IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 2 ISSUE : 11 Print / Issue Publication Date: 27-Jan-2021

ISSN : 2455-2143

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

PHISHING URL DETECTION USING PARTICLE SWARM OPTIMIZATION

Mr. Nayyar Ahmed Khan
Research Scholar
Department of Computer Science and Engineering
Himalayan University, Arunachal Pradesh, India

Dr. Sanjay Pachauri
Director Research, CSE/IT
IIMT College of Engineering
Greater Noida, U.P.

Abstract - Fake websites is the process of attracting people to visit fraudulent websites and making them to enter confidential data like credit-card numbers, usernames and passwords. We present a novel approach to overcome the difficulty and complexity in detecting and predicting fake website. There is an efficient 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 behaviour in the presence of phishing websites. This will improve the correctly classified phishing websites. The experimental results demonstrated the feasibility of using PSO technique in real applications and its better performance. This project employs the JAVA technology.

Keywords: Fake Website, Association and Classification Technique, ACO, PSO

I. INTRODUCTION

1.1 Phishing

In the computer field stealing information is a simple trick in which the hacker creates the duplicate page of a site and asks you to enter your details or credentials there. When you enter credentials such as username, password or credit card number the whole data goes to the hacker which he/she later use[4]. Phishing is an illegal process and there are many hackers who are behind the bars because of this fraudulent process called "Phishing". Phishing is nowadays a major problem for the users of social networking sites and net banking. There are many ways through which you can identify a phishing website. Some ways to detect phishing sites are:

- **Uniform Resource Locator (URL)**: Check always the web address or URL of the site you are visiting. Fake websites have different addresses than those of the genuine websites.
- **Install Anti-Phishing Software:** Use anti-phishing software to detect phishing sites.
- Browsers with inbuilt Software: Always work with the browsers who have anti-phishing software inbuilt in them. you may use opera, mozilla firefox, safari, google chrome, internet explorer 8 etc[5].
- **Slashes missing**: A fake website sometimes doesn't have slashes in-between its URL address. For example "http://www.scramable.com:login&mode= secured" but the original one is" http://www.scramable.com/wp_admin"[13]
- **Using fake passwords**: Always type your fake password on the websites that you are visiting first time and later you can change it.
- **Searching More:** Always use to search on good search engines like Google.
- **More information:** Always get more and more information about the site as phishing websites don't hold for too long.

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

- **Totally avoid Pop-Ups**: Never login with your original password in the pop-up windows as a hacker may ask you to enter the password in the pop up and after he redirects you to the exact page[17].

1.2 Phishing Nature and Detection

Phishing websites work by impersonating legitimate websites, and they have a very short life time. On average a phishing website lasts 62 hours, and users rarely visit the phishing website prior to the attack. Secondly, phishing attacks always generate mismatches between a user's perception and the truth. In successful web based phishing attacks, victims have believed they are interacting with websites which belong to legitimate and reputable organizations or individuals. Thus the crucial mismatch that phishers create is one of real *versus* apparent identity. Phishing attacks can be detected if we can detect such a mismatch. One approach is to predict a user's perception and then compare it with the actual fact understood by the system. CANTINA is an example of this approach [18].

The main issue with this approach is that the data the system relies on is under the control of attackers, and there are so many techniques that attackers can apply to manipulate the data to easily evade the detection. For CANTINA, attackers could use images instead of text in the body of the webpage, they could use iframes to hide a large amount of content from the users while computer programs can still see it; they could use Java script to change the content of the page after the detection has been done. The authentication credentials, which phishers try to elicit, ought to be shared only between users and legitimate organizations. Such (authentication credential, legitimate website) pairs are viewed as the user's binding relationships. In legitimate web authentication interactions, the authentication credentials are sent to the website they have been bound to [21]. In a phishing attack the mismatches cause the user to unintentionally break binding relationships by sending credentials to a phishing website. No matter what spoofing techniques or deception methods used, nor how phishing WebPages are implemented, the mismatch and violation of the binding relationships always exists. So, one can discover the mismatch by detecting violation of users' binding relationships. Hence phishing websites can be detected when both of the following two conditions are met:

1) The current website has rarely or never been visited before by the user;

2) The data, which the user is about to submit, is bound to website other than the current one. This detection process flow shown in Fig.1.1.

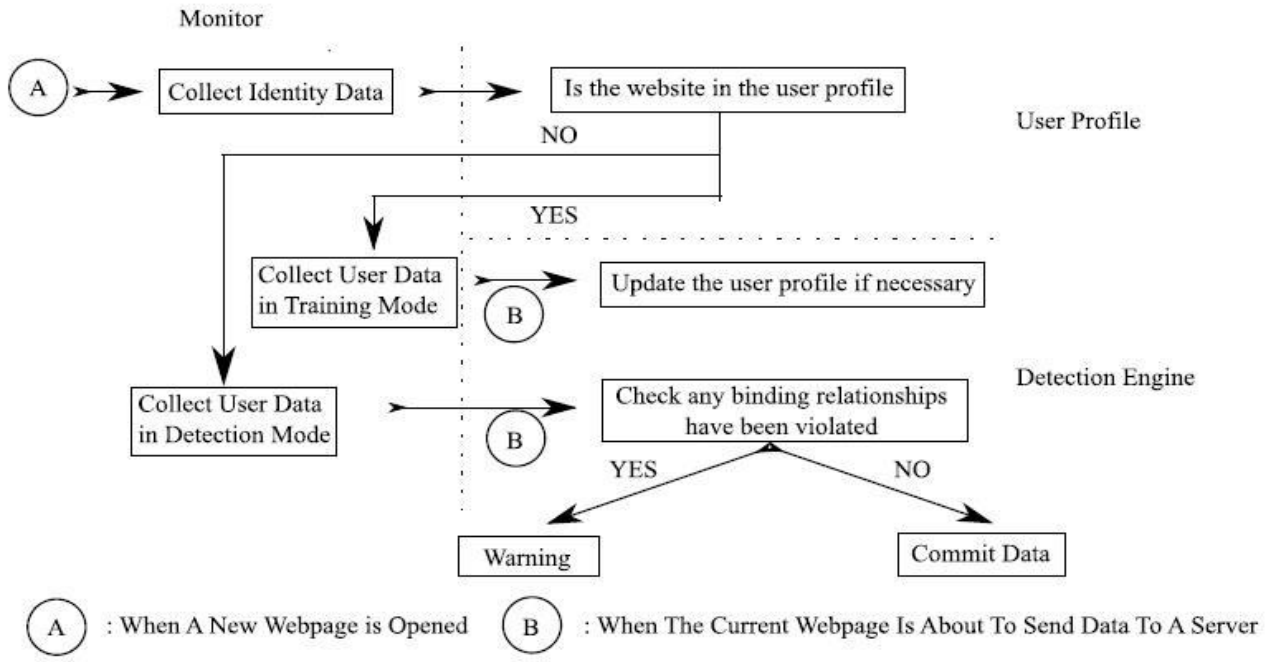

FIGURE 1.1 Detection Process Work Flow [20]

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

13. Ant Colony Optimization

The Ant Colony System or the basic idea of an ant food searching system is illustrated in Fig.1.2. In the left picture, the ants move in a straight line to the food. The next picture shows the situation soon after an obstacle is inserted between the nest and the food. To avoid the obstacle, first each ant chooses to turn right or left at random. Let us assume that ants move at the same speed depositing pheromone in the trail equivalently. However, the ants that, by chance, choose to turn right will reach the food sooner, whereas the ants that go around the obstacle turning right will follow a longer path, and so will take long time to circumvent the obstacle. As a result, pheromone gathered faster in the shorter path around the obstacle. Since ants prefer to follow trails with larger amounts of pheromone, ultimately all the ants congregate to the shorter path around the obstacle [1].

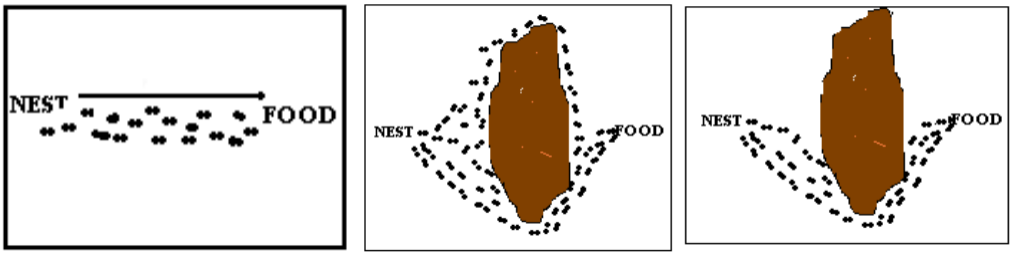

FIGURE 1.2- Illustrating the behaviour of real ant movements.

This new heuristic, called Ant Colony Optimization (ACO) has been found to be both robust and versatile in handling a wide range of combinatorial optimization problems. The main idea of ACO is to model a problem as the search for a minimum cost path in a graph. Artificial ants as if walk on this graph, looking for cheaper paths. Each ant has a rather simple behaviour capable of finding relatively costlier paths. Cheaper paths are found as the emergent result of the global cooperation among ants in the colony. The behaviour of artificial ants is inspired from real ants: they lay pheromone trails (obviously in a mathematical form) on the graph edges and choose their path with respect to probabilities that depend on pheromone trails. These pheromone trails progressively decrease by evaporation. In addition, artificial ants have some extra features not seen in their counterpart in real ants. In particular, they live in a discrete world (a graph) and their moves consist of transitions from nodes to nodes.

The ACO differs from the classical ant system in the sense that here the pheromone trails are updated in two ways. Firstly, when ants construct a tour they locally change the amount of pheromone on the visited edges by a local updating role. Secondly, after all the ants have built their individual tours, a global updating rule is applied to modify the pheromone level on the edges that belong to the best ant tour found so far [2].

1.5 Particle Swarm Optimization

Particle Swarm Optimization (PSO) technique is a modelled algorithm on Swarm intelligence, that gets a solution to an optimization problem in a search place, or model and predict social behaviour in the presence of objectives. The PSO is a stochastic, population-based computer algorithm modelled on swarm intelligence. Swarm intelligence is based on social-psychological principles and provides insights into social behaviour, as well as contributing to engineering applications.

The particle swarm optimization algorithm was first described in 1995 by James Kennedy and Russell C. Eberhart. The particle swarm simulates this kind of social optimization. A problem is given, and some way to evaluate a proposed solution to it exists in the form of a fitness function[9]. A communication structure or social network is also defined, assigning neighbors for each individual to interact with. Then a population of individuals defined as random guesses at the problem solutions is initialized. These individuals are candidate solutions. They are also known as the particles, hence the name particle swarm. An iterative process to improve these candidate solutions is set in motion. The particles iteratively evaluate the fitness of the candidate solutions and remember the location where they had their best success. The individual's best solution is called the particle best or the local best. Each particle makes this information available to their neighbors.

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

They are also able to see where their neighbors have had success. Movements through the search space are guided by these successes, with the population usually converging, by the end of a trial, on a problem solution better than that of non-swarm approach using the same methods. Each particle represents a candidate solution to the optimization problem. The position of a particle is influenced by the best position visited by itself i.e. its own experience and the position of the best particle in its neighborhood i.e. the experience of neighboring particles. When the neighborhood of a particle is the entire swarm, the best position in the neighborhood is referred to as the global best particle, and the resulting algorithm is referred to as the gbest PSO. When smaller neighborhoods are used, the algorithm is generally referred to as the lbest PSO. The performance of each particle is measured using a fitness function that varies depending on the optimization problem[19].
Each Particle in the swarm is represented by the following characteristics:

1. The current position of the particle
2. The current velocity of the particle

The particle swarm optimization which is one of the latest evolutionary optimization techniques conducts searches uses a population of particles.

II. OVERVIEW

2.1 Existing System

In the existing system, we implemented the Ant colony optimization technique to optimize the detected e-banking phishing websites. This will improve the correctly classified phishing websites. Ant Colony Optimization (ACO) is a paradigm for designing metaheuristic algorithms for combinatorial optimization problems. The essential trait of ACO algorithms is the combination of prior information about the structure of a promising solution with a posteriori information about the structure of previously obtained good solutions. The functioning of an ACO algorithm can be summarized as follows:
A set of computational concurrent and asynchronous agents (a colony of ants) moves through states of the problem corresponding to partial solutions of the problem to solve. They move by applying a stochastic local decision policy based on two parameters, called *trails* and *attractiveness [12]*. By moving, each ant incrementally constructs a solution to the problem. When an ant after completion or during the construction phase of a solution, it evaluates and modifies the trail value of the components used in its solution. This pheromone information will direct the search of the future ants.
Furthermore, an ACO algorithm includes two more mechanisms*:* trail evaporation and, optionally, daemon actions. Trail evaporation decreases all trail values over time, in order to avoid unlimited accumulation of trails over some component. Daemon actions can be used to implement centralized actions which cannot be performed by single ants, such as the invocation of a local optimization procedure, or the update of global information to be used to decide whether to bias the search process from a non-local perspective. More specifically, an *ant* is a simple computational agent, which iteratively constructs a solution for the instance to solve. Partial problem solutions are seen as *states*. At the core of the ACO algorithm lies a loop, where at each iteration, each ant *moves* (performs a *step*) from a state i to another one ψ, corresponding to a more complete partial solution. Trails are *updated* usually when all ants have completed their solution, increasing or decreasing the level of trails corresponding to moves that were part of "good" or "bad" solutions, respectively [3].

2.2 Objective

The objective is 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 can be very useful in predicting Phishing websites. We implement the Ant colony optimization algorithm to detect e-banking phishing websites. This will improve the correctly classified phishing websites. 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.

2.3 Proposed System

In the proposed system, we overcome the limitation of ACO like Sequences of random decisions (not independent) and Time to convergence uncertain in the phishing classification. 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.
Particle Swarm Optimization (PSO) is an evolutionary technology (evolutionary

computation). Predatory birds originated from the research PSO with similar genetic algorithm is based on iterative optimization tools. Initialize the system for a group of random solutions, through iterative search for the optimal values. However, there is no genetic algorithm with the cross-(crossover) and the variation (mutation). But particles in the solution space following the optimal particle search. The steps detailed chapter on the future of genetic algorithm, the advantages of PSO is simple and easy to achieve without many parameters need to be adjusted. It has been widely used function optimization, neural networks, fuzzy systems control and other genetic algorithm applications [20].Now we are using the same in website phishing field.

III. METHODOLOGY

3.1. Extracting Phishing Characteristics Attribute

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. We use a java program 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.

3.2. 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. We 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. The following list consists of the details of criteria and phishing indicators for each criterion [10].

URL & Domain Identity

1. Using IP address
2. Abnormal request URL
3. Abnormal URL of anchor
4. Abnormal DNS record
5. Abnormal URL

Security & Encryption

1. Using SSL Certificate
2. Certificate authority
3. Abnormal cookie
4. Distinguished names certificate

Source Code & Java script

1. Redirect pages
2. Straddling attack
3. Pharming attack
4. On Mouse over to hide the Link
5. Server Form Handler (SFH)

Page Style & Contents

1. Spelling Errors
2. Copying website
3. Using forms with Submit button
4. Using pop-ups windows
5. Disabling right-click

Web Address Bar

1. Long URL address
2. Replacing similar char for URL
3. Adding a prefix or suffix
4. Using the @ Symbols to confuse
5. Using hexadecimal char codes

Social Human Factor

1. Emphasis on security
2. Public generic salutation
3. Buying time to access accounts

3.3 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, ie support and confidence thresholds. Generally, in association rule mining, any item that passes Min-Supp is known as a frequent item. We recorded the prediction accuracy and the number of rules generated by the classification algorithms and

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

a new associative classification MCAR algorithm. 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[13]. On that matter and instead of employing an expert system, we 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 [6,7].

3.4. 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).

3.5. 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 [7, 8].

3.6. Ant Colony Optimization

The characteristic of ACO algorithms is their explicit use of elements of previous solutions. 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.

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

3.7 Particle Swarm Optimization

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.

A basic variant of the PSO algorithm works by having a population (called a swarm) of candidate solutions (called particles). These particles are moved around in the search-space according to a few simple formulae. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known position. When improved positions are being discovered these will then come to guide the movements of the swarm. The process is repeated and by doing so it is hoped, but not guaranteed, that a satisfactory solution will eventually be discovered[14].

Formally, let $f: \mathbb{R}^n \to \mathbb{R}$ be the fitness or cost function which must be minimized. The function takes a candidate solution as argument in the form of a vector of real numbers and produces a real number as output which indicates the fitness of the given candidate solution. The gradient of f is not known. The goal is to find a solution $\mathbf{a}$ for which $f(\mathbf{a}) \leq f(\mathbf{b})$ for all $\mathbf{b}$ in the search-space, which would mean $\mathbf{a}$ is the global minimum. Maximization can be performed by considering the function $h = -f$ instead. Let S be the number of particles in the swarm, each having a position $\mathbf{x}_i \in \mathbb{R}^n$ in the search-space and a velocity $\mathbf{v}_i \in \mathbb{R}^n$. Let $\mathbf{p}_i$ be the best known position of particle i and let $\mathbf{g}$ be the best known position of the entire swarm. A basic PSO algorithm is then:

- For each particle $i = 1, ..., S$ do:
  - Initialize the particle's position with a uniformly distributed random vector: $\mathbf{x}_i \sim U(\mathbf{b}_{lo}, \mathbf{b}_{up})$, where $\mathbf{b}_{lo}$ and $\mathbf{b}_{up}$ are the lower and upper boundaries of the search-space.
  - Initialize the particle's best known position to its initial position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
  - If ($f(\mathbf{p}_i) < f(\mathbf{g})$) update the swarm's best known position: $\mathbf{g} \leftarrow \mathbf{p}_i$
  - Initialize the particle's velocity: $\mathbf{v}_i \sim U(-|\mathbf{b}_{up}-\mathbf{b}_{lo}|, |\mathbf{b}_{up}-\mathbf{b}_{lo}|)$

Until a termination criterion is met (e.g. number of iterations performed, or adequate fitness reached), repeat:

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

For each particle $i = 1, ..., S$ do:
Pick random numbers: $r_p, r_g \sim U(0,1)$
Update the particle's velocity: $\mathbf{v}_i \leftarrow \omega\, \mathbf{v}_i + \varphi_p\, r_p\, (\mathbf{p}_i - \mathbf{x}_i) + \varphi_g\, r_g\, (\mathbf{g} - \mathbf{x}_i)$
Update the particle's position: $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i$ If
$(f(\mathbf{x}_i) < f(\mathbf{p}_i))$ do:
Update the particle's best known position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
If $(f(\mathbf{p}_i) < f(\mathbf{g}))$ update the swarm's best known position: $\mathbf{g} \leftarrow \mathbf{p}_i$
Now **g** holds the best found solution[16].

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

3.9 Pseudocode Web Phishing
Input: Webpage URL
Output: Phishing website identification
Step 1: Read web phishing URL
Step 2: Extract all 27 feature
Step 3: For each feature, Assign fuzzy membership degree value and Create fuzzy data set
Step 4: Apply association rule mining & generate classification rule.
Step 5: Aggregate all rule above minimum confidence.
Step 6: De-fuzzification of fuzzy values into original values [19].

IV. IMPLEMENTATION

4.1 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.

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 behavior is as follows:
1. An ant (called "blitz") runs more or less at random around the colony;
2. If it discovers a food source, it returns more or less directly to the nest, leaving in its path a trail of pheromone;
3. These pheromones are attractive, nearby ants will be inclined to follow, more or less directly, the track;
4. Returning to the colony, these ants will strengthen the route;
5. 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
6. The short route will be increasingly enhanced, and therefore become more attractive;
7. The long route will eventually disappear, pheromones are volatile;
8. Eventually, all the ants have determined and therefore "chosen" the shortest route [11].

The design of an ACO algorithm implies the specification of the following aspects.
- An environment that represents the problem domain in such a way that it lends itself to incrementally building a solution to the problem.
- A problem dependent heuristic evaluation function, which provides a quality measurement for the different solution components.
- A pheromone updating rule, which takes into account the evaporation and reinforcement of the trails.
- 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.
- A clear specification of when the algorithm converges to a solution [17].

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

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[8]. The ANTS algorithm is the following.

1. Compute a (linear) lower bound LB to the problem Initialize t*iy* ("i,y) with the primal variable values .
2. For k=1,m (m= number of ants) do repeat
2.1 compute h*iy* "(iy)
2.2 choose in probability the state to move into.
2.3 append the chosen move to the *k*-th ant's tabu list **until** ant *k* has completed its solution
2.4 carry the solution to its local optimum **end for**
3. **For** each ant move (iy), compute Dt*iy* and update trails by means of (5.6)
4. **If not** (end test) **go to** step 2.

4.2 Particle Swarm Optimization (pso)

The PSO algorithm has become an evolutionary computation technique and an important heuristic algorithm in recent years. The main concept of PSO originates from the study of fauna behavior. PSO learned from such a scenario and used it to solve the optimization problems. In PSO, each single solution is a "particle" in the search space. We refer to each solution as a "particle.[15]" All particles have fitness values, which are evaluated by the fitness function to be optimized. The particles also have velocities which direct the flight of the particles. Particles fly through the problem space by following the current optimum particles.

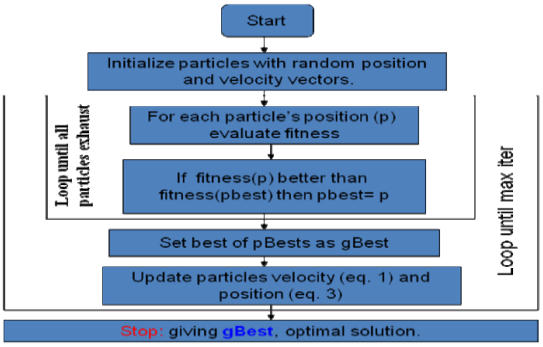

FIGURE 4.1 - Data flow for finding optimal solution

As shown in the Fig.4.1, PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. During all iterations, each particle is updated by following the two "best" values. The first one is the best solution (fitness) it has achieved so far. The fitness value is also stored. This value is called "pbest." The other "best" value that is tracked by the particle swarm optimizer is the best value obtained so far by any particle in the population. This best value is a global best and is called "gbest" [10].

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. Also found insignificant trivial influence of the *(Page Style & Content)* criteria along with *(Social Human Factor)* criteria for identifying e-banking phishing websites. Particle Swarm Optimization produces more accurate classification models than Associative classifiers. We recorded the prediction accuracy and the number of rules generated by the classification algorithm, the Ant Colony algorithm and PSO algorithm.

Table 5.1 shows that the PSO produce more accuracy and less time taken than associative classifier and ACO. Selected 802 cases randomly used for inducing rules from 1050 cases in original data set, the remaining 300 cases are used for testing accuracy of the induced rules of the proposed method by measuring the average percentage of correct predictions.

TABLE 5.1 Prediction accuracy and time taken comparison

Method	Association Classification		
Test Mode	10 Fold Cross Validation		
No. of URLs	1052	(Both)	
Correct Classified	1006		
Incorrect Classified	46		
Optimization	**Accuracy**		**Time Taken**
Association Classification		81%	12ms

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

ACO Optimization		89%	11ms

PSO Optimization		91%	9ms

The Fig.5.1 shows the comparision of fuzzy associative classifiers, ant colony optimization and PSO with the error rate.

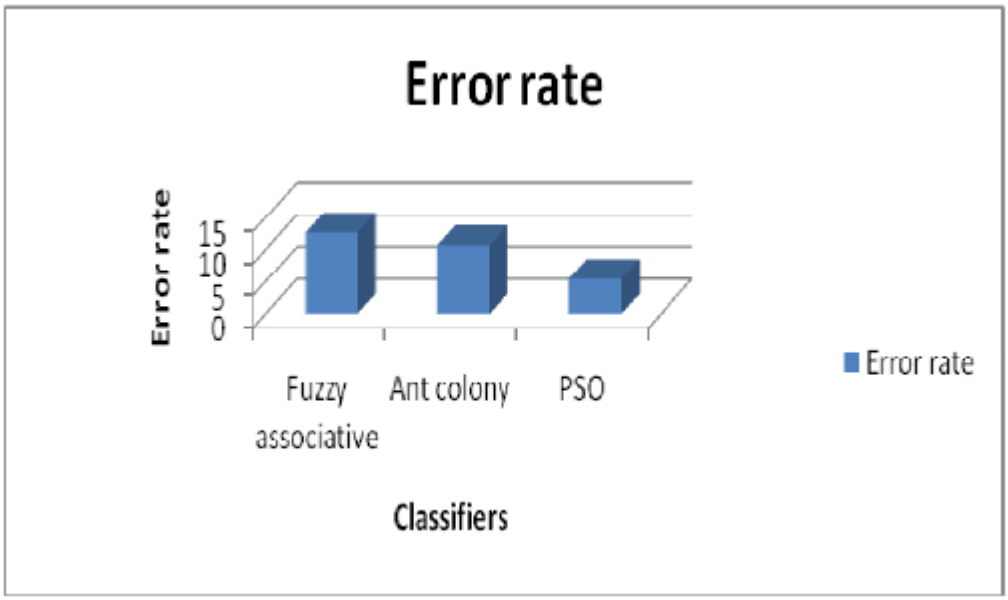

FIGURE: 5.1- Error rate comparision

The Fig.5.2 shows the comparision of ant colony algorithm and fuzzy associative algorithms in terms of prediction accuracy.

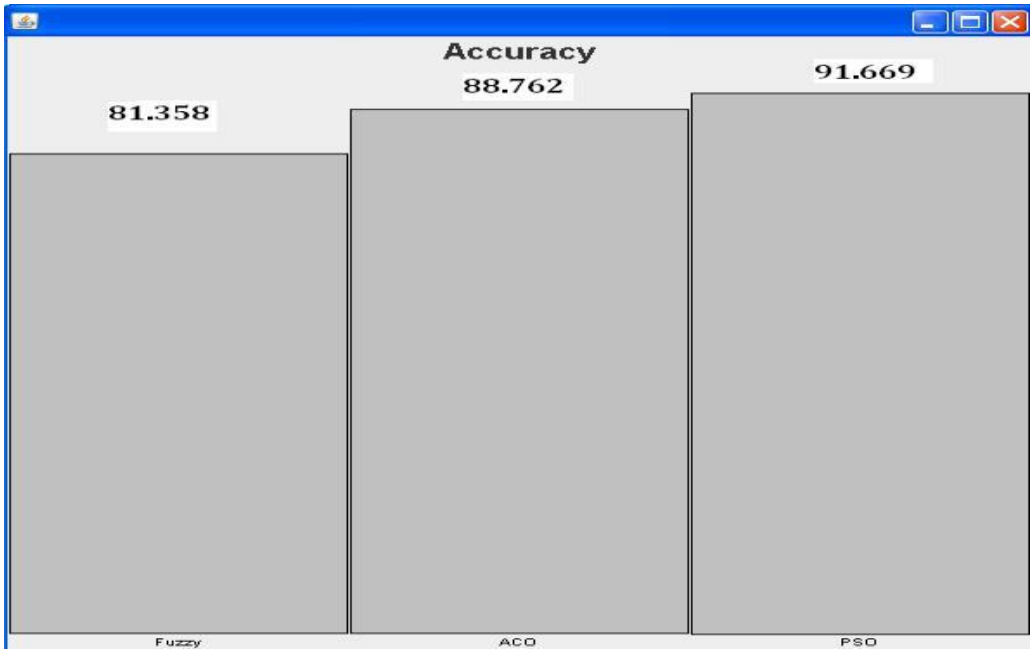

FIGURE : 5.2 - Accuracy report

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 40-50
Published Online March 2018 in IJEAST (http://www.ijeast.com)

From Fig.5.3 we can see the effectiveness of PSO implementation which shows difference from other methods.

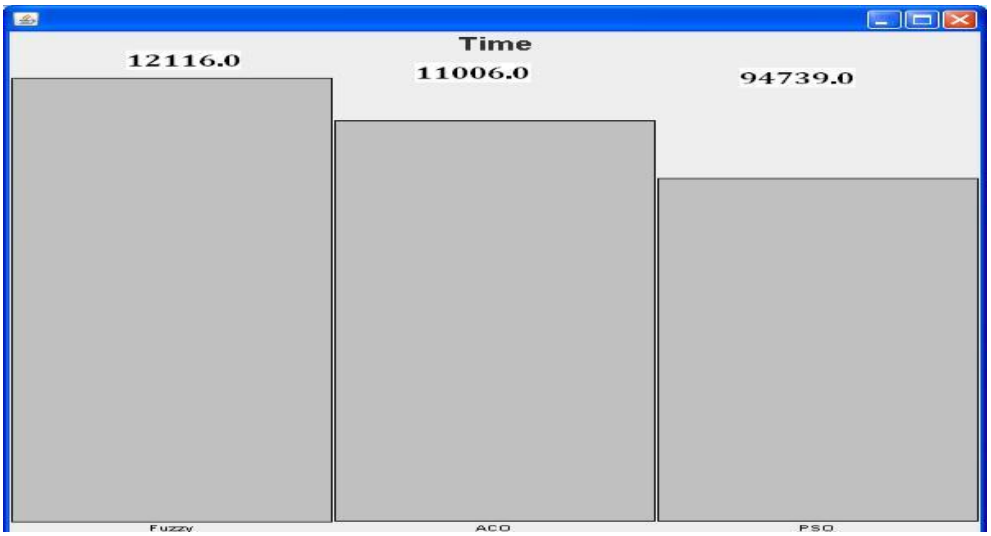

FIGURE : 5.3 – Time difference between three methods

VI. CONCLUSION

The Associative Classification Algorithm with Particle Swarm 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. Particle swarm optimization (PSO) is an algorithm modelled on swarm intelligence, that finds a solution to an optimization problem in a search space, or model and predict social behaviour in the presence of objectives. The PSO 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'. Combining these two techniques has given a fruitful result. After more than 1050 websites detection for both its application effectiveness and its theoretical groundings, PSO became one of the most successful paradigms in network security.

VII. REFERENCES

[1] A. Hossain, M. Dorigo, Ant colony optimization web page, http:// iridia.ulb.ac.be/mdorigo/ACO

[2] Ant Colony Optimization, Vittorio Maniezzo, Luca Maria Gambardella, Fabio de Luig**i.**

[3] Optimizing Large Scale Combinational Problems Using Multiple Ant Colonies Algorithm based on Pheromone Evaluation technique, Alaa Aljanaby, Ku Ruhana Ku Mahamud,

[4] Associative Classification Techniques for predicting e-Banking Phishing Websites, Maher Aburrous Dept. of Computing ,Universit y of BradfordBradford, UK.

[5] B. Adida, S. Hohenberger and R. Rivest , "Lightweight Encryption for Email," USENIX Steps to Reducing Unwanted Traffic on the Internet (SRUTI), 2005 ,

[6] 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.

[7] 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.

[8] Gartner. "UK phishing fraud losses double" STAMFORD, Conn., (April 14, 2009). "Gartner Says Number of Phishing Attacks on U.S. Consumers Increased 40 Percent in 2008".. Finextra. March 7, 2006. http:// www. finextra. com/ fullstory asp?id=15013.

[9] Jaco F. Schutte "The Particle Swarm Optimization Algorithm" EGM 6365 - Structural Optimization Fall 2005
[10] L. Bianchi, L.M. Gambardella, M.Dorigo. An ant colony optimization approach to the probabilistic traveling salesman problem. In Proceedings of PPSN-VII, Seventh InterGARTNE R, INC.
[11] M. E. Bergen, Technische Universität Berlin, Germany, 1995 Canstraint-based assembly line sequencing, Lecture Notes in Computer
[12] Miller, Rich. "Bank, Customers Spar Over Phishing Losses". *Netcraft*. http://news.netcraft.com/ rchives/ 2006/09.
[13] Mining Fuzzy Weighted Association Rules Proceedings of the 40th Hawaii International Conference on System Sciences – 2007.
[14] Particle Swarm Optimization , www.swaminteLligence.org.
[15] Particle Swarm Optimization, WIKI Pedia.
[16] Richardson, Tim (May 3, 2005). "Brits fall prey to phishing". The Register. http://www.theregister.co.uk/2005/05/03/aol_phishing/.
[17] 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)
[18] WEKA - University of Waikato, New Zealand, EN,2006: "Weka -Data Mining with Open Source Machine Learning Software in Java", 2006 ,
[19] Xun Dong,"PSO Introduction" Department of Computer Science University of York, United Kingdom.

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143