Detecting and Analyzing Social Network Based Phishing Activities over the Internet to Enhance Cyber Security

Farhana J. Shaikh\(^1\), Harun Hawaldar\(^2\), Ankita Jangade\(^3\)
Infosys Limited, India\(^2\)
School of Computer Engineering & Technology, MITAOE, Pune, India\(^1,3\)

Abstract:
Phishing is an expeditiously growing threat in web world, damaging internet user’s privacy and stealing billions of dollars every year. It is an illegitimate activity which uses a group of technical exploits and social engineering to collect an internet user’s personal information. Different phishing tools and techniques can be identified in various methods of communication like fake websites, emails, instant messages and pop-up messages. Over the time span, a number of research papers have published with different anti-phishing procedures and techniques and have failed to detect all correlated risks and provide a common solution. This paper proposes an abstract standard - ASA to study this threat in a well-ordered manner. While there is a common perception that phishing is a security attack that involves designing an identical website in order to mislead the internet users into disclosing personal, financial credentials; however, this theory has not been practically used to analyze this threat as well as research gaps systematically. ASA outlook attempts to study this cyber Attack Survey different anti-phishing method and Analyze different research perspectives. This literature survey will play a vital role to generate awareness about phishing-based web frauds in order to induce preventive actions to enhance the web security and achieve user’s trust in web world.

Keywords: Anti-phishing, Phishing, Social Engineering, Web Security.

I. INTRODUCTION
Phishing is the illegitimate fraudulent activity that aims at gaining personal credentials like usernames, passwords, credit/debit card details, OTP by attackers who pretend as a trustful entity in an online communication. Communications professing to be from famous social networking sites (YouTube, Facebook, Twitter, Messenger), online shopping sites (Amazon, Flipcart), online banks, online payment professors, internet service providers, online retailers, insurance agencies are generally used to entice the victims. Phishing is commonly carried out by sending spoofed emails or instant messages and it often redirects users to enter personal details at a forged website whose “look and feel” is exact replica of legitimate one. Phishing is one of the social engineering techniques that aims at exploiting the poor functionality of present web security technologies. Phishing, a neologism generated as homonym of fishing as it is similar to use a bait in view to catch a victim. These are attempts to handle increasing number of phishing incidents which are reported and those includes legislation, user training, awareness and technical security standards. Over the past few years, research has shown that internet users are the major victims of phishing for several reasons. Such as, users incline to judge the legitimacy of website by its GUI and feel which can be easily replicated by attackers, many users do not pay attention to security warnings, although some people are aware of phishing; they fail to use appropriate strategies for identifying phishing attacks [1]. Phishing is an internet fraud which takes many forms, phone calls offering some sale on product, to false rumors that manipulate stock prices, to fake schemes that promise huge money if the victim in foreign financial transaction through his own back account. However, phishing has become more and more complexed and advanced so that phishers can deviate the filters set by current anti-phishing methods and aim their bait to both organizations and individuals [2]. This paper, produces comprehensive research where previously researched literatures are reviewed and presented. This paper aims to explore different phishing Attacks in section II, section III Survey on different anti-phishing approaches and section IV Analyses the different research perspectives. The conclusion is described in section V.

II. ATTACK
Cyber-attack is a security exploit which is getting worse every year. It has been increasing with the increase in information on web world. Cyber attackers are becoming bold by using digital equipment and internet services as well as creating ways to deceive users into revealing personal data and sensitive financial. Attackers “unusual behavior” is harder to detect because it might not seem unexpected at all. Phishing is the most usual type of attack followed by computer viruses and hacking in which attackers pretend to be known and legitimate entity in order to dupe users into disclosing personal, financial or computer account information.

A. Phishing Attack Structure
“Phishing attack is a computer attack which send socially engineered messages via electronic communication channel to humans in order to convince them to undertake certain actions for the attacker’s benefit” [3]. Nowadays, Phishing is a scam industry. There is a scam market for information which is obtained from the victims. Phishing tools and techniques are available for free (and/or purchase) in the vast information technology world [4]. Security has been a critical issue in the
field of information technology since early 1950’s. Over the past 20 years, the different types of web-based attacks are Malicious code, Trojan Horse, Advanced worms. 10 years later there comes Botnets, Spam sites, DNS attacks. Now, cyber-attacks include bitcoin wallet stealers, banking malware, PoS attacks, ransom ware, to name a few [5]. Phishing works because phishers usually use human psychology and behavior to design and then utilize a new phishing technique [6].

- **Classification of phishing attacks**

Phishing attacks are carried out through various ways which is not limited to emails but also instant messaging, SMS, social networking sites, voice messaging and even multiplayer games [7]. Gupta et al. [8] classifies phishing attacks on the basis of the mechanism by which phisher can be able to access personal information of a victim. Either a phisher uses a way to fraud the victim or uses any malicious code to access victim personal information. A phisher may deceive any innocent user either by using fraud emails or by using spoofed websites. Malicious code, key logger, and screen capture can also use to access personal information and technique is known as technical subterfuge phishing. Phishing attacks classification is shown in Fig 1.

![Phishing Attacks](image)

**Figure.1. Classification of Phishing attacks depend on how a phisher deceive victims (Gupta et al.)**

**I] Social Engineering**

Social Engineering is basically the art of fooling a person to acquire access to system or data by making use of human psychology, instead of using technical hacking techniques or breaking in. For example, a social engineer might call an employee and pose as an IT support person, instead of trying to find software vulnerability trying to trick the employee into disclose his password.

**A] Fake Websites**

These are also known as spoofed/phishing websites, they appear identical to legitimate site visually, these fake websites are used to get personal information about the victim. Generally, the links of the fake websites are inserted in fake emails, advertisements or within malware. Phisher takes lots of trials to make visual appearance and URI pattern similar to the victim site.

**B] Spoofed Emails**

Spoofed/Phishing emails usually sent by the untrusted mail server which are constructed to appear as if it’s from trusted entity. These emails deceive its victim, so that he/she may disclose his/her personal information.

- a) **Spear phishing**- It is an email-spoofing attack that aims a specific individual or organization trying to get unauthorized access to victim’s financial credentials. Spear-phishing attacks are not generally commenced by random hackers, they are most probably to be carried by perpetrators out for financial gain, military information or trade secrets.

- b) **Whaling**- It is a phishing attack that is specifically aimed at wealthy, prominent or powerful individuals. Because of their status, if such a user becomes the victim of a phishing attack, he can be considered a “big phish” alternately, a "whale."

**II] Technical Subterfuge**

This attack installs crimeware onto victim’s computers to steal credentials directly, one can obstruct consumers online account user names and passwords, and to deceive local and remote navigational infrastructures to misdirect consumers to fake websites and to authentic websites through phisher-controlled proxies that can be used to interrupt and monitor consumer’s keystrokes by using key logging systems.

**A]Cross-Site Scripting**

XSS deals with client-side code injection attack in which an attacker can perform malicious scripts into a legitimate web application or website. XSS is one of the most uncontrolled of web application vulnerabilities and occurs when a web application makes use of unencoded or unvalidated user input within the output it generates.

**B] Session Hijacking**

The Session Hijacking attack comprises of the exploitation of the web session control mechanism, which is managed for a session token. This attack compromises the session token by predicting or stealing a valid session token to obtain unauthorized access to the Web Server.

**C] Malware Phishing**

In malware-based phishing, fake software is used to store credentials in victim computer and send it to the owner the phisher.

**D] DNS Poisoning**

In this attack the phisher has a false DNS server and somehow convince 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**

Key loggers are difficult to detect and now with screen logging the virtual keyboards have no utility at all. These capture the screen snapshots and mouse movements which are sent to the phisher who is at a remote location.

**B. Phishing Attacks Statistics Reports**

According to APWG [9] report regarding phishing attacks during the period of 2005 till 2015 shown in Fig. 2
Moreover, according to APWG 1st Quarter 2018 report total number of phish detected was 263,538. This was up 46% from the 180,577 observed in 4Q 2017. It was also significantly more than the 190,942 seen in 3Q 2017. Phishers mostly uses domain names at the largest registrars and in the largest top-level domains. APWG [9] further add that by the 2nd quarter of 2018, more than a third of phishing attacks were hosted on Web sites that had HTTPS and SSL certificates.

### C. Phishing Life-Cycle

![Figure 3. The life-cycle of phishing campaigns from the perspective of anti-phishing techniques (Mahmoud et al.)](http://ijesc.org/)

Due to the broad nature of the phishing problem, it is essential to visualize the life-cycle of the phishing attacks and based on that classify anti-phishing solutions. Fig.3 depict a flowchart describing the life-cycle of phishing campaigns from the point of view of anti-phishing techniques, which is intended to be the most comprehensive phishing solutions flowchart. When a phishing campaign is started (e.g. by sending phishing emails to users), the first protection line is detecting it. The detection techniques are broad and could incorporate techniques used by service providers to detect the attacks, end-user client software classification, and user awareness programs. Whenever a phishing campaign is detected the ability to detect phishing campaigns can be enhanced by learning from such experience. For example, it is possible to enhance the detection of future phishing campaigns by learning from previous phishing campaigns. Such learning can be executed by a software or human observer (i.e. via a machine learning algorithm).

The number of actions could be applied against the campaign, after detecting phishing attacks, the following categories of approaches exist:

- **Offensive defense** — these approaches targets to attack phishing campaigns to make them less effective. This approach is specifically useful to protect users that have submitted their personal details to attackers.
- **Correction** — these approaches mostly focus on degrading the phishing campaign. In case of phishing websites, this is achieved by removing phishing files or suspending the hosting account.
- **Prevention** — phishing prevention methods are defined diversely in the literature depending on the context. In this survey, they have mentioned how to prevent attackers from starting phishing campaigns in the future.

### D. Why People Fall For Phishing?

Dhamija et al. [10] performed earliest studies examining why people are unsafe to phishing scams. In this investigation, participants were asked to discover various web-sites as legitimate or duplicate. They observed that good — look and feel phishing web-sites fooled 90% of participants and that most of the participants neglected browser cues. Many participants have conveniently neglected fraudulent certificate popup warnings and got easily influenced by the attractive user interface. Also, Downs et al. additionally study over investigating phishing emails makes it easier to understand that people mostly use wrong heuristics in deciding how to react to emails. For example, some participants rationalized that since the business already had their information, it would be no harm to give it again.

### E. Rise in Phishing Attacks

Cyber criminals are taking advantage of the expanding phishing areas it creates for new disruptive opportunities to attack. Despite of various phishing detection and prevention mechanisms available there are several reasons of why the phishing attack is still successful. The server-side anti-phishing mechanism uses backlisting method. In one of the previous studies, the author has stated that back listing suffers from several inherent drawbacks in terms of correctness, completeness, and timeliness which gives a favorable opportunity for attackers. On the other hand, the client-side phishing detection approaches mainly rely on page contents or URLs, which are potentially easier to manipulate by attackers [12]. In addition, a recent study conducted by Abbasi et al. [13] to examine the impact of anti-phishing tools, where 400 participants were involved with either a high accuracy tool (90%) or a low accuracy tool (60%) to make a decision about phishing or legitimate website has leaked the result that participants which are using high accuracy tools outtrace the others in terms of their ability to distinguish authorized websites from suspicious, avoid making transactions on phishing websites and avoid visiting phishing websites. But these users often ignored the correct recommendations of the tool, resulting in 15% lower phishing detection rates compared with anti-phishing tool used. Even though the accuracy of the tool is critical, the authors suggest that decreasing the success rates of phishing attacks is needed and also require improving designing of warning tool and user’s awareness to eliminate this threat and secure the cyber world.
III. SURVEY ON ANTI-PHISHING APPROACHES

Survey on Anti-phishing Approaches Researchers be it from industry or academia are working very hard to develop anti-phishing techniques to fight against phishing threats. However, no one antiphishing solution is a bullet in silver against phishing attack. With time, phishing threat is increasing rapidly [8]. This section presents a comprehensive detail about different research uses different techniques to achieve the maximum accuracy result and improve the whole system.

- **Educate users about phishing attacks**
  One of the effective solutions for phishing is educate users and describe the thing that not to follow links blindly to any web sites where they have to submit confidential information like passwords [14] [15]. In one of the studies, the author has found that educating users is useful and help people to identify the fake website provided the users read the phishing material seriously and understand this crime [16]. However, it also been noted that delivering anti-phishing material and educating users are sometime not effective since users are more familiarized of receiving such cautions and thought they already knew how to protect themselves [17].

- **Layout-Similarity-Based approach to detect phishing websites**
  Rosiello et al. [18] working on phishing detection techniques for quite some time. Their previous phishing prevention system prevent sensitive information from entering the phishing sites. However, the technique was successful only if the user cooperated as well and rarely gave false responses. In this research, the authors proposed an improved form of their previous model called DOM AntiPhish. To reduce the false alarms, they have reduced the user involvement in this model. Their approach uses layout similarity information to differentiate between legitimate and malicious web pages. After evaluation, the research has demonstrated the solution is practical. This research has given a solid detection device for the users which is browser plug-in based. It is specifically designed to protect users who does not tech savvy from phishing attacks and who are inexperienced.

- **Hybrid Features for detecting phishing email**
  Hybrid Features for detecting phishing email Lew et al. [19] presented a comprehensive research on developing tools and techniques to recognize phishing emails via hybrid features like content based, URL based, behavior-based features. Phishing has become so harmful, complicated and sophisticated that it is master in avoiding filters and anti-phishing systems. Email servers now can be established with malicious detection tools since phishing emails have explored a lot of researchers to work on creating these techniques, but these efforts have not proved to be worthy enough to stop phishing emails. The approach adopted basically includes taking out feature vectors from the emails that well denote the instances. To create the phishing detection model following four elements have been used: Feature Generator, Method Selection, Machine Learning, Feature and Inductor Evaluation. A feature will be more valuable when the induction generates large amount of information. These elements have demonstrated an effectiveness of the phishing email detection and provide evidence.

- **Neural Network for detecting phishing attacks**
  Zhang and Yuan [20] discovered yet another phishing detecting approach which uses neural network with machine learning algorithm. Users do not recognize the spoofed emails as the emails seem to be coming from legitimate entity. The detection model makes use of large number of emails to detect phishing attacks, it consolidates multilayer Feed forward Neutral Networks (NNs) by defining and selecting a set of features relating to the email structure and external links. This model uses one input layer, one hidden layer, and one output layer. The methodology first extracts feature at the pre-processing stage then implements Neural Network to classify them. The result shows that Neural Network can reach a 95% accuracy level with a little misclassification.

- **Neuro-Fuzzy Logic Approach for phishing**
  Fehringer et al. [21] proposed an innovative approach intelligent phishing security (IPS) take the form of a zero order Sugeno type. This consists of four main components that include: features sources, features, adaptive neuro-fuzzy inference system. ANFIS composed of a rule base and a feature base that makes knowledgebase and a decision-making unit. ANFIS is used because it can learn and validate given features using Fuzzy IF...THEN rules. Fuzzy modelling and Gbell shape membership function used for its efficiency. A two-fold cross-validation method is used which randomly split the features into two parts, training set and testing set. First, training is carried out on a training-set only once and then testing is done on a test-set only once. After that the roles of training and testing sets are reversed. At last results are assembled to get the average errors. This approach boosts internet user’s confidence.

- **Feature Selection and Classification Approach**
  Isredza et al. [22] have studied many different types of content-based filters approaches. The research discovered that many researchers content-based email classifications focused on more sophisticated machine learning algorithms. The author presented an exclusive approach of Classification Model with the method of extracting the content-based features of phishing email by analyzing the message- ID tag and sender email with view to extract the attacker’s behavior. Based on a weighting of message header and message content and selecting the feature according to the priority ranking. The author discovered that phishing email that has ability to come from more than one domain that could indicate abnormal activity. This is done by understanding email header which is generally neglected by others. The experiment approach produces a promising result using 8 features with 94% accuracy.

- **Data Mining Approach**
  Neda Mining Approach
  Isredza et al. [23] compared two phishing detection approaches Associative Classification (AC) based on data mining and Multi-label Classifier based Associative Classification (MCAC). Experimental results using real-world data collected from various kinds of sources shows that the AC particularly MCAC recognizes phishing websites with higher accuracy as compared to other intelligent algorithms. MCAC further generates new hidden rules that other algorithms are unable to explore. This has improved the performance of its classifiers predictive. According to research conducted by author AC usually mines classifiers containing simple “If-Then” rules which has a high degree of
predictive accuracy. However, MCAC is able to mine the rules which represents correlations among the website’s features. MCAC was able to generate multi-label rules from the phishing data generating rules that are associated with a new class called “Suspicious” that was not initially in the training data set.

- **Fuzzy Data Mining Approach**

In 2010 Aburrous et al. presented a distinctive method to deal with phishing website by using Fuzzy logic combination and Data mining to secure Internet users while performing online transactions. The approach practiced to assess e-banking phishing website risk on 27 features to build a model to predict phishing websites based on fuzzy data mining. Working of this model was based on multi-layered approach in which each layer has its own rule set to define the websites in five different ways i.e. suspicious, legitimate, very legitimate, phishy or very phisy. Although the method has obtained 83.7% but, this research presents an effective method using fuzzy data mining algorithms and tools to detect e-banking phishing websites in an automated manner.

- **Computer Vision Technique**

Rao and Ali [25] propose a solution for phishing attacks by presenting a mixture of whitelist and visual similarity-based techniques. SURF detector: a computer vision technique extracts an important set of features from targeted and suspicious websites. Firstly, all legitimate URLs are saved in a whitelist and then they are compared with the URL accessed. If the URLs are same, the website is considered innocent, but if different then present in the white list and then further SURF detection comes into picture. A similarity degree is then calculated to compare between the legitimate and suspicious pages. If similarity rate is greater the web-page is considered suspicious and vice versa. This is an essential method of detecting phishing and gives a low false positive rate.

- **Anti-Phishing Tools**

Dhamija and Tygar [26] presents three important anti-phishing tools. They have given a strong impact on the aspect of controlling phishing attacks. One of these important anti-phishing tools is eBay toolbar which is essential “a browser plug-in that eBay offers to its customers to help keep track of auction sites”. Account Guard is an integral feature of this toolbar, which monitors web pages which are visited by users, keeping a track of these pages, and providing a warning in the form of a coloured tab on the toolbar. The tab is usually grey, but it turns green if the user is on a eBay or PayPal site, turns red if the user is on a site that has been granted a spoof by eBay and it thereby facilitates the users to immediately identify and detect spoofs. Spoof Guard is another important tool, which is an Internet Explorer browser’s plug-in. This plug-in fundamentally examines and inspects the web pages and generates a warning for the user. This happens when certain page is expected to have a high probability of being a spoof. “This calculation is performed by evaluating the images, domain name, links and comparing them to the stored history and by detecting common characteristics of spoofed websites” [25]. Finally, SpoofStick, another useful anti-phishing tool is an extension of the toolbar extension for the Mozilla Firefox or Internet Explorer browsers. It is actually, providing basic information about the domain name of the website being visited. For instance, when a user is visiting eBay, the toolbar will display a message saying “You’re on ebay.com”, otherwise if the user is visiting a spoofed website, the message will display something like “You’re on 10.19.32.4”. Hence, the user will immediately be able to detect possibilities of a phishing attack and take appropriate actions before any issues arise. “This toolbar can help internet users to detect phishing attacks where the fake website has a domain name that semantically or syntactically similar to a legitimate site” [26]. Cranor et al. [27] also suggest that the CloudMark is a host of important anti-phishing tools. It is an Anti-Fraud Toolbar which depends on user ratings for given websites. The toolbar assigns colour to icons as per user feedback and displays a coloured icon for each visited site. For example, sites rated as legitimate will display a green colour icon, for sites rated as fraudulent the icon turns red colour, while yellow colour icons necessarily indicate the lack of enough information to establish the credibility of the site. Similarly, EarthLink Toolbar uses combination of heuristics, user ratings, and manual verification procedures to find authenticity of the website. “Users can be able to report suspected phishing sites to EarthLink; these sites are then verified and added to a blacklist” [27]. The verified legitimate sites are indicated by green thumbs up on the EarthLink Toolbar; thumb up with greycolour indicates that the site is not suspicious but not verified either. Similarly, a yellow colour thumb down will indicate that the site is quite questionable while fraudulent websites are indicated by a red colour thumb down.

IV. ANALYSIS OF RESEARCH GAP

We understand that the phishing is a specialized social engineering attack where the attackers are very intelligent who uses spoofed emails or fake websites to deceive the victims into sharing their confidential and sensitive information, or by installing malware into the victim’s computer. Phishing attacks clearly bypass the security measures adopted by both individuals as well as business organizations at large. However, there are many issues that deal with the gap between academic literature and practical evidence. Firstly, online consumers do understand the importance of security in web world; they tend to classify other factors such as ease of use and other aspects of security concerns. Therefore, it will be very useful if further research is conducted on understanding the psychology of online consumers as to whether they do understand the significance of security warnings popping up in web browsers when they are focused on a stressful deadline. Also, the study must be directed at examining whether the user keep browser’s security features as enabled when they are given the choice to disable them. Hence, there is noticeable research gap in terms of understanding the online consumer psychology towards cyberspace security issues [28]. A further and thorough research is required in terms of examining the effectiveness of counter-measures for phishing. Large number of literatures are available on phishing, proposing diverse content whereby researchers have attempted to understand the psychology of the process behind phishing attacks and how to simulate such attacks through efforts such as developing sensible and strong servers that can automatically detect fraudulent sites and blocking spam email from reaching the end-user. However, while research on anti-phishing methodologies has led to outstanding deployments on the internet, presenting different approaches, they have all been of preventive nature. The Phisher’s continues to go ahead by
attacking the delicate link in the security chain and launch successful attacks at them. A very well said statement by previous literature [29] “Various usability studies have showed that neither server-side security indicators nor client-side toolbars and warnings are successful in preventing vulnerable users from being deceived”. Hence, even though the large-scale positive impact of modern anti-phishing and anti-spoofing efforts, still more than 11% users fall for phishing attacks by clicking on the fake link or by reading spoofed messages. All what is required by is consumer education in terms of cyber world behaviour as well as the implementation of the appropriate anti-phishing techniques and measures so that consumers can themselves minimize their risk of falling victim to phishing attacks. There is a need for significant work on this aspect. Also, a previous research has studied in the effort to combat online attacks, are hindered by the rapidly changing technology behind the digital communications. One positive factor is the commonality of using text messages in online communications. However, the way in which the text is written and transmitted are different. For example, an instant message traffic that must essentially be generated immediately and communicated quickly to be effective; while weblog users have all the time available to formulate entries. Hence, there are inherent characteristics of malicious texts in online data communications text. More experimentation is required on this aspect. Similarly, the study carried out by Barracough et al. [30] using novel approach based on the Neuro-Fuzzy system require a further study. This inventive technique of detecting phishing websites is a one-of-its-kind mechanism offering an effective solution to phishing by protecting online users while executing online transactions. However, there is a research gap in terms of embedding additional features to the scheme and also in terms of parameter optimization for achieving a 100% accuracy levels to develop a plug-in toolbar for real-time application. The authors of this paper state that there is no one solution to phishing at any one level. The stakeholders at every level of the phishing lifecycle must take greater efforts to detect such attacks and launch inventive new practices to prevent the danger. One aspect where research gap exists which deals with stakeholder’s working together with each other to identify potential phishing attacks and to build new cross-level solutions which involves all key stakeholders [31]. Further, as per the suggestions of other researches, this study also brings a close which refers that anti-phishing technique of blacklists fails to generalize to new phishing attacks which might come in future and suggest of further research gap in terms of finding practical solutions.

V. CONCLUSION

This paper has presented three major components of the study, a description of phishing attack, a survey of the anti-phishing approaches offered by different research and analysis of the research gaps. Phishing will never be completely vanished but it is important to understand this attack before proposing any solution. The ASA outlook offers guidance for understanding each of its parameters and relationship between these parameters to generate phishing literature more suitable and valuable to its reader to understand the phishing attack, survey on anti-phishing approaches and analysis on the research gaps in much improved way. This research aims to develop a practice to understand the growing threats of phishing and discover a new literature. The future work will be to get the research into the development of phishing detection system particularly against phishing websites since it is considered the most typical way of attack.

VI. REFERENCES


