A Review of Exposure and Avoidance Techniques for Phishing Attack

Kanchan Meena
Department of Computer Science & Engineering
VNS Faculty Of Engineering, Bhopal (M.P)

Tushar Kanti
Department of Computer Science & Engineering
VNS Faculty Of Engineering, Bhopal (M.P)

ABSTRACT
Phishing is a novel kind of website/network attack which makes a deceitful attempt and influences the amenities or information security instead of stealing personal, financial and transactional data, etc. To preclude users or network from the phishing different techniques has been proposed and implemented.This paper, present the review of literature about the techniques offered by different researchers for exposing and avoiding from the phishing attack also discusses the advantages and limitation of the approaches.

Keywords
Information security, Network, Phishing, Website

1. INTRODUCTION
Now a days Internet is a broadly used technology for the different application such as online reservation, for shopping, transaction of amount online from the bank, sending email, etc. are becoming more widespread day by day but there are abundant chances to hack the data. Henceforth, security over the internet is a significant issue over the internet cyber security if the major issue influencing the security. Nevertheless, in the context of internet security, phishing is one of the internet attacks. Phishing is a considerable dilemma concerning deceitful email and web sites that trick gullible users into enlightening private information. Phishing has turned out to be more and more intricate and sophisticated attack can bypass the filter set by anti-phishing techniques. Most phishing emails aim of withdrawing money from financial institutions or getting access to private information and is a severe threat to worldwide security and economy. Phishing filters are essential and extensively used to increase communication security [1]. The Phishing scams have been receiving extensive press reporting because such attacks have been getting higher in number and superiority. A number of website service providers understand that their reputation is at venture and alarm that users will lose self-assurance in electronic commerce. Emails' pose a serious threat to electronic commerce because they are used to defraud both individuals and financial organizations on the Internet. An assessment of phishing attacks [3] shows that approximately 3.6 million clients in the US alone had lost money to phishing attacks and total losses had reached approximately US$ 3.2 billion Dollar. The number of victims increased from 2.3 million in2006 to 3.6 million in 2007, an increase of 56.5%. Among all complaints received by the Federal Trade Commission in2009 from Internet users, identity theft attributed to a phishing email ranked first. It accounted for 21% of the complaints and cost consumers over 1.7 billion US dollars. According to an E-crime trends report, phishing attacks are increasing at a rapid rate. For example, phishing in Quarter 1 (Q1) of 2011 grew by 12% over that in Quarter 1 (Q1) of 2010. Phishing emails range from very simple to very complicated messages and are capable of deceiving even the clever Internet users. Fraudulent emails can steal secret information from the victims, resulting in loss of funds. As a consequence, these attacks are damaging electronic commerce in the Internet world, resulting in the loss of trust and use of the Internet. This threat has led to the development of a large number of techniques for the detection and filtering of phishing emails.

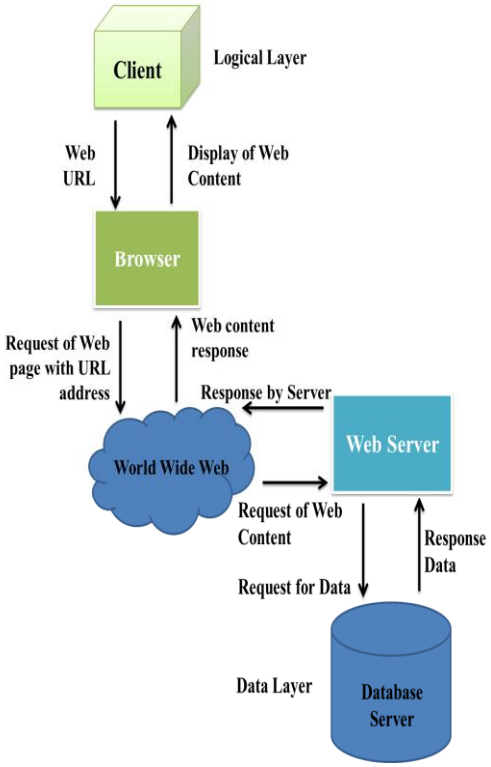

Figure 1: The Client-Server architecture over the World Wide Web[2]

The many approaches proposed in the literature to filter phishing emails, may be classified according to the different stages of the attack flow, e.g. network level protection, authentication, client side tool, user education, server side filters and classifiers, etc. this paper discuss the advantages and limitation of these approaches. An organization of the research paper is done accordingly: section II presents explanation about the literature of the approaches presented by different researchers. Section III outlines the different techniques to avoid from the phishing and last section presents conclusion of the paper.

2. RELATED WORK
This section describes the overview of techniques proposed and implemented by different researchers to prevent the website or network of the phishing attack.

A.Sarannia et al. [4] presented how to avoid the phishing scams, how it is attacked and intend a new end-user based on anti-phishing algorithm which we call “Link Guard” algorithm. Link Guard can detect not only notorious but also

unfamiliar phishing attacks. We had implemented Link Guard in windows XP. Our experiment verified that Link Guard is effective in detecting and preventing attacks.

Mahmoud Khonji et al. [5] describe a novel framework to mitigate spear phishing attacks via the use of document authorship techniques — Anti-Spear phishing Content-based Authorship Identification (ASCAI). ASCAI informs the user of possible mismatches between the writing styles of a received email body and of trusted authors by studying the email body itself (i.e. the write print), as opposed to traditional user ID-based authentication techniques which can be spoofed or abused. As a proof of concept, we implemented the proposed framework using Source Code Author Profiles (SCAP).

Maher Aburrous et al. [6] presented novel approach to overcome the 'fuzziness' in traditional website phishing risk assessment and proposed an intelligent resilient and effective model for detecting phishing websites. The proposed model is based on FL operators which is used to characterize the website phishing factors and indicators as fuzzy variables and produces six measures and criteria's of website phishing attack dimensions with a layered structure. It showed the significance and importance of the phishing website criteria (URL & Domain Identity) represented by layering one, and the varying influence of the phishing characteristic layers of the final phishing website rate.

Gaurav Kumar Tak et al. [7] proposed a knowledge base compound approach which is based on query operations and parsing techniques to counter these internet attacks using the web browser itself. In this approach we propose to analyze the web URLs before visiting the actual site, so as to provide security against web attacks mentioned above. This approach employs various parsing operations and query processing which uses many techniques to detect the phishing attacks as well as other web attacks. The aforementioned approach is completely based on operation through the browser and hence only affects the speed of browsing. This approach also includes Crawling operation to detect the URL details to further enhance the precision of detection of a compromised site. Using the proposed methodology, a new browser can easily detect the phishing attacks, SSL attacks, and other hacking attacks. With the use of this browser approach, we can easily achieve 96.94% security against phishing as well as other web based attacks.

M. Madhuri et al. [8] proposed a new end-host based anti-phishing algorithm, which we call Link Guard, 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). Because it is based on the generic characteristics of phishing attacks, Link Guard can detect not only known but also unknown phishing attacks. We have implemented Link-Guard in Windows XP. Our experiments verified that Link-Guard is effective to detect and prevent both known and unknown phishing attacks with minimal false negatives. Link-Guard successfully detects 195 out of the 203 phishing attacks. Our experiments also showed that Link-Guard is light weighted and can detect and prevent phishing attacks in real time.

M. Topkara et al. [9] proposed a novel system 'ViWiD' which is a watermarking based method and its implementation for extenuating phishing attacks. ViWiD is the trustworthiness ensures method based on visible watermarking of logo images. ViWiD executes all of the calculation on the company's web server and it does not entail installation of several devices or storage of any data such as keys or history logs on the user's machine. The watermark letter is designed to be exceptional for every user and carries a shared secret among the company and the user in order to put a stop to the 'one size fits all' attacks.

W. Y. Liu et al. [10] Proposed an effective approach to phishing Web page detection, which uses Earth Mover's Distance (EMD) to measure Web page visual similarity. The authors first convert the involved Web pages into low resolution images and then use color and coordinate features to represent the image signatures, and use EMD to calculate the signature distances of the images of the Web pages.

Arun Vishwanath et al. [11] presented an integrated information processing model of phishing susceptibility grounded in the prior research in information process and interpersonal deception. We refine and validate the model using a sample of intended victims of an actual phishing attack. The data provide strong support for the model's theoretical structure and causative sequence. Overall, the model explains close to 50% of the variance in individual phishing susceptibility. The results indicate that most phishing emails are peripherally processed and individuals make decisions based on simple cues embedded in the email. Interestingly, urgency cues in the email stimulated increased information processing, thereby short circuiting the resources available for attending to other cues that could potentially help detect the deception. Additionally, the findings suggest that habitual patterns of media use combined with high levels of email load have a strong and significant influence on individual's likelihood to be phished. Consistent with social cognitive theory, computer self-efficacy was found to significantly influence elaboration, but its influence was diminished by domain specific-knowledge.

Venkatesh Ramanathan et al. [12] proposed a novel methodology to detect phishing attacks and to discover the entity/organization that the attackers impersonate during phishing attacks. The proposed multi-stage methodology employs natural language processing and machine learning. The methodology first discovers (i) named entities, which includes names of people, organizations, and locations; and (ii) hidden topics, using (a) Conditional Random Field (CRF) and (b) Latent Dirichlet Allocation (LDA) operating on both phishing and non-phishing data. Utilizing topics and named entities as features, the next stage classifies each message as phishing or non-phishing using AdaBoost. For messages classified as phishing, the final stage discovers the impersonated entity using CRF. Experimental results show that the phishing classifier detects phishing attacks with no misclassification when the proportion of phishing emails is less than 20%. The F-measure obtained was 100%. Our approach also discovers the impersonated entity from messages that are classified as phishing, with a discovery rate of 88.1%. The automatic discovery of impersonated entity from phishing helps the legitimate organization to take down the offending phishing site. This protects their users from falling for phishing attacks, which in turn leads to satisfied customers. Automatic discovery of an impersonated entity also helps email service providers to collaborate with each other to exchange attack information and protect their customers.

Maher Aburrous et al. [13] presented novel approach to overcome the 'fuzziness' in the e-banking phishing website assessment and proposed an intelligent resilient and effective model for detecting e-banking phishing websites. The

proposed model is based on fuzzy logic combined with data mining algorithms to characterize the e-banking phishing website factors and to investigate its techniques by classifying the phishing types and defining six e-banking phishing website attack criteria's with a layered structure. Our experimental results showed the significance and importance of the e-banking phishing website criteria (URL & Domain Identity) represented by layer one and the various influences of the phishing characteristic on the final e-banking phishing website rate.

P.A. Barraclough et al. [14] proposed new inputs (Legitimate site rules, User-behavior profile, Phish-Tank, User-specific sites, Pop-Ups from emails) which were not considered previously in a single protection platform. The idea is to utilize a Neuro-Fuzzy Scheme with 5 inputs to detect phishing sites with high accuracy in real-time. In this evaluation 2-Fold cross-validation is applied for training and testing the proposed model. A total of 288 features with 5 inputs were used and has so far achieved the paramount performance as compared to all beforehand reported results in the field.

V. Shreeram et al. [15] proposed genetic algorithm to evolve rules that are used to differentiate phishing link from legitimate link. Evaluating the parameters like evaluation function, crossover and mutation, GA generate a rule-set that matches only the phishing links. This rule-set is stored in a database and a link is reported as a phishing link if it matches any of the rules in the rule based system and thus keeps it safe from fake hackers. Preliminary experiments show that this approach is effective to detect phishing hyperlink with minimal false negatives at a speed adequate for online application.

Sadia Afroz et al. et al. [16] proposed phishing detection approach—PhishZoo—that uses profiles of trusted websites' appearances to detect phishing. Our approach provides similar accuracy to blacklisting approaches (96%), with the advantage that it can classify zero-day phishing attacks and targeted attacks against smaller sites (such as corporate intranets). A key contribution of this paper is that it includes a performance analysis and a framework for making use of computer vision techniques in a practical way.

Mallikka Rajalingam et al. [17] presented an effective image-based anti-phishing method based on discriminative key point attributes in WebPages. They use an invariant content descriptor, the contrast context histogram (CCH), to figure the resemblance degree flanked by mistrustful pages and trustworthy pages. To conclude, whether two images are analogous, a common method involves extracting a vector of prominent features from each image and computing the distance between the vectors which is taken as the degree of illustration difference between the two images. The experimental results make obvious that the proposed method attains high precision and low error rates.

Radha Damodaram et al. [18] presented a novel method to prevail over the impenetrability and complication in detecting and predicting fake website. There is proficient model which is based on using Association and classification Data Mining algorithms optimizing with PSO algorithm. These algorithms were used to characterize and identify all the factors and rules in order to classify the phishing website and the relationship that correlate them with each other. It also used MCAR classification algorithm to extract the phishing training data sets criteria to classify their legitimacy. After classification, those results have been optimized with Ant Colony Optimization (ACO) algorithm. But, this work has limitations like Sequences of random decisions (not independent) and Time to convergence uncertain in the phishing classification. So to overcome this limitation, we enhance Particle Swarm Optimization (PSO) which finds a solution to an optimization problem in a search space or model and predict social behavior in the presence of phishing websites. This will improve the correctly classified phishing websites. The experimental results inveterate the practicability of using the PSO method in real applications and its enhanced performance.

Michael Atighetchi et al. [26] described a set of innovative attribute based checks for defending against phishing attack and also explain a number of anti-phishing algorithms implemented as plug-ins and highlight which attributes of phishing sites they consider. Hence, to estimate an efficacy and applicability of this system and performed widespread experimental testing. We also presented a fully automated crawling outline that we developed for testing.

Isao Echizen et al. [27] presented content-based phishing detection extracts keywords from a target Web page, uses these keywords to retrieve the corresponding legitimate site, and detects phishing when the domain of the target page does not match that of the retrieved site. It often misidentifies a legitimate target site as a phishing site, however, because the extracted keywords do not characterize the legitimate site with sufficient accuracy. Two methods are described for extracting keywords: domain keyword extraction, which extracts keywords from not only the page on the browser, but also from pages linked from this page, and time-invariant keyword extraction, which extracts keywords from the page and previous versions of the page. Experiments using 172 legitimate sites demonstrated a reduction in the false detection rate from 14.0% to 7.6%, while experiments using 172 phishing sites demonstrated no change in the rate of overlooking phishing pages.

3. ANTI PHISHING TECHNIQUES

Phishing is the process of sending an email to a user falsely claiming to be an established legitimate enterprise in an attempt to scam the user into surrendering private information that will be used for identity theft. Phishing is being combated through user education, legislation, and integrated anti-phishing measures in modern Web browsers.

3.1 Types of Phishing Attacks

Phishing has spread beyond email to include VOIP, SMS, instant messaging, social networking sites, and even multiplayer games. Below are some major categories of phishing.

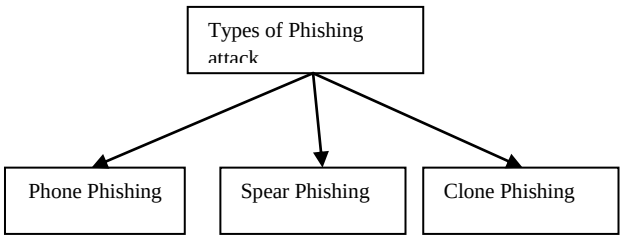

Fig. 2 Types of Phishing attacks

3.1.1 Spear Phishing

Spear phishing targets at a specific group. So instead of casting out thousands of emails randomly, spear phishers target selected groups of people with something in common, for example people from the same organization [23].

3.1.2 Phone Phishing

This category of phishing refers to messages that assert to be from a bank asking users to dial a phone number concerning problems with their bank accounts. Conventional phone equipment has devoted lines so Voice over IP, being straightforward to manipulate becomes a good choice for the phisher. Once the phone number owned by the phisher and provided by a VoIP service, is dialed, voice prompts tell the caller to enter her account numbers and PIN. Caller ID spoofing which is not proscribed by law can be used along with this so that the call appears to be from a trusted source [24].

3.1.3 Clone Phishing

In this type phisher creates a cloned email. He does this by getting information such as content and recipient addresses from a legitimate email which was delivered previously, and then he sends the same email with links replaced by malicious ones. He also employs address spoofing so that the email appears to be from the original sender. The email can claim to be a re-send of the original or an updated version as a trapping strategy [22].

To protect users against phishing, various anti-phishing techniques have been proposed in which some the techniques are described below:

3.1.4 Content Based Approach

This method is used to measure the similarity between two given web pages by calculating the similarity between the content elements (text, image, and layout) contained in the web pages. Algorithms are used to compute visual similarity to detect the phishing web pages which have higher similarities to phishing targets. It requires finding the phishing target prior to the similarity comparison computation. It also combines TFIDF retrieval algorithm to determine the likelihood that a given web page is a phishing webpage. Words with highest TF-IDF weight on a given webpage can be used to classify the webpage as legitimate or not [19].

3.1.5 Black Listing

A list of non-trusted URLs or more simply a list of banned websites is called blacklist [20]. In context of phishing that are known to have malicious intensions. This method is most frequently used by the web browsers to identify phish. This method is to check URLs against a blacklist of known phishing websites. There are more than 20 spam blacklists used today. These blacklists contain IP address or domain names of phishing websites, proxies [20]. A blacklist of known URLs or domain names is used to chunk the phishing websites. A few phishing websites are hosted on hacked domains so it is consequently not good to chunk the whole domain. Therefore instead of domains a blacklist of phishing URLs is a better solution. By adding a phish URL is a multi-step process but once the phish is confirmed, it is added to the innermost blacklist. In some cases, the blacklist is downloaded to local computers. For example, in Firefox-3, blacklists of phishing URLs or domains are downloaded to browsers, subsequent to every 30 minutes [21]. This is the most regularly used method to block Phishing websites.

3.1.6 Heuristics Based Approach

This technique rates the phishing possibility of a given webpage using reputation scores either obtained from the anti-phishing community or computed from the given webpage. However the reliability of the reputation scoring is a great challenge [19].

3.1.7 Community Information

This is the most popular anti-phishing detection technique. This relies on a blacklist of known phishing sites that blacklist is maintained by an anti-phishing community. If the user attempts to visit a URL and that URL belongs to the blacklist then the user is prevented from visiting that site or warned as the site is a known phishing site. Other techniques rely on a community of users where users mark the site as phishing or not. A site's popularity may also be indicated by community information. Net craft's anti-phishing toolbar and Google's Pagerank technique both rank websites based on popularity and risk [25].

Table 1: Prons & Cons of Anti-phishing Techniques

Methods	Prons	Cons
Community Information	-Pages are ranked on the popularity basis -Userreport whether site is good or bad	-A popular phishing site may be on top -selection of the right site is more important
Content Based Approach	-Low false alarm rate -Much aacurate	It can be tracked using invisible text
Heuristic based Approach	It easy to evaluate and easy to manage	-High probability of false and failed alarm -easy to avoid heuristic
Blacklisting based Approach	-Easy to manage -Download -Install and update quickly	-Creates false alarm rate -update is insignificant

4. CONCLUSION

The majority of the straightforward and un-experienced internet users are mostly influenced by email spoofing or tracking of personal information such attack is known as Phishing.This paper presents different anti-phishing approaches to avoid the stealing of personal information online and also shows the pros and cons of these approaches. In future work need to develop such technique which reduces false alarm rate and easy to protect the information through website.

5. REFERENCES

[1] Karthikeyan R.G, Sethuraman R, "Phishing Website Detection and Prevention of Phishing Attacks: a Field Experiment", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 2, February 2014

[2] Gaurav Kumar Tak, Gaurav Ojha, "Multi-Level Parsing Based Approach Against\ Phishing Attacks With The Help Of Knowledge Bases", International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.6, November 2013

[3] Ammar Almomani, B. B. Gupta, Samer Atawneh, A. Meulenberg, and Eman Almomani, "A Survey of Phishing Email Filtering Techniques", IEEE

Communications Surveys & Tutorials, Vol. 15, No. 4, Fourth Quarter 2013

[4] A. Sarannia, U.R. Padma, "Prevention Model for Phishing Attacks In Web Applications Using Linkguard Algorithm", International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014

[5] Mahmoud Khonji, Youssef Iraqi, Andrew Jones, "Mitigation of Spear Phishing Attacks: A Content-Based Authorship Identification Framework", 6th International Conference on Internet Technology and Secured Transactions, 11-14 December 2011, Abu Dhabi, United Arab Emirates.

[6] Maher Aburrous, M. A. Hossain, Fadi Thabatah, Keshav Dahal, "Intelligent Phishing Website Detection System using Fuzzy Techniques".

[7] Gaurav Kumar Tak, Gaurav Ojha, "Multi-Level Parsing Based Approach Against\ Phishing Attacks with the Help of Knowledge Bases", IJNSA Vol.5, No.6, November 2013

[8] M. Madhuri, K. Yeseswini, U. Vidya Sagar, "Intelligent Phishing Website Detection and Prevention System By Using Link Guard Algorithm", International Journal of Communication Network Security, ISSN: 2231 – 1882, Volume-2, Issue-2, 2013.

[9] M. Topkara, A. Kamra, and M. J. Atallah, et al, "ViWiD: Visible Watermarking Based Defense against Phishing", Lecture Notes in Computer Science, Vol.3710, 2005, pp.470-483.

[10] A. Y. Fu, W. Y. Liu, and X. T. Deng, "Detecting Phishing Web Pages with Visual Similarity Assessment Based on Earth Mover's Distance (EMD)",IEEE Transactions on Dependable and Secure Computing, Vol.3, No.4, 2006, pp.301-311.

[11] Arun Vishwanath , Tejaswini Herath , Rui Chen, Jingguo Wang , H. Raghav Rao, "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model", Decision Support Systems 51 (2011) 576–586.

[12] Venkatesh Ramanathan, Harry Wechsler, "Phishing detection and impersonated entity discovery using Conditional Random Field and Latent Dirichlet Allocation", computers & security 3 4 (2 0 1 3) 1 2 3 e1 3 9, journal homepage: www.elsevier.com/locate/cose.

[13] Maher Aburrous, M.A. Hossain, Keshav Dahal, Fadi Thabtah, "Intelligent phishing detection system for e-banking using fuzzy data mining", Expert Systems with Applications 37 (2010) 7913–7921.

[14] 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 (2013) 4697–4706.

[15] V. Shreeram, M. Suban, P. Shanthi, K. Manjula, "Anti-Phishing Detection Of Phishing Attacks Using Genetic Algorithm", ICCCCT'10-978-1-4244-7770-8, in proceeding IEEE.

[16] Sadia Afroz, Rachel Greenstadt, "PhishZoo: Detecting Phishing Websites By Looking at Them".

[17] Mallikka Rajalingam, Saleh Ali Alomari, Putra Sumari, "Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages", International Journal of Computer Science and Security (IJCSS), Volume (6) : Issue (1) : 2012.

[18] Radha Damodaram, M.L.Valarmathi, "Phishing Website Detection and Optimization Using Particle Swarm Optimization Technique", International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (5): 2011

[19] C. Emilin Shyni and S. Swamynathan, "Protecting the online user's information against phishing attacks using dynamic encryption techniques", Journal of Computer Science, 9 (4): 526-533, 2013, ISSN 1549-3636.

[20] J. Jung and E. Sit "An empirical study of spam traffic and the use of DNS black lists", In IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pages 370-375, New York, NY, USA, 2004. ACM.

[21] F. Schneider, N. Provos, R. Moll, M. Chew, and B. Rakowski, "Phishing protection: Design documentation". https://wiki.mozilla.org/Phishing_Protection:_Design_Documentation.

[22] Clone Phishing - Phishing from Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Phishing

[23] Bimal Parmar, Faronics, "Protecting against spear-phishing",http://www.faronics.com/assets/CFS_201201_Jan.pdf

[24] Phone spoofing From Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Phishing#Phone_phishing

[25] Namrata Singh, Nihar Ranjan Roy, "A Survey of Phishing Website Detection Technique", IRAJ International Conference-Proceedings of ICRIEST-AICEEMCS, 29th December 2013, Pune India. ISBN: 978-93-82702-50-4

[26] Michael Atighetchi, Partha Pal, "Attribute-based Prevention of Phishing Attacks", 2009 Eighth IEEE International Symposium on Network Computing and Applications.

[27] Shinta Nakayama,Hiroshi Yoshiura, Isao Echizen, "Preventing False Positives in Content-Based Phishing Detection", International Conference on Intelligent Information Hiding and Multimedia Signal Processing in 2009, proceeding in IEEE.