A Review of Various Techniques for Detection and Prevention for Phishing Attack

Nirmala Suryavanshi, Department of Computer Science and Engineering
Radharaman Institute of Technology and Science, Madhya Pradesh, Bhopal, India
Anurag Jain, Department of Computer Science and Engineering
Radharaman Institute of Technology and Science, Madhya Pradesh, Bhopal, India

Abstract—Use of computer network or internet for the transmission of data is growing rapidly. But more use of internet severe kind of attack may steal our personal information or any kind of information which flows through it, due to which the security can break. Among various kind of attack, one of attack is phishing. It is a kind of network attack which theft the identity of user’s online and steals some useful information such as password or ATM and financial information. The phishing is classified into two categories deceptive phishing and malware-based phishing. Various anti-phishing techniques have been developed and so many algorithms have also been proposed by various researchers to thwart the network from such serious attacks. In this, literature study of some of the previously work done to prevent network from phishing attack is described with their merits and demerits.

Keywords—Computer network, Deceptive, Malware, Phishing

I. INTRODUCTION

With due to rapid increase in the use of internet technology for communication different kind of attacks can be possible on the network such as DOS (denial of service attack), masquerade, replay and phishing etc. It is one of the most serious attacks which steals our personal information or hack the website. The word ‘Phishing’ originally emerged in 1990s. The early hackers frequently use ‘ph’ to reinstate ‘f’ to fabricate new words in the hacker’s community, as they typically hack by phones. Phishing is a novel word produced from ‘fishing’, it refers to the act that the attacker fascinates users to visit a counterfeits website by sending them counterfeit e-mails (or instant messages), and stealthily get victim's personal information such as user name, password, and national security ID, etc. This information then can be used for future target advertisements or even identity theft attacks (e.g., transfer money from victims’ bank account). The recurrently used attack process is to send e-mails to prospective victims, which appeared to be sent by banks, online organizations, or ISPs. In these e-mails, they will make up several reasons, e.g., the password of your credit card had been mis-entered for several times, or they are offering upgrading services, to allure you visit their Website to conform or amend your account Number and password through the hyperlink made available in the e-mail. We will then be linked to a counterfeited Website after clicking those links. The style, the functions performed, sometimes even the URL of these faked Websites is similar to the real Web site. It's very difficult for you to know that you are actually visiting a malicious site. If you input the account number and password, the attackers then successfully collect the information at the server side, and is able to perform their next step actions with that information (e.g., withdraw money out from your account). Phishing itself is not a new concept, but it's increasingly used by phishers to steal user information and perform business crime in recent years. Within one to two years, the number of phishing attacks increased dramatically. According to Gartner Inc., for the 12 months ending April 2004, "there were 1.8 million phishing attack victims, and the fraud incurred by phishing victims totaled $1.2 billion" [1]. APWG provides a solution directory at (Anti-Phishing Working Group) [2] which contains most of the major anti-phishing companies in the world. However, an automatic anti-phishing method is seldom reported. The typical technologies of anti-phishing from the User Interface aspect are done by [3] and [4]. The process of phishing the website is shown through fig. 1.

II. RELATED WORK

The rest of the paper is organized as follows: In section II literature of various approaches proposed by different author is described. In section III gives classification of phishing. In section IV explaining various anti-phishing technique with their merits and demerits. Section V presents conclusion of whole paper with future development of method for preventing the link or website from phishing attack.

Various approaches have been proposed for preventing the website or link from phishing attack. In this section, describes literature study of earlier work done for detecting the phishing attack.
In [5] proposed a novel system to identify phishing attacks and to ascertain the entity/organization that the attackers impersonate all through phishing attacks. The anticipated multi-stage process makes use of natural language processing and machine learning. The process primary discovers (i) named entities, which embraces names of people, societies, and locations, and (ii) concealed topics, using (a) Conditional Random Field (CRF) and (b) Latent Dirichlet Allocation (LDA) operating on together phishing and non-phishing data. Using matters and named entities as attributes, those resulting phrase categorizes every message similarly as phishing or non-phishing using Adaboost. For communication classified as phishing, the ultimate stage determines the impersonated entity using CRF. Experimental results showed that the phishing classifier identifies phishing attacks with no misclassification when the amount of phishing emails is less than 20%. The F-measure achieved was 100%. Their methodology likewise uncovered those impersonated entity starting with message that would arranged similarly as phishing, with a detection rate of 88.1%. The involuntary detection of impersonated entity from phishing helps the justifiable organization to take down the offending phishing site. This shelters their users from falling for phishing attacks, which in turn escorts to satisfied customers. Involuntary detection of an impersonated entity also helps email service suppliers to collaborate with each other to substitute attack information and defend their customers.

In [6] presented a novel approach to surmount the ‘fuzziness’ in the e-banking phishing website evaluation and proposed an intelligent resilient and successful model for detecting e-banking phishing websites/ links. The proposed system is based on fuzzy logic mingled with data mining algorithms to exemplify the e-banking phishing website factors and to scrutinize its procedures by classifying the phishing categories and defining six e-banking phishing website attack criterion’s with a layer structure. Their experimental consequences showed the implication and significance of the e-banking phishing website criterion (URL & Domain uniqueness) represented by layer one and those different impact of those phishing characteristic on the final e-banking phishing website rate.

In [7] introduced new contributions (Justifiable site rules, User-behavior outline, PhishTank, User-specific sites, Pop-Ups from emails) which were not reflected on beforehand in a single protection platform. The suggestion was to exploit a Neuro-Fuzzy method with 5 inputs to distinguishing phishing sites with high precision in real-time. In this, 2-Fold cross-validation is practical for training and testing the anticipated model. A sum of 288 features with 5 inputs was used and has so far achieved the preeminent performance as compared to all formerly reported outcomes in the field.

In [8] projected a genetic algorithm which is used to develop rules that are used to discriminate phishing link from legitimate link. Evaluating the constraints like estimation function, crossover and mutation etc. The GA spawns a rule-set that counterparts simply the phishing links. This rule-set is stored in a database and a link is reported as a phishing link if it counterparts every of the rules in the rule based system and therefore it keeps protected from forged hackers. Beginning experiments showed that this approach is efficient to perceive phishing hyperlink with minimal false negatives at a speed sufficient for online application.

In [9] proposed a novel approach to surmount the complexity and intricacy in detecting and predicting counterfeit website. There is a proficient model which is based on using association and categorization Data Mining algorithms optimizing with Particle Swarm Optimization algorithm. These algorithms were used to distinguish and recognize all the rules and factor in order to categorize the phishing website and the association that correlate them with each other. It also utilized MCAR classification algorithm to take out the phishing training data sets criterion to categorize their authenticity. After classification, those results have been optimized with Ant Colony Optimization (ACO) algorithm. But, this work has limitations like Sequences of random decisions (not independent) and Time to convergence uncertain in the phishing classification. So to overcome this limitation we enhance Particle Swarm Optimization (PSO) which finds a solution to an optimization problem in a search space, or model and predict social behavior in the presence of phishing websites. This will improve the correctly classified phishing websites. The experimental outcomes demonstrated the practicability of using PSO system in genuine applications and its improved performance. This project utilizes the JAVA technology.

In [10] discussed a Knowledge Base Compound scheme which is based on inquiry operations and parsing methods to counter these internet attacks by means of the web browser itself. In this system, they projected to scrutinize the web URLs prior to visit the authentic site, therefore, while to offer security adjacent to web attacks revealed above. This method makes use of different parsing operations and query processing which used various methods to distinguish the phishing attacks as well as other web attacks. Therefore mentioned method is absolutely based on operation through the browser and therefore merely influences the speed of browsing. This method also embraces crawling operation to perceive the URL details to supplementary augment the precision of discovery of a compromised site. By means of the proposed methodology, a novel browser can simply perceives the phishing attacks, SSL attacks, and other hacking attacks. By means of the use of this browser method, they could merely achieved 96.94% security next to phishing as well as other web based attacks.

In [11] proposed a phishing detection approach—PhishZoo—that uses profiles of trusted websites appearances to detect phishing. Our approach offers similar accuracy to blacklisting approaches (96%), with the advantage that it can categorize zero-day phishing attacks and targeted attacks against smaller websites (such as corporate intranets). Significant contribution of this paper is that it comprises a performance analysis and a framework for making use of computer vision techniques in a practical way.

In [12] proposed is a novel outline called phishing dynamic evolving neural fuzzy framework (PDENF), which acclimates the evolving connectionist system (ECoS) based on a crossbreed (supervised/ unsupervised) learning technique. PDENF adaptive online is enhanced by offline learning to distinguish vigorously the phishing email included unknown zero-day phishing e-mails prior to it get the user account. PDENF is recommended to work for high-speed “life-long” learning with small memory footprint and minimizes complication of the rule base and configuration with not many number of rules creation for email.
classification. They suppose to accomplish high performance, including elevated level of true positive, true negative, sensitivity, exactness, F-measure and complete accuracy compared with other techniques.

In [13] proposed using a trusted mechanism to execute mutual authentication that abolishes reliance on wonderful user behavior, towards Man-in-the-Middle attacks subsequent to setup, and defends a user’s account even in the existence of key loggers and most forms of spyware. They demonstrated the practicality of our system with a prototype implementation.

In [14] illustrated a novel framework to diminish spear phishing attacks by the use of document authorship methods — anti-spear phishing content-based authorship recognition (ASCAI). ASCAI enlightens the user of probable mismatches among the writing styles of body of a received email and of trusted authors by reading the email body itself (i.e. the write print), because opposed to conventional user ID-based authentication techniques which can be spoofed or abused. As a proof of concept, they implemented the proposed framework by source code author profiles (SCAP), and the assessment consequences are presented. ASCAI aims at augmenting security usability by defending trusted author’s identities from being announced by other senders through typo-squatting, cousin-naming or identity theft attacks, which are common problems with spear phishing attacks. The approach that ASCAI follows to protect trusted authors is by studying the email body itself, as opposed to conventional user ID-based methods which have many weaknesses.

III. CLASSIFICATION OF PHISHING ATTACK

The phishing attack are classified into various categories as per there way of stealing information, various researchers call them by different names such as deceptive, malware based, content-injection phishing etc [15, 16].

A) Deceptive Phishing

In this technique the phished webpage will ask the user to enter details to verify account information, fictitious account charges, undesirable account changes, system malfunction requiring users to re-enter their information, fresh free services requiring rapid action, and numerous other exciting offers so as to extend interest in users mind with the expect that the fatality will click on the link as will contribute the confidential personal information to the counterfeit webpage which can be additional used to execute scams.

B) Web Spoofing

Web Spoofing is a security attack that consents to an antagonist to scrutinize and transform all web pages sent to the user machine, and scrutinize all information entered into forms by the user. Web Spoofing works on both of the chief browsers and is not prohibited by secure correlation. The attacker can scrutinize and transform all web pages and form capitulations, even when the browser's "secure connection" indicator is designated. The user sees no suggestion that anything is erroneous. Once this information is collected, the attacker can use it to purchase things with the victims' credit cards, access their bank accounts, and establish false identities. Website spoofing is a growing phenomenon, and puts consumers at considerable risk for individuality theft and credit card deception. The attack is instigated when the casualty visits a malevolent web page, or accepts a malevolent email message (if the fatality uses an HTML enabled email reader).

C) E-mail spoofing

Email spoofing is email commotion in which the sender address and further parts of the email header are distorted to appear as though the email originated from a different source. Since core SMTP doesn't provide any authentication, it is uncomplicated to pretend and false emails. Distributors of spam frequently use spoofing in an endeavor to get recipients to open and possibly even respond to their solicitations. Spoofing are been used legitimately. Classic e.g. of senders who might desire to masquerade the source of the e-mail comprise a sender reporting mistreatment by a spouse to a welfare society or a “whistle-blower” who fears retribution.

D) Malware Based Phishing

This technique involves making run a malicious code on user’s machine which is capable of performing tasks which will provide details of the confidential data entered by the user. Malware can be introduced in the user’s machine as an attachment, by exploiting security vulnerabilities, as a downloadable file from a web site.

E) Tabnabbing

This is one of the more recent types of phishing that takes benefit of people who have manifold tabs open at any one time. Phishers misuse this propensity to repossess information of their popular websites through cookies. The hacker then plays with small favicons and creates a looks like page of the original website, asking for login credentials compromising their accounts.

F) Session Hijacking

Session Hijacking is a kind of phishing attack where user's activities are scrutinized noticeably until they log into an object account like the bank account and ascertain their credentials. At that point, the malevolent software takes control and can commence unauthorized actions, such as transferring funds, exclusive of the knowledge of the user.

G) Man-in-the-Middle Phishing

In these attacks phisher positions themselves flanked by the user and the justifiable website or system. They document the information being entered but prolong to pass it on so that user’s transactions are not exaggerated. Afterward, they can sell or employs the information or credentials collected when the user is not active on the system.

H) Search Engine Phishing

This happens when phishers generate websites with attractive (often too attractive) sounding suggests and have they indexed justifiably with search engines. Users discover the websites in the ordinary course of searching for goods or services and are deceived into giving up their information. For example, scammers have set up false banking sites offering subordinate credit costs or improved interest rates than other banks. Fatality that employ these websites to save or make more from interest charges are encouraged to reassign current accounts and tricked into giving up their particulars.
I) DNS-based Phishing
Domain Name System (DNS)-based phishing or hosts file amendment is known as pharming. The requests for Uniform Resource Locators or name service return a counterfeit address and succeeding communications are directed to a counterfeit site when the hackers interfere a company’s host files or domain name. As a result, users remain unaware about the deceit website proscribed by hackers.

IV. ANTI PHISHING TECHNIQUES
Various anti-phishing techniques have been evolved to protect our website/ link and personal information against phishing attacks.

A) List Based Approach
This is possibly the most straightforward solution for anti-phishing. A white list contains URL’s of known legitimate sites. Many current anti-phishing techniques rely on the combination of white list and blacklist. The representative blacklist/white list based systems include Phish Tank Site Checker, Google Safe Browsing, Fire Phish and Calling ID Link Advisor. This anti-phishing result would generally deploy similarly as toolbars or extension of web browsers should remind those clients if they would scan a sheltered websites. Blacklist undergo from a window of vulnerability between the time a phishing site is launched and the site’s addition to the blacklist as it requires frequent updating which is the case for white list also.

B) Ant Colony Optimization
The Ant Colony System or the basic idea of an ant food searching system is illustrated in Fig. 2. In the left picture, the ants shift in a straight row to the food. The subsequent picture illustrates the circumstances rapidly after an obstacle is inserted among the nest and the food. To evade the obstacle, initial each ant selected to turn right or left at random. Let us presuppose that ants shift at the identical speed depositing pheromone in the trail equivalently. Though, the ants that, by possibility, prefer to turn right will reach the food sooner, although the ants that go around the obstruction turning right will pursue a longer path, and hence will take long time to circumvent the impediment. As a consequence, pheromone gathered quicker in the shorter path around the impediment. Ever since ants desire to pursue tracks with better amounts of pheromone, eventually all the ants congregate to the shorter path around the impediment [19].

This novel heuristic known as Ant Colony Optimization (ACO) has been found to be mutually vigorous and multipurpose in handling an extensive range of combinatorial optimization problems. The major suggestion of ACO is to model a predicament as the search for a least cost path in a graph. Artificial ants as if walk on this graph, gazing for cheaper paths. Each ant has a somewhat uncomplicated behavior accomplished of finding comparatively costlier paths. Cheaper pathways are found as the growing consequence of the universal cooperation among ants in the colony. The behavior of artificial ants is stimulated from real ants: they put down pheromone trails (noticeably in a mathematical outline) on the graph edges and prefer their path with reverence to probabilities that depend on pheromone tracks. These pheromone tracks progressively abridged by evaporation. In addition, artificial ants have a few superficial attributes not seen in their counterpart in real ants. In meticulous, they persist in a discrete world (a graph) and their progresses consist of conversions from nodes to nodes. The ACO fluctuates from the conventional ant system in the intellect that here the pheromone tracks are updated in two ways. Initially, while ants build an excursion they nearby transform the quantity of pheromone on the visited edges by a narrow updating role. Subsequently, after all the ants have fabricates their personage tours, a global updating rule is applied to transform the pheromone level on the boundaries that belong to the preeminent ant tour found so far [20].

C) PhishZoo
It can detect current phishing sites if they look like legitimate sites by matching their content against a saved profile. In order to avoid detection, a phishing site must gaze fundamentally unique in relation to a genuine website. Our working assumption is that such different-looking sites have a better chance of catching users’ attention about their phishiness. Branding is an issue that is well-studied in the marketing literature, and, with PhishZoo, it can be used to improve security as opposed to the current case, when this branding is co-opted by attackers to mis use client trust [11].

D) K-NearestNeighbor (k-NN)
This Classifier proposed for phishing email filtering. Using this classifier, the decision is made as follows: based on k-nearest training input, samples are chosen using a pre-defined similarity function; after that, the email x is labeled as belonging to the same class as the bulk among this set of k [20].

E) Information-flow-based approaches
PwdHash is a well-known anti-phishing solution in literature [21]. It generates domain-specific passwords that are rendered unusable if they are submitted to another domain (e.g., a password for www.hotmail.com will be different if submitted to www.phisher.com). In comparison, Antiphish takes an alternate methodology and stay with track about the place sensitive data is, no doubt submitted [22]. That is, if it detects that confidential information such as a password is being entered into a form on a fake web site, a warning is generated and the pending operation is canceled. The main disadvantage of AntiPhish is that it requires user interaction to specify which sensitive information should be captured and monitored. Later, the author significantly improves the original idea of AntiPhish by eliminating the necessary user interaction with an extra comparison step that analyzes the DOM structure of the pages [23]. They present an extension of AntiPhish, called DOMAntiPhish, which leverages design similitude majority of the data should recognize between pemicious furthermore favorable pages.
F) Attribute Based Anti-Phishing Techniques

Attribute-based anti-phishing strategy uses both reactive and proactive anti-phishing. This technique has been implemented in Phish Bouncer [8] tool. The Image Attribution technique does a comparison of images of accessing site and the sites already being registered with phish bouncer. The HTML Crosslink checks and looks at the responses coming from nonregistered sites and counts the number of links the page has to any of the registered sites. A high number of cross-links indicate that it is a phishing site. In false info feeder checker, false information is provided and if that information is accepted by the site then probably that link is phished one. It checks for suspicious certificates and validates site certificates presented during SSL handshake and extends the typical Usage by looking for Certification Authority (CA). As multiple checks are performed to authenticate the site this results in slow response time [24].

F) Fuzzy Logic

Fuzzy logic has been exercised for decades in the engineering sciences to entrench specialist input into computer models for a wide range of applications. It suggests a promising unusual for measuring operational risks [25]. The fuzzy logic techniques presents more information to help risk managers successfully manage assessing and ranking website phishing risks than the existing qualitative approaches as the risks are quantified based on a amalgamation of historical data and practiced input. The benefit of the fuzzy system is that it enables processing of indistinctly defined variables, and variables whose relationships cannot be defined by mathematical relationships. Fuzzy logic can integrate expert human judgment to describe those variable and their relationships.

G) Genetic Algorithm

Genetic algorithms can be used to develop simple rules for preventing phishing attacks. These rules are used to differentiate normal website from anomalous website. These anomalous websites refer to events with probability of phishing attacks [8]. The rules saved in the rule base are usually in the following form:

```
if { condition }
then
 { act }
```

For the problems we presented above, the condition generally refers to a match between the URL of the current site link in the e-mail and the rules in PADPS (Phishing Attack Detection and Prevention System), which indicates the probability of phishing attack. The act field usually refers to an action defined by the security policy such as reporting an alert to the browser, through the status field. For example, a rule can be defined as:

```
if
 | The IP address of the URL in the received e-mail
 | finds any match in the Ruleset
 | then
 | }
```

This rule can be explained as follows: if there exists an IP address of the URL in e-mail and it does not match the defined Rule Set for White List then the received mail is a phishing mail; so the status is phishing e-mail. The final objective of applying GA is to generate rules that match only the anomalous URLs of websites. These rules are tested on historical URLs and are used to filter new URLs to find suspicious phishing attacks.

### Table 1 Advantages & Disadvantages of Anti-phishing techniques

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| List Based Approach   | - This approach is 100% accurate on decision for blacklisting of website
 | - This approach also produce less false positive rate
 | - It also requires less computational cost and easy to use. |
|                       | - If produce much memory overhead
 |                       | - If the websites are not in the list of blacklist then the accuracy is nil |
| Ant Colony Optimization| - This approach is accurate by determining the best rules or features
 | - can be used in dynamic environment
 | - retain memory of entire colony |
|                       | - It enhance false negative rate as compare to other ones. |
| PhishZoo              | - It can classify zero-day phishing and targeted attacks
 | - This approach also able to detect new attack
 | - Reduces false positive rate |
|                       | - It less robust for detection of phishing
 |                       | - It requires matching image site |
| k-nearest neighbor    | - It is much capable to achieve a true positive rate
 | - Capable to achieve high accuracy |
|                       | - huge number of feature
 |                       | - Higher cost
 |                       | - High memory requirement |
| Fuzzy Logic           | - It requires less memory
 | - Its inference speed is also very high |
|                       | - It is not 100% effective
 |                       | - It is complex to design |
| Genetic Algorithm     | - It is better in classifying the email message as phishing mails
 | - It produce less |
|                       | - It requires more domain specific knowledge
 |                       | - They are not easy to handle it. |
**V. CONCLUSION & FUTURE WORK**

Various kinds of attacks found in networks which can counterfeite our personal information such as masquerade, replay, denial of service (DoS). Phishing attack is one of the serious threats of network which stole the user’s secret or confidential information. In this paper, we study different types of anti-phishing techniques and analyses that some are more accurate in detecting such attack but they can only detect known list of attack and also more costly, increases memory overhead but this study provide us solution to combat the phishing attack. In future work, develop such technique which can detect such serious threat accurately and reduces the memory overhead together with decrease the false positive rate.

**REFERENCES**


[16]. http://www.innovateus.net/print/science/what-are-different-types-phishing-attacks


[19]. Ant Colony Optimization, Vittorio Maniezzo, Luca Maria Gambardella, Fabio de Luigi.


