As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 7(09): 01-06, 2021

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0709001
Available online at: http://www.ijmtst.com/vol7issue09.html

Invasive Weed Optimization Based Ransom-Ware Detection in Cloud Environment

Adil Hussain Mohammed

DC Access System (DCAS) Linux Engineer &amp; Department of Health Care Finance (DHCF) DHCF's headquarters 955 L'Enfant Plaza, SW, 3rd Floor, Washington, DC 20024.

To Cite this Article
Adil Hussain Mohammed. Invasive Weed Optimization Based Ransom-Ware Detection in Cloud Environment. *International Journal for Modern Trends in Science and Technology* 2021, 7, 0709003, pp. 01-06. https://doi.org/10.46501/IJMTST0709001

Article Info
Received: 09 August 2021; Accepted: 31 August 2021; Published: 01 September 2021

ABSTRACT

Cloud provide support to manage, control, monitor different organization. Due to flexible nature f cloud chance of attack on it increases by means of some software attack in form of ransomware. Many of researcher has proposed various model to prevent such attacks or to identify such activities. This paper has proposed a ransomware detection model by use of trained neural network. Training of neural network was done by filter or optimized feature set obtained from the feature reduction algorithm. Paper has proposed a Invasive Weed Optimization algorithm that filter good set of feature from the available input training dataset. Proposed model test was performed on real dataset, have set sessions related to cloud ransomware attacks. Result shows that proposed model has increase the comparing parameter values.

KEYWORDS: Cloud computing, Intrusion detection, Machine learning, Feature reduction.

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 brought down an entire Cloud platform or even many within minutes. New records are breached each year by attacker. Recently a destructive DDoS attack have 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 an 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.

RELATED WORK

Kabir et al. [8] has develop a 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 depend on selected session which increase 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 teach 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.

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 IWORD model. Data Collection of different sessionwas done in the first phase of model. To increase the machine learning capability Invasive Weed genetic algorithm reduces the features of model. At last when features get optimize then neural network get trained from the selected feature set.

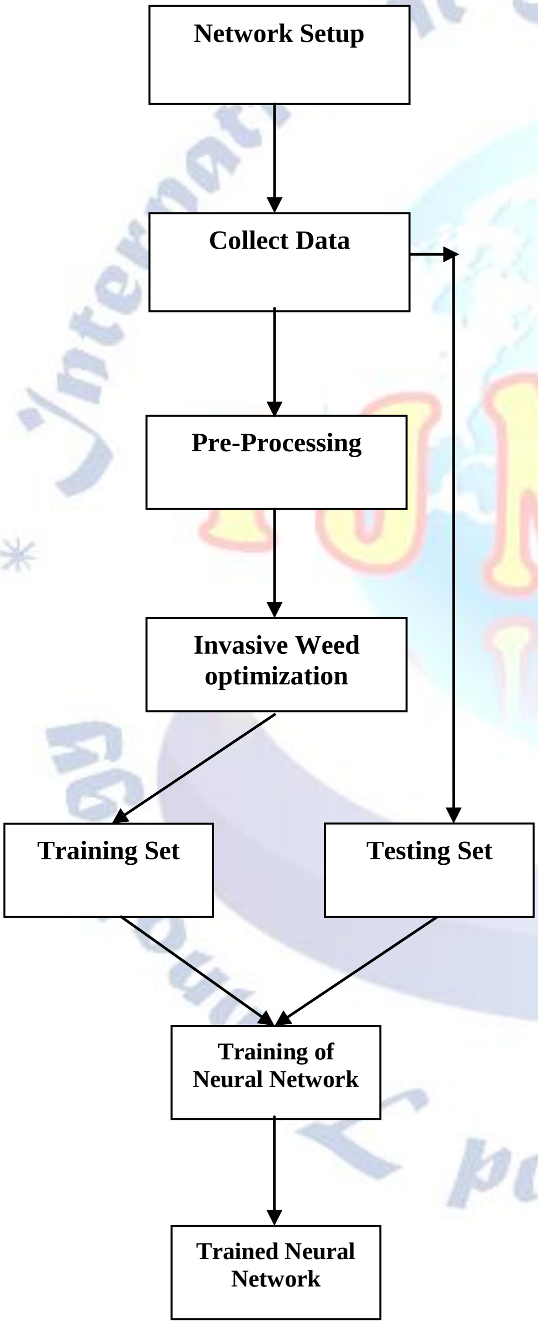

Fig. 1 Invasive weed optimization Ransomware Detection (IWORD)

Data Processing

Input raw session dataset of cloud environment need some pre-processing step before training of model. This step remove 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 divide 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.

Invasive Weed Optimization Algorithm

In order to reduce the feature set for learning some of columns present in processed dataset need to be removed. This feature updates was done by Invasive Weed optimization (IWOA). IWOA was proposed in [14] with an objective to get better seed set by a balanced weed optimization. Paper has utilizes this seed and weed optimization approach to filter features.

Generate Seeds

Each column in dataset act as seed element in the algorithm. Chromosomes in the population are seeds. Seed is a vector of 0, 1 where 1 means that feature present in the seed, while 0 means feature was not select. Size of seed is |D| let t features and population S_P have s number of seeds.

S_P←Gaussian_Distribution(t, s) --------Eq. 1

Reproduction

In this step fitness of the seed was checked by bowing in a area and check its production. Production of seed was test by fitness function. Chromosome having good fitness value are promote while other are replace or transform into other type of seed.

$$F_s = \sum_{i=1}^{t} Tf_i \times S_{Pi,s} \text{--------Eq. 2}$$

Spatial Dispersal

Fitness value of seed helps to find the better seed set. To improve the solution seed quality some changes need to be done randomly in low quality seeds. This operation act as crossover operation done in genetic algorithm. As solution need improvement in each iteration hence seed element get change from 1 to 0 or 0 to 1.

Optimize Features

After m number of max iteration best fitted seed act as final or optimize term set for the training of neural network. Seed vector having position element value 1 act as selected feature and other or element value 0 acts as rejected feature (Weed). Terms were select as final training feature set S_f. This paper has found that use of IWOA has increased the work efficiency of the ransomeware detection.

Input session S are further select or reject as per final seed vector S_f to get optimized feature set O_f.

```
Input: S, $S_f$
Output: $O_f$
Loop 1:RT
  Loop 1:t
    If $S_f[t]$ is Nonzero
      $O_f \leftarrow S[t]$
    EndIf
  EndLoop
EndLoop
```

Error Back Propagation Neural Network Training:

In this step Error Back Propagation neural network was consider for the training of intrusion detection system, here training vector consist of probability feature value set.

- A layer neural network is assume which have Four layers.
- Input layer neuron were identified as first layer, while network has two hidden layer neuron. Final or last layer is output layer.
- Weights between neuron is represent by w_{ij}, where i and j are neuron layers.

$$X_j = \sum x_i \,.\, w_{ij} \text{ ------Eq. 2}$$

where, $1 \leq i \leq n$; n is the number of inputs to node j, and b_j is the biasing for node j. Hence network will learn the weights between layers. This error need to be correct by adjusting the weight values of each layer. So estimation of error was done by eq. 3 [13].

$$e_k(n) = d_k(n) - y_k(n) \text{---Eq. 3}$$

This error need to be correct by adjusting the weight values of each layer. So here forward movement of the neural network is over and error back propagation starts. In back propagation weight adjust for learning of neural network as per input values.

Proposed IWORD algorithm

```
Input: RD // Raw Dataset
Output: TNN // Trained Neural Network

1. PD<--Pre-Process(RD) PD//Preprocessed Dataset
2. GS<-- Generate_Population(m, n) // m: number of features, n: number of solutions, GS: Generate Seeds
3. Loop 1: ITR //ITR: Iteration
4. F<--Spatial_Distribution_Function(GS, PD) // F: Fitness value
4. GS<--Seed_Reproduction(F, GS)
5. EndLoop
6. S<--Spatial_Distribution_Function(GS, PD) // S: Slected Features
7. ND<--Normalization(PD, S) // ND: Normalized Features
9. D<-- Desired_Output(ND)// D: Desired Output
10. NN<--Intialize_Netowrk() // NN Neural Netowrk
11. TNN<--Train(T, D, NN)
```

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. Comparison of model is done with work proposed in [16].

Evaluation parameters:

$$Precision = \left(\frac{True_{positive}}{(False_{positive} + True_{positive})}\right)$$

$$Recall = \left(\frac{True_{positive}}{False_{negative} + True_{positive}}\right)$$

$$F - Measure = \left(\frac{2xPrecisionxRecall}{(Recall + Precision)}\right)$$

$$Accuracy = \left(\frac{Correct\_class}{(Correct\_class + InCorrect\_class)}\right)$$

Dataset

This dataset is collection of ransomware attack session taken from [17].This dataset contains the dynamic analysis of 582 samples of ransomware and 942 of good applications (goodware), i.e. 1524 samples in total.The dataset was retrieved and analysed with Cuckoo Sandbox at the end of February 2016.

RESULTS

Comparison of proposed IWORD model was done with existing model proposed in [16].

Table 1 Precision based comparison of ransomware detection models.

Dataset Size	Previous Model []	IWORD
1000	0.7904	0.945
1100	0.781	0.945
1200	0.79	0.934
1300	0.7904	0.934
1400	0.7904	0.945

Table 1 shows that proposed model has improve the precision parameter of the ransomware detection by using filtered features for the training of neural network. It was obtained that UWORD model enhance the detection precision value by 16.17% as compared to previous model proposed in [16].

Table 2 Recall based comparison of ransomware detection models.

Dataset Size	Previous Model []	IWORD
1000	0.5743	0.7324
1100	0.5198	0.6927
1200	0.4742	0.6587
1300	0.4368	0.6243
1400		

Table 2 shows that size of dataset do not impact the work of ransomware detection. It was obtained that proposed model has use invasive weed optimization for feature selection is better as compared to other genetic algorithm used in previous model [16]. Use of linear regression for ransomware detection is tough as compared to neural network learning.

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

Dataset Size	Previous Model []	IWORD
1000	0.6652	0.8252
1100	0.627	0.7994
1200	0.5928	0.7763
1300	0.5627	0.7519
1400	0.8829	0.9717

Table 3 shows that proposed model has improve the F-measure parameter of the ransomware detection by using filtered features for the training of neural network. It was obtained that IWORD model enhance the detection f-emasure value by 19.24% as compared to previous model proposed in [16].

Table 4 Accuracy based comparison of ransomware detection models.

Dataset Size	Previous Model []	IWORD
1000	0.5375	0.7672
1100	0.5032	0.7493
1200	0.4738	0.7361
1300	0.4504	0.7210
1400	0.832	0.9687

Table 4 shows that size of dataset do not impact the accuracy value of work for ransomware detection. It was obtained that proposed model has use invasive weed optimization for feature selection is better as compared to other genetic algorithm used in previous model [16]. Use of linear regression for ransomware detection is tough as compared to neural network learning.

CONCLUSIONS

Detection od ransomware in clud environment is important factor for security of user data safety. This paper has proposed a ransomware detection machine learning model. Neural network was used for the detection of ransomware based on selected features of sessions. Feature selection was done by invasive weed optimization algorithm that do not need any training or guidance. Experiment was done on real dataset with different testing size. Result shows that proposed model has increased the precision value by and accuracy by 29% as compared to previous existing model. In future scholar can use different other dataset for detecting other type of attacks.

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.

[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] A.R. Mehrabiana,b, C. Lucasc. "A novel numerical optimization algorithm inspired from weed colonization". Science Direct, cological Logistics 2006.

[15] Daniele Sgandurra, Luis Muñoz-González, Rabih Mohsen, Emil C. Lupu. "Automated Dynamic Analysis of Ransomware: Benefits, Limitations and use for Detection." arXiv preprint arXiv:1609.03020, 2016.

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

[17] https://github.com/PSJoshi/Notes/wiki/Datasets.