The Survey Analysis Of Phishing Attack Methods

Oniyide Sakiru Adelakun
Computer Department
University of Ilorin
Ilorin, Nigeria

Awotunde Joseph Bamidele
Computer Department
University of Ilorin
Ilorin, Nigeria

Fatai olawale Waheed
Computer Department
University of Ilorin
Ilorin, Nigeria

***Abstract*--Phishing involves the act of tricking individuals into divulging their sensitive financial information to unauthorised user and the need to survey and analyse different methods that has used is paramount so as to help the incoming researchers. This will help the researchers since some literatures has been reviewed and this will motivate them to work under phishing since the methods has been analysed. The analysis is based on the different approaches, the result and conclusions of each work.**

Keywords:Phishing attacks, Hyperlink, LinkGuard algorithm, echololation,loudness, frequency.

I. INTRODUCTION

Recent developments in information technology have led to a renewed interest in internet security. Information technology has seen as a key factor in the economy development of a nation. The threat of technology-based security attacks is well understood and Information Technology (IT) organizations have tools and processes in place to manage this risk to sensitive corporate and personal data.

Whilea variety of definitions of the term phishing have been suggested, the definitionsuggested by [18] define Phishing as a form of social engineering in which an attacker, also known as a phisher, attempts to fraudulently retrieve legitimate users' confidential or sensitive credentials by mimicking electronic communications from a trustworthy or public organization in an automated fashion to “fish" for passwords and financial information from the sea of Internet users.

The impact of phishing on the global economy has been quite significant: RSA estimates that worldwide losses from phishing attacks cost more than $1.5 billion in 2012, and had the potential to reach over $2 billion if the average uptime of phishing attacks had remained the same as 2011[21].

The academic work on phishing has been diverse, with a useful starting point being the book by Jacobson [18]. Researchers have tried to understand the psychology of the process [3], how to block the spam email containing the initial enticement [6], and how server operators might automatically detect fraudulent sites [14].

Phishing attacks affect millions of internet users and are a huge cost burden for businesses and victims of phishing [2]. Gartner research conducted in April 2004 found that information given to spoofed websites resulted in direct losses for U.S. banks and credit card issuers to the amount of $1.2 billion [15].

According to the Russell Kay [5], up to 20% of unsuspecting recipients may respond to them, resulting in financial losses, identity theft and other fraudulent activity against them. Financial institutions are at risk for large numbers of fraudulent transactions using the stolen information [12].

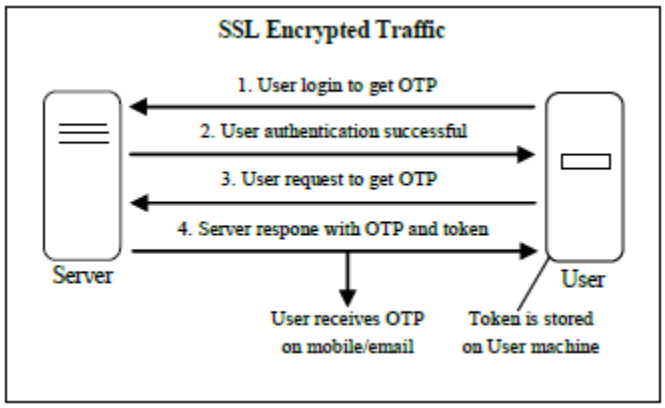
II. MATERIALS AND METHODS

The material used is the previous work that has been done on phishing attack, their methods used as well as the output so that it can serve as a guide for the incoming researchers on the same area. Each method was well analysed and the result was

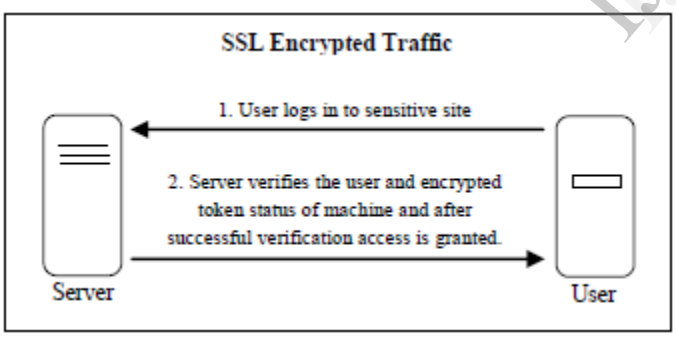
There are many research works to characterise or model phishing attack [22]. In many cases, phishers attract users by e-mail spoofing [8]. The acquisition trick which employs some phishing sites is called web-spoofing [9]. Web-spoofing can be categorised into three: downloading, cross siting scripture (XSS) and deceit. Downloading attracts the user to download and install free software out of which may be crime-ware. The phisher can steal user’s financial information through this software. XSS exploits the vulnerabilities of a legitimate site to forward personal information to a phishing site.

Ahmad Alamgir Khan (2013) presented a novel approach for Preventing Phishing Attacks using One Time Password and User Machine Identification. This system called Anti-Phishing Prevention Technique (APPT) is based on the concept of preventing phishing attacks by using combination of one time

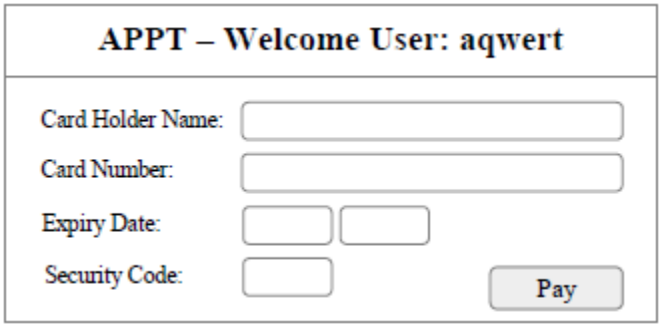
random password and encrypted token for user machine identification. The first step is to retrieve the password by SMS or by alternate emails, during that process encrypted token is created which have user specific data and is stored in the user machine. Second step is to access the required website with the password and valid token which are required for successful authentication. The diagram below illustrates how it works



The user will go to one time password (OTP) retrieval site to receive the random password which they can receive through SMS or e-mail, after authenticating the OTP, token will also require to gain access to the website. Any disparity between the two will lead to access not granted. The token at each log in is different from the other so as to check the fraudsters. The figure below demonstrate how the user and machine authentication is performed



After the login is successful, the financial transaction can then made in the format below



He concluded that by using APPT it can be assured that attack like Phishing can be prevented to a large extent. However, future work has to be done to provide more secure encryption technology that will be difficult to break. Users also have to be made aware of the risks they face on internet, and they should responsibly use the internet for their benefit

M.Madhuri, K.Yesewini and U. VidyaSagar (2013) designed an Intelligent Phishing Website Detection and Prevention System by using Link Guard Algorithm. They proposed a new end-host based anti-phishing algorithm by utilizing the generic characteristics of the hyperlinks in phishing attacks. These characteristics are derived by analyzing the phishing data archive provided by the Anti-Phishing Working Group (APWG). LinkGuard algorithm works by analyzing the differences between the visual link and the actual link (it is based on the characteristics of phishing hyperlinks and has a verified very low false negative rate). It uses the string pattern matching by classifying the hyperlink in the previous attack to determine new ones. The algorithm used were given below:

v_link: visual link;

a_link: actual_link;

v_dns: visual DNS name;

a_dns: actual DNS name;

sender_dns: sender'sDNS name.

intLinkGuard (v_link, a_link} {

1 v_dns = GetDNSName (v_link);

2 a_dns = GetDNSName (a_link);

3 if ((v_dns and a_dns are not

4 empty) and (v_dns! = a_dns))

5 return PHISHING;

6 if (a_dns is dotted decimal)

7 return POSSIBLE_PHISHING;

8 if (a_link or v_link is encoded)

9 {

10 v_link2 = decode (v_link);

11 a_link2 = decode (a_link);

12 return LinkGuard (v_link2, a_link2);

13}

14 / analyze the domain name for*

*15 possible phishing */*

16 if (v_dns is NULL)

17 return AnalyzeDNS (a_link);

```
}
intAnalyzeDNS (actual_link) {
/* Analyze the actual DNS name according
to the blacklist and whitelist*/
18 if (actual_dns in blacklist)
19 return PHISHING;
20 if (actual_dns in whitelist)
21 return NOTPHISHING;
22 return PatternMatching(actual_link);
}
intPatternMatching (actual_link){
23 if (sender_dns and actual_dns are different)
24 return POSSIBLE_PHISHING;
25 for (each item prev_dns in seed_set)
26 {
27 bv = Similarity(prev_dns, actual_link);
28 if (bv == true)
29 return POSSIBLE_PHISHING;
30 }
31 return NO_PHISHING;
}
float Similarity (str, actual_link) {
32 if (str is part of actual_link)
33 return true;
34 intmaxlen = the maximum string
35 lengths of str and actual_dns;
36 intminchange = the minimum number of
37 changes needed to transform str
38 to actual_dns (or vice verse);
39 if (thresh<(maxlen-minchange)/maxlen<1)
40 return true
41 return false;
}
```

They concluded that LinkGuard is effective, light-weighted and can detect up to 96% unknown phishing attacks in real time. It can also be useful to shield users from malicious or unsolicited links in web pages and instant messages. Its limitation is that it only implemented on Window XP operating system.

Joshua S. White, Jeanna N. Matthews and John L. Stacy (2012) designed a Method for Automated Detection of Phishing Websites through both Site Characteristics and Image Analysis. The method relies on real-time gathering and analysis of URLs posted on social media sites. The pages pointed to by each URL were characterized with a set of easily computed values such as page title text and number of links, images, forms, iframes and metatags. They take the screen shot of the image using cutycapt and computed the hash function of the resulting image. They compared the images on the websites by calculating the Hamming distance (Aggarwal et al, 1999) between their hash values. Hamming distance detects the similarity of two equal strings that the exact same length. The results demonstrated the feasibility of these techniques by comparing legitimate sites to known fraudulent versions from Phishtank.com, by actively introducing a series of minor changes to a phishing toolkit captured in a local honeypot and by performing the same process on a set of over 2.8 million URLs posted to Twitter over 4 days in August 2011. The representation of the method used was given below

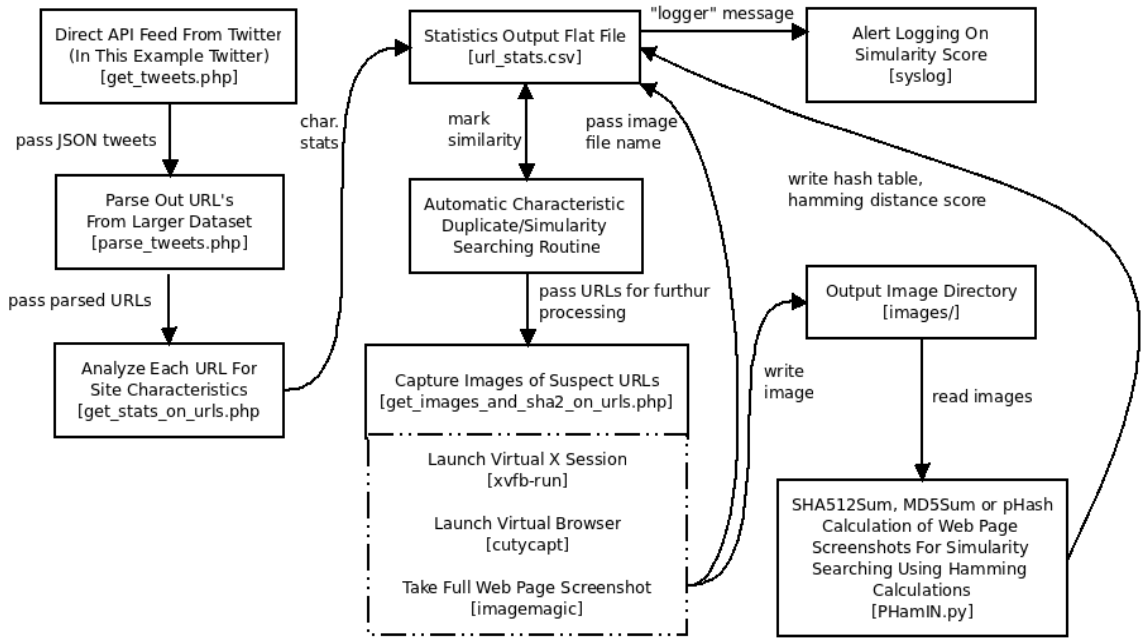


Fig 1: Phishing detection process overview

The result of their research using the model above weregiven in the table below

Table 1: Known Phishing/Non-Phishing Site Characteristics

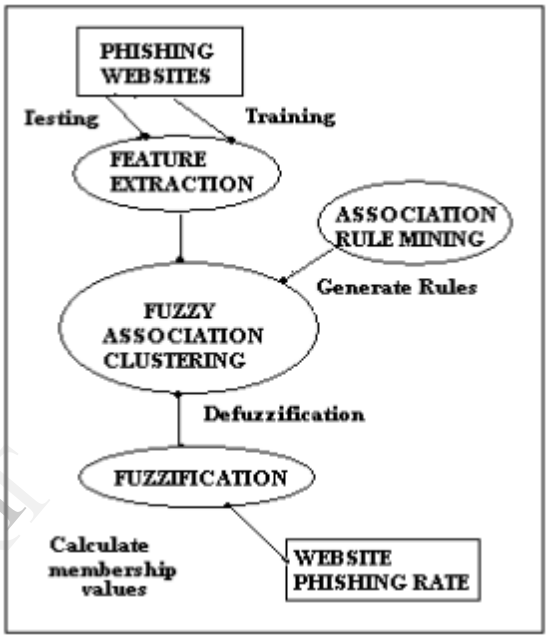
File Name	File pHash	File MD5Sum	File SHA512Sum	Hamming Score (pHash)	Hamming Score (MD5Sum)	Hamming Score (SHA512Sum)
clarkson-paypal-page-phishing-site-original.png	18446710817257652609	8b0914d7d2544f97b6e7a36adad92da1	95552b3e553530f5f00e476d1a64316d0f08fcc6a4d415fa77aafcd9de248afe9613363056acdf0aa40db4a3f5b19ac0eb2cce5b3a808d794fb575aad000057c	0	0	0
clarkson-paypal-page-phishing-site-rev2.png	18446710817257652609	d2b931f6b29bf7dcb99601f7a7c7d12b	4d39160800e9250ded72108cdb14a73cd015f7a30d2d9e01de8836ae5073ee179cd2800cd32ea9e190ffebe1dacd11ec364e26dd5147faa77eb4a12191c67b69	0	8	38
clarkson-paypal-page-phishing-site-rev3.png	18446710817257652609	b9858446a3f38dd9e9116f545f31500f	c9d09715010886682aec120a29e2135b0b3ea49474b9ff99504026406ab7ac7ab73c3f9c570ee274d014257e5ad95f5bf5d9cb18585f45c080da400a4daff632	0	8	9
clarkson-paypal-page-phishing-site-rev4.png	18446728408906826113	94b551da15320573cd51b3ee7dcbb7fa	8029e65c7294f6c8c317b8d136ab399f04b318782c885dd16ab74b20843eec0401b99b68154fbaf5798844408414db255f7154e0569590286a711c32137fc094	2	9	8
clarkson-paypal-page-phishing-site-rev5.png	18446710816720798145	df78813010280c2fde8f5e4270b928ff	55e15dae72375e9e537aabccad76d04f85a13465846b750f85a2ae5b9d95b78c255182f57a52a51d2aab2b736b0179e59d4b2a601e291d43ff73e16b29e963b8	3	10	24
ezine-screenshot-same-characteristics-as-paypal.png	15835713822348909969	8e434d107b43941c39ac4bda2c806fbb	f3c52d2d4fe608cc6f500d661260204a9c4fff9a06c24ad39513d67dacff3ed08f5a4bdf93997f00dc7b3f4d3775ee81cbed8eba961219fcaf1c5ec67b338886	19	1	9

They concluded that this approach can compare known phishing sites and their legitimate counterparts. This method has been able to find the a number of phishing sites from a live dataset and also gained critical insight into how to effectively and efficiently gather format and analyse this social data. The future work can be done on how to use o routines that automatically find correlations between potential phishing pages and known trusted sites. A cluster analysis can also be used to identify clusters of pages with similar characteristics.

MallikkaRajalingam, Saleh Ali Alomari and Putra Sumari (2012), developed a system on Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages. They present an effective image-based anti-phishing scheme based on discriminative key point features in WebPages. They use an invariant content descriptor, the Contrast Context Histogram (CCH), to compute the similarity degree between suspicious pages and authentic pages. The method used is in three phases namely:
1) Web page snapshot
2) Image wizard
3) Comparison of web pages.

The first phase was to take the snapshot of the authentic pageand save in the file system in the required image format. It should be saved only in any of the image format because the web page has to be compared with the actual image. The second phase is to create a wizard to compare the images.s. This Wizard is designed in such a way that everything appears in the wizard is clear and systematic. Separators are used to clearly distinguish each one. The last one is the comparison, the image wizard module user can give the ratio of the accuracy they need while comparing images [11]. The image was then converted to its equivalent Red, Green and Blue image format. The mean average and standard deviation of each was then determined. The absolute value of the two standard deviations was used to determine the authenticity of that message. If the absolute value is zero, it shows that the message is from a reliable source, else it is a phish message. They concluded that developing an image based comparison method which compares the images based on the color values give an accurate result and only the company which created that website knows about the color range of the images present in the web page. None can design a fake web page similar to the original page with that same color range. They suggested that in future, one can develop a fully automated crawling framework by using attribute-based phishing attacks that developed for testing, along with main experimental results.

RadhaDamodaram and M.L.Valarmathi(2012) presentedPhishing website detection and optimization using Modified bat algorithm (MBAT). Bat Algorithm is an intelligent resilient and effective model that is based on using association and classification Data Mining algorithms. These algorithms were used to characterize and identify all the factors and rules in order to classify the website using the relationship that correlate them with each other also compared their performances, accuracy, number of rules generated and speed. Even though the rules generated from the associative classification model showed the relationship between some important characteristics like URL and Domain Identity, and Security and Encryption criteria in the final phishing detection rate, there is no optimal solution. The MBAT is a metaheuristic algorithm to get an optimal solution for the search of fake websites. The MBAT algorithm optimizes a problem by iteratively trying to improve a solution with regard to a given measure of quality. The Modified Bat Algorithm is based on the echolocation behaviour of micro-bats with varying pulse rates of emission and loudness with Doppler Effect. All bats use echolocation to sense distance, and they also determine the difference between food/prey and background barriers in some magical way. Bats fly randomly with velocity vi at position xi with a fixed frequency fmin, varying wavelength λ and loudness A0 to search for prey. Doppler Effect is the change in frequency of a wave for an observer moving relative to the source of the wave. The received frequency is higher (compared to the emitted frequency) during the approach, it is identical at the instant of passing by, and it is lower during the recession. The workflow and the algorithm used were stated below



Association and Classification Rule
Input: Webpage URL
Output: Phishing website identification
Step 1: Read web phishing URL
Step 2: Extract all 27 feature
Step 3: For each feature, Assign fuzzy membership degree value and Create fuzzy data set
Step 4: Apply association rule mining & generate classification rule
Step 5: Aggregate all rule above minimum confidence.
Step 6: Defuzzification of fuzzy values into original values.
Step 7: Apply rule on test data and find whether the site is phishing or not and these steps are shown in Fig.2

MBATAlgorithm
Objective function f(**x), x**=(**x1….xd**)T
Initialize the bat population xi = 1,2…n) and **Vi** Define Pulse frequency fi at**xi**
Initialize the rates **ri**and the loudness **Ai**
While (t < Max number of iterations)
Generate new solutions by adjusting frequency, Apply equation (1)
And updating velocities and locations /solutions [Equations (2) and (4)]
If (rand >**ri**)
Select a solution among the best solutions

Generate a local solution around the selected best solution End if
Generate a new solution by flying randomly If (rand <**Ai** &**f(xi) < f(x*)**)
Accept the new solutions
Increaseri and reduce **Ai** end if
Rank the bats and find the current best **x***
end while
where,

$$f = f_{min} + (f_{max} - f_{min})\beta, \quad \text{.....(1)}$$

$$v_i^t = v_i^{t-1} + (x_i^t - x^*)f_i, \quad \text{.....(2)}$$

$$x_i^t = x_i^{t-1} + v_i^t, \quad \text{.....(3)}$$

Where C is the velocity of waves in the medium, V_r is the velocity of the receiver relative to the medium; positive if the receiver is moving towards the source, V_s is the velocity of the source relative to the medium; positive if the source is moving away from the receiver.**β**[0, 1] is a random vector drawn from a uniform distribution. Here x***** is the current global best location (solution) which is located after comparing all the solutions 4 among all the n bats.

These characteristics were tested using the same association and classification algorithm of previous bat algorithm. They concluded that using MBAT, it is more accurate in combating phishing activities than ordinary Bat Algorithms. Also, the time consuming is lesser and the error rate is minimal than the previous ones.

Ch.sonika and D.RaagaVamsi(2012) proposed a system on Adaptive Classifier and Associative Algorithms for phishing detection. They presented a novel approach to overcome the challenge and complexity in detecting and predicting offline phishing data. They proposed an intelligent effective model that really based on using improved classification like improvedC4.5, PRISM, PART and association mining algorithms MCAR. This strategy uses different classification algorithm and techniques to extract the phishing training dataset to sort out their legitimacy. The proposed framework was given below

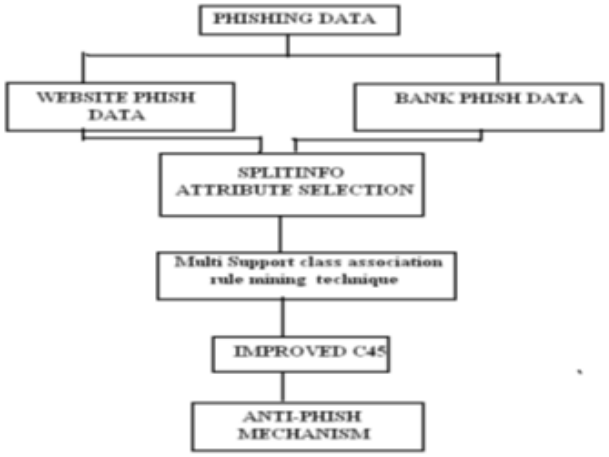


Fig 3: the proposed system

The phishtank from the phishtank.com were used to test the implementation of this system. Phishtank.com was considered to be the primary phishing report collections which consist of the time of report and further details about the screenshots of the webpages. Data mining algorithms require an offline training phase, but the testing phase requires much less time

PART algorithm is based on account that it combines both approaches to generate a set of rules. PRISM is naturally a classification rule that may only trot out nominal attributes and doesn't do any pruning. MCAR algorithm involves two phases: rules generation and a classifier builder. In the initial phase, MCAR scans the training data set to discover frequent single items, after which recursively combines their products generated to produce items involving more attributes. MCAR then generates ranks and stores the rules. Supplied in the second phase, the foundations are utilized to build a classifier by considering their effectiveness on the training data set. They also compared their performances, accuracy, range of rules generated. They concluded that this research work mainly identifies several new and generic features for identifying phishing URLs. Proposed Improvedc45 classifier and MCAR achieves a very high accuracy. One of the major contributions of this work is a large scale measurement study conducted on phishtank web urls.the future work future work could investigate how well it can be adapted to perform online phishing web classification.

PonnurangamKumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong (2007) carry out research on Lessons from a Real World Evaluation of Anti-Phishing Training. This study was conducted at a large Portuguese company and allemails and training materials were translated into Portuguese. Allparticipants in the study worked in the same floor of an officebuilding. They used PhishGuru methodology to train users aboutspear phishing and test it in a real world setting with employees of a Portuguese company. The results demonstrated that the findings of PhishGuru laboratory studies do indeed hold up in a real world deployment. Specifically, the results from the field study

showed that a large percentage of people who clicked on links in simulated emails proceeded to give some form of personal information to fake phishing websites, and that participants who received PhishGuru training were significantly less likely to fall for subsequent simulated phishing attacks one week later. They concluded thatTargeted spear phishing attacks have been more successful than generic phishing attacks in coning people and causing damages to companies and individuals.

CONCLUSION

In this work, different papers have been reviewed, the methods and their result were well analysed and this will encourage new researcher most especially on phishing attack. This will enable the researcher to know how to review paper in any area of their research which will definitely give them the focus in their respective researches.

This work covers the different methods that has been used on phishing attacks future works can be focused on another area of computer security.

REFERENCES

[1] Ahmad A.A. (2013). Preventing Phishing Attacks using One Time Password and User Machine Identification. International Journal of Computer Applications, Volume 68– No.3, pp. 7-11 April 2013

[2] Anti-Phishing Working Group (2006). Phishing Activity Trends Report. http://www.antiphishing.org/reports/apwg_report_mar_06.pdf

[3] Christine E. D., Jonathan J. O. & Eugene J. K. (2004). Anatomy of a Phishing Email. First Conference on Email and Anti-Spam (CEAS), Mountain View, CA, USA, 2-3 Aug 2004.

[4] Ch-sonika&Raaga D.V. (2012). Adaptive Classifier and Associative Algorithms for phishing Detection. ISSN: 2231-5381, Volume 3, Issue 5, pp. 610-615.

[5] Computerworld QuickStudy: Phishing By Russell Kay, http://www.computerworld.com/s/article/89096/Phishing *Accessed: 27 March 2013*

[6] Daisuke M., Hiroaki H., &Youki K. (2005). A simple altering algorithm to thwart phishing attacks. Asian Internet Engineering Conference (AINTEC), 13-15 Dec 2005.

[7] Damodaram R. &Valarmathi M.L. (2012). Phishing website detection and optimization using Modified bat algorithm. ISSN: 2248-9622Vol. 2, Issue 1, pp. 870-876

[8] Drake C.E., Oliver J.J. & Koontz, E.J. (2004). Anatomy of a Phishing Email. In: Proceedings of CEAS 2004.

[9] Felten E.W., Balfanz D., Dean D. & Wallach D.S. (1997). Web spoofing: An internet con game. In: Proceedings of NISSC 1997.

[10] Joshua S. W., Jeanna N. M. & John L. S. (2012). A Method for the Automated Detection of Phishing Websites through both Site Characteristics and Image Analysis.

[11] Kannan A., V. Mohan and N. Anbazhagan. "Image Clustering and Retrieval using Image Mining Techniques". IEEE International Conference on Computational Intelligence and Computing Research, vol.2, 2010

[12] Koprowski, Gene J., "Beware of 'Spoofing' Scams," UPI Technology News, January 2004.

[13] Kumaraguru P., Rhee Y., Acquisti A., Cranor L.F., Hong J. &Nunge E. (2007). Protecting People from Phishing: The Design and Evaluation of an Embedded Training E-mail System. In Computer/Human Interaction 2007, San Jose, USA, pp. 905-914.

[14] Liu W., Guanglin H., Liu X., Zhang M. &Xiaotie D. (2005). Detection of Phishing Webpages based on Visual Similarity. Proc. 14th International World Wide Web Conference, ACM Press, pp. 1060-1061.

[15] Litan, A. (2004). Phishing Attack Victims Likely Targets for Identity Theft. Gartner Research (2004)

[16] Madhuri M., Yesewini K. &Vidya U.S. (2013). Intelligent Phishing Website Detection and Prevention System by using Link Guard Algorithm. ISSN: 2231 – 1882, Volume 2, Issue 2, pp 9-16

[17] Mallikka R., Saleh A.A. & Putra S. (2012). Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages. International Journal of Computer Science and Security (IJCSS), Issue 1, volume 6, pp. 1-18

[18] Markus J. & Steven M. (2007). Phishing and Countermeasures: Understanding the Increasing Problems of Electronic identity Theft. Wiley and Sons Inc.

[19] Markus J. & Steven M. (2006). Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft. Wiley, Nov 2006, ISBN: 978-0-471-78245-2.

[20] Mining Fuzzy Weighted Association Rules Proceedings of the 40th Hawaii International Conference on System Sciences – 2007.

[21] RSA's January 2013 Online Fraud Report, http://brianpennington.co.uk/2013/01/30/rsas-january-online-fraud-report-2013-including-an-excellent-summary-of-phishing-in-2012/

[22] Van der Merwe A., Loock M. &Dabrowski M. (2005). Characteristics and Responsibilities Involved in a Phishing Attack. In Proceedings of ISICT 2005.