Certain Investigation on Web Application Security: Phishing Detection and Phishing Target Discovery

R.Aravindhan,
Sri Eshwar College of Engineering, Coimbatore
Coimbatore,Tamilnadu, INDIA
contact2aravind@gmail.com

Dr.R.Shanmugalakshmi ,
Goverment College of Technology
Coimbatore,Tamilnadu, INDIA
drshanmi@gct.ac.in

K.Ramya
Hindustan College of Engineering,
Coimbatore,Tamilnadu, INDIA
rythmramya@gmail.com

Dr. Selvan C,
Sri Eshwar College of Engineering, Coimbatore
Coimbatore,Tamilnadu, INDIA
dr.selvan.c@gmail.com

Abstract – With the rapid development of web applications, and with the comfort provided by these web applications , internet users utilize this benefits to a great extent that they make almost all their day to day activities such as news paper reading, shopping, electricity bill payment, ticket booking and entertainment with the help of the internet. This phenomenon forces the users of the internet to get connected with the internet for a prolonged time and hence it increases the chances of the users to get caught in the web of phishing – an attack crafted by hackers to steal sensitive information by tempting the users with lucrative offers initially and then redirecting them to a fraudulent website(which the user may not suspect) where they can deceive the user by asking them to submit their credentials(usually users submit their credentials without knowing that these are fake offers created with a sole intention of stealing sensitive information).In spite of the alert and awareness given by the web community in this regard, more and more phishing artist succeed in their attack. Also these phishing artist develop novel attacks such as tab nabbing, website impersonation etc that attracts more and more internet user to be caught in the web of phishing. However many tools and methodologies have been developed to prevent phishing and to alert users orally and visually. But still the success rates of the phishing attack remains high and also the approaches related to phishing detection suffers high false positive and false negative ratio. In this paper various tools and methodologies used to prevent phishing has been analyzed and an efficient mechanism has been proposed to prevent phishing.

Index Terms— phishing, Anti-phishing, profile cloning, Phish Tank, Gold Phish.

I INTRODUCTION

Phishing is a trap where any targeted individual, is communicated by someone impersonating as a legitimate and a reputed organization to entice the individual into providing sensitive information such as banking information, credit card details and passwords. The personal information is then used to access the individual's account which results in identity theft and financial loss. The first phishing lawsuit was filed in 2004 against a Californian teenager who created a mockery of the website "America Online". With this imitated website, sensitive information was earned from the deceived users and credit card details were accessed to withdraw money from their accounts.

[978-1-4673-9206-8/16/$31.00 ©2016 IEEE]

Communications purporting to be from popular social web sites, auction sites, banks, online payment processors or IT administrators are commonly used to lure the unsuspecting public. Apart from the internet phishing, phishing artist also make phone calls and send text messages with lucrative offers and entice the targeted individual to provide sensitive information.

Phishing can be performed in different ways:
a) email-to-email: In this approach, an email is sent to an targeted user, prompting sensitive information which when the user provides will reach the phishing artist.
b) email-to-website: Here an email is sent to the web users that contains a link, which when clicked will redirect the user to a fraudulent website.
c) website-to-website:
This type of phishing occurs, when a phishing website is reached, when a user clicks on an advert link that appears on a legitimate website.
d) browser-to-website:
This occurs due to the carelessness of the web user, who may misspell a legitimate website that ultimately leads to a phishing website.
The table shows the statistics regarding the phishing in first quarter of 2013[1]

Title	January	February	March
Number of unique phishing e-mail reports (campaigns) received by APWG from consumers	28850	25385	19892
Number of unique phishing websites detected	46066	35024	36983
Number of brands targeted by phishing campaigns	402	348	405
Country hosting the most phishing websites	USA	USA	USA
Contain some form of target name in URL	50.03%	0.75%	55.89%
No hostname: just IP address	1.84%	1.92%	5.24%

Percentage of sites not using port 80	1.36%	2.33%	0.64%

Table 1

There are different phishing strategies developed by the attacker with a sole intention-deceiving the targeted web user and gaining access to sensitive information. Few of the phishing strategies are given below:

a) Luring mail

Phishing scams often include lucrative offers and attention-grabbing statements, for example-"you have won a lottery prize" in the emails. The mails are scrupulously crafted in such a way to ensure that the targeted users trust the identity of the mail and the validity of the offer and hence many people fall prey to these luring phishing emails and eventually end up providing their sensitive information out of which the attackers make a monetary gain.

b) Urgent emails

A favorite phishing strategy is to urge the targeted users without giving them time to analyze whether the mail is a legitimate mail or just a trap. They urge the user by insisting that the fake deals are only for a limited time. Sometimes, the targeted users will be threatened that their account will be suspended unless they update their personal details immediately. Most reliable organizations give ample time before they terminate an account and they never ask patrons to update personal details over the Internet. So the targeted user will panic and half heartedly disclose their personal information which will be then used to make monetary gain.

Figure 1

c) Fake website links

Another tactics implemented by the attackers is to send a link in an email, which redirects them to a fake phishing website which looks exactly like an official website, where fields are provided for entering personal information like credit card numbers, SSN, PIN, password etc and then the phishers gain access to the personal information. A link may not be all it appears to be. For instance, the link in the mail appear as http://www.facebook.com/login but it may instead be directed to another site like http://www.faceboook.com/login and hence the user may submit their login credentials in the space provided in the fake website and then the user may suffer the consequences. The intention of the attackers is to move the web user from a trusted domain to a phishing domain through hyperlinks from where they can gain access to the sensitive information

d) Spam Mails

In phishing, bulk mails are usually sent to a great number of users. Spam mails use the drawbacks of current security techniques to access sensitive information.

Though phishers are always coming up with new phishing strategies, there are some ideas or tips that can be followed to fight against phishing to some extent.

1) Spam filters can be used to escape form scam fails. These filters analyze the origin of the message, the software used to send the message, and the appearance of the message to determine if it is a spam mail or a legitimate one. However, occasionally, spam filters may even block emails from legitimate sources and so it cannot accurately filter phishing mails.

2) The browser settings can be changed to prevent fraudulent websites from opening. Before navigating to a website, the browser checks the requested URL against the phishing database sources such as phish tank [14] which keeps a list of phishing websites and blocks the access to the website or alerts the user with an alert message. But, it cannot be assured that all the phishing websites has been included in these databases.

3) Browsing habits needs to be changed to prevent phishing, for instance one should not get lured into fake deals. If any sensitive information is prompted in a website it is always safe to contact the organization personally, verify the validity of the message, and then disclose the credentials.

4) The links in an email which redirects to a website must be checked. Mostly safe website addresses begins with "https" and if the website address does not contain "https", it can be a fake email.

Though many approaches and methodologies, such as i) detecting suspicious websites with heuristics ii) educating and training web users to avoid phishing attacks iii) compiling whitelist and blacklists iv) filtering emails that prove to be suspicious v) using visual aids to alert web users about phishing websites, have been developed, phishing still remains as a threat to both the web users and the web owners. In the remaining sections of this paper, the short comings of the contemporary phishing detection tools and methodologies have been analyzed and a novel phishing detector has been proposed.

II ANTI-PHISHING TOOLS

Anti-phishing tools are developed with intent to identify phishing content in website and email. It is usually integrated as a tool bar with the web browser that may display the real domain name of the website that the user is viewing, there by alerting the users to thwart phishing.

Many anti-phishing tools and services are emerging which prevents the web users from being a victim of phishing.

a) Web of Trust (WOT)

This is just an Internet website reputation rating tool. When the WOT is installed the browser add-on shows the users the reputations of websites, which are calculated through a combination of user ratings and data from other sources. The WOT browser add-on first sends the user rating to the WOT site and then shows the computed results as a color-coded icons which appears in the user's browser tool-bar and near the external links of leading search engines, email-servers and social network sites[2].

b) Phish tank

It is a user community based anti-phishing site. It maintains a database of the phishing sites which will be used by web browser and other anti-phishing tools to detect phishing. Suspected website are reported to the phish tank and based on the ratings of the web user community it is evaluated and added to the phish tank databases, if it is found to be a phishing site.

c) *ECS (Envelope content splitting) based tools*

ECS makes mail spoofing virtually impossible by authenticating the sender, who sends the link of the fraudulent website in e-mail, as a part of mail transaction, thereby preventing the phishing artist from impersonating any reputed organization.ECS virtually protects the mail recipient from phishing and mail spoofing.

d) Earth link anti spam tool

Earth link is a client based anti spam tool that finds the latest mail spam circulating among the mail boxes of different mail server and blocks the spam mail.

e) Pineapp mail-secure

This is a platform independent anti spam tool that can be integrated with email servers, workstation e-mail clients, email relays, security appliances routers etc

f) Mozilla thunder bird

It is another junk filtering tool which incorporates a Bayesian spam filter. This filters mail based on the white list included in the address book of the e-mail user.

g) Geo trust

It is an organization that validates SSL certificates. Web browsers deny the web users to access those websites which do not have SSL certificate, thereby preventing the web users from falling into the trap of phishing.

h) Spam Assassin

Spam assassin is a computer program that is used for e-mail spam filtering. Spam Assassin uses spam detection techniques such as a)DNS based spam detection b)fuzzy-checksum based spam detection c)blacklist based detection d) Bayesian filtering based approach. This tool can be integrated with mail server of any organization to prevent incoming spam mails or with the mail box of an individual user.

III APPROACHES TO COMBAT ANTI-PHISHING

[A] List Based Anti Phishing Approaches

In this approach a website is classified as phishing or trusted by a mere database look up. These list based approaches are further broke down into blacklist and whitelist.

1) Blacklists

Blacklist is a database that contains the list of websites that have been proven to be fraudulent. Before navigating to a website, the browser checks the database to see, if the requested URL is a legitimate or a phishing website. Some online databases such as Phish tank [3] provides list of phishing websites, however it takes time for a phishing website to be analyzed and added to the black list. Within this time frame, the phishing website would have done most of the damage to the targeted user and the organization that is being impersonated. Another issue is that blacklist approach cannot aid in detecting targeted phishing attacks such as spear phishing [4] since they target small groups or even sometimes an individual web user.

2) White lists

This approach is not as prevalent as blacklists. The idea is that a site must be explicitly trusted before granting access to the website. Here the user manually builds a whitelist by adding the trusted website to the white list. In this case either a) user is denied access to the websites other than the one included in the white list – static implementation or b) user is prompted to add the website to the whitelist before granting access to the website. The problem with this approach is that, every time the user visits a new website, he has to add that to the blacklist and eventually the user will be annoyed and disable this feature. Also, most of the website that the user navigates will be new and the web user has to frequently and manually analyze the trustworthiness of the website which the user has not already included in the whitelist.

Both these type of the list based approach, fails to detect zero day attack (newly created phishing site) as there will not be any information recorded in any blacklist databases. This is because considerable amount of time and analysis is required before confirming the trustworthiness of the website and adding it to the phishing list.

[B] Heuristics based anti phishing approaches

This heuristics based anti phishing approaches uses one or more characteristics of the phishing/legitimate website to reckon phishing, instead of looking it in a list. Few of the characteristics that the detector uses are a) uniform resource locator (URL) b) hyper text markup language (HTML) code c) the page content. The characteristics /parameters are determined such that it reflects the nature of the website accurately. These approaches use machine learning techniques to train the phishing detectors.

One of the heuristic approaches used the composition of the URL to identify the phishing website [5] and some other approaches used DOM properties which are believed to contain the true identity of the website [6].The main strength of the heuristic based phishing detectors is that they are capable of detecting Zero day phishing websites. However the

irrelevant heuristics increase the false positive and false negative ratios, whereas redundant heuristics decreases the performance of the detectors.

[C] Content-Based Phishing Analysis

[9] Suggested a content based image analysis technique to combat zero-day phishing attacks. The heuristic based approach is primarily used to detect text based zero day phishing attacks and it cannot detect the phishing e-mail which contains images. However the suggested a method is capable of detecting image based phishing attacks with high accuracy. It captures the image of a page, then it uses optical character recognition (OCR) to convert the image to text, then the content of the text is analyzed and finally it uses the Google Page Rank algorithm to make a decision on the validity of the site. A tool called GoldPhish is developed which is browser plugin that is used to detect and report phishing sites. The dynamic and adaptive GoldPhish tool consist of three main steps

a) Image capturing
GoldPhish tool utilizes an internal web browser and captures the image in a predefined format and an optimal resolution, which increases the accuracy of the OCR as well as minimizes the time required for OCR.

b) Optical character recognition
OCR process the captured image using Microsoft office Document Imaging (however other commercial adhoc OCR software may increase the accuracy of the OCR results thereby increasing the performance of GoldPhish tools).The OCR is also capable of recognizing the text from the logos contained in the image. Then the text that is obtained from the image is submitted as input to the Google search engine

c) Google search
The GoldPhish tool enters the text into Google search API. The text is submitted line by line to preserve the layout and since the Google search is limited to 50 keywords the most relevant and important keyword is submitted to the Google API. Then the Google search API is implemented to return the first four result as it is sufficient, because a legitimate site will generally come up within the first four results due to its high page rank. on the other hand a phishing site will linger around the web only for a limited period, the page rank will be low thereby will indexed in the top four search results and hence the GoldPhish tool considers it as phishing site and alerts the user via the toolbar.

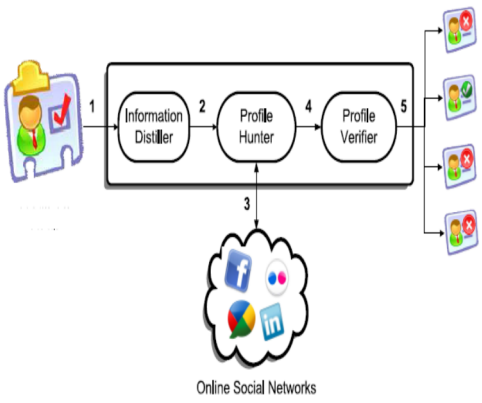
The GoldPhish tool efficiently detects the phishing site. GoldPhish tool accurately classifies the phishing site even without using complex machine based learning algorithm. It gives 0% false positive and negligible 2% false negative. The time taken for detecting the legitimacy of a website is a crucial factor for any anti-phishing tools as it annoys the web users if the tool bar takes a considerable amount of time before making a decision. The time taken by the GoldPhish toolbar to verify a web page is determined by subtracting the time taken to load a page without GoldPhish from the time taken to render a decision using GoldPhish and the mean time is found to be 4.31 seconds which is reasonably acceptable.

In spite of the accuracy and the minimal time provided by the GoldPhish toolbar, it has the following disadvantage: it has to depend on the Google search engine and page rank algorithm, so if the attackers can compromises the search engine and page rank algorithm, it may jeopardize the GoldPhish tool.

[D] Detection of social profile cloning

[10] Suggested a method to prevent social profile cloning. A profile cloning is a form of impersonation where an attacker creates a new profile similar to a profile of the targeted individual by copying their personal information which characterizes the targeted individual. By doing so a form of trusted entity is created in one or more social networks. As a result, unsuspected users and other attackers tend to trust this fake profile and hence the attackers utilize this trust to drive the social network users, especially those who trust this fake profile into a phishing site and lure them into leaking their sensitive information and sometimes cause repercussions to the owner of the original profile .In the suggested method a tool is designed and implemented which detects social profile cloning which consists of three main components:
a) Information distiller:
This analyses the user information and extracts the pieces of information that are more user specific and then these user-identifying terms are used to create a user-record.

b) Profile hunter
This component processes user records and uses the user-identifying terms to locate social network profiles that may potentially belong to the user. Different profiles are harvested from the search results that are obtained and then a profile record is created which contains a link to the user's legitimate profile and all other profiles returned in the results.



c) Profile verifier:
Each profile contained in the profile record is analyzed on the grounds of similarity and a similarity score is calculated. After that all the harvested profiles along with its associated similarity score is presented to the owner of the legitimate profile.

According to the experiment conducted with ten different linked in profile , the suggested method were capable of detecting the cloned profile with 100% accuracy i.e. zero false positive and zero false negative. However it cannot detect the cloned profiles with deliberately injected mistakes .Also, in spite of the widely available social networks, the tool is experimented only on a single social network (LinkedIn).

[E] URL based Phishing detectors
[11] Proposed a URL based phishing detectors particularly used to reduce the false positive ratio of the phishing detectors. The main idea of this approach is to extract the possible domain name from the victim URL and then compare the page rank of the extracted domain names with the actual domain. If, there is a considerable difference in the ranks then the extracted domain name will be reported as phishing. This approach based on the domain rank can effectively detect phishing, because the phishing campaigns live only for a short period of time (63% of the phishing lives only for only two hours) [7].The process of the heuristic URL based classifier is shown in fig.

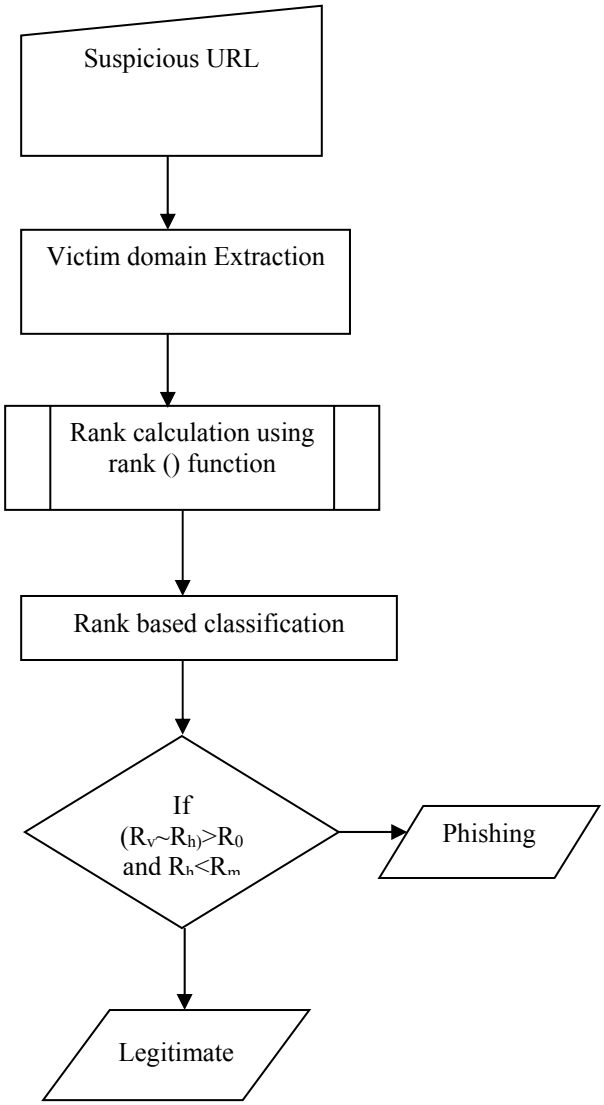
It primarily consists of three process and they are

a) Domain name extraction
In this step the domain name is extracted by lexically analyzing a given URL where a list of Top Level Domain (TLD) is used as a reference.

b) Domain rank extraction
Here the domain rank algorithm based on the citation counting is used to extract the rank of a given domain. The rank is calculated using the formula(A) = n(A), where the rank of the document A R(a) is equal to the number of other documents n that links back to the document A.

c) Classification
In this step, first the rank R_v and R_h of the victim domains v and host domain h respectively are calculated using the rank () function. Then if the difference between the domain rank is greater than the offset R_0 and if the rank of the victim is below the minimum rank threshold R_m(prevents the legitimate site to fall under phishing category)

This approach effectively classifies phishing based on the features found in the URL, but a considerable delay is introduced to extract the possible domains and to calculate the citation count for each of the domains.



[F]Behavior model approach for testing phishing sites

[12] Suggested a behavior model based approach for testing phishing sites. This method addresses the problem faced by the conventional phishing detector which is based on testing against a known set of inputs and matching the actual output with the expected ones. This approach checks the behavior of the website using a number of heuristics and then the behavior model is characterized using a notion of finite machine. There are totally eight different heuristics that falls under three main categories which is used to test the behavior model. They are

a) State based heuristics

i)No loop (H1): If a website traverses more than one state, it will be considered as a heuristic to indicate whether the website is phishing or legitimate. ii) Single loop (H2). If requests and the corresponding responses result in a website to remain the same state more than once, a loop is formed. A test case having a loop with respect to a state can represent either a phishing or a legitimate website.iii) Multiple loops (H3). If requests and the corresponding responses result in the formation of more than one loop, then multiple loops is formed which is again used as heuristics to detect advanced attacks

b) Response and form based heuristics

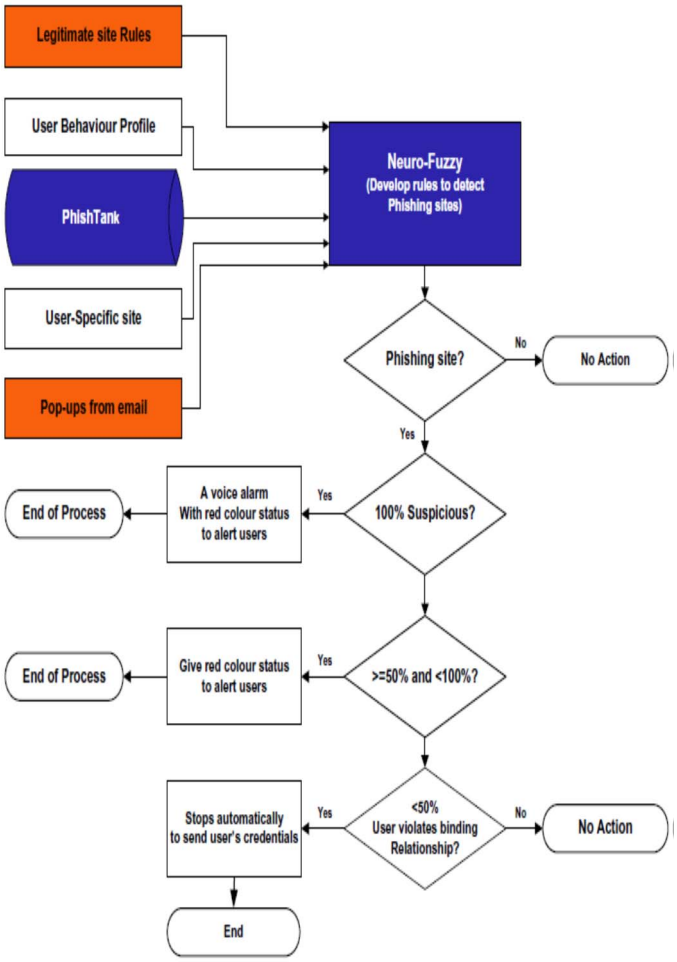
i) Maximum form submission (H4): a legitimate website uses a limited number of forms, whereas a phishing website is designed to accept numerous form submissions and hence Maximum form submissions are used as a heuristics. ii) Maximum error message (H5): Phishing websites rarely verify the provided information for login functionality, whereas a legitimate website rejects random inputs and generates an error messages in response pages. Thus number of error message is considered as a heuristic.iii) Presence of supplied input (H6): A legitimate website often greets a user after a successful login or registration with the supplied name, user id, or an email, whereas a phishing website does not generate a response page IV) No form (H7). This heuristic criterion checks whether a form submission results in a page that has no input form (or no hyperlink to proceed further). V) Common form (H8). This criterion is satisfied, if a current form being submitted with random inputs matches with any of the forms that have been submitted before.

A tool called Phish Tester is implemented and tested against 33 phishing and 19 legitimate websites and it has been found that the approach is capable of accurately detecting all phishing(zero false positive) and legitimate website(zero false negative).however this method is not efficient in testing phishing sites that contains embedded objects such as images and flash content

[G] Hybrid neuro-fuzzy phishing detectors

[13] Suggested a robust method based on hybrid neuro-fuzzy scheme to detect phishing sites. Five inputs such as a) Legitimate site rules b) User-behavior profile c) Phish Tank d) User-specific sites e) Pop-Ups from emails are used to improve the accuracy of the detector in real time. The main process in this hybrid methodology includes a) identifying and extracting the features based on the five inputs b) developing a hybrid neuro-fuzzy model c) training the fuzzy inference model in real time d) detecting the phishing sites based on the output of the neuro-fuzzy model. The overall process is explained in the fig

A 2-fold cross validation is applied to train and test the model, which uses 288 extracted features to accurately detect the phishing sites. However more features are required to improve the accuracy further and parameter optimization is required to improve the performance of the detector



[H] Ant phishing through Phishing Target Discovery

[14] Defines a method to discover the target phishing site. A target phishing site is a legitimate website that is impersonated by the attackers to lure the legitimate website users into providing their website information. Generally a phishing site not only cause serious threat to the public, but also hazard the reputation of the imitated legitimate website and hence target discovery is ineludible with the case of the anti-phishing approaches. The proposed method has taken effort to discover the target phishing site and inform the owners of the targeted site so that necessary counter measures may be initiated to avert further consequences. In this method a term called “parasitic relationship” is introduced which exhibits the relationship between the phishing site and the targeted entity.

This term plays a vital role in target discovery due to the fact that 42% of the phishing sites contains link that is directed to the targeted entity [8].Given a web page P, the method employs the following steps

a) Finding the initial associated page set
Here the web pages that are directly associated with P (directly associated pages) and the web pages which contain similar text and visual information (indirectly associated web pages) is extracted to form an initial set of associated pages.

b) Building a web graph
A web graph is built by expanding the initial associated lists which requires two steps a) new web pages, that are directly associated (both forward and backward link) with the web pages contained in the initial associated lists are collected) Then the newly collected web pages is taken as reference and the web pages associated (only forward) with these web pages are collected to form a web graph

1) Partitioning the web graph
   A minimum-cut technique is employed to partition the web graph which removes irrelevant, unrelated and redundant web pages.

2) Reinforcing the associated page set:
   Finally the associated page set is reinforced by adding more active pages and removing inactive pages. Then from the reinforced associated set parasitic-coefficient is calculated which aids in phishing target discovery.

A total of 10,005 phishing sites which targeted 154 reputed organization is tested for accuracy and it is found that the phishing detection accuracy was 99.2% and phishing target discovery accuracy was 92.1%.Also the mean false alarm rates was recorded as 2.2% which is quite acceptable, but further effort needs to be taken to speed up the process of target phishing discovery.

[I] Phishing detection and impersonated entity discovery using Conditional Random Field and Latent Dirichlet Allocation

[15] Proposed multi-stage methodology employing machine leaning and natural language processing for building a robust tool which detects phishing and discovers impersonated entity (target phishing site).The robust detector has different stages as follows:

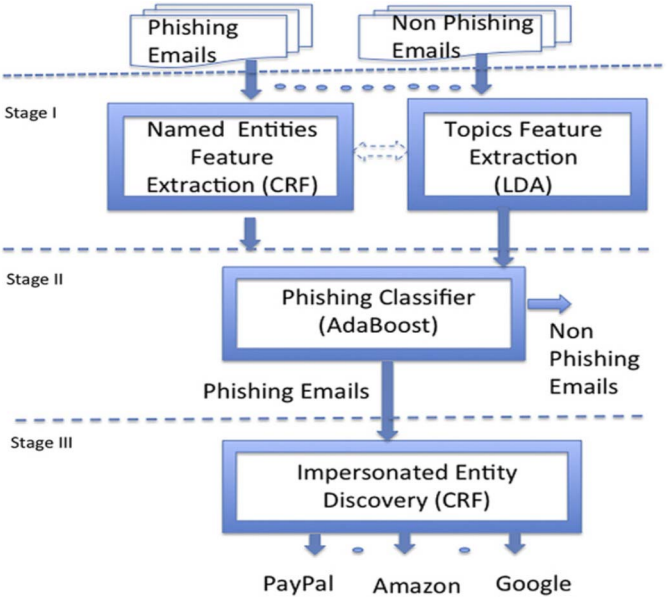
Stage 1: Feature extraction

1) Named entity feature extraction
The first stage makes use of Conditional Random Field (CRF) to extract named entities such as name of the phishing e-mail recipient, name of the organization which is impersonated in the e-mail and the location specified in the mail. The named entity plays a vital role in classifying the phishing mail and the legitimate one

2) Feature extraction
Latent Dirichlet Allocation (LDA) is used to extract the content of the messages by analyzing group of words or phrases which constitute a vital parameter that is used to discriminate phishing and legitimate e-mail

Stage 2: phishing classifier
In this stage a hybrid robust and accurate classifier (AdaBoost) is built by combining many moderately accurate classifiers. AdaBoost is employed on combined feature sets of topic distribution probabilities obtained using LDA and named entities obtained using CRF to build a strong classifier for phishing detection.

Stage 3: impersonated entity discovery
After classifying the phishing mail using AdaBoost, impersonated entities are automatically discovered by using the features extracted by CRF and LDA and finally the organization/company which is impersonated is alerted to keep their customers safe and secure.

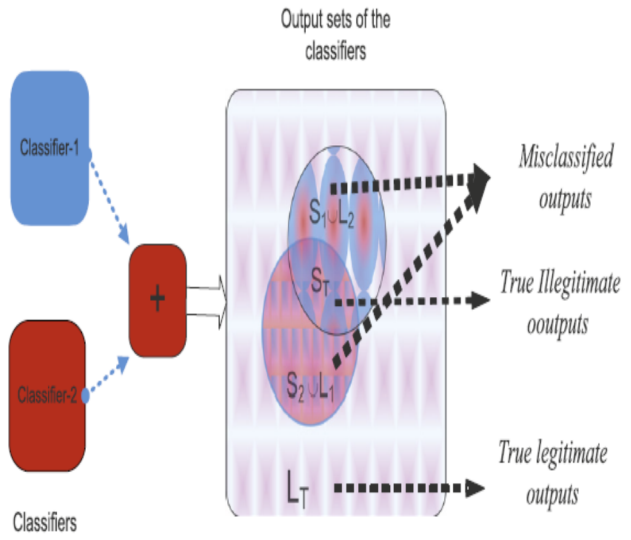


A 10-fold cross validation is employed to validate the phishing classifier. The classifier shows 100% accuracy when the percentage of phishing mail is less than 20%.The impersonated entity discovery rate for phishing e-mails was observed to be 88%, but on the other hand the discovery rate of phishing e-mail was considerably low.

[J] Multi-tier phishing detectors

[16] Uses a multi-tier classification for phishing e-mail. An innovative method is proposed to select the relevant and optimum feature based on message content and message header. In the first stage of classification two different

classifiers T_1 and T_2 parallely classifies the suspicious e-mail. If both the classifier classifies the suspicious e-mail as phishing(true phishing S_T) then it is considered as phishing e-mail and on the other hand if both the classifiers classifies the suspicious e-mail as legitimate(true legitimate L_T).However if the first classifier classifies as phishing and second classifier classifies as legitimate($S_1 U L_2$) or vice versa ($S_2 U L_1$), it is considered as misclassified output and it is sent to the next stage classifier T_3 , which then classifies the misclassified e-mail accurately.



The proposed method classifies the phishing e-mails with 97% accuracy; however it exhibits an average 2% false positive and 9% false negative.

Phishing detectors/ parameters	Can detect Zero Day attack?	Whether it aids target discovery?	Ability to find social profile cloning?	Whether detects phishing sites containing	False positive rate	False negative rate	Whether Detects image based phishing?	Whether depends on Google page rank algorithm or phishing database?
[A]	NO	NO	NO	NO	NA	NA	NO	YES
[B]	YES	NO	NO	NO	Between 0.4% and 2.26%	Between 6.69% and 16.9%	NO	YES
[C]	YES	NO	NO	YES	2%	2%	YES	YES
[D]	NA	YES	YES	NO	0%	0%	NO	NO
[E]	YES	NO	NO	NO	1.39%	16.69%	NO	YES
[F]	YES	NO	NO	NO	0%*	0%*	NO	NO
[G]	YES	NO	NO	NO	1.5%	1.5%	NO	YES
[H]	YES	YES	NO	NO	2.2%	2.2%	NO	YES
[I]	YES	YES	NO	NO	0%*	0%*	NO	NO
[J]	YES	NO	NO	NO	2%	9%	NO	NO
* → limited data set is used for testing								

IV CONCLUSION

The credibility of a phishing detector lies in protecting the individuals from falling prey to phishing. Obviously we cannot claim that every web users are security experts and hence a sophisticating phishing tool is required to protect the web users. The list based and heuristic based approaches, though can detect the phishing website the observed false alarm rate with this approach is quite unacceptable. On the other hand the machine learning and multi-tier classification approaches improves the detection accuracy, but it is inefficient in detecting image based phishing content. Also, when on one side ,the web users suffers privacy and economic loss ,due to the unintended disclosure of sensitive information, on the other side the phishing sites jeopardize the impersonated site which is usually a reputed organization and hence they tend to lose valuable customer and money. However, only a few researches have taken effort to discover the targeted phishing site. Recently, a more sophisticated form of phishing based on flash content is prevalent, which bye passes the strategy of most phishing. So, we propose a phishing detector, which can detect phishing content containing embedded objects and aids in target discovery. The proposed phishing tool will be designed in such a way that it will have a perfect trade-off between time complexity and detection accuracy.

V REFERENCES

[1] Phishing activity trend report, "http://docs.apwg.org/reports/ apwg_trends_report_q1_2013".

[2] Features of phishing e-mail, "http://www.phishing.org".

[3] Phish tank. Available at: http://www.phishtank.com/ accessed on 26 Oct 2009.

[4] B. Kesler, H. Drinan, and N. Fontaine. News briefs. IEEE Security and Privacy, 4(2):8–13, 2006.

[5] S. Garera, N. Provos, M. Chew, and A. D. Rubin. A framework for detection and measurement of phishing attacks. In WORM '07: Proceedings of the 2007 ACM workshop on Recurring malcode, pages 1–8, New York, NY, USA, 2007. ACM.

[6] Y. Pan and X. Ding. Anomaly based web phishing page detection.In ACSAC '06: Proceedings of the 22nd Annual Computer Security-Applications Conference, pages 381–392, Washington, DC, USA, 2006.IEEE Computer Society.

[7] Steve sheng, Brad wardman, Gary warner, Lorrie faith Cranor, Jason Hang, and Chengshan Zhang, "An emprical analysis of phishing blacklists" http://ceas.cc/2009/papers/ceas2009-paper-32.pdf, 2009,Accessed july 2010.

[8] M. Cova, C. Kruegel, and G. Vigna, "There is No Free Phish: An Analysis of 'Free' and Live Phishing Kits,"*Proc. 2nd Usenix Workshop Offensive Technologies* (WOOT 08), Usenix Assoc., 2008, pp. 1–8.

[9] Matthew Dunlop, Stephen Groat, and David Shelly " GoldPhish: Using Images for Content-Based Phishing Analysis " *the Fifth International Conference on Internet Monitoring and protection* ,Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA, 2010.

[10] Georgios Kontaxis, Iasonas Polakis, Sotiris Ioannidis and Evangelos P. Markatos "Detecting Social Network Profile Cloning**"** *Foundation for Research and Technology Hellas - Institute of Computer Science*

[11] Mahmoud khonji, Andrew Jones, Youssef Iraqi, "A novel phishing classification based on url features" 2011 IEEE GCC Conference and exhibition(GCC), Dubai,United Arab Emirates, 2011

[12] Hossain Shahriar, Mohammad Zulkernine "Trustworthiness testing of phishing websites: A behavior model-based approach" Future Generation Computer Systems School of Computing, Queen's University, Kingston, Canada -2010

[13] P.A. Barraclough, M.A. Hossain, M.A. Tahir, G. Sexton, N. Aslam, "Intelligent phishing detection and protection scheme for online transactions" Expert Systems with Applications 40 4697–4706(2013)

[14] Liu Wenyin, Gang Liu, Bite Qiu, and Xiaojun Quan, "Antiphishing through Phishing Target Discovery" Website Security -City University of Hong Kong.

[15] Venkatesh Ramanathan, Harry Wechsler "Phishing detection and impersonated entity discovery using Conditional Random Field and Latent Dirichlet Allocation" Department of Computer Science, George Mason University, Fairfax, VA 22030, USA - 2012

[16] Rafiqul Islam, Jemal Abawajy "A multi-tier phishing detection and filtering approach" Journal of Network and Computer Applications 36 324–335 - (2013)

2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS -2016), Jan. 22 – 23, 2016, Coimbatore, INDIA