Rakesh Jha, 2021, 9:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Moth Flame Based Feature Selection for Ransom Ware Detection Training Model

M.Tech. Scholar Rakesh Jha, Asst. Prof. Sumit Sharma
Department of Computer Science and Engineering,
Vaishnavi Institute of Technology and Science,
Bhopal

Abstract- Technology increases the comfort and dependency for many of daily work. Cloud services taken by different organization to control, monitor various activities. As people gathering attracts different attacks that can harm service, product, data, etc. So protection from such kind of attacks needs to be done. This paper has developed a ransom ware detection model that can provide alarm the network when it detects such kind of pattern by trained neural network. In this work detection features were reduced by moth flame optimization genetic algorithm. Selected features were further processed to get training dataset for neural network learning. Paper has used error back propagation neural network. Testing and training was done on real ransom ware dataset. Result shows that Moth Flam Optimization Ransom ware Detection (MFORD) has increase true false alarm and reduces false alarm.

Keywords: Cloud computing, Intrusion detection, Machine learning, and Feature reduction.

I. INTRODUCTION

The appealing features of Cloud computing continue to fuel its integration in many sectors including industry, governments, education, entertainment, to name few [1]. Cloud computing aims to provide convenient, on-demand, network access to a shared pool of configurable computing resources, which can be rapidly provisioned and released with minimal management effort or service provider interactions [2].

The pay-as-you-go and the on-demand elastic operation Cloud characteristics are changing the enterprise computing model, shifting on-premises infrastructures to off premises data centers, accessed over the Internet and managed by cloud hosting providers. However, many security issues arise with the transition to this computing paradigm including intrusions detection.

Regardless the important evolution of the information security technologies in recent years, intrusions and attacks continue to defeat existing intrusion detection systems in Cloud environments [3, 4]. Attackers developed new sophisticated techniques able to bring down an entire Cloud platform or even many within minutes. New records are breached each year by attacker.

Recently a destructive DDoS attack has brought down more than 70 vital services of Internet including Github, Twitter, Amazon, Paypal, etc. Attackers have taken advantages of Cloud Computing and Internet of Things technologies to generate a huge amount of attack traffic [5, 6].

Employing effective IDS in the cloud is a challenge from different aspects. One aspect is the complication of the security problem due to the cloud's deep stack of dependent layers. The functionality and security of a higher layer depend on its lower layers. This aspect is further augmented by the sophistication of modern attacks.

Another aspect is the new requirements stemming from the unique characteristics of the cloud environment such as scalability and elasticity [7]. These requirements pose additional challenges on the traditional IDSs in many ways. Hence, the development of robust cloud-oriented IDSs must

identify and accommodate such unique cloud requirements. The last aspect is the deployment architecture selection as each choice has its own advantages and limitations with respect to the effectiveness of the IDS.

II. RELATED WORK

Kabir et al. [8] has developed an OALSSVM model (optimum allocation least square support vector machine). In this paper optimum allocation term select session from the whole dataset either from training or testing section of dataset. These selected session or samples were used to train the support vector machine model. So, output of proposed OALSSVM is depending on selected session which increases its accuracy of intrusion detection.

Chuanlong Yin [9] In this article, author examine how to present an interruption recognition framework in light of thoughtful learning, and this exertion offer a thoughtful knowledge approach for interruption recognition using recurrent neural networks (RNN-IDS). In addition, this exertion inspects the execution of the model in balancing categorization and multiclass arrangement, and the amount of neurons and characteristic learning rate impacts on the implementation of the planned show. This effort compares it and those of J48, artificial neural network, arbitrary woodland, bolster vector machine, and further machine knowledge approach planned by history analysts on the standard information directory index.

Moukhafi et al. [10] has proposed a feature reduction model for increasing the detection accuracy of intrusion in the network. This paper has utilized a particle swarm optimization genetic algorithm for the selection of features form the input dataset as per number of class for detection. Selected feature from the training dataset were used to train support vector machine. This hybrid genetic and SVM model work well to detect DOS attacks.

Kaiyuan et. al. in [11] propose a network intrusion detection algorithm combined hybrid sampling with deep hierarchical network. Firstly, we use the one-side selection (OSS) to reduce the noise samples in majority category, and then increase the minority samples by synthetic minority over-sampling technique (SMOTE). In this way, a balanced dataset can be established to make the model fully learn the features of minority samples and greatly reduce the model training time. Secondly, we use convolution neural network (CNN) to extract spatial features and Bi-directional long short-term memory (BiLSTM) to extract temporal features, which forms a deep hierarchical network model.

In [12] have proposed to utilize information mining system, order tree and bolster vector machine for intrusion discovery. Information mining system have made valuable strides towards arrangement of different issues in various issues, use information digging for tackling the issue of intrusion as a result of following reasons: It can process expansive measure of information.

Client's subjective advancement isn't vital, and it is more appropriate to find the disregarded and shrouded data. Machine learning is a logical teaches that enables PCs to learn in light of information and naturally figures out how to perceive complex examples and to settle on keen choice in light of information. ID3 and C4.5 two basic arrangement tree calculation utilized as a part of information mining. Bolster vector machines are an arrangement of related administered learning techniques utilized for grouping and expectation.

Author said C4.5 calculation is smarter to SVM in recognizing system intrusions and FAR (false caution rate) in KDD CUP 99 dataset.

Subramanian1 et. Al. in [13] shape the future generation of cloud security using convolution neural network because CNN can provide automatic and responsive approaches to enhance security in cloud environment. Instead of focusing only on detecting and identifying sensitive data patterns, ML can provide solutions which incorporate holistic algorithms for secure enterprise data throughout all the cloud applications.

III. PROPOSED METHODOLOGY

Cloud session need to be continuously monitor for identifying any unfair activity. Explanation of this monitoring model was proposed by this section of paper. Fig. 1 shows flow of proposed MFORD model. Data Collection of different session was done in the first phase of model. To increase the machine learning capability Invasive Weed genetic algorithm reduces the features of model.

International Journal of Science, Engineering and Technology

An Open Access Journal

At last when features get optimize then neural network get trained from the selected feature set.

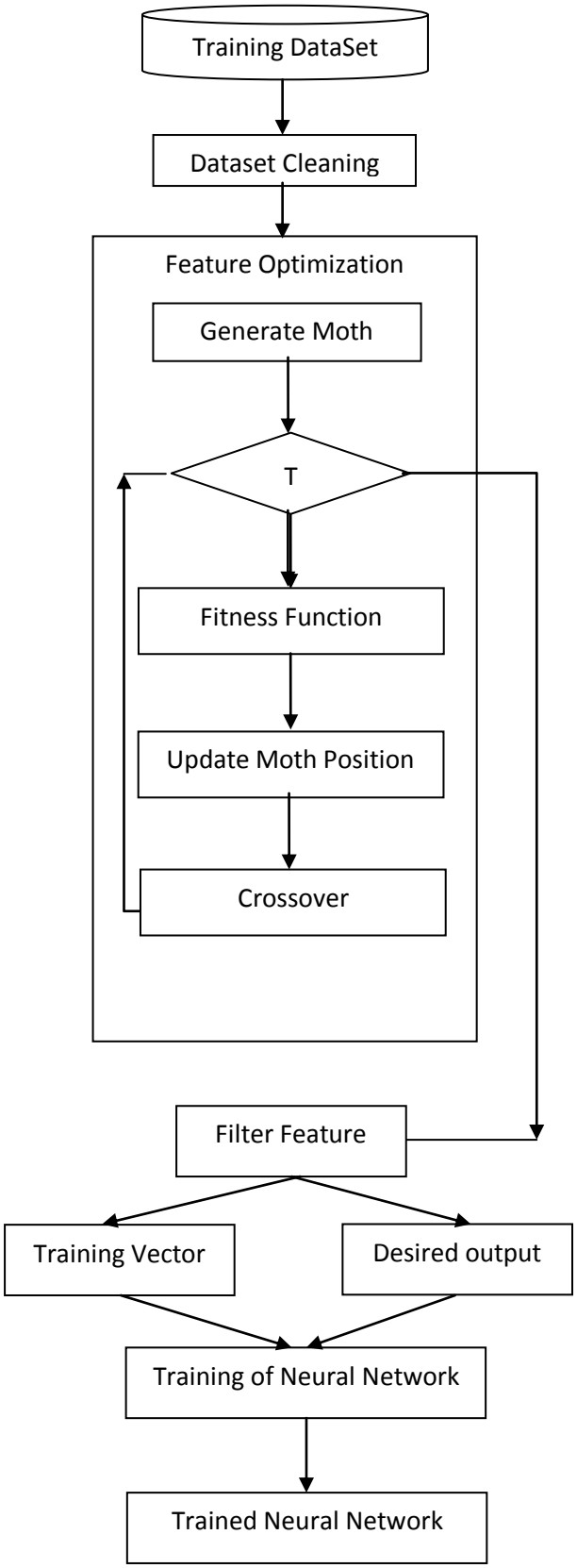

Fig 1. Moth Flame optimization Ransomware Detection (MFORD).

1. Dataset Cleaning:

Input raw session dataset of cloud environment need some pre-processing step before training of model. This step removes some of feature columns from the dataset. Such as session unique Id, some of textual information which not help in classification such as protocol layer name, DNS etc [14]. Whole dataset was dividing into training and testing session.

Each session class is separate into vector. Removal of such noise kind of information has increases the work efficiency by getting better training model.

2. Feature optimization:

Input processed matrix was further process by Moth Flame Optimization Algorithm to reduce values of training vector and increases learning accuracy.

3. Moth Flame Optimization Algorithm:

In this algorithm paper has consider each chromosome as a Moth. Objective of this algorithm was to find a Moth Flame belonging to path towards moon. Moth Flame is chromosome of this work.

4. Generate Moth Flames:

Moth Flames are group of chromosome and a chromosome is possible solution of optimized feature set. So a Moth Flame is a vector of n number of elements, where n is number of column in CD. Each element is a binary value in Moth Flame vector.

One show that a feature is considers for the training and zero shows that feature is not selected for the population. So if p number of Moth Flames generate then M is Moth Flame population matrix having pxn dimension. Selection of f number of feature in vector done by random value generator function Gaussian.

M←Generate_Moth Flame (p, n, f) ----------Eq. 2

5. Fitness Function:

Each Moth Flame was rank as per distance. So evaluation of distance done by fitness value. Moth Flame feature vector pass training vector to the neural network for training and measure the detection accuracy of the work [11]. This detection accuracy value is distance parameter in the work.

6. Update Moth Flame position:

Once F value obtain by fitness function then sort f in deseeding order and find best Moth Flame out of all chromosomes available in the population.

International Journal of Science, Engineering and Technology

An Open Access Journal

7. Crossover:

Genetic algorithm success depends on change of chromosomes, hence as per changing parameter X, number of random position value of Moth Flames were modified. This operation was not done in best local Moth Flame. In this step each Moth Flame X number of positions was modified randomly from zero to one or one to zero as per best local Moth Flame feature set.

This Moth Flame were further test for path distance and compared its fitness value with parent Moth Flame if child Moth Flame has better values then remove parent otherwise parent will continue.

After this step if maximum iteration steps occur then jump to filter feature block otherwise evaluate fitness value of each Moth Flame Moth Flame.

8. Filter Feature:

Once iteration gets complete then find best Moth Flame from the last updated population. Feature having value one in chromosome consider as selected feature for training vector and other consider as unselected. Desired output matrix was also prepared in this section.

Input: M, CD
Output: F

- Loop w=1:W // for w Moth Flames
- Loop s=1:CD // for s training session
- TV[s]←Training_Vector(W[w], CD[s])
- DO[s]←Desired_Output(W[w], CD[s])
- EndLoop
- TNN←Train_Neural_Network(TV, DO)
- Loop s=1:CD // for s training session
- TV←Training_Vector(W[w], CD[s])
- O←Predict(TV,TNN)
- If DO[s] equals O
- F[w]←Increment F by 1
- EndIf
- EndLoop
- Endloop

In above algorithm TV is training vector, DO is desired output.

9. Training of Neural Network:

Neural network consider takes input training vector and desired output during training. For each set of training vector neuron weight value adjust for e number of epochs.

Trained neural network was directly used for predicting the session class as attack or normal.

IV. EXPERIMENTS & RESULTS ANALYSIS

Implementation of proposed model was done on MATLAB software. Experimental work was done on machine having I3 processor with 4 GB RAM.

1. Dataset:

This dataset is collection of ransom ware attack session taken from [14].This dataset contains the dynamic analysis of 582 samples of ransom ware and 942 of good applications (good ware), i.e. 1524 samples in total. The dataset was retrieved and analyzed with Cuckoo Sandbox at the end of February 2016.

2. Results:

Comparison of proposed MFORD model was done with existing model DNAact-Ran proposed in [15].

Table 1. Precision based comparison of ransom ware detection models.

Dataset Size	DNAact-Ran	MFORD
800	0.7474	0.9674
950	0.866	0.9811
1100	0.866	0.9811
1250	0.866	0.9811
1400	0.866	0.9811

Precision values of ransom ware detection model shows that proposed model has increases the work detection value by 13.9% as compared to existing model. It was found that elected features for training the neural network has increases the work performance. Table 1 values shows that use of moth flame for feature selection algorithm is highly effective.

Table 2. Recall based comparison of ransom ware detection models.

Dataset Size	DNAact-Ran	MFORD
800	0.3991	0.4522
950	0.6199	0.6474
1100	0.539	0.571
1250	0.4741	0.5103
1400	0.4228	0.4597

Recall values of the ransom ware detection models for different dataset size were shown in table 2. It

International Journal of Science, Engineering and Technology

An Open Access Journal

was obtained that proposed model has increases the performance by 7.03% as compared to other existing models. Moth flame based feature optimization has increases the performance of the work. Learning of neural network with less and effective number of features increases the work performance.

Table 3. F-Measure based comparison of ransomware detection models.

Dataset Size	DNAact-Ran	MFORD
800	0.5203	0.6163
950	0.7226	0.7801
1100	0.6645	0.7219
1250	0.6128	0.6714
1400	0.5682	0.6261

F-measure values of ransom ware detection model shows that proposed model has increases the work detection value by 9.58% as compared to existing model. It was found that elected features for training the neural network has increases the work performance. Table 1 values shows that use of moth flame for feature selection algorithm is highly effective.

Table 4. Accuracy based comparison of ransomware detection models.

Dataset Size	DNAact-Ran	MFORD
800	42.76	49.96
950	59.31	66.14
1100	53.77	60.04
1250	49.08	55.32
1400	45.32	51.32

Accuracy values of the ransom ware detection models for different dataset size were shown in table 4. It was obtained that proposed model has increases the performance by 11.5% as compared to other existing models. Moth flame based feature optimization has increases the performance of the work. Learning of neural network with less and effective number of features increases the work performance.

V. CONCLUSIONS

Ransomwarre in cloud makes serious losses for service provider and user both. So detection of such attacks needs to be done at each session in the network. This paper has proposed a genetic and neural network hybrid model for ransom ware detection. It was obtained that proposed model has reduced the input training dataset by use of math flame optimization genetic algorithm and further paper has trained the neural network from selected feature set. Experiment was done on real ransom ware dataset different testing sessions.

Result shows that model has increases the precision by % as compared to existing model. In future scholars can train a model that can predict the class of the ransom ware as well.

REFERENCES

[1] D. A. Fernandes, L. F. Soares, J. V. Gomes, M. M. Freire, P. R. Inacio, Security Issues In Cloud Environments: A Survey, International Journal ´ Of Information Security 13 (2) (2014) 113–170.

[2] P. Mell, T. Grance, the Nist Definition of Cloud Computing.

[3] S. Iqbal, M. L. M. Kiah, B. Dhaghighi, M. Hussain, S. Khan, M. K. Khan, K.-K. R. Choo, On Cloud Security Attacks: A Taxonomy and Intrusion Detection and Prevention as A Service, Journal of Network and Computer Applications 74 (2016) 98–120.

[4] Wikipedia, 2016 Dyn Cyberattack [Online; Accessed 10-November-2017)].

[5] The Guardian, Ddos Attack That Disrupted Internet Was Largest Of Its Kind In History, Experts.

[6] Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, M. Rajarajan, "A Survey of Intrusion Detection Techniques in Cloud", Journal of Network and Computer Applications 36 (2013), Pp. 42–57.

[7] R. Vijayanand, D. Devaraj, and B. Kannapiran, ''A Novel Intrusion Detection System for Wireless Mesh Network with Hybrid Feature Selection Technique Based On GA and MI,'' J. Intell. Fuzzy Syst., Vol. 34, No. 3, Pp. 1243–1250, 2018.

[8] E. Kabir, J. Hu, H. Wang, and G. Zhuo, ''A Novel Statistical Technique for Intrusion Detection Systems,'' Future Gener. Comput. Syst., Vol. 79, Pp. 303–318, Feb. 2018.

[9] Chuanlongyin, Yuefei Zhu, Jinlong Fei, And Xinzheng He. "A Deep Learning Approach For Intrusion Detection Using Recurrent Neural Networks" Current Version November 7, 2017.

[10] M. Moukhafi, K. El Yassini, and S. Bri, ''A Novel Hybrid GA and SVM with PSO Feature Selection for Intrusion Detection System,'' Int. J. Adv. Sci. Res. Eng., Vol. 4, Pp. 129–134, May 2018.

International Journal of Science, Engineering and Technology

An Open Access Journal

[11] Kaiyuan Jiang, Wenya Wang, Aili Wang, and Haibin Wu. "Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network". IEEE Access February 24, 2020.

[12] Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao. "Predicting Students' Performance Using Id3 and C4.5 Classification Algorithms". International Journal Of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.5, September 2013.

[13] E. K. Subramanian, Lathatamilselvan. "A Focus on Future Cloud: Machine Learning-Based Cloud Security". Service Oriented Computing And Applications, 12 August 2019.

[14] Andronio, Nicoló, Stefano Zanero, and Federico Maggi. "HelDroid: dissecting and detecting mobile ransom ware." International Workshop on Recent Advances in Intrusion Detection. Springer International Publishing, 2015.

[15] Firoz Khan, Cornelius Ncube, and R.Lakshmana Kumar, Seifedine Kadry, Yunyoung Nam. "A Digital DNA Sequencing Engine for Ransom ware Detection Using Machine Learning". IEEE Access 2020.