Web Phishing Detection

A

dissertation

submitted in partial fulfillment of the
requirements for the award of degree of
Master of Technology
in
Computer Science and Engineering

Submitted By

Akshat Mishra
(16535001)

Under the guidance of

Prof. Manoj Misra
Dept. of Computer Science and Engineering

INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE

Department of Computer Science

May, 2018

CANDIDATE'S DECLARATION

I hereby declare that the dissertation entitled ”Web Phishing" submitted by me in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Computer Science and Engineering to the Department of Computer Science and Engineering, Indian Institute of Technology Roorkee is my original work carried during May 2017 to May 2018 under the guidance of prof. Manoj Misra, Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee.

The content presented in this dissertation has not been submitted by me for the award of any other degree of this and any other institute.
Date:

Place: Roorkee

Akshat Mishra

CERTIFICATE

This is to certify that the statement made by the candidate in the declaration is correct to the best of my knowledge and belief.

Date:

Place: Roorkee

prof. Manoj Misra

(Professor)

(Department of Computer Science and Engineering)

(Indian Institute of Technology Roorkee)

ACKNOWLEDGEMENTS

First and foremost,I would like to extend my heartfelt gratitude to my guide and mentor **Dr. Manoj Misra,** professor, Department of computer Science and Engineering, Indian Institute of Technology Roorkee, for his invaluable advices, guidance, and encouragement and for sharing his knowledge. His wisdom, knowledge and commitment to the standards inspired and motivated me. He has been very generous in providing the necessary resources to carry out my research. He is an inspiring teacher, a great advisor, and most importantly a nice person.

I am greatly indebted to all my friends, who have graciously applied themselves to the task of helping me with ample moral supports and valuable suggestions.

On a personal note, I owe everything to the Almighty and my father. The support which I enjoyed from my father provided me the mental support I needed.

Akshat Mishra

ABSTRACT

Phishing is a way of obtaining personal information through illegitimate websites that are exactly same as legitimate website. There are many techniques available to detect phishing websites, but current techniques leave much to be obtained. A primary problem is that web browsers rely on a blacklist to detect phishing sites, but phishing sites have a very short lifespan only few hour to few days . A faster recognition system is needed to identify zero day phishing sites which are new phishing sites that have not yet been discovered.

This research introduces a new method of detecting illegitimate websites using Search Engine and content of the E-mail. Current phishing detection techniques are examined and the proposed detection method is implemented and evaluated against of known phishing sites while the results were analyzed.

Chapter 1

OVERVIEW OF PHISHING

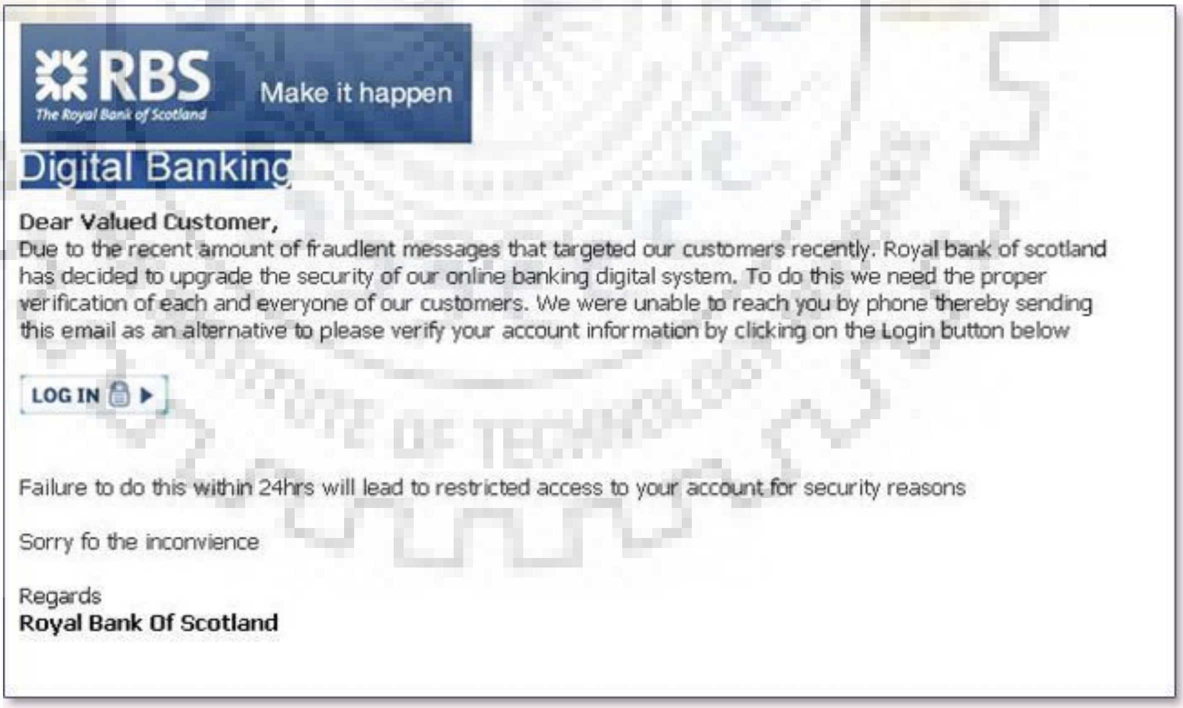
On January 2, 1996, the term phishing was used for the first time by Usenet newsgroup to denote the fraud with the America Online (AOL) users in which phishers used an algorithm to create randomized credit card numbers. While lucky hit is few but enough to cause a lot of damage to AOL and its users [1]. The term "phishing refers to the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money), often for malicious reasons, by disguising as a trustworthy entity in an electronic communication"[2].

1.1 Introduction

Generally in phishing there are three phases. In first phase, the phisher creates a phishing website and goes for phishing by sending out a large number of emails to random email addresses. The phisher tries to convince the reader to visit the phishing website by clicking on the content present in the email. When the user clicks on the link, the link in the email directs the user to the phishing website which appears exactly same as the legitimate target website. The phishing is considered successful when user enters confidential information on the phishing page and shares its users confidential information to the phisher. Afterward, the phisher tries to use the confidential information by opening accounts, making purchases, or transferring money using the captured information, or the phisher merely acts as a middleman and sells the information to other criminals persons.

According to the Microsoft Computing Safety Index, released on Feb 11, 2014, phishing cause an impact of US$5 billion worldwide annually by various forms of identity theft and if the cost of damage caused due to reputation of peoples from online included then the overall cost becomes nearly USD $6 billion, or an estimated average of USD $632 per loss. According to the survey, 20 percent Indians were victims of online phishing attacks, of which 12 percent Indians on an average cost Rs 7,500 because of identity theft [5].

Initially, the phisher uses electronic mail messages which are designed to look like an e-mail from a trusted agent such as a bank or online commerce site generally. These messages use a sense of urgency, for example, a threat to account holder to suspend the account if he/she does not take the required action within mentioned time in the email generally a very short span of time is given to the account holder to take action so that it motivates the account holder to take action immediately without any second thought. And this urgency leads the account user to fall into the trap of phisher without suspecting. But nowadays many new social engineering approaches is been developed by phishers to trick unsuspecting users [6]. These include filling out a survey for a banking site, e-mail messages claiming rewards and asking the victim to verify the credit/debit card information or message containing a URL to redirect the victim to a website that mimics the look-alike of the legitimate site. Over the time these fake websites and emails become more technically advanced to deceive casual investigation.



Dear Valued Customer,
Due to the recent amount of fraudlent messages that targeted our customers recently. Royal bank of scotland has decided to upgrade the security of our online banking digital system. To do this we need the proper verification of each and everyone of our customers. We were unable to reach you by phone thereby sending this email as an alternative to please verify your account information by clicking on the Login button below

Failure to do this within 24hrs will lead to restricted access to your account for security reasons

Sorry fo the inconvience

Regards
Royal Bank Of Scotland

Fig 1.1. Phishing email example [7]

Today internet has dominated many sectors including e-healthcare and e-commerce. Internet use has increased the comfort of human life but it has also increased the need for security measures. All the web browsers and servers adopt several measures to make a guarantee of safe business through internet. But even then browsers and servers are vulnerable to attacks like phishing. Phishing is a type of online theft. Phishing aims to steal the personal sensitive information of online banking users. But last a few years phishing has received huge press coverage because phishing attacks have escalated in number as well as in sophistication. According to Anti Phishing Working Group (APWG) Feb. 23, 2017 report on phishing attacks, the total number of phishing attacks in 2016 was 1,220,523 around 65 percent of increase compared to 2015. According to the report, the most infected country by malware is China, where 47.09% of machines are infected, followed by Turkey 42.88% and Taiwan 38.98%[8]. The list of top 10 countries is given on the table.

Ranking	Country	Infection Rate
1	China	47.09%
2	Turkey	42.88%
3	Taiwan	38.98%
4	Guatemala	38.56%
5	Ecuador	36.54%
6	Russia	36.02%
7	Peru	35.75%
8	Mexico	35.13%
9	Venezuela	34.77%
10	Brazil	33.13%

Table 1.1**:** Malware infection rate according to APWG report Feb 2017[8]

In recent years a lot of anti-phishing techniques have been proposed and implemented. With each new anti-phishing technique phishers find a new way of phishing, hence it created a race between phishers and working anti-phishing organizations. According to the APWG October 17,2017 report “On average, each malware targeted three companies, and each pharming incident targeted two targets. The maximum number of targets for a single piece of malware was 23, while one pharming attack targeted six companies. The targeted companies were usually from the financial sector: banks and credit card companies .”[21] The number of phishing attacks identified in the first half of this year is given in table 2.

	January 2017	February 2017	March 2017	April 2017	May 2017	June 2017
Number of unique phishing websites detected	42,889	50,567	51,265	50,328	45,327	50,720
Number of unique phishing e-mail reports (campaigns)	96,148	100,932	121,860	87,453	93,285	92657
Number of brands targeted by phishing campaigns	424	423	444	460	457	452
Number of domain names used in attacks	13,977	15,877	17,397	21,652	21,373	18,404

Table 1.2 Different type of phishing attacks reported in first half of 2017. APWG report october 17 2017[20]

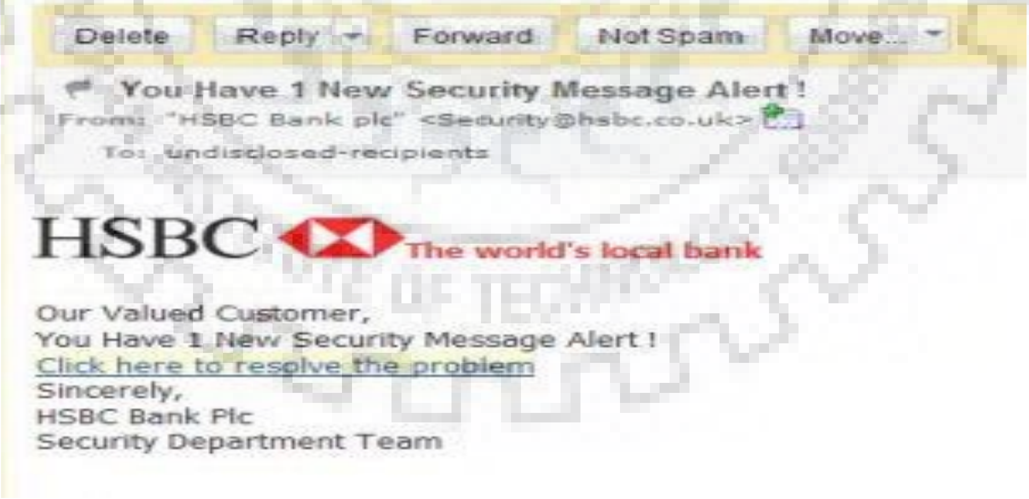
The one of the most successful phishing attack of history unfolded this year named as “WannaCry” ransomware which began on 12 May 2017 and within two days it has infected more than 230,000 computers in over 165 countries. Initially it was thought that WannaCry propagates through emails like most of other malwares of its kind but later on it was discovered that it uses emails only to transfer itself from one network to another but within the local network it exploiting EternalBlue, a service of Windows Server Message Block (SMB) protocol. The vulnerability in this protocol is known to National Security Agency (NSA) of U.S and even to Microsoft about the flaw in their operating systems months before the attack, even on March 14, 2017, Microsoft issued security bulletin MS17-010, which detailed the flaw and its patch that been released for all Windows versions that were currently supported at that time, that are Windows 7, Windows 8.1, Windows 10, Windows Server 2008, Windows Server 2012, and Windows Server 2016[21][22]. These kind of flaw in existing systems and protocols make a need of software that can scan emails and verify the genuinity of the mail so that spreading of such malware can be stopped

1.3 Problem statement :

To propose a technique for detecting phishing websites through phishing emails that contain only an image, a redirect link to phishing website attached to the image or contains an image a redirect link phishing website attached to image and some text data

DESCRIPTION :

With the advancement in technology nowadays emails support html5 which can be exploited by phisher by sending an email that contains the only image in which text data is written and a click button is incorporated.



You Have 1 New Security Message Alert !
From: "HSBC Bank plc" <Security@hsbc.co.uk>
To: undisclosed-recipients

HSBC The world's local bank

Our Valued Customer,
You Have 1 New Security Message Alert !
Click here to resolve the problem
Sincerely,
HSBC Bank Plc
Security Department Team

fig 1.2 Image of an phishing email containing only an image in which text is written and a redirect link to phishing website is attached to it.

Most phishing detection techniques in order to detect phishing emails or phishing websites use the text present in the email or website detect phishing emails or website and completely ignores the images present in the email. This gives an advantage to phishers by sending a phishing email which only contains an image to which a redirect link is attached. The currently used techniques fail to detect these emails if they are zero-day attack as current implemented model uses a blacklist to detect phished emails or websites[24].

1.4 Organization of thesis :

From now onwards, this thesis is organized as follows: Chapter 2 discusses background of Phishing Techniques; Chapter 3 narrates various works done till date in detecting Phishing; Chapter 4 elaborates proposed algorithm and its' applicability while Chapter 5 explains and analyzes the experimental results; finally this thesis is concluded in Chapter 6 with some touch of possible future works.

Chapter 2

LITERATURE SURVEY:

2.1 Introduction:

web browsers currently in use detect phishing websites in a same way of detecting viruses. A database of black list of known phishing websites is maintain and the list in the web browser is being updated on regular interval just like a list of known virus signatures is updated on antivirus software.

2.2 Literature survey :

Some of the anti-phishing techniques is examined and a description of the techniques that have been previously used to detect phishing along with their proposed merits and demerits are in table:

Name of the Paper	**Proposed Methodology**	**Merits**	**Demerits**
Detecting Phishing Websites using Automation of Human Behavior [9](Rau et al., CPSS, 2017)	In this paper the author proposed a technique for checking the login page & if not found then it considered the website as legitimate otherwise it feeds the website fake login credential to the websites, if login is successful then it is considered to be as an phishing website or otherwise two heuristics check (Zero links in the body of source code and Common Page Redirection ratio) is apply to check whether the site is an phishing or a legitimate one.	The technique third party independent i.e it does not require external database or search engine support.	The technique checks for login page and if not found it declare it as legitimate which can protect user account hack but there are other ways. For eg. a survey in which a phisher ask for personal information from the unaware user and use it for identity theft.

Utilisation of website logo for phishing Detection[10](Chiew et al.,ICCS 2015)	In this paper the author proposed a technique which consists of two phase to detect phishing. In the first phase the phishing detector extract the logo images from the webpage then in order to detect the right image it uses a machine learning technique and based on the result in the second phase google image search is used to depict identity. Then the url of the image found in google image search is compared with the url of query website to determine whether the given website is legitimate one or the phishing.	This technique does not maintain any kind of database for whitelist and blacklist hence zeroth day phishing site detection is possible.	In this technique image is processed to compare hence it required higher computation cost compared to other techniques.
Phish Shield: A Desktop Application to Detect Phishing Web Pages through Heuristic Approach[11](Rao et., PCS 2015)	In this paper the author proposed a technique which can be implemented as a desktop application. The application takes the url of the suspected website as input and apply five heuristics features that are Use of whitelist ,Zero links in body portion of HTML , Footer links pointing to NULL (#) , Use of copyright and title content and Website identity to detect. and based on the result it decided whether the website is phishing or legitimate one.	The technique does not require any kind of database servers of whitelist or blacklist and also does not require heavy computation and also able to detect zero day phishing attacks.	The technique is implemented as an application on this machine hence it is possible for a phisher to disable by use malwares.

An efficacious method for detecting phishing web pages through target domain identification[12](Ramesh et al.,DSS 2014)	In this paper the author proposed a technique that collects the direct and indirect links of the web pages and creates two sets S1and S2 . From these two set the target domain set is created by eliminating irrelevant links. Then this new set is given as the input to the TID (Target Identification) algorithm the algo perform 3rd party DNS lookup for suspicious and target domain to detect whether the webpage is legitimate or a phishing page .	The technique is able detect zero day phishing websites and also detect pharming attacks.	The technique is not third party independent as it really on third party DNS for lookups and there is a possibility that third party DNS is also compromised with the same pharming attack in that case this technique is unable to detect.
Bait Alarm: Detecting Phishing Sites Using Similarity in Fundamental Visual Features[13] (Li et al., ICINCS 2013)	In this paper the author proposed a technique that uses the visual features for comparison as the phishing pages generally have same visual appearance of the target page. The proposed technique utilize by quantify the similarity between the Web Pages layout by considering css as the base for detecting visual similarities.	In this technique maintains a database (kind of a whitelist)of legitimate site but in this technique cascading style sheet (css) is compared hence it make possible for this technique to detect zero day phishing websites.	This technique required a database of legitimate site to compare with large number of phishing sites, hence its computation cost become higher compared to whitelist database.

Phish Guard: A Browser Plug-in for Protection from Phishing[14](Joshi et al.,IMSAA 2008)	In this paper the author proposed a technique called phish guard that feeds large number randomly generated login credentials to the suspecting site and based on the response from the server after taking these fake credential that is after taking these credential whether the login is successful or not it is decided the website is legitimate or not and if not then a warning message is generated.	This technique lightweight as random string generation cost is considerably low and also it does not requried any kind of database.	This technique feeds large number randomly generated login credentials because of that website server might consider user as bot and may temporarily block services to that user
CANTINA: A Content Based Approach to Detecting Phishing Web Site”[15] (Y.Zhang et ai.,WWW 2007)	In this approach the authors proposed a technique which uses TF-IDF information retrieval algorithm. According to this approach on a given webpage TF-IDF is applied to calculate score and based on the score a lexical signature is generated of five items then these five items is feed to the google search engine and if the domain name of the current matches with any of the top n search then it is considered as legitimate one else illegitimate one.	This approach does not require any kind of Data based to be maintained for phishing websites.	The major drawback of this antiphishing technique is that the web sites which are newly launched can have low ranking on search engines due to this they are classified as phishing websites as they appear lower in results.
Client-side defense against web-based identity theft[16](N.Chou et al., NDSS 2004)	In this approach a toolbar (spoof guard) applies a series of heuristics to identify phishing pages the toolbar first checks the current domain name and checks the website that has been currently visited to catch fraudulent website .If two identical images are spotted on different website there is a	Unlike the other toolbars based approaches Spoof Guard ,does not uses whitelists or blacklists.	The password tracking feature generates interrupt when user tries to use the same username and password for more

	chance that a fraudulent site has copied the image from legitimate site.		than one site .
A phish detector using lightweight search features[17](G. Varshney et al., ECS 2016)	This approach is designed to work on client side in the browser as extension. When user visits any website its URL is copied and the domain name from the URL is extracted and search on google via google web search API in background then the top n search results are extracted and compared with the URL currently visited by user. If there is a match then it is a legitimate page else illegitimate page.	This technique is resource effective in terms of computational and communication cost and also gives a pop up alert, that displays the actual domain name of the visited page and a text message about its authenticity.	The drawback of the this technique is that the web sites which are newly launched can have low ranking on search engines due to this they are classified as phishing websites.
Fighting Phishing Attacks: A Lightweight Trust Architecture for Detecting Spoofed Emails [18](B.Adida et al., DIMACS 2005)	In this approach the authors proposed a key distribution architecture and identity based on digital signature for making email trustworthy and detecting spam mails by detecting email spoofing	The technique is lightweight but require a pre-establish public key infrastructure and cooperation between email domains is also required. All legitimate uses of email remain fully functional after the changes required by the scheme.	Real time implementation requires considerable changes in the email service on provider's side.

Detection and Prevention of Phishing Attack Using Dynamic Watermarking [19](A.P.Singh et al., ITMC 2011)	In this approach the authors proposed a technique which asks user is additional enter some extra information at the time of registration on watermark image by fixing points on image ,these fixing position is used as secret key. These credentials for a particular user can be changed every time user login. During each login phase a user will verifies authentic watermark with its position and decide the authenticity of website.	In this technique no external device is needed. it is easy to implement and cheaper compare to other technique.	During each time user logout he/she is aks to re entering new position in the image which is annoying and it is also possible phisher can get secret key through phishing.

Chapter 3

PROPOSED METHOD:

3.1 Proposed method:

In the proposed solution, we extract the information from the image, and also any plain text or HTML text (if any present in the email) and based on that information we check the authenticity of the email as described in the algorithm below.

3.1.1 Algorithm:

Step 1 Retrieving key inputs:

To retrieve the input (text data) from the image we are using Optical character recognition (also known as optical character reader or OCR). They are the various technique for retrieving the text data from an image. Some of them are matrix matching, fuzzy logic, feature extraction, structural analysis and neural networks[25]. The technique used for this research is OCR Tesseract which is based on feature extraction method.[26] we also retrieve any other text (if any) in form of plain text or in HTML text present in email and store that text data in a file.

Step 2 Eliminating contemporary English words:

By removing the most common words of English like bigrams, trigrams etc. from the file generated in the previous step help us to generate the smaller string with an important keyword like account, organization name which is been targeted etc. To do so we are maintaining a dictionary of most common words of English language.

Step 3 Make a string of "n" most frequent words:

From the file generated in first step we count the occurrence of each word by using a hash table and then create a string of "n" most frequent words as it is found heuristically that the important keywords used in phishing emails of like loan, account, credit card debit card organization name ect. are used more frequent.[15] In this research the value used is for n is 15 found heuristically.[15]

Step 4 Feeding searching:

The string generated in the previous step is feed to search engine and top “N” results (links) been retrieved and stored in a new file. The search engine used for this research is google search engine and the value of N is 30, which is been find heuristically.[15][17]

Step 5 Determining the final link from the redirect link:

From the redirect link attached to the image we get the final link of the website by using a python library urllib.request.

Step 6 Determining website is phishing or legitimate :

From the file generated in step 4, the links are matched one by one, with the final link found in the previous step. If we found a match then the email and the website is considered as the legitimate, otherwise phishing, which is found heuristically that the search engines place popular sites on top of their results[17] and as phishing site are up for only a few hours to few days they can't be on the top results, like the legitimate websites.

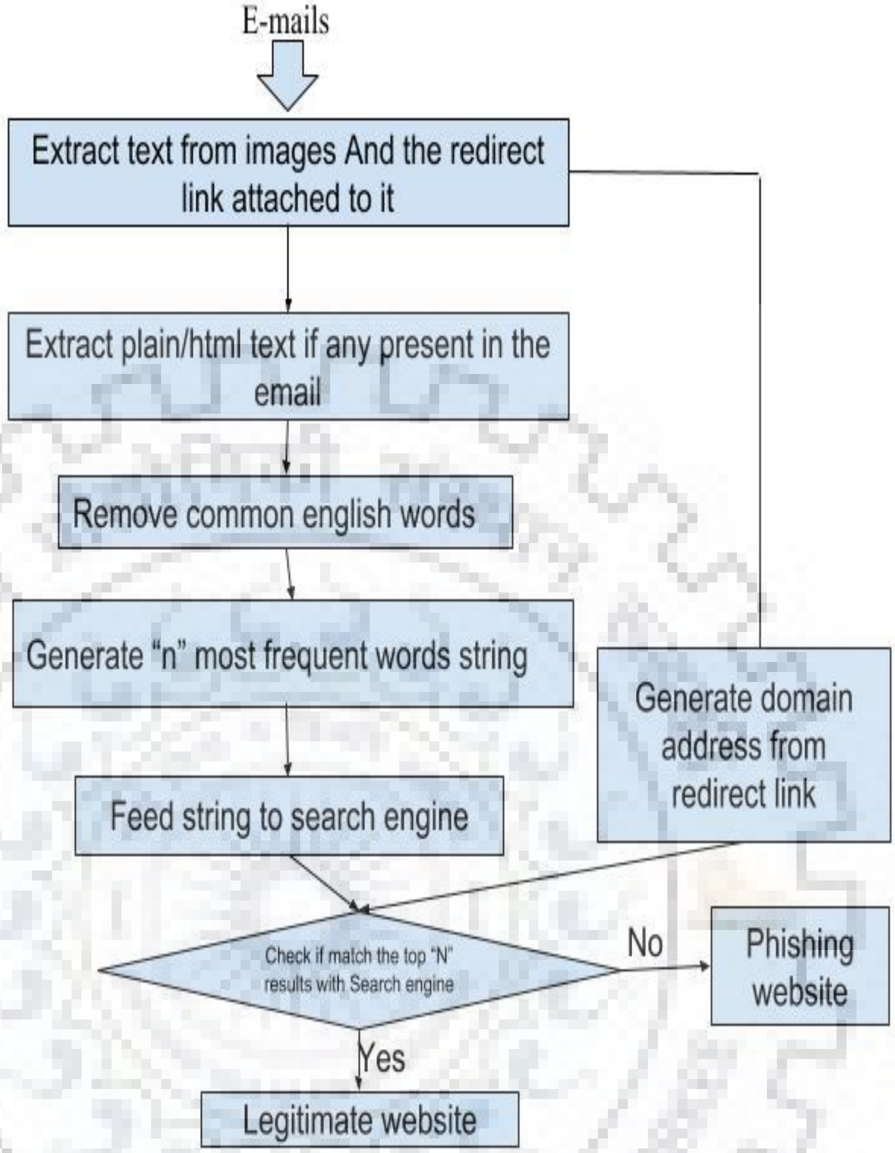

E-mails
Extract text from images And the redirect link attached to it
Extract plain/html text if any present in the email
Remove common english words
Generate "n" most frequent words string
Feed string to search engine
Generate domain address from redirect link
Check if match the top "N" results with Search engine
No
Phishing website
Yes
Legitimate website

Fig 3.1 Flow chart of proposed technique

CHAPTER 4

EXPERIMENTS AND RESULTS

4.1 Introduction:

Conducting experiments on phishing website is sometimes dangerous as these websites contain vulnerable code scripts which can install malware on systems without the knowledge of the user and difficult because of their short-lived nature these websites. For this recherche phishtank was chosen as the source URL of phishing sites and emails. In order to be sure of the data experimenting on one can manually the site by accessing the .eml through any .eml readers like Thunderbird, Outlook express etc.

4.2 Evaluation Criteria:

The evaluation criteria of the model comprising detection accuracy and overall accuray Table 4.1 provides the confusion matrix that explains True-Negative (TN), False-Negative (FN), False-Positive (FP) and True-Positive (TP).

	Legitimate	Illegitimate
Legitimate	TN	FP
Illegitimate	FN	TP

Table 4.1 Confusion Matrix

Overall Accuracy:

$$\text{Overall Accuracy} = \frac{TP+TN}{TP + TN + FP + FN}$$

In order to evaluate the performance of the proposed algorithm the following metrics are used for evaluation:

4.2.1 False positive measure:

This measure refers to those phished emails that are been falsely accepted by the algorithm. False positive is basically the ratio of number of suspected phishing emails which are accepted as legitimate to the total number of emails which are used for evaluation.

$$\text{False positive percent} = \frac{\text{no. of phishing emails accepted as legitimate}}{\text{total number of emails are used}} *100$$

4.2.2 False negative measure:

This measure refers to those legitimate emails that are been falsely rejected by the algorithm. False negative percent is calculated by the following formula.

$$\text{False negative percent} = \frac{\text{no.of legitimate emails detected as phished}}{\text{total number of emails are used}} *100$$

4.2.3 Acceptance rate:

Acceptance rate measures the performance of the proposed solution in terms of emails accepted identify correctly to the total number of emails which are used for for the experiment. Acceptance rate is calculated as follows.

$$\text{Acceptance rate} = \frac{\text{no.of emails identified correctly}}{\text{total number of emails are used}} *100$$

4.3 Experiments and Results:

The purpose of this experiment was to determine what percentage of phishing and legitimate website would be detected using phishing emails by the proposed technique alone. The number of emails in the dataset is 3521 of which 1417 are phishing emails and 2104 are legitimate emails. The overall accuracy is 72.22%, false positive measure is 10.28% , false negative measure is 17.9%, acceptance rate is 72.22 %and the performance and false error obtain are presented below.

	Numbers	Percent %
Correctly identified mails	2543	72.22%
Incorrectly identified mails	978	27.77%

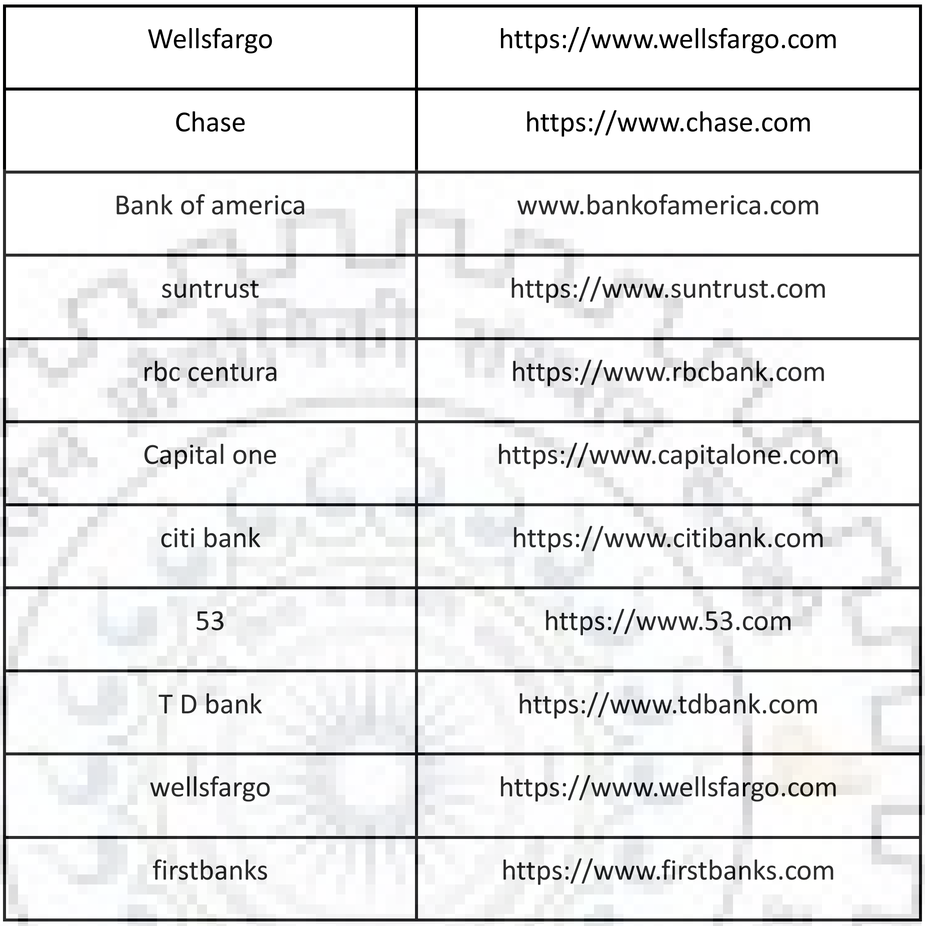
Table 4.2 Performance of proposed technique

	Identified as legitimate	Identified as Phishing
Legitimate	1488	616
Phishing	362	1055

Table 4.3 Confusion matrix of proposed technique

4.4 Examining the false positives and false negatives:

The next step we manually examine the list of legitimate sites that are detected as phishing and also the list of phishing sites that get pass through the detection. In the manual verification, it is found that the most common target of phishers is financial institutions. The list of some banks names and their URL’s are found in the dataset (table 4.4) and also most of the phishing websites that detected as legitimate are hosted on a free website hosting websites some of the free website hosting websites found in the dataset are (table 4.5) below examining the legitimate site.



Wellsfargo	https://www.wellsfargo.com
Chase	https://www.chase.com
Bank of america	www.bankofamerica.com
suntrust	https://www.suntrust.com
rbc centura	https://www.rbcbank.com
Capital one	https://www.capitalone.com
citi bank	https://www.citibank.com
53	https://www.53.com
T D bank	https://www.tdbank.com
wellsfargo	https://www.wellsfargo.com
firstbanks	https://www.firstbanks.com

Table 4.4 list of bank name and there urls found in dataset

Altervista	https://en.altervista.org
ripway.com	ripway.com.websiteoutlook.com
lycos	www.tripod.lycos.com
freewebs	www.freewebs.com

Table 4.5 free web hosting sites and there urls detected in datasets

By examining the dataset we can heuristically determine that if we maintain a blacklist for free website hosting websites and whitelist for most commonly targeted financial institution. and if the name of any financial institution come in our string of most frequent words the we verify the start of the final URL address of redirect link attached to the image with the respective URL of that financial institution in our list. If it matches to whitelist we can identify it as a legitimate one, otherwise illegitimate as it's fair to assume that all the links of that financial institution is hosted with same website name assumed, and if it matches with the blacklist it can be considered as illegitimate one as it is also fair to assumed that a financial institution never hosts its website of a free web hosting website.

4.5 Modified Algorithm:

Step 1 Retrieve key inputs.

Step 2 Determining the final link from the redirect link of image

Step .3 Check for whitelist keyword present in the retrieved input.

Step 3.1 If yes then match the start of final link found in step 2 with the link associated with that keyword.

Step 3.1.1 If match found declare it as legitimate

Step.4 Trim the link found in step 2 for words by considering special characters as start and end of word.

Step 5 Check the words found in step 4 in blacklist. If found declare it as illegitimate.

Step 6 Eliminating contemporary English words.

Step 7 Make a string of "n" most frequent words.

Step 8 Search engine feeding.

Step 9 Determining website is phishing or legitimate.

4.6 Experiments Results after modification:

The number of emails in the dataset is 3521 of which 1417 are phishing emails and 2104 are legitimate emails. The overall accuracy of the system is 79.97% false positive measure is 9.96% , false negative measure is 10.05%, acceptance rate is 79.97% and the performance and false error obtain are presented below.

	Numbers	Percent %
Correctly identified mails	2816	79.97%
Incorrectly identified mails	705	20.02%

Table 4.6 Performance of modified proposed technique

	identified as legitimate	identified as illegitimate
Legitimate	1750	354
Illegitimate	351	1066

Table 4.7 Confusion matrix of modified proposed technique

Chapter 5

CONCLUSION AND FUTURE WORK:

5.1 Conclusion:

Phishing detection methods are rapidly changing to keep up with new techniques used by phishers. Combating phishing is an ongoing battle that will probably never end much like the ongoing battle with spam emails.

Because of the broad nature of the subject, this researcher provides only one implementation method of phishing detection which is a through email which is the most common way phishing by running a small software just like an antivirus for checking emails. There is still much work to be done in both the detection algorithms and the implementation of the software and requires more research, features, and testing before becoming a commercially acceptable product.

5.2 Future work:

By maintaining few blacklist and whitelist for the free web hosting site(as they are the most common place for hosting phishing website) and financial institutions (as they are common phishing targets), the accuracy of the software can be increased.

REFERENCE:

[1] Phishing 2017, *History of Phishing*, phishing org, n.d, viewed 9 may 2017 <http://www.phishing.org/history-of-phishing>

[2] Wikipedia 2017 *Phishing* Wikimedia Foundation, Inc. n.d viewed 9 may 2017<https://en.wikipedia.org/wiki/Phishing>

[3] Abad, C. "*The economy of phishing: A survey of the operations of the phishing market*". First Monday volume 10, number 9 (September 2005). <http://firstmonday.org/issues/issue10_9/abad/index.html>

[4] McGrath, D. K. and Gupta, M. 2008. Behind phishing: an examination of phisher modi operandi. In "*Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats"* San Francisco, California, April 15, 2008.

[5] Microsoft 2014 *20% Indians are Victims of Phishing Attacks: Microsoft Computing Safety Index* Microsoft Computing Safety Index (MCSI) viewed 21 may 2017 <https://news.microsoft.com/en-in/20-indians-are-victims-of-phishing-attacks-microsoft-computing-safety-index/>

[6] I.Fette, N.Sadeh, A.Tomasic, "Learning to Detect Phishing Emails",16th international conference on World Wide Web Pages 649-656,May 2007 .

[7] Royal Bank of Scotland 2018, *Scam E-mails*, Royal Bank of Scotland Plc,n.d viewed 8 may 2017 <https://www.rbs.co.uk/microsites/global/phishing_demo/index.htm>

[8] Anti-Phishing Working Group 2017, *Phishing Attack Trends Report 4th Quarter 2016,* viewed 23 may 2017 <http://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf>

[9] R.S Rou, A.R Pais ,"Detecting Phishing Websites using Automation of Human Behavior " Cyber Physical System Security'17 , page no 33-44,april 2017

[10] K. L. Chiew, E. H. Chang, W. K. Tiong, et al. "Utilisation of website logo for phishing detection." Computers & Security, 54:16-26, 2015.

[11] R. S. Rao,S. T. Ali. Phishshield: "A desktop application to detect phishing web pages through heuristic approach". Procedia Computer Science, 54:147-15 6, 2015.

[12] G. Ramesh, I. Krishnamurthi, K. S. S. Kumar." An efficacious method for detecting phishing webpages through target domain identification"n. Decision Support Systems, 61:12 -22, 2014.

[13] J. Mao, P. Li, K. Li, T. Wei , Z. Liang, "Baitalarm: Detecting Phishing Sites using Similarity in Fundamental Visual Features",*5th International Conference on Intelligent Networking and Collaborative Systems, 2013, IEEE,* pp. 790–795, September 2013.

[14] Y. Joshi, S. Saklikar, D. Das and S. Saha, "PhishGuard A Browser Plug-In for Protection from Phishing",2nd International Conference on Internet Multimedia Services Architecture *and Applications, 2008, IEEE,* pp. 1–6, December 2008.

[15] Y.Zhang, J.Hong, L.Cranor, "CANTINA: A Content-Based Approach to Detecting Phishing Web Site", 16th international conference on World Wide Web pp 639-648 May 2007.

[16] N.Chou,R.Ledesma,Y.Teraguchi,D.BonehJ.C.Mitchell,"Client-side defense against web-based identity theft"NDSS Symposium 2004.

[17] G.Varshney,M.Misra,A.Pradeep, "A phish detector using lightweight search features", Elsevier Computer and Security, Vol-62,PP-213-228,September 2016

[18] B.Adida,S.Hohenberger,R.L.Rivest, "Fighting Phishing Attacks: A Lightweight Trust Architecture for Detecting Spoofed Emails",DIMACS Workshop on Theft in E-Commerce, 2005.

[19] A.P.Singh,V.Kumar,S.S.Sengar,M.Wairiya "Detection and Prevention of Phishing Attack Using Dynamic Watermarking",Information Technology and Mobile Communication, pp 132-137,January 2011.

[20] Anti-Phishing Working Group 2017,*Phishing Attack Trends Report 1st half 2017,* viewed 9 nov 2017< http://docs.apwg.org/reports/apwg_trends_report_h1_2017.pdf>

[21] Wikipedia 2017, *WannaCry Ransomware Attack, Wikimedia Foundation, Inc.* 9 Nov 2017, viewed 10 Nov 2017 <https:// en.wikipedia.org/ wiki/WannaCry _ransomware _attack>

[22] Microsoft 2017, *Microsoft Security Bulletin MS17-010 - Critical,* Microsoft, n.d, viewed 23 may 2017, <https://technet.microsoft.com/en-us/library/security/ms17-010.aspx>

[23] J.Milletary *“Technical Trends in Phishing Attacks*” Software Engineering Institute, May 2005

[24] Mozilla 2018, *How does built-in Phishing and Malware Protection work* Mozilla, n.d, viewed 23 march 2018 <https://support.mozilla.org/en-US/kb/how-does-phishing-and-malware-protection-work>

[25] N . V Rao, A.S.C.S. Sastry,*”Optical Character Recognition Technique Algorithm”* Journal of Theoretical and Applied Information Technology, Vol.83. No.2, January 2016.

[26] R.Smith *“An Overview of the Tesseract OCR Engine”* IEEE Ninth International Conference on Document Analysis and Recognition, DOI: 10.1109/ICDAR.2007.4376991, 05 November 2007.