaraguru et al. (2010), Khonji et al. (2013) have shown that anti-phishing training is most effective when it is done at the time when user is dealing with a phishing attack, and when it is done periodically.

Most training programs have focused on adults. Lastdrager et al. (2017) explored training of school going children against phishing. Majority of children in USA and Europe access the internet daily. Attackers can get information about a target from their social networks and children in their social network can be used to get information of the target (e.g.: of his/her parents). Their results showed that training children against phishing works only for the short term. Although, they also suggest that such security programs in school curriculum can be helpful in making future generations aware of security threats.

Oliveira et al. (2017) mentioned that demographic-tailored training and prevention approach will increase the effectiveness of security measures because a demographic-targeted solution will impose lesser requirements on people and will match their specific vulnerabilities.

The best way of training against phishing attacks is "at the moment of attack periodic demographic tailored" training.

Research on User Training against phishing is mainly done on developing Micro Games or Embedded Training.

2.2.1 Micro Games

Sheng et al. (2007) developed Anti-Phishing Phil, a game which educates users about various parts and cues to identify phishing and legitimate URLs. They used leaning sciences to build intervention based designs, and found it to be more effective than security notices emailed by companies to users. The mouse pointer is visualized as a fish, and asks users to hover the fish over worms which then show URLs. The game educates users about URLs and the fish has to eat (safe) or reject (phishing). Their evaluation with more than 4500 people showed 61% improvement in users' ability to identify phishing URLs and also decrease in false positives. Though the game teaches about URLs, it does not give real time experience of detecting a phish or about the social engineering techniques used in attacks.

Denning et al. (2013) let users interact with cards in a board game, which teaches users about different social engineering techniques attackers use. But this game is not meant for teaching users to identify phishing attacks; it makes users more aware about tricks used by attackers.

Recently, Wen et al. (2017) developed a game called What.Hack (pronounced what dot hack). They gave example of DNC hackings during the 2016 US presidential election, where staff were tricked into sharing passwords which granted access to confidential information by fake Google security emails. They mentioned that vulnerabilities such as these are due to insufficient and tiresome training when it comes to information security, and a potential solution is the introduction of more engaging training methods, which teach information security in an active and entertaining way. The authors mentioned that existing games teach users about specific aspects of phishing, such as teaching for phishing URLs, but not for malicious attachments. The authors incorporated such combined phishing techniques in their game. What.Hack presents a sequence of puzzles in a story-based game context, to teach users about social engineering threats. The player is provided with a rulebook that tells the players which emails are safe or unsafe. The player is asked to correctly identify phishing emails else it will have negative consequences. The authors have yet to test the effectiveness of their game.

Several commercial offerings are available, but their details are not available in public literature, for example Email Security or Anti-Phishing Phyllis by Wombat Security (Email Security or Anti-Phishing PhyllisTM | Wombat Security, 2017). It teaches users to identify phishing attacks, using interactive training and character-driven training game.

2.2.2 Embedded Training

In this type of training, the educational material is embedded, i.e. integrated, into the daily primary tasks of users.

An approach where simulated phishing emails are sent to users to train them, have been used with Indiana University students (Jagatic et al., 2007), West Point cadets (Ferguson, 2005), and New York state office employees (New York State Office of Cyber Security & Critical Infrastructure Coordination). The approach led to an improvement in users' phishing detection ability.

Kumaraguru (2009) developed PhishGuru, which sends simulated phishing emails to users in their normal use of emails, and uses interventional educational messages to educate them. When the users fall victim to the attack, such as clicking on a link in the phishing email, the system teaches users about the attack. They use intervention type design to show educational messages. They tested two different design types, text-graphics and comic strip, to show their interventional educational messages. They found comic strip design type to be more effective than text-graphics.

2.3 Better Interfaces

This line of research aims to have better design of phishing warnings in browsers or tools or services so that users can take better informed decisions against such attacks.

Security indicators in browsers depict connection states such as Secure HTTPS, HTTP, Broken HTTPS, and EV. Security indicators are found to be ineffective in changing users' behavior; and many users don't even look at them as these indicators fail to catch users' attention, and many who look at them don't understand what they mean (Schechter and Dhami, 2007; Wu et al., 2006).

Another major issue with warnings is that users get habituated to frequent warnings. Habituation causes users to ignore warnings (Wu et al., 2006; Hong, 2012); Egelman et al. (2008). Users were found to close warnings as soon as they see them, due to the habit developed over time due to frequent appearance of warnings, or because people don't understand the warnings, or the warnings are interrupting what users are trying to do, or frequent misclassification causing users' distrust in the warnings (Hong, 2012).

Lin et al. (2011) stated that domain highlighting (domain name of the address is highlighted in the address bar) was somewhat effective for few users, but not for most.

Herley et al. (2009) found that frequent assessment of indicators consume considerable cognitive energy of users.

Two-Factor Authentication (2FA) can be used to verify a user's identity using two different components related to the user. This may increase cost of attackers, but man-in-the-middle techniques are also used by attackers to circumvent 2FA.

In case of emails, some email services show warnings to users on top of the message and through popups, if some discrepancies are found in Gmail: This message may not have been sent by...GmailHelp. (Thunderbird's Scam Detection | Thunderbird Help, 2017). But a study of their effectiveness is yet to be done.

Egelman et al. (2008) and Wu et al. (2006) showed active warnings to be more effective than passive warnings. Users pay attention to warnings only when forced to do so and using a user-understandable warning message. Also, results from the active warning indicators depend on their design and implementation, as noted that warnings in one browser gave better results than other browser. Interruptions have shown to increase effectiveness of warning messages, but they should not be too frequent such that the user gets habituated to them and starts ignoring them.

Felt et al. (2016) mentioned that mobile indicators are lacking due to lack of space and inappropriate use of this space.

2.3.1 Passive Indicators

A passive indicator does not interrupt the user. These indicators contain text and/or icons of some specific colors. An example of passive indicators is browser security indicators, which was discussed earlier. Research has shown that many users don't look at or ignore passive indicators (Wu et al., 2006).

Some techniques help users verify whether they are on the correct site. Extended Validation (EV) certificates are used by browsers to verify the legitimacy of the certificate purchasing organization. Browsers use colors to depict the status of this certificate. In SiteKey, users are shown a secret picture when signing in, that they have pre-chosen. But like other passive browser security indicators, both EV certificates and SiteKey were found to be ineffective in changing users' behavior during phishing attacks (Jackson et al., 2007), and SiteKey suffers from Man In The Middle (MITM) and Security image attacks Information Security.

Felt et al. (2016) proposed three security indicators from a survey of 1329 people, which are now used with Google Chrome 53. They mentioned different design constraints faced by security indicators, such as scale, shape, interpretation. Also, most of the research on security indicators was done in between 2002-08, so it was necessary to understand current needs, and develop indicators that can be interpreted well by normal internet users.

Bravo-Lillo et al. (2013) designed "Inhibitive Attractors". These are User Interface (UI) elements which draw user's attention to important part of the message in a security dialog to help users make better informed decision. And these UI elements appear only when user wants to perform a potential unsafe action so as to prevent user from performing such action. The authors developed five such inhibitive attractors. Their results showed that correspondents using systems with inhibitive attractors could made significantly more number of informed decisions than those using systems without inhibitive indicators.

2.3.2 Active Indicators

Active indicators interrupt user's activities. Egelman et al. (2008) showed that an active warning message's effectiveness depends on its design. An active warning message had different performance in one browser than another due to different designs. Nicholson et al. (2017) proposed use of social saliency nudges in emails, for users to take better informed decisions in case of phishing emails. They used Signal Detection Theory to evaluate sender saliency (highlighting sender name, email, send time) and receiver saliency (showing numbers of other people in the organization receiving the same email). They used 18 emails (6 phish, 12 genuine) with 281 participants. Sender saliency was effective in improving detection of phishing emails, but receiver saliency did not have significant effect. The authors state that people with more impulsive behavior are more susceptible to phishing.

Conclusion

An organized study consisting of all important aspects of phishing detection is provided in this paper, which the authors found to be inadequate in the existing literatures. Phishing attacks have shown to be evolving over time, and continue to be a threat to corporates and individuals. Different points on phishing as addressed in literature were discussed, namely, why users fall for phishing, what are the different software techniques to mitigate phishing, what are the different user training approaches, how interfaces can be designed better to guide users to make better informed decisions, and what are the current trends in phishing. Countering phishing attacks requires work to be done in various fields. Technological solutions to counter phishing attacks are not foolproof. More research needs to be done on reducing false positive rates, when using machine learning algorithms. Training of users is effective if they retain the learning and are able to apply it in case of an attack. It is important to train users at-the-moment of the attack, periodically, and considering various user demographics. Design of user interfaces and indicators such that they guide users to take better informed decisions against such attacks, can help users in not falling for these attacks. Finally, different directions in which one can pursue research in Phishing are mentioned. No single way can mitigate phishing; hence it requires collaborative effort in all directions.

Acknowledgment

The authors would like to thank anonymous reviewers for their valuable comments.

References

[1]. **Afroz, S., & Greenstadt, R. (2011).** Phishzoo: Detecting phishing websites by looking at them. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on* (pp. 368-375). IEEE.

[2]. Agari dmarc adoption report Open Season for Phishers. Retrieved from https://www.agari.com/wp-content/uploads/2017/08/Agari_DMARC_Adoption_Report_PR1.pdf

[3]. **Aggarwal, A., Rajadesingan, A., & Kumaraguru, P. (2012).** PhishAri: Automatic realtime phishing detection on twitter. In *eCrime Researchers Summit (eCrime)*, 2012 (pp. 1-12). IEEE.

[4]. **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.

[5]. **Alnajim, A., & Munro, M. (2009).** An anti-phishing approach that uses training intervention for phishing

websites detection. In *Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on* (pp. 405-410). IEEE.

[6]. Ardi, C., & Heidemann, J. (2016). Auntietuna: Personalized content-based phishing detection. In *NDSS Usable Security Workshop (USEC)*.

[7]. Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2010). New filtering approaches for phishing email. *Journal of Computer Security*, 18(1), 7-35.

[8]. Steve. (2014). *DKIM replay attacks Word to the Wise* [Blog Post]. Retrieved from https://wordtothewise.com/2014/05/dkim-replay-attacks/

[9]. Bravo-Lillo, C., Komanduri, S., Cranor, L. F., Reeder, R. W., Sleeper, M., Downs, J., & Schechter, S. (2013). Your attention please: designing security-decision UIs to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security* (p. 6). ACM.

[10]. Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors*, 58(8), 1158-1172.

[11]. Cao, Y., Han, W., & Le, Y. (2008). Anti-phishing based on automated individual white-list. In *Proceedings of the 4th ACM Workshop on Digital Identity Management* (pp. 51-60). ACM.

[12]. Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, 12(1), 28-38.

[13]. Cardinal , D. (2012). Diving into DMARC: Can it really end spam, or at least phishing? *ExtremeTech.* Retrieved from https://www.extremetech.com/or-at-least-phishing

[14]. Chen, K. T., Chen, J. Y., Huang, C. R., & Chen, C. S. (2009). Fighting phishing with discriminative keypoint features. *IEEE Internet Computing*, 13(3), 56-63.

[15]. Chen, T. C., Dick, S., & Miller, J. (2010). Detecting visually similar web pages: Application to phishing detection. *ACM Transactions on Internet Technology (TOIT)*, 10(2), 5:1–5:38.

[16]. Chou, N., Ledesma, R., Teraguchi, Y., & Mitchell, J. C. (2004). *ClientSide Defense Against Web-Based Identity Theft* (pp. 1-15). In NDSS.

[17]. Cimpanu, C. (2016). *Toy Maker Mattel Loses $3M in BEC Scam, Then Fights for it and Gets It Back.* Retrieved from https://news.softpedia.com/news/toy-maker-mattel-loses-3m-in-bec-scam-then-fights-for-it-and-gets-it-back-502401.shtml

[18]. Cobb, M . (2011). *The fight against phishing: Utilizing SPF and DKIM authentication technology.* Retrieved from http://searchsecurity.techtarget.com/answer/The-fight-against-phishing-Utilizing-SPF-and-DKIM-authentication-technology

[19]. Comparison of DNS blacklists. (2017). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Comparison_of_DNS_blacklists&amp;oldid=795659445

[20]. Conway, D., Taib, R., Harris, M., Yu, K., Berkovsky, S., & Chen, F. (2017). A qualitative investigation of bank employee experiences of information security and phishing. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)* (pp. 115-129). *USENIX Association.*

[21]. Crowe, J. (2016). *Phishing by the Numbers: Must-Know Phishing Statistics 2016* [Blog Post]. Retrieved from https://blog.barkly.com/phishing-statistics-2016

[22]. Cui, Q., Jourdan, G.V., Bochmann, G. V., Couturier, R., & Onut, I.V. (2017). Tracking phishing attacks over time. In *Proceedings of the 26th International Conference on World Wide Web WWW '17* (pp. 667-676). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

[23]. Denning, T., Lerner, A., Shostack, A., & Kohno, T. (2013). Control-Alt Hack: The design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security CCS '13* (pp. 915– 928). New York, NY, USA: ACM.

[24]. Dewan, P., & Kumaraguru, P. (2015). Detecting Malicious Content on Facebook. *arXiv preprint arXiv:1501.00802*.

[25]. Dewan, P., & Kumaraguru, P. (2017). Facebook Inspector (FbI): Towards automatic real-time detection of malicious content on *Facebook. Social Network Analysis* *and Mining*, 7(1), 15.

[26]. Dewan, P., Bagroy, S., & Kumaraguru, P. (2016). Hiding in plain sight: Characterizing and detecting malicious Facebook pages. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* 193-196.

[27]. Dewan, P., Kashyap,A.,& Kumaraguru, P. (2014). Analyzing social and stylometric features to identify spear phishing emails. In *2014 APWG Symposium on Electronic Crime Research (eCrime)* (pp. 1-13).

[28]. DNSBL. (2017). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=DNSBL&oldid=800548089

[29]. Duman, S., Kalkan-Cakmakci, K., Egele, M., Robertson, W., & Kirda, E. (2016). Email Profiler: Spearphishing filtering with header and stylometric features of emails. In *Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual* (Vol. 1, pp. 408-416). IEEE.

[30]. Durumeric, Z., Adrian, D., Mirian, A., Kasten, J., Bursztein, E., Lidzborski, N., & Halderman, J. A. (2015). Neither snow nor rain nor MITM...: An empirical analysis of email delivery security. In *Proceedings of the 2015 Internet Measurement Conference* (pp. 27-39). ACM.

[31]. Egelman, S., Cranor, L. F., & Hong, J. (2008). You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1065-1074). ACM.

[32]. Email Security or Anti-Phishing PhyllisTM | Wombat Security. (2017). Retrieved from https://www.wombatsecurity.com/training-modules/email-security-or-anti-phishing-phyllis

[33]. Equifax or Equiphish? — Krebs on Security. (2017). Retrieved from https://krebsonsecurity.com/2017/09/equifax-or-equiphish/

[34]. Felt, A. P., Reeder, R. W., Ainslie, A., Harris, H., Walker, M., Thompson, C.,... & Consolvo, S. (2016). Rethinking Connection Security Indicators. In *SOUPS* (pp. 1-14).

[35]. Ferguson, A. J. (2005). Fostering E-Mail Security Awareness: The West Point Carronade. *EDUCASE Quarterly*, 1. Retrieved March 22, 2006 from http://www.educause.edu/ ir/library/pdf/eqm0517.pdf

[36]. Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to detect phishing emails. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 649-656). ACM.

[37]. Fox-Brewster, T. (2016). *Who's Better at Phishing Twitter, Me Or Artificial Intelligence?* Retrieved from https://www.forbes.com/sites/thomasbrewster/2016/07/25/artificial-intelligence-phishing- twitter-bots/

[38]. Gaffney, G. (2011). *The Myth of the stupid user. Information & Design.* Retrieved from http://infodesign.com.au/usabilityresources/articles/themythofthestupiduser/

[39]. Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM workshop on Recurring malcode* (pp. 1-8). ACM.

[40]. Google Chrome Privacy Whitepaper. (2017). Retrieved from https://www.google.co.in/chrome/browser/privacy/whitepaper.html

[41]. Google Safe Browsing. (2017). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Google_Safe_Browsing

[42]. Gorling, S. (2006). The Myth of User Education. In *Proceedings of the 16th Virus Bulletin International Conference*.

[43]. Hara, M., Yamada, A., & Miyake, Y. (2009). Visual similarity-based phishing detection without victim site information. In *Computational Intelligence in Cyber Security, 2009. CICS'09. IEEE Symposium on* (pp. 30-36). IEEE.

[44]. He, M., Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., ...& Sutanto, A. (2011). An efficient phishing webpage detector. *Expert Systems with Applications*, 38(10), 12018-12027.

[45]. Herley, C. (2009). So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 Workshop on New* *security paradigms workshop* (pp. 133-144). ACM.

[46]. Ho, G., Sharma, A., Javed, M., Paxson, V., & Wagner, D. (2017). Detecting Credential Spearphishing in Enterprise Settings. In *26th Security Symposium* (pp. 469-485). USENIX Association.

[47]. Hong, J. (2012). The state of phishing attacks. *Communications of the ACM*, 55(1), 74-81.

[48]. Internet Crime Complaint Center (IC3) (2016). Business E-mail Compromise: The 3.1 Billion Dollar Scam. Retrieved from https://www.ic3.gov/media/2016/160614.aspx

[49]. Jackson, C., Simon, D. R., Tan, D. S., & Barth, A. (2007). An evaluation of extended validation and picture-in-picture phishing attacks. In *International Conference on Financial Cryptography and Data Security* (pp. 281-293). Springer, Berlin, Heidelberg.

[50]. Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, 50(10), 94-100.

[51]. Jain, A. K., & Gupta, B. B. (2017). Phishing detection: Analysis of visual similarity based approaches. *Security and Communication Networks*, 2017.

[52]. Kennedy, M. (2017). *After Massive Data Breach, Equifax Directed Customers to Fake Site*. Retrieved from http://www.npr.org/sections/thetwo-way/2017/09/21/552681357/after-massive-data-breach-equifax-directed- customers-to-fake-site

[53]. Khonji, M., Iraqi, Y., & Jones, A. (2011). Mitigation of spear phishing attacks: A content-based authorship identification framework. In *Internet Technology and Secured Transactions (ICITST), 2011 International Conference on* (pp. 416-421). IEEE.

[54]. Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. *IEEE Communications Surveys & Tutorials*, 15(4), 2091-2121.

[55]. Kumaraguru, P. (2009). *Phishguru: A system for educating users about semantic attacks*. Carnegie Mellon University.

[56]. 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* (p. 3). ACM.

[57]. 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* (pp. 905-914). ACM.

[58]. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10(2), 7.

[59]. Lastdrager, E., Gallardo, I. C., Hartel, P., & Junger, M. (2017). How Effective is Anti-Phishing Training for Children? In *Symposium on Usable Privacy and Security (SOUPS)*.

[60]. Lin, E., Greenberg, S., Trotter, E., Ma, D., & Aycock, J. (2011). Does domain highlighting help people identify phishing sites? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2075-2084). ACM.

[61]. Ludl, C., McAllister, S., Kirda, E., & Kruegel, C. (2007). On the effectiveness of techniques to detect phishing sites. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (pp. 20-39). Springer, Berlin, Heidelberg.

[62]. Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1245-1254). ACM.

[63]. Muncaster, P. (2017). *Social Media Phishing Attacks Soar 500%.* Retrieved from https://www.infosecurity-magazine.com/news/social-media-phishing-attacks-soar/

[64]. New York State Office of Cyber Security & Critical Infrastructure Coordination. Gone Phishing. A Briefing on the Anti-Phishing Exercise Initiative for New York State Government. Aggregate Exercise Results for public release.

[65]. Nicholson, J., Coventry, L., & Briggs, P. (2017). Can we fight social engineering attacks by social means?

Assessing social salience as a means to improve phish detection. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)* (pp. 285-298). USENIX Association.

[66]. Oliveira, D., Rocha, H., Yang, H., Ellis, D., Dommaraju, S., Muradoglu, M., ... & Ebner, N. (2017). Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6412-6424). ACM.

[67]. Phishing Activity Trends Report. Retrieved from http://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf

[68]. Phishing Scamsat All-TimeHigh, Employee Training NotKeeping Pace | Wombat Security. (2017). Retrieved from https://www.wombatsecurity.com/about/news/phishing-scams-all-time-high-employee-training-not-keeping-pace

[69]. Phishing threatens today's economy. (2017). In *NY Times*. Retrieved from https://cdn2.hubspot.net/hub/372792/file-1519503800-pdf/ Phishing Threatens Todays Economy NY Times FINAL.pdf

[70]. PhishTank > Frequently Asked Questions (FAQ). (2017). Retrieved from http://www.phishtank.com/faq.php#howisphishtankdiffer

[71]. PhishTank> Friends of PhishTank. (2017). Retrieved from https://www.phishtank.com/friends.php

[72]. Postmaster Tools – Google. (2017). Retrieved from https://gmail.com/ postmaster/

[73]. Prakash, P., Kumar, M., Kompella, R. R., & Gupta, M. (2010). Phishnet: Predictive blacklisting to detect phishing attacks. In *INFOCOM, 2010 Proceedings IEEE* (pp. 1-5). IEEE.

[74]. Ramesh, G., Krishnamurthi, I., & Kumar, K. S. S. (2014). An efficacious method for detecting phishing webpages through target domain identification. *Decision Support Systems*, 61, 12-22.

[75]. Robertson, A. (2017). *Google Docs users hit with sophisticated phishing attack.* Retrieved from https://www.theverge.com/2017/5/3/15534768/google-docs-phishing-attack-share-this-document-with-you-spam

[76]. Rosiello, A. P., Kirda, E., & Ferrandi, F. (2007). A layout-similarity-based approach for detecting phishing pages. In *Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on* (pp. 454-463). IEEE.

[77]. Sachs, D. (2013). *How to Take Down a Phishing Site: 5 Crucial Steps.* Retrieved from http://info.brandprotect.com/blog/blog/bid/88212/how-to- take-down-a-phishing-site-5-crucial-steps

[78]. Schechter, S. E., Dhamija, R., Ozment, A., & Fischer, I. (2007). The emperor's new security indicators. In *Security and Privacy, 2007. SP'07. IEEE Symposium on* (pp. 51-65). IEEE.

[79]. 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 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373-382). ACM.

[80]. 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* (pp. 88-99). ACM.

[81]. Sheng, S., Wardman, B., Warner, G., Cranor, L. F., Hong, J., & Zhang, C. (2009). An empirical analysis of phishing blacklists. In *CEAS 2009 - Sixth Conference on Email and Anti-Spam*.

[82]. Social engineering (security) page Version ID: 800193757. (2017). *In Wikipedia*. Retrieved from https://en.wikipedia. org/wiki/Webserver_directory_index

[83]. Stringhini, G., & Thonnard, O. (2015). That ain't you: Blocking spearphishing through behavioral modelling. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (pp. 78-97). Springer, Cham.

[84]. The Spamhaus Project Frequently Asked Questions (FAQ). (2017). Retrieved from https://www.spamhaus.org/

faq/section/Spamhaus20DBL#371

[85]. **Thunderbird's Scam Detection | Thunderbird Help (2017).** Retrieved from https://support.mozilla.org/en-US/kb/thunderbirds-scam-detection#wthunderbirds-automatic-scam-filtering

[86]. **Toolan, F., & Carthy, J. (2009).** Phishing detection using classifier ensembles. In *eCrime Researchers Summit, 2009. eCRIME'09* (pp. 1-9). IEEE.

[87]. **Vaas, L. (2016).** *How hackers broke into John Podesta, DNC Gmail accounts – Naked Security.* Retrieved from https://nakedsecurity.sophos.com/2016/10/25/how-hackers-broke-into-john-podesta-dnc-gmail-accounts/

[88]. **Vishwanath, A. (2014).** Habitual Facebook use and its impact on getting deceived on social media. *Journal of Computer-Mediated Communication*, 20(1), 83-98.

[89]. **Wang, J., Li, Y., & Rao, H. R. (2016).** Overconfidence in phishing email detection. *Journal of the Association for Information Systems*, 17(11), 759-783.

[90]. **Weinberg, N. (2013).** *How to blunt spear phishing attacks?.* Retrieved from https://www.networkworld.com/article/2164139/network-security/how-to-blunt-spear-phishing-attacks.html

[91]. **Welcome to APWG & CMU's Phishing Education Landing Page (2017).** Retrieved from http://phish-education.apwg.org/r/en/index.htm

[92]. **Wen, Z. A., Li, Y., Wade, R., Huang, J., & Wang, A. (2017).** What.Hack: Learn Phishing Email Defence the Fun Way. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 234-237). ACM.

[93]. **Wenyin, L., Liu, G., Qiu, B., & Quan, X. (2012).** Antiphishing through phishing target discovery. *IEEE Internet Computing*, 16(2), 52-61.

[94]. **Whittaker, C., Ryner, B., & Nazif, M. (2010).** Large-scale automatic classification of phishing pages. In *NDSS* (Vol. 10, p. 2010).

[95]. **Wu, M., Miller, R. C., & Garfinkel, S. L. (2006).** Do security toolbars actually prevent phishing attacks?, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 601-610). ACM.

[96]. **Zhang, H., Liu, G., Chow, T. W., & Liu, W. (2011).** Textual and visual content-based anti-phishing: A Bayesian approach. *IEEE Transactions on Neural Networks*, 22(10), 1532-1546.

[97]. **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* (pp. 639-648). ACM.

ABOUT THE AUTHORS

Aniket Bhadane is pursuing M.Tech in the Department of Computer Engineering at Government College of Engineering, Pune, Maharashtra, India. He received his Bachelor of Engineering Degree in Computer Engineering from Savitribai Phule Pune University (formerly University of Pune), India. His research interests are in the field of Cyber Security, User Authentication, and Usable Security.

Dr. Sunil B. Mane is currently working as an Associate Professor in the Department of Computer Engineering and Information Technology at Government College of Engineering Pune (An Autonomous Institute of Govt. of Maharashtra), India. He has more than 15 years of teaching experience. He has over 25 research publications in various National/International Journals and Conferences. He is a Board of Studies member in Computer Engineering/Information Technology of various autonomous engineering institutes. He has delivered lectures on information and cyber security domain as invited speaker. He is serving as Co-Chief Investigator for the Information Security Education and Awareness (ISEA) project, Ministry of Information Technology, Govt. of India. His areas of research are Data Privacy and Cyber Security.