ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Artificial Immune System Based Classification Approach for Detecting Phishing Mails

Dr.A.Vijaya Kathiravan, B.Vasumathi

Asst. Prof of Computer Application, Dept of Computer Science, Government Arts College, Salem, India

M.Phil Research Scholar, Dept of Computer Science, Government Arts College, Salem, India

ABSTRACT: Phishing/Spam is an attack that deals with social engineering methodology to illegally acquire and use someone else's data on behalf of legitimate website for own benefits. Phishing emails are messages designed to fool the recipient into handing over personal information, such as login names, passwords, credit card numbers, account credentials, social security numbers etc. Fraudulent emails harm their victims through loss of funds and identity theft. They also hurt Internet business, because people lose their trust in Internet transactions for fear that they will become victims of fraud. Filtering approaches using blacklists are not completely effective as about every minute a new phishing scam is created. It has been investigated that the statistical filtering of phishing emails, where a classifier is trained on characteristic features of existing emails and subsequently is able to identify new phishing emails with different contents. This paper deals with the phishing detection problem and how to auto detect phishing emails. The proposed phishing detection model is based on the extracted email features to detect phishing emails, these features appeared in the header and HTML body of email. The developed model introduces Artificial Immune System methodology to classify whether the tested email is phishing or not.

KEYWORDS: Phishing Email, Swarm Intelligence, AIS Classification, Spam Detection.

I. INTRODUCTION

As people increasingly rely on the Internet for business, personal finance and investment, Internet fraud becomes a greater and greater threat. One interesting species of Internet fraud is phishing. Phishing is an online identity theft technique used to lure consumers into disclosing their personally identifiable information including Social Security numbers (SSN), account names and passwords, credit card information and any other personal information.

In recent years, phishing has become an enormous problem and threat for all big internet based commercial operations. The term covers various criminal activities which try to fraudulently acquire sensitive data or financial account credentials from internet users.

Phishing attacks use both social engineering and technical means in order to get access to such data. Phishing attack begins with a spoofed email masquerading as trustworthy electronic correspondence that contains hijacked brand names of banks, credit card companies, Social networking sites or ecommerce sites.

The persuasive inflammatory language of the email combined with a legitimate looking Web site is used to convince recipients to disclose sensitive information. In the end, consumers are lured in by these seemingly legitimate communications into providing sensitive information, often resulting in credit card fraud; identify theft, and even financial loss.

This paper presents a new approach using swarm intelligence to quickly detect phishing emails. This approach is based on some characteristics that are present in phishing emails. A set of features are extracted from tested email for phishing detection purpose. Then, the proposed algorithm is used to classify each email depending on existences flags of the adopted features.

II. RELATED WORKS AND LITERATURE REVIEW

Most anti-phishing tools employ email filtering techniques to classify legitimate emails and suspected spam in the mail inbox. The user is left to decide whether to open or discard such emails. If no anti-phishing tool is installed or the

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

user has not updated the anti-phishing program, then there is no layer of protection. This is referred to as passive anti-phishing [16]. It is because the approach only locally protects the user from a phishing attack but does not make any effort to stop or remove the Phisher at the source. The Phisher then continues with the phishing operation to further increase its victims.

While there are several email filters, browser tools, anti spyware and anti-virus software, very few research efforts have been entirely focused to protect online users from phishing attacks in the past. Existing phishing and spam techniques suffer from one or more limitations and they are not 100% effective at stopping all spam and phishing attacks [17]. Phishers are able to find ways to bypass existing rule-based and statistical based filters without much difficulty. Major e-mail service providers such as Yahoo, Hotmail, Gmail, and AOL filter all incoming emails separating them into Inbox (legitimate email) and junk (illegitimate email) email folders. However, these e-mail service providers do not actually attempt to remove the phishing page associated with the illegitimate email.

Furthermore, Phishers have readily available tools to bypass such spam filters [18]. We refer to this as a passive anti-phishing approach. This is because the approach only attempts to locally protect an individual from a phishing attack, but does not actively make any effort to remove or shut down the Phisher at the source. In effect, the Phisher is free to continue with the fraudulent operation and can potentially accrue further victims.

Phishing emails filtering methods are classified features-based many techniques. The classification can be done via many methods, such as by features extraction, machine learning technique or by clustering methods. Other innovative approaches have been devised for the purpose of detecting phishing e-mails. These approaches are based on the principle of distinguishing between phishing and ham emails. However, email filtering method has been considered one of the practical approaches in detecting phishing email. Its mechanism is based on defining a sender reliance cost by the Domain Name Server (DNS) inquiry and on analyzing message contents. However, this approach is not void of shortages. It, for instance, depends only on the cost of the DNS, which analyzes the address of the sender by the DNS [15]. This feature is considered unpractical due to the fact that phishing emails might appear in various shapes and have different features. That is, a phisher might use many techniques other than DNS; a matter that increases the probability of error in detecting such threats.

Recent research depends on machine learning technique for detecting phishing emails. There are three types of machine learning usually used in field of phishing email, some of them used supervised learning and some of them used unsupervised learning while some of them used hybrid (supervised/unsupervised) learning technique depend on classifiers .The main rule of the classifiers depend on learning several inputs or features to expect a desirable output.
For detecting phishing emails, many approaches have been proposed.

In [1], Phishing Detection by determining Reliability factor using Rough set theory. It uses thirteen basic factors directly responsible for phishing, which are grouped into four strata. Reliability factor is determined on the basis of the outcome of these strata using Rough set Theory. The limitation of this approach is that it only determines the probability of a site to be reliable or unreliable.

Phishing mail detection based on structural properties. The approach proposed in this paper demonstrates the ability to identify phishing via appropriate identification and usage of structural properties of the email. The experiments performed by employing SVM as the classification technique show promising results in classifying phishing emails with minimum errors. However, the experiment base used in this work is not large enough to draw a broader conclusion [2].

Detection of phishing emails using feature decisive values. In [3], they evaluate and compute the weight of each feature, and then we use the most effective features for classifying the emails. New algorithm is used to classify emails into phish or ham email based on the existence and the weight of features appeared in the email using a new equation to compute the features weight.

In [4], A survey of learning based techniques of phishing email filtering. The current paper focuses on machine learning applications used to detect and predict phishing emails. The current approaches include many filters based on various classification techniques used in different parts of email messages. More and more existing techniques for filtering phishing emails have limitations, technology techniques still have many limitation on

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

accuracy or performance because they are "time consuming", costly, and the huge number of rules created from learning techniques have increased, many algorithms have been adapted, but still there is no standard technique that is able to stop phishing attacks in general, or phishing emails as a special case.

Evolving fuzzy neural network for phishing email detection [5]. This paper seeks to Detection and Prediction of unknown "zero-day" phishing Emails by provide a new framework called Phishing Evolving Neural Fuzzy Framework (PENFF) that is based on adoptive Evolving Fuzzy Neural Network (EFuNN). PENFF does the process of detection of phishing email depending on the level of features similarity between body email and URL email features. The totality of the common features vector is controlled by EFuNN with evolving hyper-sphere technique to create rules that help predict the phishing email value in online mode.

In [6], Proposed phishing mail detection using fuzzy classification method. In this paper, it builds a fuzzy rule generation system to detect phishing mail. It classifies email into different category like very legitimate, legitimate, suspicious, phishy, very phishy etc. A motivation behind using fuzzy rule is soft decision boundaries provide sharp transition between classes.

Detecting phishing attacks in purchasing process through proactive approach [7] presents a framework for multilevel monitoring of service systems. It uses a proactive method to shut down a Phisher's operation by using a Pguard. This effectively stops a phishing attack at its source thereby protecting a significant number of other innocent users from being duped in the future.

Feature selection for improved phishing detection. In this paper, they have evaluated two common feature selection techniques: correlation based and wrapper based feature selection techniques for phishing website detection. They also evaluated two search methods: genetic search and greedy forward selection. Applying the techniques on real-world data sets, they experimentally demonstrated that feature selection technique can improve classification results [8].

In [9], Phishing attack detection, classification and proactive prevention using fuzzy logic and data mining algorithm. Initially the system assesses and classifies phishing emails using Fuzzy Logic and the RIPPER Data Mining algorithm. In assessing the Phishing email, Fuzzy Logic linguistic descriptors are assigned to a range of values for each key phishing characteristic indicators. The Data Mining RIPPER algorithm is used to characterize the Phishing emails and classify them based on both content-based and non-content based characteristics of Phishing emails.

This paper [10], An optimized feature selection technique for email classification presents a particle swarm optimization based feature selection technique, capable of searching for the optimal parameter values for SVM to obtain a subset of beneficial features. PSO is applied to optimize the feature subset selection and classification parameters for SVM classifier. It eliminates the redundant and irrelevant features in the dataset, and thus reduces the feature vector dimensionality drastically. Optimal subset of features is then adopted in both training and testing to obtain the optimal outcomes in classification.

Model and Algorithm in artificial immune system for spam detection present the self and non-self in a way to create efficiency of detector generation through equation. The novelty of this paper is to generate a new self (system) that randomly create antibody, introducing a new self detector method, with respect to self and non-self producing advance antibody. Also self and non-self matching algorithm is also presented. Mathematical model for effective matching of self and non-self for effective detector has been proposed [11].

Efficient spam filtering based on artificial immune system [12], briefly introduce the recent advances in immune based spam filtering methods and put emphasis on combining immune theory with statistical methods. It is shown that combining immune ideas with classical statistical methods can effectively improve the performance of a spam filter. In addition they present a framework of DTE method & also spam filtering using SOM based systems.

In [13], Detecting HTTP Botnet using artificial immune system, the group of hosts that show similar communication pattern in one step has been monitored and also performing malicious activities in another step and try to find common hosts in them. There is no need for prior knowledge of Botnets such as Botnet signature and other details about Botnets and only requires positive examples, which are readily available before an exploit.

This paper SMS Spam filtering technique based on artificial immune system [14], proposed a mobile agent system for detecting SMS-Spam based on AIS. This system contains dataset, tokenizer, analysis engine, stop word filter, AIS engine, and training process. The system used AIS features to building the antibodies (detectors), by initial training

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

phases. The generation, updating, and elimination of detector based on the AIS engine, the content of spam and non-spam SMS Messages used in training.

Most of the detection techniques use decision tree, machine learning algorithms, genetic algorithm, clustering techniques. In these techniques crisp logic is used. They classify email as spam and Not-spam email. Crisp logic is often failed because it does not provide sharp boundaries. Several Artificial Intelligence (AI) techniques including neural networks and fuzzy logic are successfully applied to a wide variety of decision making problems in real world. Up to our knowledge, there was not developed any system to the phishing mail detection based on swarm intelligence method. In this work we would like to build a swarm intelligence based system to detect phishing mail. The current approach is highly compacted framework.

III. PROPOSED DESIGN OF WORK

The proposed architecture explained clearly in Fig. 1 which provides ordered steps of how to distinguish between phishing emails and ham emails.

A. *AIS Based Architecture*

This architecture is divided into three stages, first stage is pre-processing of the data set, second stage is email object similarity and third stage is integrated with Swarm intelligence approach for detection of phishing emails. All of this stages will work after determine the features of phishing email which used in our framework. Phishing Email Features are used for classification. Based on these features phishing email is separately collected and filtered sequentially.

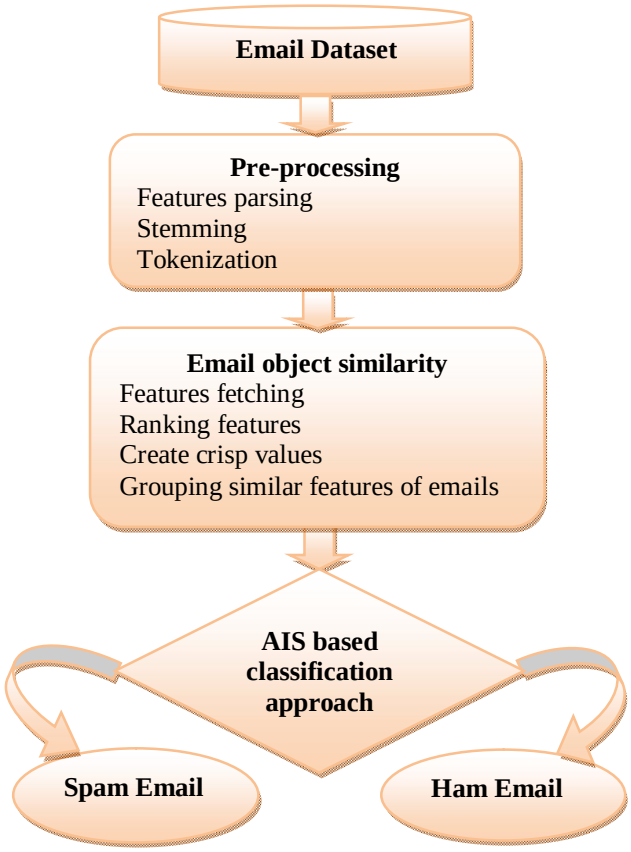

Fig. 1: AIS Based Architecture

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

B. ***Check Measures for Phishing Email***

All of this stages will work after determine the features of phishing email which used in our framework. Swarm intelligence based classification algorithm separately collects and filters email sequentially, which depends on sixteen features which represent the most effective features of phishing email. The features represented as binary value (0, 1) "1" to include the features and "0" otherwise. The sixteen features are as follows:

1. Using IP address (ipaddress)
2. Difference between sender domains with the domain of embedded links (diffsenlindom)
3. Number of links (nulinks)
4. Nonmatching between target and text of URLs (Tardiflink)
5. Number of different domains (nudiffdomain)
6. Number of dots in a domain (nodot)
7. Click here (clickhere)
8. Pictures used as links (NoPicLinks)
9. HTML e-mail (html e-mail)
10. Use of JavaScript (jascript)
11. Non-standard port in the URL (nonstport)
12. URL containing hexadecimal characters or @ symbol (hexorat)
13. Message size (messize)
14. Faking a secure connection (facksecon)
15. HTML form (htmlform)

Using IP address (ipaddress)

A number of phishers depend on their PCs as hosts for a phishing Web site. However, these PCs sometimes do not have DNS entries. Therefore, the easiest way to hide the normal form of a URL is to use IP addresses. Legitimate companies rarely use an IP address as a link page. We take "http://218.56.77.130/paypal.com" as example. For this feature, if an e-mail message has a link similar to an IP address, the probability of the e-mail being a phishing e-mail is increased. This is a binary feature that takes a value of 1 if the e-mail contains a URL similar to an IP address and 0 otherwise.

Difference between sender domains with the domain of embedded links (diffsenlindom)

When the link embedded in the HTML does not equal the sender's domain, it is most likely a phishing e-mail. For example, an e-mail may contain the following information: From: "identdep_op720@southtrust.com",
URL link: "http://accounts.keybank.com".
This is a binary feature. Therefore, if the domain name in the "from "field does not equal the domain name in the URL (embedded HTML), the value of this feature is 1 and 0 otherwise.

Number of links (numlinks)

One of the features of a phishing e-mail is a number of links embedded in HTML parts. In the proposed framework, links are distinguished based on tags <a> with HREF. This feature includes "mail to:" links. After analyzing the data set, we suggest this binary feature takes a value of 1 if there are more than three embedded links and 0 otherwise.

Nonmatching between target and text of URLs (Tardiflink)

If they have different host a value "1" and"0" otherwise.

Number of different domains (numdiffdomain)

The main part of a domain name which starts with http:// or https:// is extracted for all URLs starting with http:// or https://. In the present study, the main part of a link is assumed to include the section after the first dot up to the first slash ("/") if the link has a long domain name. For example, the "main" part of www.sg.echool.edu is sg.echool.edu and the "main" part of "www.jordan.com" is jordan.com. After analyzing 4,000 phishing and ham e-mails, many phishing e-mails were found to have more than three domains. Therefore, we suggest this feature takes a value of 1 if the number of different domains is more than three and 0 otherwise.

Number of dots in a domain (numdot)

Attackers utilize many methods to stage a phishing attack. One method depends on the inclusion of a sub-domain. We take "http://www.may-bank.update.data.com" as example. This link appears to be hosted by Maybank, but

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

it is actually not. There are four dots in the domain. Generally, a legitimate company will have no more than three dots on its domain name. Therefore, this feature depends on determining the maximum number of dots. We suggest it takes a value of 1 if there are more than three dots in the domain and 0 otherwise.

Click here (clickhere)

Many phishers use words like "click here," "click," or "here" in the text portion of their links in order to hide a suspicious domain name. When users click on such words, they are redirected to a phishing Web site. We take <a HREF="http://61.119.228.47/.eBay/" > click here </A> as example. If an e-mail message has one of the three words mentioned above, it is flagged as a phishing e-mail takes a value of 1 and 0 otherwise.

Pictures used as links (NoPicLinks)

Some attackers use an image as a link to hide fraudulent URLs. We take <imgsrc=http:// www.paypalobjects.com/en _US as example. The maximum number of images used as a link is calculated based on the tag "<img src=URL" embedded in the HTML. After analysis, we suggest this is binary feature takes a value of 1 if there are more than two pictures used as links and 0 otherwise.

HTML e-mail (html e-mail)

At present, creating a phishing e-mail is difficult without using an HTML code because an HTML code enables an embedded link to connect directly to other Web sites. The presence of an HTML code in an e-mail can be determined using MIME types. If the MIME type is either text/html or a multipart/alternative, an HTML code is embedded in the message. This is a binary feature takes a value of 1 if no HTML code is embedded and 0 otherwise.

Use of JavaScript (jascript)

One of the primary methods used by an attacker to build a phishing e-mail is the use of java script because with this simple language, the phisher can program pop-up windows. The phisher can then change the status bar of a Web browser, enabling him/her to build a complex attack using an embedded script code inside a link. An e-mail message can be determined to have a java script by the tag "JavaScript" or <script>. This binary feature takes a value of 1 if the message has a java script code and 0 otherwise.

Non-standard port in the URL (nonstport)

A server accesses Web pages using ports and a few phishers use non-standard ports to hide their identity and location. Web pages use port 80 as default and some normal ports such as 443 are used by legitimate companies. The port number in a URL link comes after a colon. For example, in http://www.paybankonline.com:ac@50.28.170.70:8030/,:8030 represents the port number. This is a binary feature that takes a value of 1 if the e-mail message uses a port other than 80 or 443 and 0 otherwise.

URL containing hexadecimal characters or @ symbol (hexorat)

Some attackers use hexadecimal character codes to hide embedded URLs. Attackers can write an IP address using the "%" symbol to build a hexadecimal number. Sometimes, they use the "@" symbol to confuse users. This binary feature takes a value of 1 if the message URL contains either the "%"or @ symbol and 0 otherwise.

Message size (messize)

Message size refers to the size of an e-mail in bytes. Most phishing e-mails have a size of less than 25 kb. However, based on a semantic report SYMANTIC, 2010, more than 90% of phishing e-mails have a size of less than 20 kb. Therefore, we suggest this binary feature takes a value of 1 if the message size is less than 25 kb and 0 otherwise.

Faking a secure connection (facksecon)

One of the most fraudulent applications used by phishers utilizes URLs that begin with "https://" (instead of using "http://") to trick users into believing that the link is a legitimate URL supported by a Secure Sockets Layer certificate. We take https://www.maybank.com%01 [string of ~ 60 —%01 elided]@203.172.185.20/f/ as example.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Clicking on this link will redirect the user to "http:// 203.172.185.20/f/" which tries to mimic a secure connection. This binary feature takes a value of 1 if the embedded URL starts with https:// and 0 otherwise.

HTML form (htmlform)

One of the earliest features used to collect user information directly by e-mail utilizes the FORM feature. This feature is a simple code written using HTML, which allows a form requiring the entry of user information such as usernames and passwords to be built. The FORM feature uses a button to submit this information to a phisher account. This feature can be detected if the HTML code has a <FORM>tag. This binary feature takes a value of 1 if the message has a <form> tag and 0 otherwise.

C. ***Swarm Intelligence Technique***

Swarm Intelligence can be described by considering five fundamental principles.

1) Proximity Principle: The population should be able to carry out simple space and time computations.

2) Quality Principle: The population should be able to respond to quality factors in the environment.

3) Diverse Response Principle: The population should not commit its activity along excessively narrow channels.

4) Stability Principle: The population should not change its mode of behavior every time the environment changes.

5) Adaptability Principle: The population should be able to change its behavior mode when it is worth the computational price.

Artificial Immune System Algorithm (AIS)

AIS is one of the Swarm Intelligence Technique that has been used in this paper to detect phishing mails. Proposed by Dasgupta in 1999 [19]. Artificial Immune algorithm is based on clonal selection principle and is a population based algorithm. AIS is inspired by the human immune system which is a highly evolved, parallel and distributed adaptive system that exhibits the following strengths: immune recognition, reinforcement learning, feature extraction, immune memory, diversity and robustness.

The artificial immune system (AIS) combines these strengths and has been gaining significant attention due to its powerful adaptive learning and memory capabilities.

The main search power in AIS relies on the mutation operator and hence, the efficiency deciding factor of this technique. The steps in AIS are as follows:

1. Initialization of antibodies (potential solutions to the problem). Antigens represent the value of the objective function f(x) to be optimized.

2. Cloning where the affinity or fitness of each antibody is determined. Based on this fitness the antibodies are cloned that is the best will be cloned the most. The number of clones generated from the n selected antibodies is given by:

$$Nc = \sum \text{round}(\beta * j/i) \quad i = 1,2\ldots\ldots n,$$

Where Nc is the total number of clones, β is a multiplier factor and j is the population size of the antibodies.

3. Hypermutation: The clones are then subjected to a hyper mutation process in which the clones are mutated in inverse proportion to their affinity; the best antibody's clones are mutated lesser and worst antibody's clones are mutated most. The clones are then evaluated along with their original antibodies out of which the best N antibodies are selected for the next iteration. The mutation can be uniform, Gaussian or exponential.

IV. EXPERIMENTAL RESULT

All the experiments has been conducted on a PC with Intel(R) Core(TM) i5-4200U CPU @1.6 GHz and 4GB RAM using MATLAB. The experiment is designed to test the performances of the AIS, which is evaluated via its true positive (TP) rate and false positive (FP) rate.

Our experiment has been conducted on four benchmark corpora PU1, PU2, PU3, PU4. The corpora are pre-processed with elimination of HTML tags, attachments, and header fields. In all PU corpora, the duplicates were separated because it might cause over-optimistic results in experiments. In PU1 total 1099 messages are considered out of which, 481 messages are spam and remaining 618 are legitimate. In PU2 total 721 messages are considered out of which, 142 messages are spam and 579 are legitimate. In PU3 total 4139 messages are considered out of which, 1826 messages are spam and 2313 are legitimate. In PU4 total 1142 messages are considered out of which, 572 messages are

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

spam and 570 are legitimate. All the messages are available in pre-processed form and also available in English [15] respectively.

In this section, A comparison between our proposed work with Support Vector Machine has been presented. For all the methods our main focus of comparison is on accuracy. Comparisons of results are shown in Table I.

METHOD → / ↓ DATASET	SVM		PROPOSED SYSTEM	
	Spam	Ham	Spam	Ham
PU1 (Total of 1099 Message)	434	665	453	646
PU2 (Total of 721 Message)	149	572	141	580
PU3(Total of 4139 Message)	1828	2311	1804	2335
PU4(Total of 1142 Message)	570	572	569	573

Table I. Result Comparison of Proposed System with SVM Classifier

V. CONCLUSION AND FUTURE WORK

Swarm Intelligence Algorithms are going to be a new revolution in computer science. In this paper, it has been proposed a phishing detection approach that classifies the phishing mail by checking the phishing characteristics. All the experiment with existing data sets has been conducted in a controlled environment. It shows that it is possible to detect phishing emails with high accuracy by using one of the Swarm Intelligence approach called Artificial Immune System (AIS), using features that are more directly applicable to phishing emails than those employed by general purpose spam filters. Future extension of this work may include a real time application of our proposed work for effective spam filtering.

REFERENCES

1. Anugrah kumar and Sarvesh SS Rawat, “Phishing Detection by determining Reliability Factor using Rough Set Theory”, 2002.
2. Madhusudhanan Chandrasekaran, Krishnan Narayanan and Shambhu Upadhyaya, “Phishing E-mail Detection Based on Structural Properties”, 2005.
3. Noor GhaziM.Jameel and Loay E.George, “Detection Phishing Emails using Feature Decisive Values”, july 2013.
4. Ammar Almomani, Tat-Chee Wan and Ahamad Manasrah, “A Survey of Learing Based Techniques of Phishing Email Filtering”, 2012.
5. Ammar Almomani, Tat-Chee Wan, Eman Almomani and Ahamad Manasrah, “Evolving Fuzzy Neural Network for Phishing Emails Detection”, 2012.
6. Ami K. Trivedi and G.J. Sahani, “Proposed Phishing Mail Detection using Fuzzy Classification methods”, 2013.
7. S.Arun, D.Anandan, T.Selvaprabhu, B.Sivakumar and P.Revathi, “Detecting Phishing Attacks in purchasing process through Proactive Approach”, 2012.
8. Ram B. Basnet, Andrew H. Sung and QuingZhong Liu, “Feature Selection for Improved Phishing Detection”, 2011.
9. Rosana J. Ferolin, Bobby D. Gerardo, Yung-Cheol Byun and Chul-Ung Kang, “Phishing Attack Detection, Classification and Proactive Prevention using Fuzzy Logic and Data Mining Algorithm”, 2011.
10. Olaleye Oludare, Olabiyisi Stephen, Olaniyan Ayodele, Fagbola Temitayo, “An Optimized Feature Selection Technique For Email Classification”, 2014.
11. Ismaila Idris, “Model and Algorithm in Artificial Immune System for Spam Detection”, Jan-2012.
12. Athare Sharayu S and Prabhudev Irabashetti, “Efficient Spam Filtering Based on Artificial Immune System”, 2014.
13. Amit Kumar Tyagi and Sadique Nayeem, “Detecting HTTP Botnet using Artificial Immune System”, May-2012.
14. Tarek M Mahmoud and Ahmed M Mahfouz, “SMS Spam Filtering Technique Based on Artificial Immune System”, 2012.
15. Inomata A, M. Rahman, T. Okamoto and E. Okamoto,. “A novel mail filtering method against phishing”, 2005.
16. Shah R, Trevathan J, Read W, Ghodosi H, ”A Proactive Approach to Preventing Phishing Attacks Using the Pshark Model”, March 2009.
17. Rokach, Lior, Oded, Maimon, “Data mining with decision trees: theory and applications”, 2008.
18. Afroz S, Greenstadt R, “PhishZoo: Detecting Phishing Websites by Looking at Them”, 2011.

D. Dasgupta, “Artificial Immune Systems and Their Applications”, Springer, Berlin, 1999.