ISSN No. 0976-5697

Volume 2, No. 1, Jan-Feb 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Fake Website Detection: Association Classification Algorithm with Ant Colony Optimization Technique

Radha Damodaram*
Asst. Professor,
Department of BCA, SS & IT,
CMS College of Science & Commerce
Coimbatore.
radhabalaji10@gmail.com

Dr.M.L.Valarmathi
Asst. Professor,
Dept. of Computer Science & Engg
Government College of Technology,
Coimbatore.
ml_valarmathi@rediffmail.com

Abstract: Phishing website is the process of enticing people to visit fraudulent e-banking websites and persuading them to enter identity information such as user names and passwords. This paper presents a novel approach to overcome the difficulty and complexity in detecting and predicting e-banking phishing websites. The proposed system is an intelligent resilient and effective model that is based on using association and classification Data Mining algorithms combining with Ant Colony Optimization technique. These classification 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. The Ant colony optimization algorithm implemented to detect e-banking phishing websites. The experimental results demonstrated the feasibility of using Associative Classification technique and Ant Colony Optimization in real applications and its better performance.

Key words: Association and Classification, Ant Colony Optimization, Fuzzification, Defuzzification.

I. INTRODUCTION

A. *Phishing*

Phishing is an e-mail fraud method in which the perpetrator sends out legitimate-looking email in an attempt to gather personal and financial information from recipients. Typically, the messages appear to come from well known and trustworthy Web sites. Web sites that are frequently spoofed by phishers include PayPal, eBay, MSN, Yahoo, Best Buy, and America Online. A phishing expedition, like the fishing expedition it's named for, is a speculative venture: the phisher puts the lure hoping to fool at least a few of the prey that encounter the bait[1].

Phishers use a number of different social engineering and e-mail spoofing ploys to try to trick their victims. In one fairly typical case before the Federal Trade Commission (FTC), a 17-year-old male sent out messages purporting to be from america Online that said there had been a billing problem with recipients' AOL accounts. The perpetrator's e-mail used AOL logos and contained legitimate links. If recipients clicked on the "AOL Billing Center" link, however, they were taken to a spoofed AOL Web page that asked for personal information, including credit card numbers; personal identification numbers (PINs), social security numbers, banking numbers, and passwords [2]. This information was used for identity theft.

B. *Data Mining*

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data [3]. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the

relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

C. Existing System

[a] The approach described here is to apply data mining algorithms to assess e-banking phishing website risk on the 27 characteristics and factors which stamp the forged website[4].
[b] Associative and classification algorithms can be very useful in predicting Phishing websites.
[c] It can give us answers about what are the most important e-banking phishing website characteristics and indicators and how they relate with each other.
[d] The choice of PART algorithm is based on the fact that it combines both approaches to generate a set of rules.
[e] Associative classifiers produce more accurate classification models and rules than traditional classification algorithms[5].

D. Objective:

The motivation behind this study is to create a resilient and effective method that uses Data Mining algorithms and tools to detect e-banking phishing websites in an Artificial Intelligent technique. Associative and classification algorithms with Ant Colony Optimization can be very useful in predicting Phishing websites.

E. Proposed system:

This proposed system implements the ant colony optimization algorithm for predicting e-Banking Phishing Websites. This paper presents a novel approach to overcome the 'fuzziness' in the e-banking phishing website assessment and propose an intelligent resilient and effective model for detecting e-banking phishing websites. There is a significant relation between the two phishing website criteria's *(URL & Domain Identity)* and *(Security & Encryption)* for identifying e-banking phishing website. Also found insignificant trivial influence of the *(Page Style & Content)* criteria along with *(Social Human Factor)* criteria for identifying e-banking phishing websites.

II. LITERATURE SURVEY

A. Introduction

"Phishing" is the term for an e-mail scam that spoofs legitimate companies in an attempt to defraud people of personal information such as logins, passwords, credit card numbers, bank account information and social security numbers. For example, an e-mail may appear to come from PayPal claiming that the recipient's account information must be verified because it may have been compromised by a third party. However, when the recipient provides the account information for verification, the information is really sent to a phisher, who is then able to access the person's account. The term phishing was coined because the phishers are "fishing" for personal information. Phishing e-mails are sent to both consumers and companies, trying to gain either personal information from an individual or confidential information about an enterprise. In phishing e-mail messages, the senders must gain the trust of the recipients to convince them to divulge information. The phishers attempt to gain credibility through mimicking or "spoofing" a legitimate company through methods such as using the same logos and color scheme, changing the "from" field to appear to come from someone in the spoofed company, and adding some legitimate links in the e-mail.

One approach is to stop phishing at the email level, since most current phishing attacks use broadcast email (spam) to lure victims to a phishing website. Another approach is to use security toolbars. The phishing filter in IE7 is a toolbar approach with more features such as blocking the user's activity with a detected phishing site. A third approach is to visually differentiate the phishing sites from the spoofed legitimate sites. Dynamic Security Skins proposes to use a randomly generated visual hash to customize the browser window or web form elements to indicate the successfully authenticated sites. A fourth approach is two-factor authentication, which ensures that the user not only knows a secret but also presents a security token [6]. Many industrial anti-phishing products use toolbars in Web browsers, but some researchers have shown that security tool bars don't effectively prevent phishing attacks. Another approach is to employ certification, e.g., Microsoft spam privacy. A variant of web credential is to use a database or list published by a trusted party, where known phishing web sites are blacklisted. The weaknesses of this approach are its poor

scalability and its timeliness. The newest version of Microsoft's Internet Explorer supports Extended Validation (EV) certificates, coloring the URL bar green and displaying the name of the company. However, a recent study found that EV certificates did not make users less fall for phishing attacks.

B. Filtering Approaches for Phishing Email

Phishing emails usually contain a message from a credible looking source requesting a user to click a link to a website where she/he is asked to enter a password or other confidential information. Most phishing emails aim at withdrawing money from financial institutions or getting access to private information. Phishing has increased enormously over the last years and is a serious threat to global security and economy [7]. There are a number of possible countermeasures to phishing. These ranges from communication-oriented approaches like authentication protocols over blacklisting to content-based filtering approaches. The first two approaches are currently not broadly implemented or exhibit deficits. Therefore content-based phishing filters are necessary and widely used to increase communication security. A number of features are extracted capturing the content and structural properties of the email. Subsequently a statistical classifier is trained using these features on a training set of emails labeled as ham (legitimate), spam or phishing. This classifier may then be applied to an email stream to estimate the classes of new incoming emails. This paper describes a number of novel features that are particularly well-suited to identify phishing emails. These include statistical models for the low-dimensional descriptions of email topics, sequential analysis of email text and external links, and the detection of embedded logos as well as indicators for hidden salting. Hidden salting is the intentional addition or distortion of content not perceivable by the reader. For empirical evaluation there is a large realistic corpus of emails pre-labeled as spam, phishing, and ham (legitimate). In experiments our methods outperform other published approaches for classifying phishing emails.

C. Security Toolbars

The first attempts specifically designed to filter phishing attacks have taken the form of browser toolbars, such as the Spoofguard and Netcraft toolbars. Most toolbars are lucky to get 85% accuracy identifying phishing websites. Accuracy aside, there are both advantages disadvantages to toolbars when compared to email filtering.

The first disadvantage toolbars face when compared to email filtering is a decreased amount of contextual information. The email provides the context under which the attack is delivered to the user. An email filter can see what words are used to entice the user to take action, which is currently not knowable to a filter operating in a browser separate from the user's e-mail client [8]. An email filter also has access to header information, which contains not only information about who sent the message, but also information about the route the message took to reach the user. This context is not currently available in the browser with given toolbar implementations. Future work to more closely integrate a user's email environment with their browser could alleviate these problems, and would actually provide a potentially richer context in which to make a decision.There are some pieces of information available in the web browser and website itself that could help to make a more informed decision, especially if this information could be combined with the context from the initial attack vector, such as the email prompting a user to visit a given website.

The second disadvantage of toolbars is the inability to completely shield the user from the decision making process. Toolbars usually prompt users with a dialog box, which many users will simply dismiss or misinterpret, or worse yet these warning dialogs can be intercepted by user-space malware. By filtering out phishing emails before they are ever seen by users, avoid the risk of these warnings being dismissed by or hidden from the user. Also prevent the loss of productivity suffered by a user who has to take time to read, process, and delete these attack emails.

D. Html emails

Most emails are sent as either plain text, HTML, or a combination of the two in what is known as a multipart/alternative format. The email is flagged with the HTML email feature if it contains a section that is denoted with a MIME type of text/html. (This includes many multipart/alternative emails). While HTML email is not necessarily indicative of a phishing email, it does make many of the deceptions seen in phishing attacks possible. For a phisher to launch an attack without using HTML is difficult,

because in a plain text email there is virtually no way to disguise the URL to which the user is taken. Thus, the user still can be deceived by legitimate-sounding domain names, but many of the technical, deceptive attacks are not possible. This is a binary feature.

E. Contains Javascript

JavaScript is used for many things, from creating popup windows to changing the status bar of a web browser or email client. It can appear directly in the body of an email, or it can be embedded in something like a link. Attackers can use JavaScript to hide information from the user, and potentially launch sophisticated attacks. An email is flagged with the "contains javascript" feature if the string "javascript" appears in the email, regardless of whether it is actually in a <script> or <a> tag. This might not be optimal, but it makes parsing much simpler, especially when dealing with attacks that contain malformed HTML. This is a binary feature.

F. Blacklists

Blacklists hold URLs (or parts thereof) that refer to sites that are considered malicious. Whenever a browser loads a page, it queries the blacklist to determine whether the currently visited URL is on this list. If so, appropriate countermeasures can be taken. Otherwise, the page is considered legitimate. The blacklist can be stored locally at the client or hosted at a central server. Obviously, an important factor for the effectiveness of a blacklist is its coverage. The coverage indicates how many phishing pages on the Internet are included in the list. Another factor is the quality of the list. The quality indicates how many non-phishing sites are incorrectly included into the list. For each incorrect entry, the user experiences a false warning when she visits a legitimate site, undermining her trust in the usefulness and correctness of the solution.

Finally, the last factor that determines the effectiveness of a blacklist-based solution is the time it takes until a phishing site is included. This is because many phishing pages are short-lived and most of the damage is done in the time span between going online and vanishing. Even when a blacklist contains many entries, it is not effective when it takes too long until new information is included or reaches the clients. The study attempted to measure the effectiveness of popular black-lists. In particular, study about the blacklists maintained by Microsoft and Google. These blacklists are the ones that are most wide-spread, as they are used by Internet Explorer and Mozilla Firefox, respectively.

Page analysis techniques examine properties of the web page and the URL to distinguish between phishing and legitimate sites. Page properties are typically derived from the page's HTML source. Examples of properties are the number of password fields, the number of links, or the number of unencrypted password fields (these are properties used by SpoofGuard). The effectiveness of page analysis approaches to identify phishing pages fun-dementedly depend on whether page properties exist that allow to distinguish between phishing and legitimate sites. Thus, the aim is to determine whether these properties exist, and if so, why they might be reasonable candidates to detect phishing pages.

In a first step, This approach defines a large number of page properties that can be extracted from the page's HTML code and the URL of the site. Then, it analyzes a set of phishing and legitimate pages, assigning concrete values to the properties for each page. Finally, using the collected data as training input, applied machine-learning techniques to create a web page classifier. The resulting classifier is able to distinguish well between phishing and legitimate classifiers, with a very low false positive rate. This indicates that the aforementioned page properties that allow one to identify malicious pages do indeed exist, at least for current phishing pages. It seems that Microsoft has drawn a similar conclusion, as the new Internet Explorer browser also features a phishing page detection component based on page properties. This component is invoked as a second line of defense when a blacklist query returns no positive result for a visited URL.

G. Two Way Factor Authentication

An authentication factor is a piece of information and process used to authenticate or verify the identity of a person or other entity requesting access to online resources. User authentication for most web sites and services today is accomplished by means of a single authentication factor: a password [9]. Where a higher level of assurance is required (e.g. for access to on online banking service), a second factor is typically employed in addition to the password – hence "two factor authentication" (also called "multi factor authentication" or "strong authentication").

There are three main types of authentication factor:

[a] knowledge factors – e.g. passwords, PINs;
[b] Possession factors – e.g. ID cards, tokens;
[c] Human factors (aka biometrics) – e.g. fingerprints, iris scans.

Some security practitioners argue that "true" two factor authentication requires two distinct types of factor; however, this is just a matter of semantics. There is nothing inherently less secure about using two factors of the same type.

H. Needs for Two Factor Authentication:

Passwords alone provide very poor security. They can be guessed, phished and hacked and are clearly inadequate to protect high value online services such as Internet banking. Indeed, the Federal Financial Institutions Examination Council (FFIEC – the body responsible for promoting uniformity in the supervision of US financial institutions) has mandated two factor authentication for consumer online banking services.

Compliance is also driving adoption of two factor authentication in other areas – for example, the Health Insurance Portability and Accountability Act (HIPAA) in healthcare, where the important issue is the confidentiality of user data (patient records). And as more and more of our personal information goes online, privacy – and the threat of identity theft – is increasingly an issue in applications as diverse as gaming and dating and as common as Facebook. Further requirements for two factor authentication include: protection of company confidential data (e.g. customer information on salesforce.com), controlling access paid-for content (e.g. music/video downloads from iTunes) and, perhaps most importantly, demonstrating due care to customers and users.

III. METHODOLOGIES

A. Extracting phishing characteristics

Two publicly available datasets were used to test our implementation: the "phishtank" from the phishtank.com which is considered one of the primary phishing report collators. The PhishTank database records the URL for the suspected website that has been reported, the time of that report, and sometimes further detail such as the screenshots of the website, and is publicly available. A java program is used to extract the above features, and store these in database for quick reference.

Our goal is to gather information about the strategies that are used by attackers and to formulate hypotheses about classifying and categorizing of all different e-banking phishing attacks techniques. The following Table.1 consists of the details of criteria and phishing indicators for each criteria [10].

Table 1: Phishing Indicators and their criteria

Criteria	Phishing Indicators
URL & Domain Identity	Using IP address
	Abnormal request URL
	Abnormal URL of anchor
	Abnormal DNS record
	Abnormal URL
Security & Encryption	Using SSL Certificate
	Certificate authority
	Abnormal cookie
	Distinguished names certificate
Source Code & Java script	Redirect pages
	Straddling attack
	Pharming attack
	On Mouse over to hide the Link
	Server Form Handler (SFH)
Page Style & Contents	Spelling Errors
	Copying website
	Using form s with Submit button
	Using pop-ups windows
	Disabling right-click
Web Address Bar	Long URL address
	Replacing similar char for URL
	Adding a prefix or suffix
	Using the @ Symbols to confuse
	Using hexadecimal char codes
Social Human Factor	Emphasis on security
	Public generic salutation
	Buying time to access accounts

[a] Fuzzification

In this step, linguistic descriptors such as High, Low, Medium, for example, are assigned to a range of values for each key phishing characteristic indicators. Valid ranges of the inputs are considered and divided into classes, or fuzzy sets. For example, length of URL address can

range from 'low' to 'high' with other values in between. This cannot specify clear boundaries between classes. The degree of belongingness of the values of the variables to any selected class is called the degree of membership; Membership function is designed for each Phishing characteristic indicator, which is a curve that defines how each point in the input space is mapped to a membership value between [0, 1].

Linguistic values are assigned for each Phishing indicator as Low, Moderate, and high while for e-banking Phishing website risk rate as Very legitimate, Legitimate, Suspicious, Phishy, and Very phishy (triangular and trapezoidal membership function). For each input their values ranges from 0 to 10 while for output, ranges from 0 to 100. An example of the linguistic descriptors used to represent one of the key phishing characteristic indicators (URL Address Long)[11].

[b] Rule Generation using Associative Classification Algorithms

To derive a set of class association rules from the training data set, it must satisfy certain user-constraints, i.e support and confidence thresholds. Generally, in association rule mining, any item that passes MinSupp is known as a frequent item. The recorded prediction accuracy and the number of rules generated by the classification algorithms and a new associative classification MCAR algorithm [12]. Error rate comparative having specified the risk of e-banking phishing website and its key phishing characteristic indicators, the next step is to specify how the e-banking phishing website probability varies. Experts provide fuzzy rules in the form of **if…then** statements that relate e-banking phishing website probability to various levels of key phishing characteristic indicators based on their knowledge and experience. On that matter and instead of employing an expert system, the utilized data mining classification and association rule approaches in our new e-banking phishing website risk assessment model which automatically finds significant patterns of phishing characteristic or factors in the e-banking phishing website archive data

[c] Aggregation of the rule outputs

This is the process of unifying the outputs of all discovered rules. Combining the membership functions of all the rules consequents previously scaled into single fuzzy sets (output).

[d] Defuzzification

This is the process of transforming a fuzzy output of a fuzzy inference system into a crisp output. Fuzziness helps to evaluate the rules, but the final output has to be a crisp number. The input for the defuzzification process is the aggregate output fuzzy set and the output is a number. This step was done using Centroid technique since it is a commonly used method. The output is e-banking phishing website risk rate and is defined in fuzzy sets like '**very phishy**' to '**very legitimate'**. The fuzzy output set is then defuzzified to arrive at a scalar value [13, 14].

[e] Ant Colony Optimization (Enhancement)

The original idea comes from observing the exploitation of food resources among ants, in which ants' individually limited cognitive abilities have collectively been able to find the shortest path between a food source and the nest.

[a] The first ant finds the food source (F), via any way (a), then returns to the nest (N), leaving behind a trail pheromone (b)

[b] Ants indiscriminately follow four possible ways, but the strengthening of the runway makes it more attractive as the shortest route.

[c] Ants take the shortest route, long portions of other ways lose their trail pheromones.

[f] Performance Comparisons:

The performance analysis of the proposed system is compared with the existing system with the performance metrics mentioned.

Error rate: The proposed algorithm will get the less error rate when compared to the existing algorithm.[15]

Correct prediction: the proposed algorithm predicts the phishing website more accurate than the existing algorithm.

B. Pseudo code Web Phishing

Input: Webpage URL
Output: Phishing website identification
Step 1: Read web phishing URL
Step 2: Extract all 27 feature
Step3: 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: De-fuzzification of fuzzy values into original values.

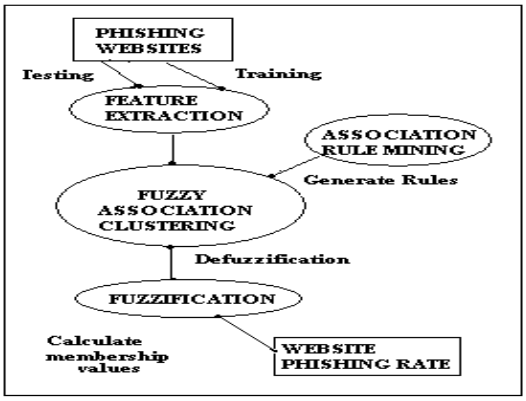

Figure. 1 detecting the website

Step 7: Apply rule on test data and find whether the site is phishing or not and these steps are shown in Fig.1.

IV. IMPLEMENTATION

A. Ant Colony Optimization:

The ant colony optimization algorithm (ACO), is a probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs. This algorithm is a member of ant colony algorithms family, in swarm intelligence methods, and it constitutes some meta-heuristic optimizations. The original idea comes from observing the exploitation of food resources among ants, in which ants' individually limited cognitive abilities have collectively been able to find the shortest path between a food source and the nest[16].

[a] The first ant finds the food source (F), via any way (a), then returns to the nest (N), leaving behind a trail pheromone (b)
[b] Ants indiscriminately follow four possible ways, but the strengthening of the runway makes it more attractive as the shortest route.
[c] Ants take the shortest route, long portions of other ways lose their trail pheromones.

In a series of experiments on a colony of ants with a choice between two unequal length paths leading to a source of food, biologists have observed that ants tended to use the shortest route. A model explaining this behaviour is as follows:

[a] An ant (called "blitz") runs more or less at random around the colony;
[b] If it discovers a food source, it returns more or less directly to the nest, leaving in its path a trail of pheromone;
[c] These pheromones are attractive, nearby ants will be inclined to follow, more or less directly, the track;
[d] Returning to the colony, these ants will strengthen the route;
[e] If two routes are possible to reach the same food source, the shorter one will be, in the same time, traveled by more ants than the long route will
[f] The short route will be increasingly enhanced, and therefore become more attractive;
[g] The long route will eventually disappear, pheromones are volatile;
[h] Eventually, all the ants have determined and therefore "chosen" the shortest route.

The design of an ACO algorithm implies the specification of the following aspects.

[i] An environment that represents the problem domain in such a way that it lends itself to incrementally building a solution to the problem.
[ii] A problem dependent heuristic evaluation function, which provides a quality measurement for the different solution components.
[iii] A pheromone updating rule, which takes into account the evaporation and reinforcement of the trails.
[iv] A probabilistic transition rule based on the value of the heuristic function and on the strength of the pheromone trail that determines the path taken by the ants.
[v] A clear specification of when the algorithm converges to a solution [17].

The ant system simply iterates a main loop where *m* ants construct in parallel their solutions, thereafter updating the trail levels. The performance of the algorithm depends on the correct tuning of several parameters, namely: a, b, relative importance of trail and attractiveness, r, trail persistence, t*ij*(0), initial trail level, *m*, number of ants, and Q, used for defining to be of high quality solutions with low cost[18]. The ANTS algorithm is the following.

[a] Compute a (linear) lower bound LB to the problem Initialize t*iy* ("i,y) with the primal variable values .
[b] For k=1,m (m= number of ants) do repeat
[i] compute h*iy* "(iy)
[ii] choose in probability the state to move into
[iii] append the chosen move to the *k*-th ant's tabu list until ant *k* has completed its solution
[iv] carry the solution to its local optimum end for

[c] For each ant move (iy), compute Dt*iy* and update trails by means of (5.6)

[d] If not (end test) **go to** step 2[19].

V. RESULTS AND DISCUSSION

There is a significant relation between the two phishing website criteria's *(URL & Domain Identity)* and *(Security & Encryption)* for identifying e-banking phishing website[20]. Also found insignificant trivial influence of the *(Page Style & Content)* criteria along with *(Social Human Factor)* criteria for identifying e-banking phishing websites. Ant colony optimization produces more accurate classification models than Associative classifiers.

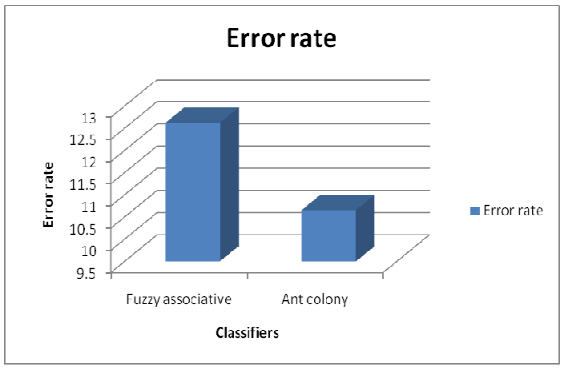

Figure. 2 Error rate comparision with fuzzy associative algorithm

The Fig.2 shows the comparision of fuzzy associative classifiers and ant colony optimization with the error rate.

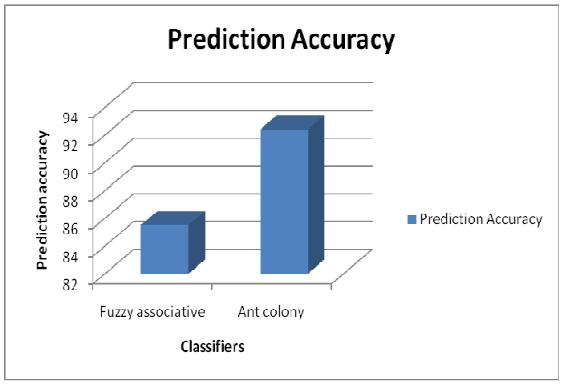

Figure. 3 Accuracy report

The results shows that the ACO produce less error rate than associative classifier. Selected 550 cases randomly used for inducing rules from 650 cases in original data set, the remaining 100 cases are used for testing accuracy of the induced rules of the proposed method by measuring the average percentage of correct predictions.

The Fig.3 shows the comparision of ant colony algorithm and fuzzy associative algorithms in terms of prediction accuracy.

VI. SCREEN SHOTS

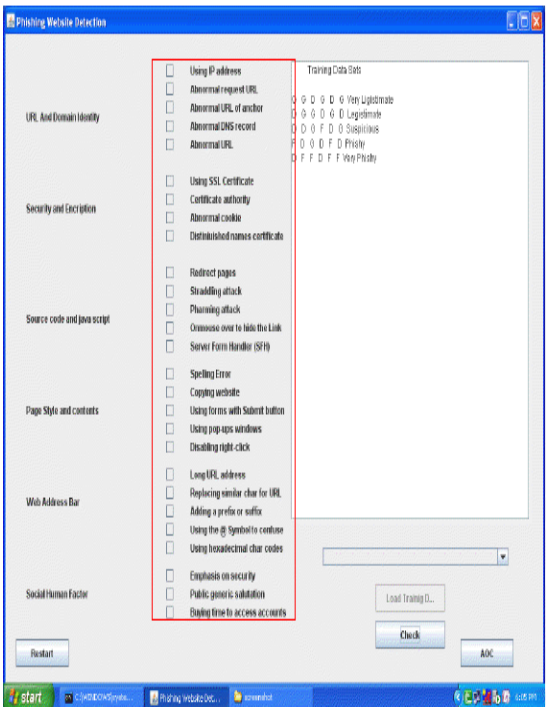

Figure. 4 loading website

Fig.4 shows the screen shot which contains 27 characteristics and factors which stamp the forged website (within red rectangular box).

Fig.5 Shows loading http:/ /graandchase .uni7.net website into the software. After loading this link the detecting process starts when check button pressed. Fig.6 shows the values entered for number of folds number of ants, rule of convergence etc. The proposed algorithm divides the ants' population into multiple colonies (folds) and effectively coordinates their works.

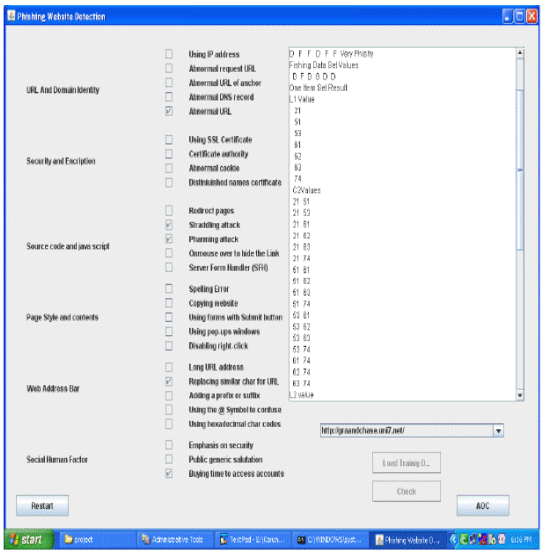

Figure. 5 Result displaying while detecting

An average and maximum pheromone evaluation functions are used in the process of the ant's decision making.

The results show that the proposed algorithm outperforms the ACO algorithm with similar number of ants. .After pressing the submit button the result will be as shown in Fig.7.

According this, the website is phishy because it cosists errors which circled in red color. Straddling attack is a noncommittal or equivocal position in networks and Pharming is a hacker's attack aiming to redirect a website's traffic to another bogus website.

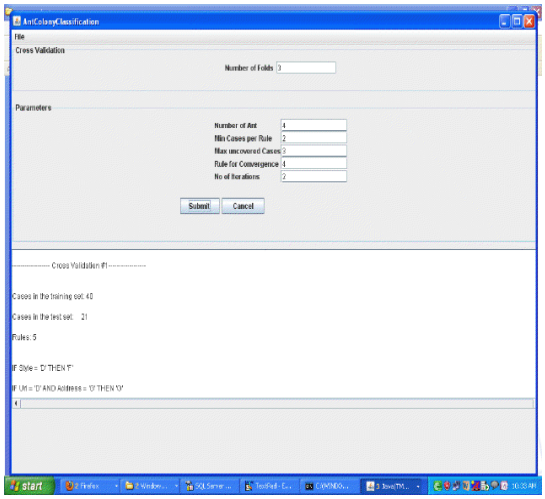

Figure. 6 Applying ant colony algorithm

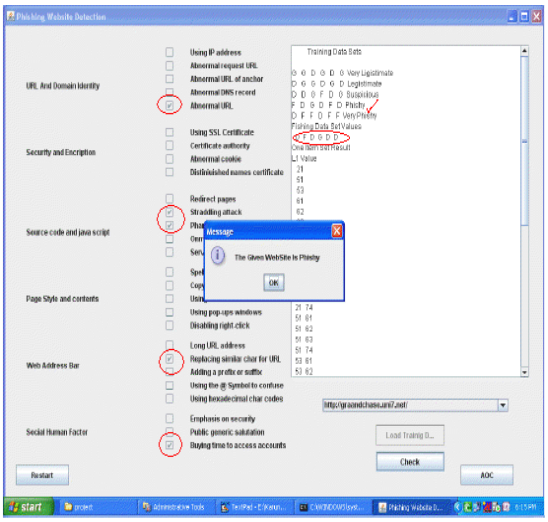

Fig. 7 Result after Associative & Ant Colony algorithms application

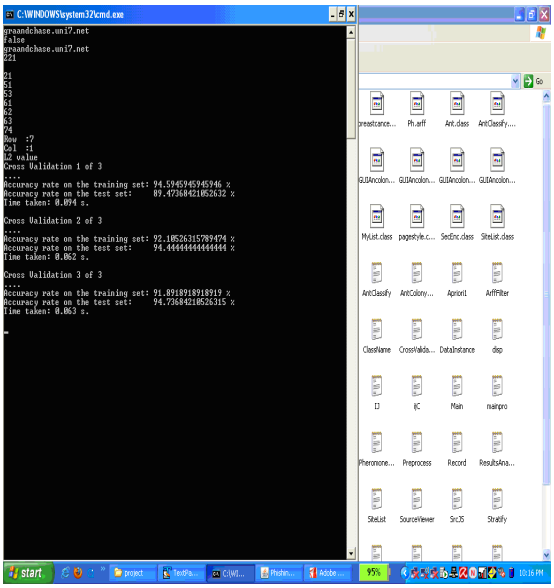

Figure.8 Accuracy and time showing after detection

Pharming can be conducted either by changing the hosts file on a victim's computer or by exploitation of a vulnerability in DNS server software. Fig.8 shows the practical part of comparative study that utilizes Fuzzy and ACO algorithms. Our choice of these methods is based on the different strategies they used in learning rules from data sets.

Finally, the experiment results showed that the proposed associative classification algorithm with ant colony optimization technique out

performed all other traditional classifications in terms of accuracy (94%) and speed (0.063seconds) since it requires only one phase to discover frequent items and rules.

VII. CONCLUSION

The Associative Classification Algorithm with Ant Colony Optimization Technique for e-banking phishing website detection model is outperformed when compared with existing classification algorithms in terms of prediction accuracy and error rate. The ant colony optimization algorithm is a probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs. The associative classification algorithm for e-banking phishing website model showed the significance importance of the phishing website in two criteria's (URL & Domain Identity) and (Security & Encryption) with insignificant trivial influence of some other criteria like 'Page Style & content' and 'Social Human Factor'[21]. Combining these two techniques has given a fruitful result. After more than 500 websites detection for both its application effectiveness and its theoretical groundings, ACO became one of the most successful paradigms in network security.

VIII. REFERENCES

1. http://commons.wikimedia.org/wiki/File:Phishing_info_graph.svg ,http:/ /www .gartner. com /it/ page
2. STAMFORD, Conn., (April 14, 2009). "Gartner Says Number of Phishing Attacks on U.S. Consumers Increased 40 Percent in 2008". Gartner. "UK phishing fraud losses double". Finextra. March 7, 2006. http:// www. finextra. com/ fullstory asp?id=15013.
3. Richardson, Tim (May 3, 2005). "Brits fall prey to phishing". The Register. http:/ /www .theregister.co.uk/2005/05/03/aol_phishing/.
4. Miller, Rich. "Bank, Customers Spar over Phishing Losses". Netcraft. http://news.netcraft .com/ rchives/ 2006/09.
5. Associative Classification Techniques for predicting e-Banking Phishing Websites, Maher Aburrous Dept. of Computing, Universit y of BradfordBradford, UK.
6. GARTNE R, INC. Gartner Says Number of Phishing Emails Sent to U.S. Adults Nea rly Doubles in Just Two Years, http //www .gartner. com/ it/pag e.jsp3.
7. T.Moore and R. Clayton, "An empirical analysis of the current state of phishing attack and defence", In Proceedings of the Workshop on the Economics of Information Security (WEIS2007)
8. Antiphishing Toolbars the comparision with existing toolbars, IJCA November, 2009
9. A. Hossain, M. Dorigo, The Two Way Authentication Approach, http:// iridia.ulb.ac.be /mdorigo/ TWA.html
10. Ant Colony Optimization, Vittorio Maniezzo, Luca Maria Gambardella, Fabio de Luig**i.**
11. Mining Fuzzy Weighted Association Rules Proceedings of the 40th Hawaii International Conference on System Sciences – 2007.
12. WEKA - University of Waikato, New Zealand, EN, 2006: "Weka -Data Mining with Open Source Machine Learning Software in Java", 2006,
13. Technische Universität Berlin, Germany, 1995 M. E. Bergen, Canstraint-based assembly line sequencing, Lecture Notes in Computer
14. A. Hossain, M. Dorigo, Ant colony optimization web page, http:// iridia.ulb.ac.be /mdorigo/ACO / ACO. html N. Ascheuer, Hamiltonian path problems
15. L. Bianchi, L.M. Gambardella, M.Dorigo. An ant colony optimization approach to the probabilistic traveling salesman problem. In Proceedings of PPSN-VII, Seventh Inter GARTNE R, INC.
17 B. Adida, S. Hohenberger and R. Rivest , "Lightweight Encryption for Email," USENIX Steps to Reducing Unwanted Traffic on the Internet (SRUTI), 2005 ,
18. Ant Colony Optimization, www. Scholar pedia.org article/AntColonyOptimization.
19. "The Techniques of ACO", www .csse .monash Edu.au/~bemdm/CSE460/Lectures./ Cse4609.pdf
20 'Particle Swarm Optimization", www.swaminte Lligence.org.
21. Bing Liu, Wynne Hsu, Yiming Ma, "Integrating Classification and Association Rule Mining." Proceedings ofthe Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98, Plenary Presentation), New York, USA.