EAI Endorsed Transactions
on Energy Web and Information Technologies Research Article EAI.EU

# Review of Various Methods for Phishing Detection

R.Sakunthala jenni[1],*, S.Shankar[2]

[1] Research Scholar, Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore, India.

[2] Professor, Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore, India.

## Abstract

In this modern world where the technology has spread rapidly and the inception of cell phones, computers and also the rate at which internet is being used has increased in all fields both commercial, financial and also individuals. The above said inception is a boon but the users are facing dangerous challenges. Hence phishing and information pilfering which are done through spams and deceptive emails. This kind of spams and deceptive emails could lead to great losses for institutions like financial and similar ones. It is understood that in the beginning it is very difficult to judge or trace the modesoprandi of these hackers. To ascertain this cryptic methods of phishing attacks can be effectively done only by developing a particular software thereby safe guard the users. To pick out and detect the operation methods of hackers, the researcher uses the Data Mining process where a number of datamining tools, which analysis the data are used. This learning, basically is Data Miming process and the informations are taken out through different outlets and sources.

**Keywords:***phishing, anti-phishing, hackers, fraud websites, legitimate websites.*

Received on 31 May 2018, accepted on 10 September 2018, published on 12 September 2018

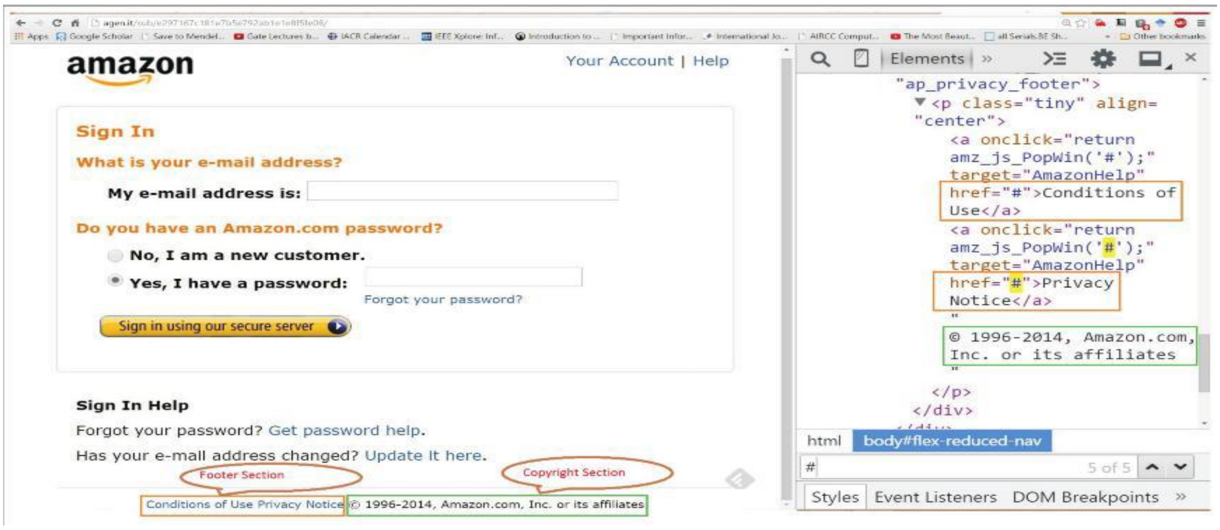
3rd International Conference on Green, Intelligent Computing and Communication Systems - ICGICCS 2018, 18.5 - 19.5.2018, Hindusthan College of Engineering and Technology, India

doi: 10.4108/eai.12-9-2018.155746

## 1. Introduction

Millions of people through out our globe, have made their lives dispensable without the use of various kind of smart phones. They have the technique of accessing into any Iinnumerable facilities. Furthermore Business houses, financial institutions, banks other facilities like online shopping, give extensive fine services through their websites. Transactions and businesses done through websites and cyberspace reduces time wastages, less traffic and the cost effective is better. So website business transactions have become very popular. But inspite of all these excellent benefits, we learn that there are dangerous repercussions the users face, by the hackers who deceive through their phishing activities in order to pilfer valuable informations [1]. The hacker's intentions are to rob their money, or some of them do it for revenge and a few other hackers do just for fun and thrill. They are in, for destructive activities. These unethical practioners are called hackers, crackers, and intruders [2]. The most important operational part in a computer is the security systems. Pilfering the details of an individual person or institution through emails is a dangerous aspect faced by a person or institution. The hacker's main intention is to steal the users information and use it in the wrong way. In this the tool spam is used by which they pose themselves as genuine to attract the people towards huge amount of lottery money or other fascinating attraction to divert and attract the users there by they can get access to their personal information's like Account numbers and financial information, where we deeply analyses and we learn that certain set patterns are fabricated and followed when they write emails and they stress on words like winner, lottery, visa etc., Certain data and information from the hackers and emails, which they send to users are useful and helpful to find out and trace them and this in turn helps to develop security tools. To extract and find out these deceptive patterns in large amounts, the best method is Data Mining. Data Mining has been researched and it has various operation procedure to trace and discover and acquire useful information like cluster and classification, Artificial network, Bayesian network, Decision tree and Machine learning [3]. The main purpose of this paper is to pinpoint and clarify the concepts of phishing attacks and moreover

*Corresponding author. Email:kalaivanijenni@gmail.com



amazon

Your Account | Help

Sign In

What is your e-mail address?

My e-mail address is:

Do you have an Amazon.com password?

○ No, I am a new customer.

● Yes, I have a password:

Forgot your password?

Sign in using our secure server

Sign In Help

Forgot your password? Get password help.

Has your e-mail address changed? Update it here.

Conditions of Use Privacy Notice © 1996-2014, Amazon.com, Inc. or its affiliates

```
"ap_privacy_footer">
  ▼<p class="tiny" align=
  "center">
     <a onclick="return
     amz_js_PopWin('#');"
     target="AmazonHelp"
     href="#">Conditions of
     Use</a>
     <a onclick="return
     amz_js_PopWin('#');"
     target="AmazonHelp"
     href="#">Privacy
     Notice</a>
     "
     © 1996-2014, Amazon.com,
     Inc. or its affiliates
     "
   </p>
  </div>
```

Figure.1 Contents and the source Code as Features for Detecting Phishing Attacks.[30]

understand the usual practice of Datamining. Next step is to elaborate on the different mining techniques to detect and clarify the different phishing attacks. In the end, conclusions are realized after the study is presented.

## 2. Phishing

The word is derived from the origin in trapping and catching a fish-'fishing'- normally such terms are not popularly located in the computer science , for the sake of security operations which are carried out on social engineering.

### 2.1. Phishing Attacks

Figure 1 shows an example Hence by one of these connection in the internet spam, the user is instructed to visit a illegitimate webpage which has resemblance to the login page of the well-known Amazon webpage. In this source code the tags are utilized as a vital feature for tracing the Phishing attacks. It is established that we find no actual authentic code for the pages foot notes, inside the source code of the page. It is noted that there are no authentic connection and this in turn helps to detect the Phishing Attack.

This method, is adopted for security in the internet has taken a vital role in recent times. It indicates to the method where hackers illegal methods use, to send across emails (spams) chat, through information of the online users like their ID, phone number, account number and numbers. The methods of phishing has many approaches, using spam, and it varies each day and here the toll of phishing multiplies. In spite of the simplicity of this method, the attacks are very destructive and absolutely effective and this is supposed to be the most dangerous threats on security, in the online operations. As we termed earlier the word 'Phishing' is like 'Fishing' only the spelling differs. In the actual fishing our aim is to catch a fish, here in phishing we trace and catch the hackers. They use spoofed websites to catch the user's personal information which in turn the hackers use for their benefits [4]. Approaches like email deceive the users. They personalize their new business house and institution or trustworthy persons to invite the attention or attract the users, get them in to spill out their personal details and they finally become a victim. To put it in to a nutshell they get trapped in some of the links, later on they realized as not genuine but illegal [5]. |The first step to trace out the phishers, is to trace out their modesoprandi in their internet links. The important features for tracing and detecting attacks include the time, that is, when, in other words they ape the site's name, IP addresses, some errors present which are fishy, the type @ character and so on. Nevertheless this above said operation is not easy due to the vast size of information, which is disorganized, sometime hidden and camouflaged low level information, which creates confusion in understanding their technologies. So to find out we could use counteraction effectively and always attack their weakness in the cyberspace which they use, to do their stealing operation. Clicking on links is one of the tactics the hackers follow, so that they have access to the user's page [6]. Till now phishing is not totally defined because of their varied approaches, so the definitions are many. In 2012 papers, the different definitions of phishing is shown and this is presented for our understanding. Papers are extracted from the work of Phish Tank company (http://www.phish-tank.com/): [7].
Phishing attacks are done by taking identity and legitimacy of a person or institution which are genuine. For example, the hackers may come across different types of websites which are used by the shoppers, and they forward details through messages to them to acquire personal details like password, account number etc. They do it in a very professional way so that users will not have any suspicion. Figure 2 shows the peak time of Phishing Attacks the second half of the year

2014 which is based on the domain. According to the figure the peak time for the info domain it shown clearly [8].

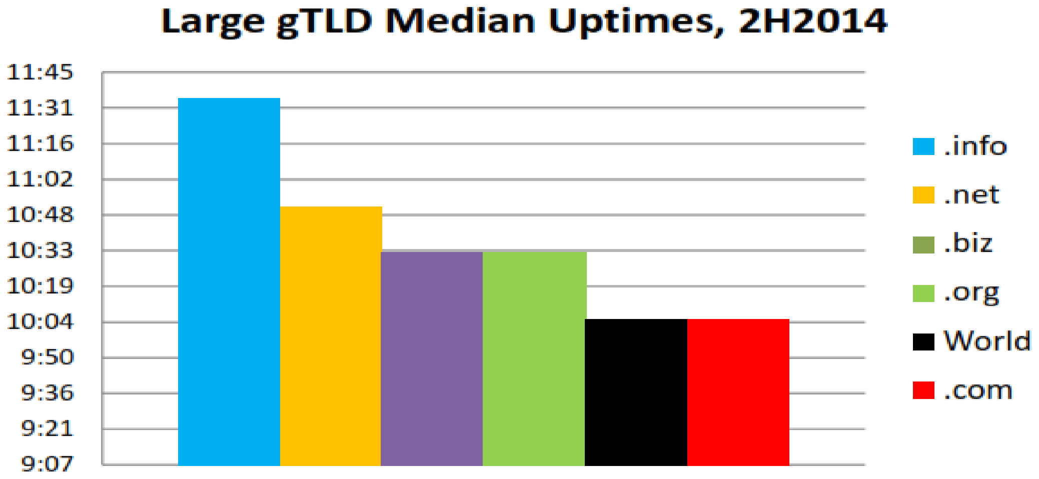


Figure. 2 Shows the peak time of Phishing Attacks in the Second Half of 2014 Based on Domains

## 2.2 Phishing Attacks: Its Range &Competence

The range and scope which the phishing attacks posses, depends on the type of the domain. Varied top level domains [TDLs] have varied methods of attacks.

- Conflicting phishing attacks

Various techniques are there to encounter phishing attack which is explained herewith [8].

1. Training the internet users (victims) for encountering phishing attacks.
2. Training the users to type messages rather than click on the links.
3. Instructing the users not to encourage emails with the public without specific names.
4. Make the users understand that institutions and banks linked with users, generally, who do not call for their account number.
5. The users should install anti-phishing extraction on the browsers.

- Stimulus and aim of phishing

[15] The below explains the stimulus and aim behind the attacks.

1. The main aim and purpose is to get financial gains. The desperation to get money induces the hackers to pilfer and steal from individuals, banks and institution.
2. They operate this by hiding their identity so that they can carry out their destructive actions. For this they use stolen user’s name and passwords –for example while shopping through internet, gamming, product sales and even child abuse is done.

## Phishing Life Cycle

The nature of phishing attacks are elaborate. As we have life cycle for many things phishing attack also has a start, develop and finishing stages.
Phishing attacks figure3 starts with the users finding of emails of hackers very genuine. The hackers send out emails which will be absolutely like a genuine one and it will be so interesting that the users are carried away, get into the hackers email and expose their personal information and all possible data. Almost 65% of phishing attacks begin by the visit of links received within an email [9]. As per the figure there are three stages. The first one is the early phishing stages which is explained above, then it leads to the mid phishing and ends with the post phishing. In the second stage the users are carried away by this unauthentic links and open and disclose details. In the post age the defrauding becomes a success. In this process the users reveal their private information’s which the hackers need.

## 3. Solution

There are 5 methods to solve the problem [10]
(1) The first one is identifying the needed data. We need a set of details which are already identified. These should have certain influences on the out i.e. the classifier. So the set of output and input should be detected.
(2) Phishing data has many sources for example the Phish Tank. Phish Tank has pairs of input instances and the derived destination class.
(3) Determining the input features. It all depends on how carefully the features are selected. In this unwanted, irrelevant and unconnected are discarded so that the magnitude of the training data set is minimized. This in turn helps in the learning process and execution of the process too.

(4) The most vital and important step to be taken is the choice of a mining algorithm. We have vast rang in the mining processes, in the literature and each of this process has its advantages and disadvantages. There are three important elements for selecting these classifications. They are
1.The input Data characteristics
2.Te accuracy rate measure of the classifier predictive power
3. The clarity and the understandability of the output
(5) Finally it is the evaluation of the process. That is the derived classifier performance of the data.

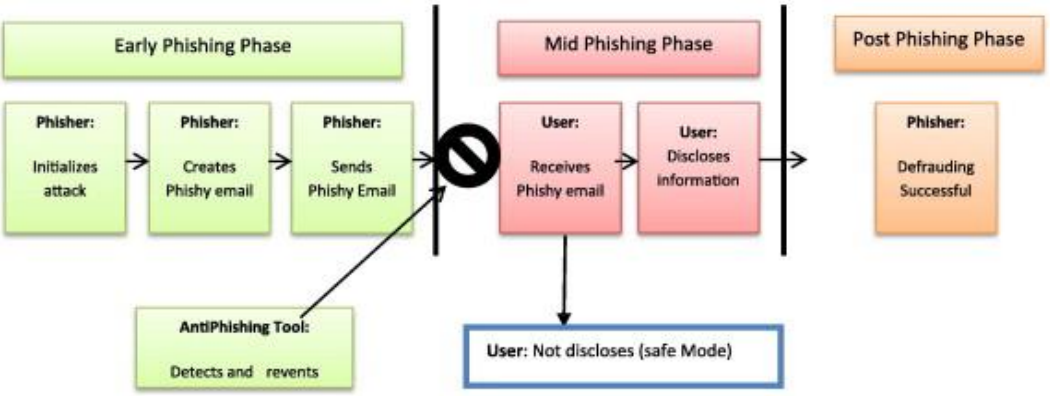

Early Phishing Phase
Mid Phishing Phase
Post Phishing Phase
Phisher: Initializes attack
Phisher: Creates Phishy email
Phisher: Sends Phishy Email
User: Receives Phishy email
User: Discloses information
Phisher: Defrauding Successful
AntiPhishing Tool: Detects and revents
User: Not discloses (safe Mode)

Figure. 3. Phishing life cycle.[16]

We cannot pinpoint and say that any particular classifier is the best because its performance over all depends on the characteristics of the training data and AC is selected because if its many outstanding features, like prognostic approach and meticulousness and also the quality of result which acquired. To minimize the phishing there are two methods.one is the technical method and the other one is Nontechnical method.

## 3.1 Non Technical Method

Legal solution to the problem is the non-technical method. This method is adopted in many countries, and under this the US was the first to bring the law against phishing activities and when the law was enforced lot of hackers were caught, arrested punished and sued for the illegal activities. The law was brought phishing under computer crime in the year 2004 by the Federal Trade Communication (FTC). It is an agency started in the interest of the consumers, for their protection and safety [13]. Later Australian and UK enforced this law, in the year 2005-6. Their law enforced and prohibition of fraud activities, i.e. phishing websites and also ordered imprisonment of the hackers (http//www.finextra.com.news). Eventually countries like Australia agreed upon and signed the papers, with the Microsoft Company. The law was brought the personal who were the trainers, were given the knowledge as how to phish the hackers and learn their modesoprandi [12]. But, this legal approaches was not 100% successful in catching the phishers, mainly because it is the task to trace the hackers because of them disappearing (leaving their track) from the cyber world, quickly.

### 3.1.1 Awareness for the user

Creating awareness for the users is also equally important. They have to be educated about these criminal, their way of operation etc., [12]. The internet users should be convinced and also learn how to trace, through the security indicators within the website, the problem will be reduced to a large extent because it is two parties tracking them. The main advantage for the hackers is the ignorance of users. They lack the basic knowledge, and not aware that such dangerous activities are done by hackers. Nevertheless it is on total, a difficult task to implement. Reason behind this is the users need substantial time and knowledge, and become more clever and talented in learning phishing techniques, which even experts in this field sometimes fail and over look.

## 3.2 Technical Method

There are two prime approaches in this technical approaches. One is the blacklist technique and the other heuristic based [15]. According to the method of operation of the blacklist the URL seems to have comparison with a pre-established phishing URL. In this approach it does not deal with all phishing websites because there are inflow of new fake websites consumes considerable amount of time before we could establish and add to existing list. Whereas in the heuristic-based technique, one could fast recognize the newly launched fake websites without considerable wastage of time [14]. Because of the certain disadvantages we saw, the above technique, and a need has arised to explore and innovate new and different approaches.

Among them MacAfee has brought about new solutions. Furthermore non profit organizations like APWG [15] Phish Tank (Phish Tank 2006) has brought in better practice and methods, to help the users against the phishers and help the users gain more experiences recognition of phishing websites and within shorter periods leads to the success of anti-phishing techniques, accurate decisions are not derived due to fluctuations in the pulse in the increase of deceptive assured

conclusion, to put it in to a nutshell branding is an authentic website , most of the time is not original.
Some technical solutions realized by researchers and scholars for dealing can be seen herewith.

### 3.2.1 Blacklist-Whitelist Techniques

Blacklist as a name suggest is considered to be baleful and which has been gathered through techniques like users vote. So whenever a new website is established, the browser guides us to check whether the new website comes under a blacklist. If it is under the blacklist the browser warns the users to stop sending personal informations like ID, bank A/c no: etc. It is noteworthy that blacklist can be recorded in the user's computer or optionally on a requested server, by the phishers as and when where is the URL request. According to [17] blacklists are noted and established at different frequencies. The estimation was 50-80% of phishing URL's which are displayed in the blacklist 12 hours after their launch besides other black lists through Google's need on an average of 7 hours for update [18]. So it is inferred and understood that a black list updated then and there in the interest of the safety of users and that they do not become victims of the blacklist hackers. The blacklist approach is embattled with respect to various solutions one of the important is harmless browsing in google. In this file or predefined phishers are used-URLs to trace out fraudulent URLs. A different technique that is followed for the protection of Microsoft IE9 which works against phishing technique and also protect site advisor which are actually data based solutions and these are created to detect and catch illegitimate attacks like Trojan horses and Spyware. These have crawler which is automatically operate and help to browse the website and also establish threat and rate the range of threat which is connected with the already entered URL. Nevertheless, site advisers cannot locate or identity newly created dangers. The third one is the anti-phishing tool called Verisign traces numerous websites which can recognize "clones" so that one could find out the illegal websites. There is always a competition between the attackers and users, so approaches are not fool proof. There is a technique tiled Net-craft which is comparatively small process that activates on a web browser. It depends on totally illegitimate website which comes under the blacklist which in turn is recognized by Net-craft and also which is induced by the users and this is verified by Net-craft. Net-craft clearly shows the location of the server where the webpage is hosted. Users who are experience accept that Net-craft is very useful to the operation to site an example webpage which has "ac.uk". it is not done other than UK.
Whitelist is not similar to blacklist and as the term explains they are genuine websites. All others cannot be called as an automated- individual-whitelist (AIWL). This AIWL is a tool which works against phishing, where the user's whitelist is used as the basic, inside a genuine, trustworthy websites. This AIWL, very efficiently can trace all log- in entered by the user by using Naïve Bays algorithms. When repeated successful log on is done in the case of the users; the specific website is received. AIWI's work is to induce to connect the website to the whitelist by the users. Yet another way to correct the mad activity was the research with reliable on the whitelist which can be seen in phish 200 [19]. Here phish200 creates profiles of genuine and trusted websites which is based on Fuzzy hashing techniques. What we mean by website profiles is, it is an amalgamation of several metrics. That can distinctly pick out that website. In this process it combines whitelist with blacklist and as we learnt that the heuristic approach is a warning process for the users abutt the hackers. The researcher's belief that the detection technique should be derived from the users view point as because 90% of users depends on how the website looks and so that the genuine of the website could be verified.

### 3.2.2 Fuzzy rule approaches

Here the technical approach followed on [20] the basis of contradicting some rules on the basis of algorithms, this is done after gathering different kinds of features which varies in features, and the capacity of website as depicted in table. There will be three uncertain values. They are the "legitimate", "genuine", "doubtful". After a series of experiments the authors evaluate it using the below mentioned algorithms in Weka, PRIMS, C4.5, JRIP and Part [21, 22] from the result they established a very distinguished connection of both "URL" features and "Domain Identity". Nevertheless they could not assess any justification on the features. Larger set features were used by authors of (Alburrous, Hossain Dahal and Thabtah) to foresee websites type based on fuzzy logic. Their developed method, in spite of giving good results, in accuracy, it is not clear as to how they established, extracting from the website and specific features associated with human factors. Everything was done and arrived at a conclusion based on human experience rather than intelligent data mining techniques. This paper has been taken with the intension of solving the above said problems. The authors divided the websites under different categories very legitimate, legitimate, suspicious phishy or very phishy. The fine line for these different legitimate categories were not earlier established.

### 3.2.3 Machine learning techniques

The different types of methods adopted, progress and established in order to control phishing with done by the support vector machine (SVM). This SVM is a popular machine through which training is given to the users, and is used to solve classification problems effectively [23]. It became popular because of its ability to bring forth accurate results from un structural problems like text categorization. It is realized that it is realized that it is possible to visualize SVM equal to a hyper plane which has the ability to split the object (points) and which belong to the group (negative objects) by the SVM algorithm while learning, where the hyper-plain is got so as to which in turn can divide positive and negative objects with maximum level. This level depicts the area between the hyper plane to the nearest positive and negative object proposed new technique which had the help of SVM. The purpose of this discovery of authentic and unusual (suspicious) operations, example is phishing, through the homepage under the company's name which seen in the domain name. The next one is titled as “Page categorizer” which shows the characters connected structural features (unusual URL, unusual DNS record, etc.,) which is difficult to copy as duplicates.

There are 6 different architectural properties which has been selected and Vapkin’s Support Vector Machine (SVM) algorithm [25] was employed for establishing to know if we can find it is a legitimate or not. Later it was established in that the “Identity Extractor” has an important characters in union with illegitimate URLs. This information was arrived after some limited data lets were done which consisted 179 URLs. From this an 84% accuracy was inferred by using other features, a solution will be arrived to make this accuracy more precise.

A comparative study was done, on the problems of email phishing, by using machine learning techniques, which included SVM, decision tree, and naïve bays, by [26]. A research work done on a random forest algorithm titled “Phishing Identification by learning on features of email received” (PILFER) in a unsystematic way. The experiment was done it 860 illegitimate emails and 695 legitimate website. It is noted that PILFER has sharp accuracy to detect illegitimate emails. IP based URLs were used in a few features in order to detect illegitimate emails. This has connection between the emails and also number of connection inside the email, and domains which appear in the email, number of spots inside the connection and the contents of java scripts and spam filter output. It was concluded by the author that is possible for PILFER inclined on to the classification of can be boosted towards the classification of emails. In the process of combination varied ten features “Spam filter output”. To assess the authors used the same data set. From the result it was inferred that PILFER it was decreased the rate of false positive.

### 3.2.4 The Cantina Technique

This is a technique reached by [27] where they used “Carnegie Mellon Anti-phishing and Network Analysis Tool” (CANTINA). From this method the type of websites which used frequency –inverse –document –frequency is established.TF-IDF [28] Cantina checks the website and what it has and then arrive at a decision as to know the nature of website is phishy which used TF-IDF. These analyses the importance of weight and also the importance of analyzing with frequency. For a given webpage, CANTINA calculates the TF-IDI, the next step is to taken TF-IDF which are higher than other and this is added to the URL to acquire the lexical signature. This is equally entered in to a search engine. A Legitimate is that which is among the current first 30 results otherwise it is called phishy. If the result is zero after the search it is established as phishy. To solve the problem the researcher use the method where the combination of TF-IDF with different character, (mistrustful URL, life of domain, There are limitations in this technique and it is because some legitimate websites uses like TF-IDF such as could be unfit.

One more technique with uses with additional attributes [29] have use of data set which has 200 websites, half of them were legitimate and the other half were illegitimate and 8 features. The features are suspicious link, domain age IF-IDF and so on. While executing the experiments some changes in the performance is noticed which is as follows. They are explained below.

1. To begin, finding the genuine website, a filter was set because the attracting fake sites for dragging the user’s to the fake website which could cause lot of harm for the users information we should have the screen to resists with in the ‘site’.
2. According to the version of the researcher both features like ‘Domain age’ and ‘Known image’ these are not very important.
3. Thirdly they researched and established a new type of fizzy webpage and primarily on the top of the domain.

### 3.2.5. Associative Classification Data Mining Technique

Neda A et al [16] studied the website phishing by using Multi-label Classifier based Associative Classification (MCAC) method was used to identify the phishing websites with accuracy. The new rule was generated for enhancement using MCAC. The websites doesn’t considered which has content based features. The MCAC technique will cannot create by previous algorithms.

MCAC method works by below mentioned Points.

1. Looking for the hidden relationship between the class attribute and the attribute values training.
2. To form the association rules by using this relationship.
3. Based on the support and confidence, the rules may be sort by using the sorting algorithm.
4. Proper rules are accepted, ignored the duplicate rules.

Various features was identified and discussed previous related to phishy and legitimate websites and collected over 1350 various websites from different sources. Some features was having categorical values like as phishy, suspicious and legitimate. These kind of values was replaced -1, 0, 1 respectively. Usually the website can be divided into two different classes like, legitimate and phishy. The AC method was used in this study can discover the rules with one class and two classes also (legitimate and phishy). MCAC method can make new kind of rules through new class which earlier not seen in the database in the name of “suspicious”. If websites are considered as suspicious, it may be either legitimate or phishy. The end-user can give an accurate solution based on the assigned weights to the data.

Various algorithms was used to evaluate the efficiency and applicably in specific MCAC method on the collected data. The major method was used in this studied besides are MCAC (CBA,MCAR and MMAC) and (C4.5, PART,RIPPER).

Table 1. Comparative and analysing study

| S.no | Author | Paper Title | Description of the work | Result |
|---|---|---|---|---|
| 1 | Sheng, S., et al. 2009 | An empirical analysis of phishing blacklists. | The effectiveness of phishing blacklists were studied.<br>In this study 191 new phish used that were lesser than 30 minutes old to run two tests on 8 anti-phishing toolbars. | According to this paper blacklists are noted and established at different frequencies. The estimation was 50-80% of phishing URL's which are displayed in the blacklist 12 hours after their launch besides other black lists through Google's need on an average of 7 hours for update. |
| 2 | Afroz, et al. 2011 | PhishZoo: Detecting phishing websites by looking at them. | Here phishZoo creates profiles of genuine and trusted websites which is based on Fuzzy hashing techniques. | PhishZoo provides accuracy of 96%like Blacklist. By using this approach can classify the zero-day phishing attack. |
| 3 | Aburrous, et al. (2010a). | Predicting phishing websites using classification mining techniques. | The technical approach followed on the basis of contradicting some rules on the basis of algorithms, this is done after gathering different kinds of features which varies in features, and the capacity of website as depicted in table. There will be three uncertain values. They are the "legitimate", "genuine", "doubtful". | The results indicating Associative classification algorithm of MCAR showed less error rate (12.622%) when compared to other traditional classification. The error rate was measured based on accuracy and speed. |
| 4 | Sadeh, et al, (2007) | Learning to detect phishing emails. | The research work done on a random forest algorithm titled "Phishing Identification by learning on features of email received" (PILFER) in an unsystematic way. | It is noted that PILFER has sharp accuracy to detect illegitimate emails. IP based URLs were used in a few features in order to detect illegitimate emails. From the result it was inferred that PILFER it was decreased the rate of false positive. |
| 5 | Thabtah, et al, (2009). | Naïve Bayesian based on chi square to categorize Arabic data. | Cantina checks the website and what it has and then arrive at a decision as to know the nature of website is phishy which used TF-IDF. These analyses the importance of weight and also the importance of analyzing with frequency. | To solve the problem the researcher use the method where the combination of TF-IDF with different character, (mistrustful URL, life of domain, |
| 6 | Neda A, et al, (2014). | Phishing detection based Associative Classification data mining | Website phishing by using Multi-label Classifier based Associative Classification (MCAC) method was used to identify the phishing websites with accuracy. The new rule was generated for enhancement using MCAC. | MCAC method can make new kind of rules through new class which earlier not seen in the database in the name of "suspicious". If websites are considered as suspicious, it may be either legitimate or phishy. The end-user can give an accurate solution based on the assigned weights to the data. |

## Conclusion

Unauthentic websites or emails manipulate and get valuable private information of the users. This illegal attempt is known as phishing. Phishing attacks have many kind of methods. That are spam/email, web based delivery, instant messaging, Trojan hosts, system reconfiguration and so on. So, some techniques are developed to chase phishing websites, to detect and arresting the activity. Especially a method called Data mining which have different anti phishing methods to catch phishing attacks. In this research paper we establish different kinds of Data mining technologies. Such as Black and Whitelist approach, Fuzzy approach, Cantina approach. Every technology that was followed as a few pros and cons. Our survey paper has a foresight visionary of phishing website and different technology of phishing attacks. From this review report clearly understood that one type of technology is not sufficient to detect phishing attacks and it also names of providing high authority in phishing detection. In detecting challenge of phishing activity with high accuracy. Our work in future has decided to take a challenge and establish better effective approaches for minimizing this phishing activity.

## References

[1] G. Aaron, "The state of phishing," *Computer Fraud & Security,* vol. 2010, pp. 5-8, 2010.

[2] E. Shein, "The gods of phishing," *Infosecurity,* vol. 8, pp. 28-31, 2011. 548(11)70023-7

[3] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications–A decade review from 2000 to 2011," *Expert Systems with Applications,* vol. 39, pp. 11303-11311, 2012. http://dx.doi.org/10.1016/j.eswa.2012.02.063

[4] M. Aburrous, M. A. Hossain, K. Dahal, and F. Thabtah, "Intelligent phishing detection ystem for e-banking using fuzzy data mining," *Expert systems with applications,* vol. 37, pp. 7913-7921, 2010. http://dx.doi.org/10.1016/j.eswa.2010.04.044

[5] T. A. Almeida and A. Yamakami, "Facing the spammers: A very effective approach to avoid junk e-mails," *Expert Systems with Applications,* vol. 39, pp. 6557-6561, 2012. http://dx.doi.org/10.1016/j.eswa.2011.12.049

[6] C. Konradt, A. Schilling, and B. Werners, "Phishing: An economic analysis of cybercrime perpetrators," *Computers & Security,* vol. 58, pp. 39-46, 2016. http://dx.doi.org/10.1016/j.cose.2015.12.001

[7] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," *Communications Surveys & Tutorials, IEEE,* vol. 15, pp. 2091-2121, 2013. http://dx.doi.org/10.1109/SURV.2013.032213.00009

[8] P. Kumaraguru, P. Dewan, and R. Clayton, "2014 APWG Symposium on Electronic Crime Research (eCrime)."

[9] KasperskyLab(2013). <http://www.kaspersky.com/about/news/spam/2013/>.

[10] Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., & Chen, R. J. (2011). An efficient phishing webpage detector. Expert Systems with Applications: An International Journal, 38(10), 12018–12027.

[11] W. D. Yu, S. Nargundkar, and N. Tiruthani, "A phishing vulnerability analysis of web based systems," in *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on,* 2008, pp. 326-331. http://dx.doi.org/10.1109/iscc.2008.4625681

[12] The Government of Australia (2011). Hackers, Fraudsters and Botnets: Tackling the problem of cyber crime. Report on Inquiry into Cyber Crime.

[13] Kunz, M., & Wilson, P. (2004). Computer crime and computer fraud report. Submitted to the Montgomery County Criminal Justice Coordinating Commission.

[14] Miyamoto, D., Hazeyama, H., & Kadobayashi, Y. (2008). An evaluation of machine learning-based methods for detection of phishing sites. Australian Journal of Intelligent Information Processing Systems, 2, 54–63.

[15] Aaron, G., & Manning, R. (2012). APWG phishing reports. <http://www.antiphishing.org/resources/apwg-reports/>.

[16] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based Associative Classification data mining," *Expert Systems with Applications,* vol. 41, pp. 5948-5959, 2014. http://dx.doi.org/10.1016/j.eswa.2014.03.019

[17] Sheng, S., Wardman, B., Warner, G., Cranor, L. F., Hong, J., & Zhang, C. (2009). An empirical analysis of phishing blacklists. In CEAS.

[18] Dede(2011).<http://blog.sucuri.net/tag/blacklisted> (accessed June 25, 2013).

[19] Afroz, S., & Greenstadt, R. (2011). PhishZoo: Detecting phishing websites by looking at them. In Proceedings of the 2011 IEEE fifth international conference on semantic computing (ICSC '11) (pp. 368–375). Washington, DC, USA: IEEE Computer Society.

[20] Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010a). Predicting phishing websites using classification mining techniques. In Seventh international conference on information technology; 2010 (pp. 176–181). Las Vegas, Nevada, USA: IEEE.

[21] WEKA (2011). Data Mining Software in Java. Retrieved December 15, 2010 from <http://www.cs.waikato.ac.nz/ml/weka>.

[22] Witten, I., & Frank, E. (2002). Data mining: Practical machine learning tools and techniques with Java implementations. San Francisco: Morgan Kaufmann.

[23] Song, M. (2009). Handbook of research on text and web mining technologies. Information science reference, IGI global.

[24] Pan, Y., & Ding, X. (2006). Anomaly based web phishing page detection. In ACSAC '06: Proceedings of the 22nd annual computer security applications conference (pp. 381–392). Washington, DC: IEEE

[25] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[26] Sadeh, N., Tomasic, A., & Fette, I. (2007). Learning to detect phishing emails. In Proceedings of the 16th international conference on world wide web (pp. 649–656).

[27] Guang, X., Jason, O., Carolyn, P. R., & Lorrie, C. (2011). CANTINA+: A feature-rich machine learning framework for detecting phishing web sites. In ACM transactions on information and system security (pp. 1–28).

[28] Thabtah, F., Eljinini, M., Zamzeer, M., & Hadi, W. (2009). Naïve Bayesian based on chi square to categorize Arabic data. In Proceedings of the 11th international business information management association conference (IBIMA) conference on innovation and knowledge management in twin track economies (pp. 930–935).

[29] Sanglerdsinlapachai, N., & Rungsawang, A. (2010). Using domain top-page similarity feature in machine learning-based web. In Third international conference on knowledge discovery and data mining; 2010 (pp. 187–190). Washington, DC: IEEE.

[30] Marjan Abdeyadan, Ali Rayat Pishes (2016) Detecting Internet Phishing Attacks Using Data Mining Methods of the 3rd International Conference on Innovative Engineering Technologies (ICIET'2016)August 5-6, Bangkok (Thailand).