INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Oppositional Cuckoo Search Based Weighted Fuzzy Rule System in Malicious Web Sites Detection from Suspicious URLs

Kotoju Rajitha[1*] **Doddapaneni VijayaLakshmi**[2]

[1]*MGIT, Telangana, India*
[2]*MGIT, Telangana, India*
* Corresponding author's Email: rajithak9408@gmail.com

Abstract: The primary intention of this research is to design malicious web sites detection from Suspicious URLs. Huge web pages are gone by every day over a system and malicious websites may contaminate client machines. In this work, we design the Malicious Web Sites Detection from Suspicious URLs based on Oppositional Cuckoo Search (OCS) algorithm and fuzzy logic classifier (FLC). The system consists of two modules such as (i) feature selection and (ii) classification. At first, we take the four kinds of features from the dataset which have totally thirty features. Among that, we select the important features using OCS algorithm. After that, we train the selected features using FLC and then we calculate the fuzzy score. Finally, in testing, the FLC is detecting the malicious URL based on the fuzzy score. The experimental results demonstrate that the proposed malicious URL detection method outperforms other existing methods.

Keywords: malicious; URL; detection; Oppositional Cuckoo Search; Fuzzy logic classifier; Web Sites Detection.

1. Introduction

The quick advancement and development internet and nearby system frameworks have changed the computing world in the most recent decade. The expense of transitory or perpetual harms was created by unapproved access of the intruders to progressively actualize different frameworks to screen information stream in their networks. These frameworks are for the most part alluded to as malicious websites detection or Intrusion Detection Systems (IDSs) from Suspicious URLs [1]. In addition, Intrusion recognition plans can be arranged into two classifications: misuse and anomaly intrusion detection [2]. Misuse alludes to known assaults that adventure the known vulnerabilities of the framework. On the off chance that the watched movement of a client goes amiss from the normal conduct, an anomaly is said to happen [3]. Be that as it may, as more data on people and organizations are put in the cloud, concerns are starting to become about exactly how safe a domain [4]. Albeit most PCs in delicate applications gather audit trails, this audit trails were, for the most part, settled for execution estimation or accounting purposes and offer little help in recognizing intrusions [5]. Conventional assurance systems, for example, client verification, information encryption, abstaining from programming mistakes and firewalls are utilized as the primary line of protection for PC security [6]. At that point frameworks are intended to identify anything that goes amiss from ordinary action [7]. Close by different procedures for avoiding interruptions, for example, encryption and firewalls, Intrusion Detection Systems (IDS) are another noteworthy technique used to shield PC frameworks [8].

Soft computing is a general term for portraying an arrangement of improvement. Handling strategies for this are Fuzzy Logic (FL) [9], Artificial Neural Networks (ANNs) [10], Probabilistic thinking (PR) [11], and Genetic Algorithms (GAs) [12]. To build the reception of Web and cloud administrations, cloud administration suppliers (CSPs) should first set up trust and security to reduce the stresses of an extensive number of clients. A solid cloud biological system ought to be free from misuse, viciousness, duping, hacking, infections, regret, smut, spam, and security and copyright violations [13].

In this paper, we explain the malicious web sites detection from Suspicious URLs using oppositional cuckoo search algorithm and fuzzy logic system. This system consists of two modules such as (i) feature selection and (ii) classification. At first, we consider the four types of features (Domain-based features, HTML and JavaScript based features, Abnormal based features and Address Bar based features) from the URLs which is present in the dataset itself. Among the features, we select the important features based on oppositional cuckoo search algorithm (OCS). Then, the reduced features are given to the input of the fuzzy logic system. The fuzzy classifier is based on the concepts of fuzzy rules, used for the classification of the malicious and non-malicious URLs. Finally, in the testing stage, we detect the malicious and non-malicious URLS. The rest of the paper is organized as follows: A brief review of some of the literature works in malicious detection techniques are presented in Section 2. The background of the research is explained in section 3. The proposed methodology is detail described in Section 4. The experimental results and performance evaluation discussion are provided in Section 5. Finally, the conclusions are summed up in Section 6.

2. Review of Related Works

In recent times, intrusion detection has received a lot of interest among the researchers because it is widely applied for preserving the security within a network. Here, we present some of the techniques for malicious web sites detection: A. Zarrabi and A. Zarrabi [14] have presented IDS as a Service in a cloud to secure client system. It misses a few attributes in system traffics that make it conceivable to extricate the required information from the client system for assessment. But in this paper, some of the scalability issues are occurred. Moreover, Misuse detection can detect known malicious web pages, but it cannot detect new ones. In contrast, anomaly detection can detect unknown malicious web pages, but it has a high false positive rate. To overcome the problem, S. Yoo and S. Kim [15] have clarified the Two-Phase Malicious Web Page Detection Scheme Using Misuse and Anomaly Detection.

Essentially, Wireless Sensor Networking was a standout amongst the most encouraging advances that have applications extending from medicinal services to the strategic military. To improve the security of wireless system, I. Butun et al. [16] recommended a study of the cutting edge in Intrusion Detection Systems (IDSs) that were proposed for WSNs was given IDSs along their groupings, outline details, and necessities were quickly presented. Moreover, M. Mabzool and M. Z. Lighvan [17] have clarified the Intrusion recognition framework taking into account web utilization mining. Here, web server access logs utilized as the info information and after pre-preparing, scanners and every distinguished assault were recognized. But this method presence of high rate of false alerts causes unnecessary interference of human analyst. In addition, U. Ravale et al. [18] have clarified the Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K-Means and RBF Kernel Function. One of the essential difficulties in interruption identification was the issue of misconception, misdetection, and absence of continuous reaction to the assault.

For the significance of an effective Intrusion Detection System, K.S. A. Kumar and V. N. Mohan [19] introduced a mix of three methods containing two machine-learning standards. K-Means Clustering, Fuzzy Logics, and Neural Network strategies were conveyed to arrange a compelling interruption location framework. To enhance the exactness and effectiveness of the Intrusion Detection System, S. R. Gaddam et al. [20] exhibited "K-Means+ID3," a strategy to course k-Means bunching and the ID3 choice tree learning strategies for characterizing peculiar and ordinary exercises in a PC system, a dynamic electronic circuit, and a mechanical mass-pillar framework.

Each of these frameworks has their merits and demerits. One of the regular issues in comparable tasks is the incapacity in covering various types of assaults. The other issue is immensity of false alerts, scalability issues. Additionally, optimization based feature selection process affects convergence problems and low search capacity To overcome the difficulty present in the above literature in our work we proposed an Oppositional Cuckoo Search Based Weighted Fuzzy Rule System in Malicious Web Sites Detection from Suspicious URLs.

3. Back Ground of the Research

❖ Cuckoo search (CS) algorithm

CS enhancement technique is a nature propelled Metaheuristic calculation, which depends on the commit brood parasitic conduct of some cuckoo flavors in a blend with duty flight conveyance of a few winged birds and organic product flies. This initial cuckoo has novel qualities that they lay an egg to the next host bird's nests. The eggs those are like the host bird's egg survive and turn into a full grown cuckoo. Other divergent eggs which are

predicted by the host birds are murdered. The developed eggs that survived uncover the reasonableness of the nests in that area. The more profit is picked up where more eggs survive the cuckoos algorithm is going to optimize in that area. CS is based on three important rules.

- Each cuckoo lays one egg at a time and places it in an arbitrarily picked nest.
- The best nests with the most noteworthy nature of eggs (solutions) continue to the next generation
- The number of accessible host nests is fixed, and a host has a probability $P_a \in (0,1)$ of finding an alien egg. In this case, the host bird either tosses out the egg or deserts the nest to construct another one in an alternate area.

Levy flight behavior, as opposed to straight forward arbitrary walk behavior, can be utilized to increase the execution of the CS. The following formula can portray Levy flight behavior while producing new solutions $s_i(t+1)$ for the i^{th} cuckoo [21].

$$s_i(t+1) = x_i(t) + \alpha \oplus Levy(\lambda) \quad (1)$$

Pseudo code of cuckoo searches algorithm

Begin
Objective function $F(s) = s = (s_1, s_2, ..., s_d)$
Generate initial population of n host nests;
While (t<Max generation) or (stop criterion)
Get a cuckoo randomly by levy flight
Evaluate its fitness F_i
Choose a nest among n (say, j) randomly
If $F_i > F_j$
Replace j by the new solution;
end if
A fraction probability (P_a) of the worse nests are abandoned and new ones are built;
Keep the best solutions/nests;
Rank the solutions/nests and find the current best;
Pass the current best solutions to the next generation;
end while

Where $\alpha > 0$ is the final size that has to be related to the problem of interest scale, and the product $\oplus$ refers to an entry –wise multiplication.

The formula that describes the Levy flight behavior in which the step lengths fit a probability distribution is:

$$Levy \oplus u = t^{-\lambda} \quad (2)$$

As per this equation, cuckoo birds' consecutive jumps or steps mainly form a random walking process that corresponds to a power-law step-length distribution with a heavy tail. The pseudo code of cuckoo search algorithm is illustrated in table 1.

Opposition-based learning is a rather simple concept that aims to improve the convergence rate and/or accuracy of computational intelligence algorithms [21]. In order to improve the search capability of CS algorithm, the purpose of this paper is to present a CS algorithm based on opposition-based learning (OCS).

4. Proposed Methodology

The main intention of this paper is malicious websites detection from suspicious URLs based on the optimal fuzzy logic system. The malicious web detection is mainly used for security of internet service provider (ISP).The detailed design detection system of websites consist of two major stages such as (i) Feature selection based on oppositional cuckoo search algorithm and (ii) detection based on the fuzzy logic system. At first, input dataset (URLs) is brought into the system, which has URLs and corresponding phishing features. Then, oppositional cuckoo search based attribute reduction method is applied to select an optimal subset of attributes which therefore reduces computational burden and enhances the performance of the fuzzy logic system. The obtained dataset with a subset of attributes is divided into two subsets: training dataset and testing dataset. Training dataset is used to build a fuzzy logic system while the testing dataset is used to test the obtained fuzzy logic system. The detail of each module is discussed in the following subsection. The overall architecture of proposed methodology is given in figure 1.

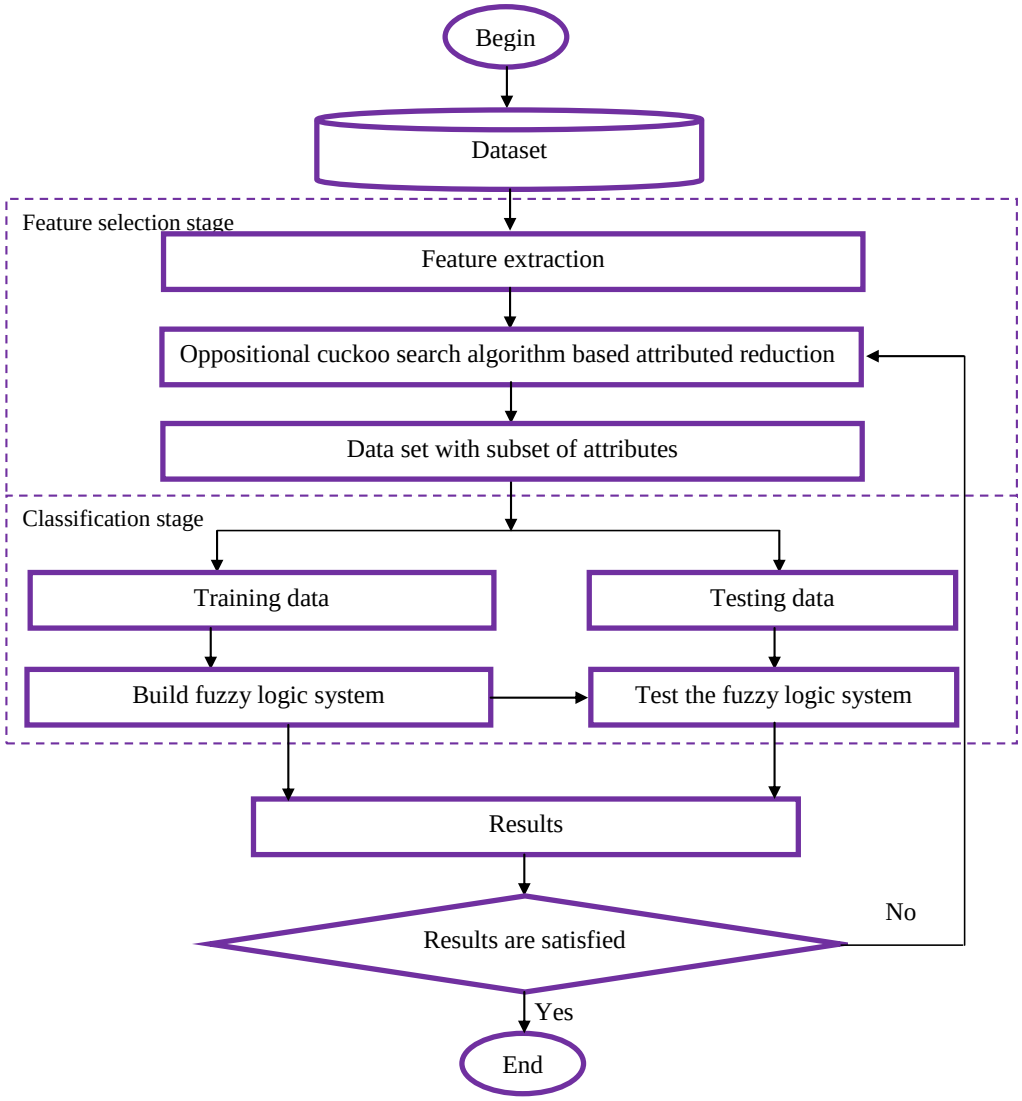

Figure.1 Overall architecture of the proposed malicious detection system

4.1 Feature selection based on Oppositional Cuckoo Search

Feature selection is an important step in the Malicious Web Sites Detection. The objective of feature selection is extract meaningful features from URLs; so that they can be classified according to their source. In this paper, we select the important features using Oppositional Cuckoo Search algorithm (OCS). Feature selection strategies are used to remove the irrelevant features. Feature selection improves the accuracy of algorithms by reducing the dimensionality and removing irrelevant features. In this section, we select the optimal features which are necessary and sufficient for solving the classification problems. The step by step process of feature selection is explained below;

To optimize the features, OCS algorithm initially creates an arbitrary population of the solution. Solution creation is an important step of optimization algorithm that helps to identify the optimal solution quickly. At first, we randomly create the solution (nest) of entire data's in the dataset. The solution has two parts such as length of the solution and selected features. The solutions are represented in the format is shown in equation 3 and solution encoding is given in figure 2.

$$P_i = \begin{bmatrix} s_{11} & s_{12} \ldots\ldots s_{1D} \\ s_{21} & s_{22} \ldots\ldots s_{2D} \\ s_{n1} & s_{n2} \ldots\ldots s_{nD} \end{bmatrix} \quad (3)$$

Step 1: Solution encoding

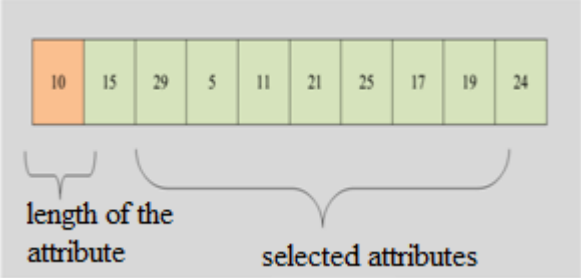

Figure.2 Solution encoding

Step 2: Calculation of opposite solution

Every solution P_i has a unique opposite P_{opi} solution. The opposite solution $OP(s'_1, s'_2, \ldots, s'_n)$ is calculated based on the equation;

$$s'_i = a_i + b_i - s_i \quad , \qquad i \in 1,2,\ldots, n \tag{4}$$

$$OP_i = \begin{bmatrix} s'_{11} & s'_{12} \ldots\ldots s'_{1D} \\ s'_{21} & s'_{22} \ldots\ldots s'_{2D} \\ s'_{n1} & s'_{n2} \ldots\ldots s'_{nD} \end{bmatrix} \tag{5}$$

Step 3: Fitness calculation

The fitness calculation is a crucial aspect in OCS. It is used to evaluate the aptitude (goodness) of candidate solutions. Here, classification accuracy is the main criteria used to design a fitness function. The fitness computation is executed for each solution. For each iteration, the fitness is calculated using equation (6),

$$Fitness = \frac{TN + TP}{(TN + TP + FN + FP)} \tag{6}$$

Where, *TP*→ True positive, *TN*→ True negative, *FP*→ False positive, *FN*→ False negative.

Step 4: Update based on cuckoo search algorithm

After, calculating the fitness value, we update the solution based on the cuckoo search algorithm. The superiority of the new solution is evaluated and a nest is selected between arbitrarily. If the superiority of new solution in the selected nest is improved than the previous resolution, it will be alternated by the new explanation (Cuckoo). Or else, the prior explanation is put to the side as the finest explanation. The levy flights utilized for usual cuckoo search algorithm is,

$$m_i^{*} = m_i^{(t+1)} = m_i^{(t)} + \alpha \oplus Levy(n) \tag{7}$$

Where, t is step size, and $\alpha > 0$ is the step size scaling limit. Here, the entrywise product $\oplus$ is comparable to those utilized in CS, $M_i^{(t+1)}$ and represents ($t+1$) the egg (feature) at nest (solution), $i=1,2 \ldots m$, and $t=1,2 \ldots d$. the levy flights utilize an arbitrary level extent which is drained from a levy allocation.

Step 5: Termination criteria

The algorithm discontinues its execution only if a maximum number of iterations is achieved and the nest which is holding the best fitness value is selected and it is given as the best rule for the classification.

4.2. Classification based on fuzzy logic classifiers

In this section, the malicious URL is detected based on the fuzzy logic classifiers (FLC). Generally, rule generation-based systems are entertainingly used to classification problem in different applications fields like error detection, biology, and medicine. In recent years, fuzzy inference system (FIS) [22] is the one of the important classifier employed to different fields among different rule-based systems. One of the most significant matters of proposing FIS is finding out the rule base or right selection of rules, which are the general problem taken by the dissimilar researchers. For different data application, it is very hard for human experts to offer enough information necessary to produce fuzzy rules. As a result, computational methods are employed to produce fuzzy rules from data routinely. Moreover, using the fuzzy system, detection of malicious URLs is performed. The detailed procedure of fuzzy system is made cleared in the beneath segment.

- **Fuzzy system**

This section explained the fuzzy system proposed for the URL detection method. The most important ideas behind a fuzzy system use the concept of linguistic variables to make decisions based on fuzzy rules and thus get a better response compared to a system by means of crisp values.

- **Design of fuzzy system**

Suggesting of the fuzzy system has three important steps a) Fuzzification b) fuzzy inference engine c) Defuzzification.

Fuzzification: adapts the crisp input to a linguistic variable with the membership function gathered in the fuzzy knowledge base.

Fuzzy inference engine: with the help of If-Then type fuzzy rules, changes the fuzzy input into

the fuzzy output.

De-fuzzification: changes the fuzzy output of the inference engine to crisp using membership function equivalent to those employed by the fuzzifier. In our work, crisp rules are fuzzified inference system through the triangular membership function. Fuzzification is essential as a degree of the membership function is précised for each member of the set. The fuzzy system forecasts the effects more accurately with the optimized membership function. While we are planning the fuzzy system, the fuzzy membership function definition and fuzzy rule base are the two significant steps.

- **Membership function**

A membership function (MF) is a curve that characterizes how each point in the input space is recorded to a membership value (or degree of membership) between 0 and 1. In addition, by selecting the proper membership gives the better result. The simplest form of the membership function is triangular membership function and it is compared with other membership function it gives better performance. So in this work, we have selected the triangular membership function to change the input data into the fuzzified value. Based on the dataset, we select the vertices of membership function. The Triangular membership function contains three vertices *i, j* and *k* of *f(x)* in a fuzzy set A (*i*: lower boundary and *k*: upper boundary where membership degree is zero, *j*: the center where membership degree is 1. membership values are described as beneath:

$$f(x)=\begin{cases}0 & if \;\; x \le i \\ \dfrac{x-i}{j-i} & if \;\; i \le x \le j \\ \dfrac{k-x}{k-j} & if \;\; j \le x \le k \\ 0 & if \;\; x \ge k \end{cases} \quad (8)$$

Figure 3 shows the plot considering all the three membership functions having overlapping values. Here, the curves for, low, medium and high are shown for a particular one attribute. The membership function of the medium required three parameters and the membership function of Low and high are required two parameters.

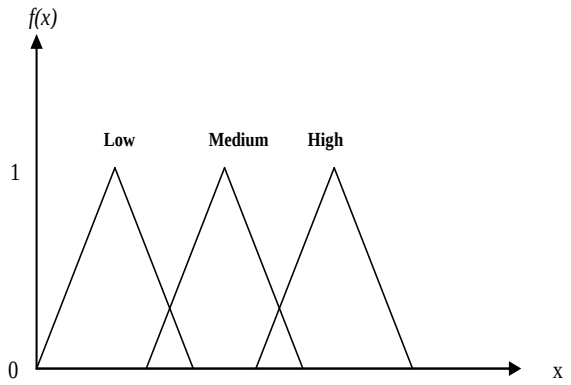

Figure.3 Triangular membership function with defined parameters and their values

- **Rule base**

In addition, the fuzzy rules are generated using subtractive clustering. Here, the rule should have two different decisions like as YES and NO. Some of the sample rules are given below;

R_1: IF (X_1 is L) and (X_2 is H) and (X_3 is M) and (X_4 is H) and . . . . (X_N is M) THEN Y_0 =1

R_2: IF (X_1 is M) OR (X_2 is H) OR (X_3 is L) OR (X_4 is M) OR . . . . (X_N is H) THEN Y_0=0.

Where; R_1, R2 are the fuzzy rules, X_1, X2, X3 are the input attribute. Y_0 is 1 means malicious and Y_0 is 0 means legitimate.

4.4 Testing process

The testing data (D^{TE}) with reduced attribute (N) is given to the fuzzy logic system, where the test data is converted to the fuzzified value based on the fuzzy membership function. Then, the fuzzified input is matched with the fuzzy rules defined in the rule base. Here, the rule inference procedure is used to obtain the linguistic value that is then converted to the fuzzy score using the average weighted method. From the fuzzy score obtained, the classification decision is produced whether the test data belongs to the malicious or not.

5. Result and Discussion

In this section, we discuss the result obtained from the proposed technique. For implementing the proposed technique, we have used Mat lab version (7.12). This proposed technique is done in windows machine having Intel Core i5 processor with speed 1.6 GHz and 4 GB RAM. For comparing the performance, we use two type of dataset such as URL Reputation Data Set and Phishing Websites Data Set.

5.1. Dataset discretion

In this paper, we used two type of datasets such as URL Reputation Data Set and Phishing Websites Data Set which are collected from the UCI Machine learning repository. Phishing Websites Data Set has totally 2456 instance and 30 attributes. Phishing websites are forged WebPages created and utilized by phishers to copy the web pages of legitimate websites by which results in a lack of faith in internet based services but also financial loss. Dataset is collected from, 'phishtank" which is one of the most crucial phishing-report collectors. The PhishTank database collects the URLs of the website that is suspected as phishing which is being reported. In addition, legitimate websites were collected from yahoo directory and starting point directory. These directories contain addresses of legitimate sites for different types of services. Like that, URL Reputation Data Set has 2396130 instance and 3231961 attributes.

5.2 Evaluation metrics

The evaluation of proposed URLs detection technique the following metrics as suggested by below equations,

$$TPR = \frac{|TP|}{|TP|+|FN|} \tag{9}$$

$$TNR = \frac{|TN|}{|TN|+|FP|} \tag{10}$$

$$FPR = \frac{|FP|}{|TN|+|FP|} \tag{11}$$

$$FNR = \frac{|FN|}{|TN|+|FN|} \tag{12}$$

$$ACC = \frac{|TP|+|TN|}{|TN|+|TN|+|FN|+|FP|} \tag{13}$$

$$EER = \frac{|FP|+|FN|}{|TN|+|TN|+|FN|+|FP|} \tag{14}$$

5.3. Performance analysis of proposed approach

The basic idea of our proposed approach is to detect the malicious URL from Suspicious URLs. In this approach, consist of three modules such as feature selection and classification. In feature selection, we used oppositional cuckoo search algorithm and classification we used fuzzy logic classifier. The based on the fuzzy score value we check whether the URL is malicious or not.

Table 1. Performance analysis of proposed method using various measures for Phishing Websites Data Set

Measures	OCS	CS	GA	WOP
TP	63	55	53	48
TN	28	31	31	31
FP	3	0	0	0
FN	6	14	16	21
TPR	0.913043	0.797101	0.768116	0.695652
TNR	0.903226	1	1	1
FPR	0.096774	0	0	0
FNR	0.086957	0.202899	0.231884	0.304348
EER	0.09	0.14	0.16	0.21
ACC	0.91	0.86	0.84	0.79

Table 2. Performance analysis of proposed method using various measures for URL Reputation Data Set

Measures	OCS	CS	GA	WOP
TP	2001	2094	2068	4
TN	2401	1818	1092	2641
FP	240	823	1549	0
FN	115	22	48	2112
TPR	0.945652	0.989603	0.977316	0.00189
TNR	0.909125	0.688376	0.41348	1
FPR	0.090875	0.311624	0.58652	0
FNR	0.054348	0.010397	0.022684	0.99811
ACC	0.925373	0.822367	0.664284	0.556023
EER	0.074627	0.177633	0.335716	0.443977

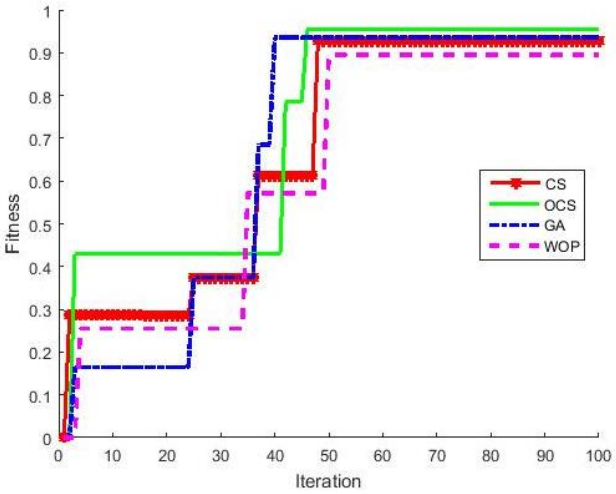

Figure.4 Performance of fitness function using different approaches

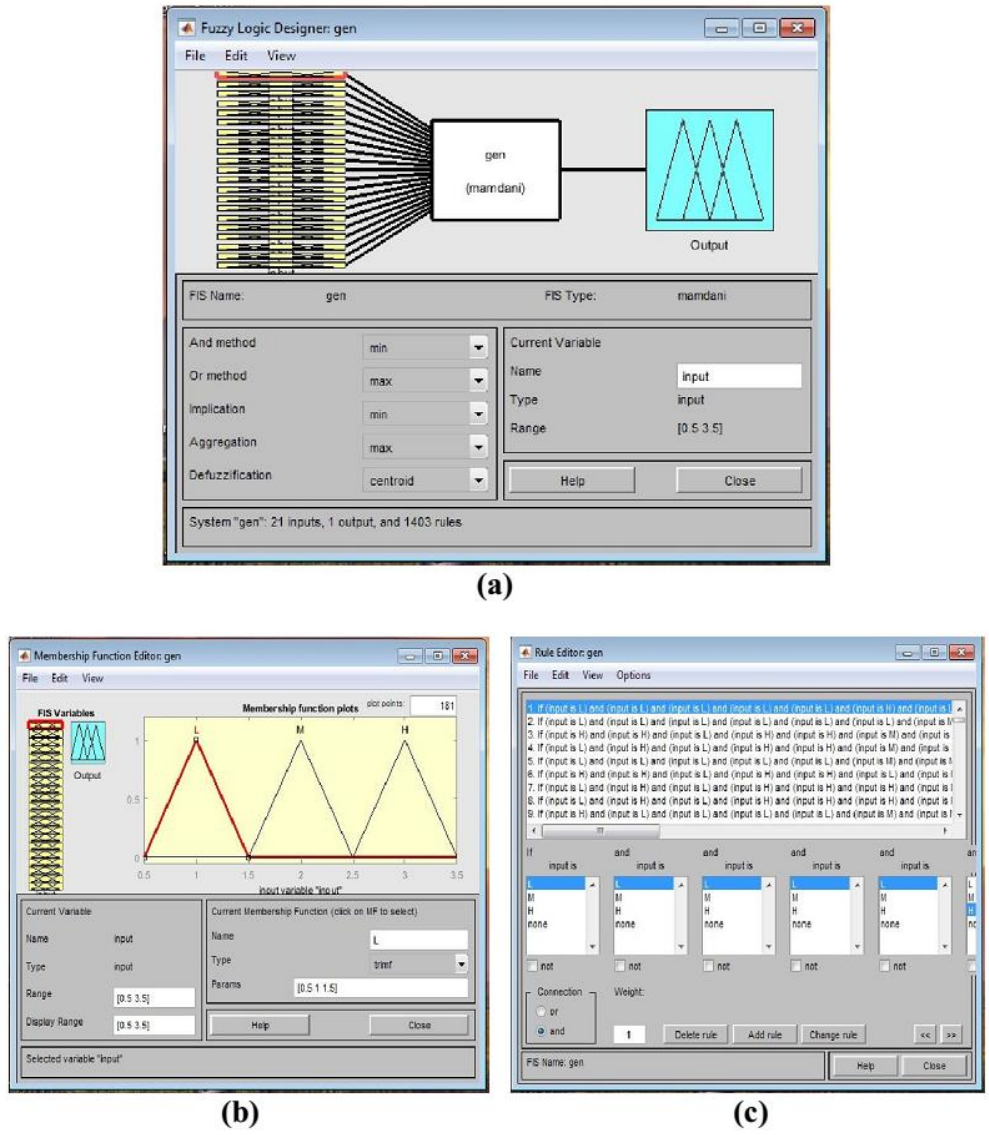

Figure.5 Visual representation of OCS algorithm with FLC (a) fuzzy structure (b) membership function (c) obtained Rules

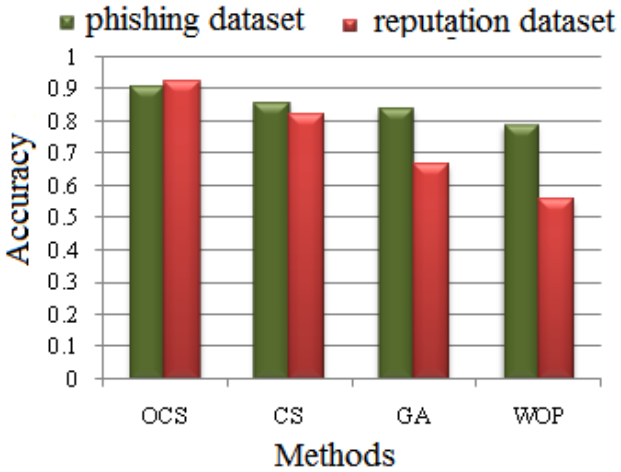

Figure.6 Performance of proposed approach based on accuracy plot

Discussion

The above table 1 shows the performance analysis of proposed method using various measures for phishing websites dataset. This dataset has totally 2456 instance and 30 attributes. If a malicious is proven present in a URL, the given test also indicates the presence of malicious, the result of the detection test is considered true positive (TP). Similarly, if a detection test is proven absent in a Malicious, the detection test suggests the malicious is absent as well, the test result is true negative (TN). Both TP and TN suggest a consistent result between the detection test and the proven condition (also called the standard of truth. If the test indicates the presence of malicious in a URL which actually has no such malicious, the test result is false positive

(FP). Similarly, if the result of the test suggests that the malicious is absent for a URL with Malicious for sure, the test result is false negative (FN). Both false positive and false negative indicate that the test results are opposite to the actual condition. In this study, we compared our proposed methodology with CS+FL based prediction, GA+FL based prediction (with optimization) and FL based prediction (without optimization). In CS+FL approach, the feature selection is carried out based on the CS algorithm and the classification is carried out using FL algorithm. Moreover, GA+FL approach, the feature selection is based on GA and classification is carried out FL algorithm. Similarly, without optimization based prediction the classification is carried out FL algorithm. Here, the features are directly given to the classifier. When analyzing table 1, we obtain the maximum TP value of 63 using proposed approach (OCS+FL), which is 55 for using CS+FL, 53 for using GA+FL and 48 for using without optimization (WOP) algorithm. The system obtained the maximum value of TP, TN and a minimum value of FP, FN means, we consider the system have good performance. From table 1, we obtain the maximum accuracy of 91% which is 86% for using CS+FL, 84% for using GA+FL and 79% for using WOP. From the result, we clearly understand our proposed approach is better than other approach. Because of, the CS+FL based prediction is well defined approach even through it has low search capacity and low convergence rate. Moreover, GA+FL also give the very slow convergence rate. Similarly, using WOP based prediction we cannot achieve the better results, because the large number of features is the great obstacles for prediction. In order to improve the search capability CS algorithm in our approach we introduce the opposition-based learning (OBL) with cuckoo search (CS) algorithm. Opposition-based learning is a rather simple concept that aims to improve the convergence rate and/or accuracy of computational intelligence algorithms. Therefore, we obtain the better result compare to other approach. Table 2 shows the performance analysis of proposed method using various measures for URL reputation dataset. Here, also our proposed approach achieves the maximum accuracy of 92% which is high compared to the existing approaches. When analyzing table 2, WO and GA approach are worst performance compared to other two approaches. Moreover, figure 4 shows the performance of fitness function using different approaches. Here, we compare our proposed OCS algorithm fitness value with other approaches. Additionally, figure 4 and 6 shows the visual results of performance of fuzzy logic classifier with CS and OCS. Moreover, figure 6 shows the Performance of proposed approach based on accuracy plot. Here, we used two types of dataset. When analyzing figure 7, we obtain the maximum of the accuracy of 92% using URL reputation dataset. From the above figures, we clearly understand our proposed approach have better performance compared to the existing approaches.

6. Conclusion

Malicious Web sites mainly encourage the expansion of Internet criminal behavior and constrain the development of Web services. To overcome this problem, this paper, we have explained the malicious URL detection based on oppositional cuckoo search (OCS) algorithm and fuzzy logic classifier (FLC). In the proposed model consist of three modules such as feature extraction, feature selection, and classification. In feature extraction module, we extract the four types of features present in each URL. After that, we reduce the feature using Oppositional cuckoo search algorithm. Finally, we classify the URL based on the Fuzzy logic classifier. Experimental results indicate that the proposed method of OCS+FLC for malicious detection framework have outperformed by having better accuracy of 92% which is high compared with existing approaches. In the future study, the creators will continue to examine and plan productive heuristic methodologies for feature reduction strategies to manage tremendous amounts of attributes and expansive quantities of URLs. In addition, the opposition cuckoo search algorithm can be utilizing a hybrid algorithm in further study. As a result, enhancing execution of feature reduction strategies as well as enhancing learning parameters of the fuzzy logic framework can be done. Finally, the parallel algorithm will be explored further in future studies to achieve a significant speedup.

References

[1] E. K. Reddy, “Neural Networks for Intrusion Detection and Its Applications”, *Proceedings of the World Congress on Engineering*, Vol. 2, No. 5, pp. 3-5, 2013.

[2] A. Anand and B. Patel, "An Overview of Intrusion Detection System and Types of Attacks It Can Detect Considering Different Protocols ", *Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2, No. 8, pp. 94-98, 2012.

[3] J. Ryan, R. Miikkulainen and M. J. Lin, "Intrusion Detection with Neural Networks", *to appear in*

Advances in Neural Information Processing Systems, pp. 943-949, 1998.

[4] S. Subashinin and V. Kavitha, "A survey on security issues in service delivery models of cloud computing", *ELSEVIER Journal of Network and Computer Applications*, Vol. 34, No. 1, pp. 1–11, 2011.

[5] Teresa F. Lunt, "A survey of intrusion detection techniques", Journal of Network and Computer Applications, Vol. 36, No. 1, pp. 42-57, 2013.

[6] S. Peddabachigari, "Modeling intrusion detection system using hybrid intelligent systems", *ELSEVIER Journal of Network and Computer Application*, Vol. 30, No. 1, pp. 114–132, 2007.

[7] O. Depren and M. Topallar, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks", *ELSEVIER Journal of Expert Systems with Applications*, Vol. 29, No. 4, pp. 713–722, 2005.

[8] J. Kim and P. J. Bentley, "Immune System Approaches to Intrusion Detection – A Review", *Journal of Natural Computing*, Vol. 6, No. 4, pp. 413-466, 2007.

[9] R. Shanmugavadivu, "Network intrusion detection system using fuzzy logic", Indian Journal of Computer Science and Engineering, Vol. 2, No. 1, pp. 101-111, 2011.

[10] Devikrishna and Ramakrishna, "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks", *International Journal of Engineering Research and Applications*, Vol. 3, No. 4, pp. 1959-1964, 2013.

[11] V. Gowadia, C. Farkas and M. Valtorta, "PAID: A Probabilistic Agent-Based Intrusion Detection system", *journal of Computers and Security*, Vol. 24, No. 7, pp. 529-545, 2005.

[12] M. M. M. Hassan, "Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, No. 7, 2013.

[13] Kaihwang, "Trusted Cloud Computing with Secure Resources and Data Coloring", *In Proceedings of IEEE Transaction of Internet Computing*, Vol. 14, No. 5, pp. 14-22, 2010.

[14] A. Zarrabi and A. Zarrabi, "Internet Intrusion Detection System Service in a Cloud", *International journal of computer science*, Vol. 9, No. 5, No.2, pp. 308-315, 2012.

[15] S. Yoo and S. Kim, "Two-Phase Malicious Web Page Detection Scheme Using Misuse and Anomaly Detection", *International Journal of Reliable Information and Assurance*, Vol. 2, No.1, 2014

[16] I. Butun, S. D. Morgera and R. Sankar, "A Survey of Intrusion Detection Systems in Wireless Sensor Networks", *IEEE communications surveys and tutorials*, Vol. 16, No.1, pp. 266-282, 2013.

[17] M. Mabzool and M. Z. Lighvan, "Intrusion detection system based on web usage mining", *International Journal of Computer Science, Engineering and Applications*, Vol. 4, No. 1, 2014.

[18] U. Ravale, N. Marathe and P. Padiya, "Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function", *Procedia Computer Science*, Vol. 45, pp.428-435, 2015.

[19] K. S. A. Kumar and Dr. V. N. Mohan, " Novel Anomaly Intrusion Detection Using Neuro-Fuzzy Inference System ", *IJCSNS International Journal 6 of Computer Science and Network Security*, Vol. 8, No. 8, pp.6-11 , 2008.

[20] S. R. Gaddam, V. V. Phoha and K. S. Balagani, "K-Means+ID3: A Novel Method for Supervised Anomaly Detection by Cascading K-Means Clustering and ID3 Decision Tree Learning Methods", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 3, pp. 345-354, 2007.

[21] S. Roy and S. S. Chaudhuri, "Cuckoo Search Algorithm using Lèvy Flight: A Review", *international journal of Modern Education and Computer Science*, Vol. 5, No. 12, pp. 10-15, 2013.

[22]A. Altaher, A. Almomani and S. Ramadass, "Application of Adaptive Neuro-Fuzzy Inference System for Information Secuirty", *Journal of Computer Science*, Vol. 8, No. 6, pp. 983-986, 2012

[23]https://archive.ics.uci.edu/ml/datasets/URL+Reputation