A Survey on Detection of Phishing Attacks

G. N. Sneha[1], M. B. Nanda[2]
[1]*Student, Dept. of Computer Science and Engineering, Sapthagiri College of Engineering, Bengaluru, India*
[2]*Assistant Professor, Dept. of Computer Science and Engg., Sapthagiri College of Engineering, Bengaluru, India*

***Abstract*: Phishing is a kind of attack in which criminals use spoofed emails and fraudulent web sites to trick people into giving up personal information. Phishing Websites is really a complex and dynamic problem. Phishing can be defined as a criminal activity using social engineering techniques. The most common example of phishing is an email asking to enter account/credit card information for Electronics Commerce websites and online banking. It is also kind of attacks in information security. It points to the act that the attacker demand users to visit a faked web site by sending them faked e-mails or instant messages, and without a sound get victim's personal information such as user name, password, national security ID, etc. By using the software phishing spam mails are detected by examining mail contents. For Detection of phishing attracts Bayesian algorithm is used. The main goal of the phishers is always to attract nation into giving up important information. Phishing is also identified as "Brand Spoofing". One of the most important aims of phishing is to dishonestly carry out fraudulent economic transactions on behalf of users using a fake email that contains a URL pointing to a fake web site concealed as an online bank or a government entity. Phishing is a rising difficulty for internet user the statement has its origin from two words - Password harvesting or-fishing for password.**

***Keywords*: Pharming, embedding and company.**

1. Introduction

Phishing is a method of trying to gather personal information using misleading Email and websites. Here what you need to know about this venerable but increasingly sophisticated form of cyber-attack. Phishing is a cyber-attacks that uses disguised email as weapon. The goal is to trick the email recipient into believing that the message is something they want or need a request from their bank, or a note from someone in their company and to click a link or download an attachment. What really distinguishes phishing is the from the message takes the attackers pretence as a trusted entity of some kind, often a real or plausibly real person or a company the victim might do business with. It's one of the oldest types of cyberattacks. With phishing messages and techniques becoming increasingly sophisticated. Phishing is pronounced just like its spelled, which is to say like the word "fish" the analogy is of an angular throwing a baited hook out there and hoping you bite. The "ph" is a piece of convention of capricious programmer spelling, and was most likely impacted by the expression "phreaking" another way to say "telephone phreaking" an ahead of schedule from of hacking that included playing sound tones into phone handsets to get free telephone calls. In phishing general information is stolen form the user are credit card information, user account number, user password and user name, Internet banking information. Anti-Phishing Simulator which is designed to prevent serious threats like this catches malicious email arriving at email addressing integrated into system. Here we use a software "Anti phishing simulator" to detect phishing emails. Classification of spam words added to data base by Bayesian Algorithm.

2. Various techniques used for phishing attack

1. Deceptive phishing.
2. Spear Phishing
3. CEO Phishing
4. Pharming
5. Dropbox Phishing
6. Goggle Docs Phishing

3. Phishing techniques used by attackers

1. Embedding a link in an email that redirects your employee to an unsecure website that requests sensitive information.
2. Installing a Trojan via a malicious email attachment or ad which will allow the intruder to exploit loopholes and obtain sensitive information.
3. Spoofing the sender address in an email to appear as a reputable source and request sensitive information.
4. Attempting to obtain company information over the phone by impersonating a known company vendor or IT department.

4. Protect itself against phishing:

1. Educate your employee and conduct training sessions with mock phishing scenarios.
2. Deploy a SPAM filter that detects viruses, blank sender, etc.
3. Keep all the system current with the latest security patches and updates.
4. Install a antivirus solution, schedule signature updates, and monitor the antivirus status on all equipment.
5. Develop a security policy that includes but isn't limited to password expiration and complexity.
6. Deploy a web filter to block malicious websites.
7. Convert HTML email into text only email messages or

disable HTML email messages.

8. Require encryption for employees that are telecommuting.

5. Literature survey

Online detection and prevention of phishing attacks:

It is a new type of attack where the attacker creates a copy of existing web page for fooling the user. They use this types of techniques to submit user personal details, financial or password. They propose a new end host based anti-phishing algorithm called Link Guard. These characteristics derived by analysing the phishing data archive provided by the anti-phishing work group. Since it depends on the conventional attributes of phishing attacks, Link Guard can distinguish referred to as well as obscure phishing attacks.

Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity:

Nowadays social network has become one of the most universal platform for user to interact with each other. The appearance of web page is among the most important factor in misleading user and thus the similarity among the web pages is critical metric for detection phishing website. In this we present algorithm to quantify the suspicious rating of web pages. Since cascading style sheet(CSS) in technique to specify page layout across browser implementation our approach uses CSS as the basis to accurately quantify the visual similarity of each page attachment.

Tracking phishing attacks over time:

Phishing attacks are one of the important threats to individual are corporations in today's Internet. Combatting phishing is this top priority, has been the focus of much work both on the academic and on the industry side. This provides several opportunities and insight for the fight against phishing.

Phishing Detection by determining Reliability Factor using Rough Set:

Phishing is a common online weapon used against. The inception of internet nearly everything raging from money transaction to sharing information is done through online only in most part of the word. This also give rise to a malicious activity. Here it approaches towards phishing detection using rough set theory. Using this it detects the suspected site is it valid or fake.

6. Proposed methodology

A. *Processing cycle of phishing attacks*

The above fig. 1 is about processing cycle of phishing attacks. This process start from phishing the email. The attacker will send the fraudulent or fake website to the user or a targeted user through the mail. User will go to that website through the system or some of the devices, then that system will be in danger because the user will click the website that website will be directed to the some of the fake website addresses so that the system will be in danger. The attacker will access the user devices or the system completely through the RAT (Remote Access Trojan) is a malware program that includes a back door for administrative control over the target computer. Through the internal network all the data of the system will be collected by the attackers. This how processing cycle of phishing attacks works.

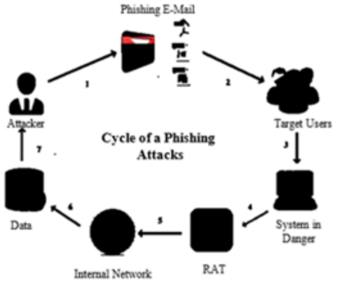


Fig. 1. The processing cycle of phishing attacks

7. Flow chart

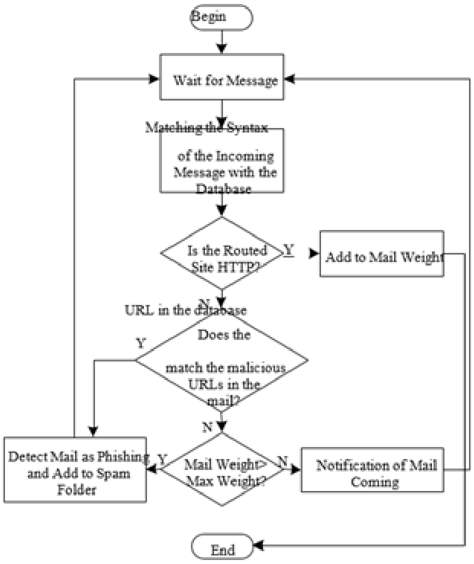

Begin
Wait for Message
Matching the Syntax of the Incoming Message with the Database
Is the Routed Site HTTP?
Y
Add to Mail Weight
N
Does the URL in the database match the malicious URLs in the mail?
Y
N
Detect Mail as Phishing and Add to Spam Folder
Y
Mail Weight> Max Weight?
N
Notification of Mail Coming
End

Fig. 2. Flow chart

Initially it begins or starts, then it waits for message that means at the starting point only it checks whether the message is phishing message or not if it is phishing message then it blocks if not user get the message. First what the system do is it check the Incoming message with the syntax matching. Then it checks whether the message from the routed site or from the trusted site. If it is from trusted site, then mail weight will be added that means it will be considered that message notification will be received by user. If it is not from the trusted site, it checks that the incoming message URL and the database URL if these both match then that is detected mail as phishing and add to spam folder. Then it checks the mail weight, if mail weight is greater than that max weight then user get the notification to user. Then mail weight is not greater than max weight then that mail will be detected as phishing and added to spam folder.

8. Logic operation cycle

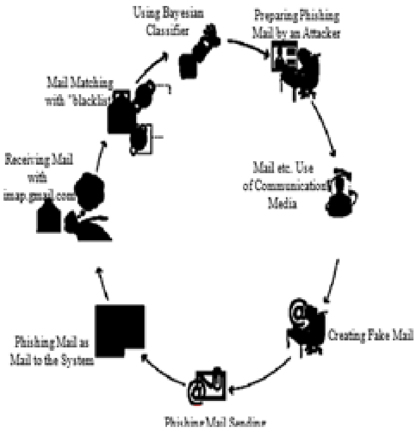
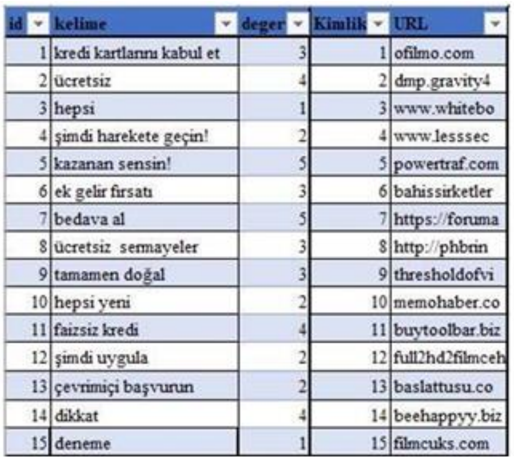

Using Bayesian Classifier
Preparing Phishing Mail by an Attacker
Mail Matching with "blacklist
Receiving Mail with imap.gmail.com
Mail etc. Use of Communication Media
Creating Fake Mail
Phishing Mail as Mail to the System
Phishing Mail Sending

Fig. 3. Logic operation cycle

9. Data set used

Table 1
Data set used



id	kelime	deger	Kimlik	URL
1	kredi kartlarını kabul et	3	1	ofilmo.com
2	ücretsiz	4	2	dmp.gravity4
3	hepsi	1	3	www.whitebo
4	şimdi harekete geçin!	2	4	www.lesssec
5	kazanan sensin!	5	5	powertraf.com
6	ek gelir fırsatı	3	6	bahissirketler
7	bedava al	5	7	https://foruma
8	ücretsiz sermayeler	3	8	http://phbrin
9	tamamen doğal	3	9	thresholdofvi
10	hepsi yeni	2	10	memohaber.co
11	faizsiz kredi	4	11	buytoolbar.biz
12	şimdi uygula	2	12	full2hd2filmceh
13	çevrimiçi başvurun	2	13	baslattusu.co
14	dikkat	4	14	beehappyy.biz
15	deneme	1	15	filmcuks.com

To estimate the results, a particular classification algorithm works on a set of qualifications and a training set containing the relevant result, often referred to as the target or estimated quality. The algorithm tries to predict the results and investigate possible relationships between qualifications. Then, the algorithm is given an unseen data set, called the set of estimates, containing the same set of attributes, with the exception of an unknown set of estimates. The algorithm analyses the input and generates an estimate. Prediction correctness indicates that the algorithm used is "good".

10. Conclusion

Email is a standout amongst the most significant specialized strategies. In this research we focus on developing detecting and preventing the phishing website. Number of spam is increasing in every year. Some effective tricks have been developed with addition of spam senders. Here it works on two consistence. One is on client side another one is on sever side phishing. The effective of these approaches is evaluated on the broad body of the simple text data and the text embedded image data set. Anti-phishing simulator collects the phishing and spam message. In the future, it is aimed to analyze mail content more thoroughly with basic text mining by increasing the spam keyword database much more. It is also aimed to obtain more accurate results and classification with artificial neural networks.

References

[1] I. R. A. Hamid and J. H. Abawajy, "Online Detection and Prevention of Phishing Attacks" 2014 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, Melbourne.

[2] M. Al-Daeef, N. Basir and M. M. Saudi, "A method to Measure the Efficiency of Phishing Emails Detection Features " 2015 International Conference on Information Science & Applications.

[3] Pradeepthi KV and Kannan A, "WebSSS Phishing Detection Using a Deep Learning Framework" 2015 Sixth International Conference on Advanced Computing (ICoAC), Chennai.

[4] Ahmet Selman Bozkir Hacettepe University Dept. of Computer Engineering Ankara, Turkey "Use of HOG Descriptors in Phishing Detection" April 2016.