An Optimized Approach to Detect Attacks in IDS using Fuzzy Methodology

C.M.Nalayini
Department of Information Technology, Velammal Engineering College.
Email: nalayinicm13@gmail.com
R.RamyaEswari
Department of Information Technology, Velammal Engineering College.
Email: ramyaeswari1098@gmail.com
R.Swedha
Department of Information Technology, Velammal Engineering College.
Email: swedharajendren98@gmail.com
P.S.Sushila
Department of Information Technology, Velammal Engineering College.
Email: sushilakannan16@gmail.com

ABSTRACT

Intrusion detection system plays an important role in assuring information security. Intruders can launch attacks from a compromised user account without being identified. The key technology is to accurately identify various attacks in the network. But Industrial process automation isundergoing an increased use of information communication technologies due to high flexibility interoperability and easy administration. However, traditional intrusion detection systems for the Information Technology domain are not entirely suitable for industrial process automation. Hence, the combined use of fuzziness-based and RNN IDS is very suitable for the classification with high accuracy and that its performance is superior to that of traditional machine learning classification methods. This model improves the accuracy of the intrusion detection by using of Machine learning methodologies and fuzziness has been used in identifying various types of attacks, and a machine learning approach used for preventing intrusions. As a result, the hypothesis of security violations can be detected by monitoring system audit records for abnormal patterns of system usage and the access controls to grant or restrict the levelof access to the network can be defined as the result by enhancing the detection rate of the intrusions which is very effective.

Keywords - **Semi supervised learning (SSL), Recurrentneural network (RNN), single hidden layer feedforward network (SLFN), fuzziness, Anova F-test.**

I. INTRODUCTION

The intrusion detection system is a software that helps to supervise a system network vicious activity and action violations. All vicious activity can be commonly described to the central authority or collected eventually using defense data and action administration system. This system especially introduced to detect the Signature Based Attacks and Anomaly Based Attacks. Here, the Machine learning methodologies have been applied to identify the various types of attacks, and can help the network administrator take the corresponding measures for preventing intrusions. An intrusion detection system is basically different from firewall. The intrusion detection system can be categorized by where the attack arises either in network or the host of the system. Intrusion detection system basically focus on analyzing achievable circumstances, logging data about the user and broadcasting the attack. In addition, companies use intrusion detection system for various purposes like classifying the problem with guaranteed policy, detailing current hazards. The intrusion detection system is a apathetic technology, it encounter and accept the problem but break the flow of the network traffic. These are the best intrusion detection system tools such as snort, Suricata etc.

1.1 Characteristics

An intrusion detection system should identify the following issues,

- It should be fault tolerant.
- It must impose minimal overhead.
- It must scope with changing system behavior.
- It must be easily able to identify different types of errors such as false positive, false negative or subversion errors.

1.2 Advantages of IDS

- The network or computer is frequently monitored for any invasion or attack.
- It effectively avoids any damage to the network.
- It provides user friendly interface

1.3 Disadvantages of IDS

- It produces false report of malicious activity.
- Encrypted packets are notprocessed.
- False positives are frequent.
- They are susceptible to protocol-based attacks.

1.4 Features of IDS

- It has rules-based traffic collection engine to perform content patternmatching.

- It is a web based administrator.
- It uses html template called themes.

II. Types of Attacks

Denial of Service (DOS): Itis an attack which is meant to shutdown a machine or network making it inaccessible to its intended users.

User to Root (U2R): It is an attack launched by an attacker to gain unauthorized access to a victim machine in the entire network.

Probe (Probing):It is an attack that creates new threat for collaborative intrusion detection system.

Remote to Local (R2L): It is an attack that targets one or network of computers which steals data illegally by introducing viruses or other malicious software.

III. Literature Survey

3.1 Quantitative Measure Of Fuzziness With The Non-Probabilistic Entropy analogous to shannon's Information Entropy

Zadeh et al [11] also generalized the probability measure of an event of a fuzzy event and suggested using entropy in information theory to interpret the uncertainty associated with a fuzzy event. They considered fuzziness to be a type of uncertainty and also defined a quantitative measure of fuzziness with the non-probabilistic entropy analogous to Shannon's information entropy. They also proposed three properties that fuzziness should hold. These properties depict that, the fuzziness degree should attain its maximum when the membership degree of every element is equal and its minimum when every element either belongs to the fuzzy set or absolutely not. In the study, they consider fuzziness as a type of cognitive uncertainty, coming from the transition of uncertainty from one linguistic term to another, where a linguistic term is a fuzzy set defined in a certain universe of discourse. Based on the fuzziness value we categorize the samples into three groups, i.e., Low fuzziness group, Mid fuzziness group, and High fuzziness.

3.2 Generalization Performance of Single Layer Feed Forward Neural Network

Schmidt et al[5]. Are the pioneers who earlier studied the impact of random initialization on generalization performance of single layer feedforward neural network. They experimentally demonstrated that single layer feedforward neural network can obtain a better performance by choosing random weights associated with the input layer and by analytically computing the weights of the output layer. This research regarding the non-iterative training of neural networks using randomization. The researchers also concluded that, in single layer feedforward neural network, the weights of the output layer are significantly most important than the weights found in the hidden layer.

[1] Intrusions Into Network Attached Industrial Control Systems

Long et.al. [21] proposed methods of mitigating denial of service attacks that originate from either local or wide area networks by modeling stochastic process of packet delay jitter and loss. An approach to detect the intrusions into network attached system Industrial Control Systems (ICS) by measuring and verifying data that is transmitted through the network but is not inherently the data used by the transmission protocol. PLC shadowing and data duplication can be used to mitigate the attacks on industrial control systems. Attack mitigation techniques include the use of firewalls, recommended by the National Institute of Standards and Technology, and some protocol modifications to prevent man-in-the-middle attacks.

3.4 Deep Learning With Network Packets

Torres et al., changed network traffic features into a sequence of characters and then used RNNs to learn their temporal features, which were further applied to detect malware traffic. The common point of the research methods is that, CNNs or RNNs alone used and learned a single type of traffic feature. Network traffic has an obvious hierarchical structure, a sequence of traffic bytes will be tracked. Based on the format of specific network protocols, multiple traffic bytes are combined to form a network packet, and multiple network packets communicated between two sides are further combined to form a network flow. Through this, traffic bytes, network packets and network flows have been used in the field of natural language processing. The training and testing times, the input data for the HAST-IDS consists of raw network traffic, the training and testing include the time required for feature extraction and feature selection manually designed features and do not require time for feature extraction and selection. But host-based IDS monitor activities associated with a particular host, and a network-based IDS listens to network traffic. But host-based IDS monitor activities associated with a particular host, and a network-based IDS listens to network traffic.

3.5 A Novel Network Architecture for DDoS Attack

Phan Van Trung et al. combines a novel network architecture for DDoS Attack Prevention based on statistical parameters of incoming traffic using the combination of a hard threshold method and Fuzzy Logic algorithm. The statistical parameters, which are examined by switches and periodically sent to the Controller in the SDN architecture, comprise distribution of inter-arrival time, distribution of packet quantity per flow and flow entries to a server. A security application running on the controller applies the hybrid algorithm with combination of the hard threshold and fuzzy logic schemes in response to analyzing statistical parameters and decides whether the corresponding server is either under a DDoS attack or in the normal state. The attack prediction and Prevention, especially for Distributed Denial of Service (DDoS) attacks is a major issue in SDN environments, during both states of normal and DDoS attack traffic.

3.6 Intrusion Detection Using Fuzzy Inference Systems

Trung et al. combines the hard thresholds of intrusions which has been detected and fuzzy inference system (FIS) to detect risk of DDoS attacks based on the real traffic characteristic under normal and attack states. A deep learning approach for flow-based anomaly detection in an SDN environment and shows strong potential to be used for flow-based anomaly detection in SDN environments. It uses Three features which are most commonly used for are chosen for detecting the attacks and they are Distribution of Inter-arrival Time, Distribution of packet quantity per flow and Flow quantity to a server. They also use many feature selection algorithms to increase the detection accuracy. The attack prediction and Prevention, especially for Distributed Denial of Service (DDoS) attacks is a challenge in SDN environments and analyzes the characteristics of traffic flows up-streaming to a Vietnamese ISP server, during both states of normal and DDoS attack traffic.

3.7 A novel semi-supervised SVM classification technique.

Maulik and Chakraborty[20] designed a novel semi-supervised SVM classification technique that exploits both labelled and unlabeled data points by considering the problem of pixel classification of remote sensing images and applied the marginal maximization principle to both labeled and unlabeled patterns. They experimentally confirmed that their learning scheme removes unnecessary points to a great extent from the unlabeled data and increases the accuracy level. In the next section, we present our proposed algorithm, which relies on fuzziness outputted by the classifier on a group of unlabeled data. The algorithm can be utilized in an effective way for the improvement of classifier performance and to have lower computational cost.

3.8 Intrusions Into Network Attached Industrial Control Systems

Long et al. [21] proposed methods of mitigating denial of service attacks that originate from either local or wide area networks by modelling stochastic process of packet delay jitter and loss. An approach to detect the intrusions into network attached Industrial Control Systems (ICS) by measuring and verifying data that is transmitted through the network but is not inherently the data used by the transmission protocol. PLC shadowing and data duplication can be used to mitigate the attacks on industrial control systems. Attack mitigation techniques include the use of firewalls, recommended by the National Institute of Standards and Technology, and some protocol modifications to prevent man-in-the-middle attacks.

IV. PROPOSED WORK

4.1 Data Pre-Processing

The NSL-KDD dataset have been used to find the intrusion detection. It consists of 38 numeric features and 3 non-numeric features. The input value of RNN-IDS should be a numeric matrix, and so nonnumeric features in the NSL KDD dataset has to be converted into numeric feature. Thus, taking the input as in the form of NSL KDD dataset. Then obtaining the output in the form of numeric features that is the non-numeric data are converted into numeric data.

4.2 Softmax Function

The SoftMax function is aconsideration of thecoordination function that "smashes" a K-dimensional vector to a K-dimensional vector of real values which lies in the range of 0 to 1.the numeric data are taken as the input and this function is then applied to those numeric data and find the min and max values of the data that ranging from 0 to 1.

4.4 Fuzziness

Fuzzy logic has become the most successful technology today for developing mature control system. The term fuzziness refers to the unclear border between two lexical terms and is based on the membership function of fuzzy sets.It is commonly used to allocate the weights for the fuzzy sets in the neural networking, It is the anticipation measure of an event to a fuzzy event and suggested using entropy in information theory to interpret the uncertainty associated with a fuzzy event.

4.5 Neural Network with Random Weights

Artificial neural networks are skilled using a stochastic optimization algorithm called stochastic gradient descent. It uses volatility in order to find a good enough set of weights for the specific mapping function from inputs to outputs in your data that is being learned.

V. METHODOLOGY

The first and most important step is to initialize the weights of network randomly and repeat the process

- Randomly select the input parameters
- Compute the hidden layer output matrix.
- Calculate the output weight.

5.1 Anova F-Test

Anova F-Test is a analytical method used to test differences between two or more means. It is an algorithm, in which all testing Ts sampleswere categorized into three groups according to the magnitude of fuzziness, and the group with highest accuracy was incorporated into the original training set Tr. Retraining was performed with the new training set Tr'. Their proposed technique is regarded as an approach to Semi-Supervised Learning in which some samples with unknown labels having low fuzziness participate in the training process.

Input: • *T d*: Labeled dataset, unlabeled dataset and test dataset are taken as the input.

- Classifier: C NNR w.
- H: Number of hidden nodes.
- Hidden-node output function: g(z) = 1 1+ e −z.
- Output: • TsAccuracy: Testing accuracy.

5.2 Detection

The intrusion detection is to improve the accuracy of classifiers in effectively identifying the intrusive behavior.

INPUT: Predictedvalues through classification.

OUTPUT: Classify and predict

A set of predicted data from the updated weights, and a set of values from classification.

Let S be the matrix of size that contains all the predictions that the data have assigned to the network packets, used for identifying the Intrusion.

- To get the prediction on a intrusion by, we can calculate the dot product of the two vectors.
- The error between the estimated prediction and the real prediction, can be calculated by the following equation for each user-item pair:

$$e^{2}_{ij=(}r_{ij}\text{-}\hat{r}_{ij)}{}^{2}=(r_{ij\text{-}\sum}{}^{k}{}_{k=1}p_{ik}q_{kj})^{2}$$

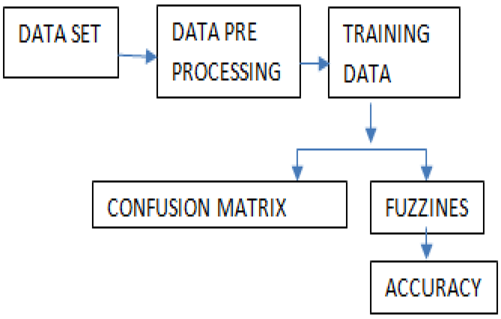
- To minimize the error, we have to modify the values of and. we need to know the gradient at the present values, and therefore we differentiate the above equation with respect to these two variables separately

$$e^{2}_{ij=-2(}r_{ij}\text{-}\hat{r}_{ij)}{}^{2}=(p_{ik})=-2e_{ij}p_{ik}$$

- Compute two matrices P and Q such that P X Q

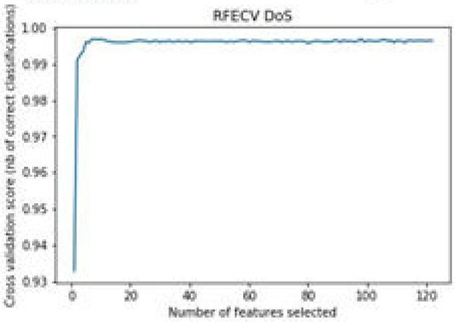
VI. System Design

The system design is given in the figure. In which the workflow of the project is explained clearly.



VII. Experimental Result:

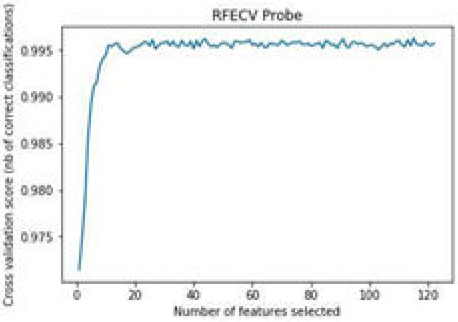
7.1 Matplotlib For Dos With Accuracy



```
Accuracy: 0.99639 (+/- 0.00341)
```

Matplotlib is the graphical representation of the python programming language where the accuracy for dos attack is 0.99639.

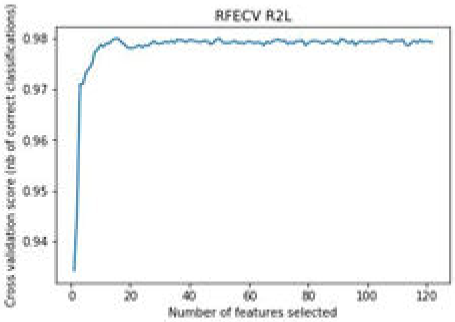
7.2 Matplotlib For Probe With Accuracy



```
Accuracy: 0.99571 (+/- 0.00328)
```

Matplotlib is the graphical representation of the python programming language where the accuracy of the probe attack is 0.99571.

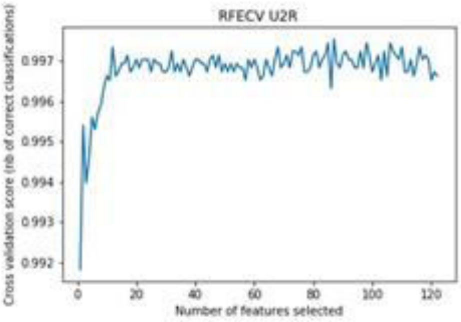
7.3 Matplotlib For R2l With Accuracy



```
Accuracy: 0.97920 (+/- 0.01054)
```

Matplotlib is the graphical representation of the python programming language where the accuracy of the r2l attack is 0.97920.

7.4 Matplotlib for U2R With Accuracy



```
Accuracy: 0.99663 (+/- 0.00259)
```

Matplotlib is the graphical representation of the python programming language where the accuracy of the u2r attack is 0.99663.

VIII. Conclusion

The RNN-IDS model not only has a strong modeling ability for intrusion detection, but also has lower accuracy in both binary and multiclass classification. Under the task of multiclass classification on the NSL-KDD dataset, when compared to the fuzziness based Intrusion Detection Systems. The Fuzziness based IDS model can effectively improve both the accuracy of intrusion detection and the ability to recognize the intrusion type. The proposed IDS is an adaptive solution which provides the capability of detecting known and novel attacks as well as being updated according to the new input from human experts in a cost-effective manner. The experiments indicated that both of the proposed methods perform slightly faster than the traditional method of full the sizeretraining when of the training set exceeds. The experiments indicated that the system was capable of detecting both known and novel traffic types while offering acceptable rates of detection and false positives. In addition, the experimental results showed that the performance of the system was improved after being updated according to the newly arrived attacks.

REFERENCES

[1] C. Chen , Y. Gong (2008), Y. Tian , Semi-supervised learning methods for network intrusion detection, in: *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics*, pp. 2603–2608 .

[2] Y. LeCun, Y. Bengio, and G. Hinton (2015), "Deep learning,"Nature, vol. 521, no. 7553, pp. 436-44.[3] A. Blum , S. Chawla (2011) , Study from semi supervised data using graph mincuts, in: *Proceedings of the Eighteenth International Conference on Machine Learning, pp.* 19–26 .

[3] W. Chen , Y. Shao , N. Hong (2013), Laplacian smooth twin support vector machine for semi-supervised classification, Int. J. Mach. Learn. Cybern. 5 (3), 459–468 .J. Schmidhuber (2015), *"Deep learning in semantic networks: an analysis of," Neural networks,* vol. 61, pp. 85-117.

[4] L. Liu, L. Shao, X. Li, and K. Lu (2016), "Learning spatio-temporal representations foraction recognition: A genetic programming approach," *IEEE transactions on cybernetics,vol*. 46, no. 1, pp. 158-170.

[5] A. Fujino , N. Ueda , K. Saito (2007), A hybrid generative/discriminative classifier design for semi-supervised learning, Trans. Jpn. Soc. Artif. Intell. 21, 301–309 .

[6] Liu, Y. T. Su, W. Z. Nie , and M. Kankanhalli (2017),"Hierarchical Clustering Multi-Task studying for Joint Human Action Grouping and Recognition, *"IEEE Transactions on Pattern Analysis and Machine Intelligence* ,vol. 39, no. 1, pp.102-114.

[7] J. Wu, Y. Zhang, and W. Lin (2016), "Good practices for learning to recognize actions using FV and VLAD," *IEEE transactions on cybernetics, vol.* 46, no. 12, pp. 2978-2990.

[8] J. Cao , J. Hao , X. Lai , C.-M. Vong and M. Luo (2016), Ensemble extreme learning machine and sparse representation classification algorithm, J. Frankl. Inst. 353 (17), 4526–4541.

[9] L, Zadeh, Probability measures of fuzzy events.J.Math.Anal.Appl.23(2)(1968)421-427

[10] D.E. Denning (2007), An intrusion-detection model, *IEEE Trans. Softw. Eng*. 13 (2) 222–232.

[11] G.B. Huang , L. Chen and C.K. Siew (2006), *Universal approximation using incrementalconstructive feedforward networks with random hidden nodes, IEEE Trans. Neural Netw.* 17 (4), 879–892 .

[12] G.B. Huang , Q.Y. Zhu and C.K. Siew (2004) , Extreme learning machine: a new learning scheme of feedforward neural networks, in: *Proceedings of the IEEE International Joint Conference on Neural Networks,* vol. 2, IEEE, pp. 985–990 .

[13] Badran, Khaled , & Rockett, Peter (2012), Mutual in- formation based feature selection for intrusion detection. Network and Computer Application, 34 , 1184– 1199.

[14] Bhuyan, M. H. , Bhattacharya, D. K. , & Kalita, J. K. (2014). *Network anomalydetection: Methods, systems and tools. IEEE Communication Surveys and Tutorials*, 16 , 303–336 .

[15] Bolon-Canedo, V. , Sanchnez-Marano, N. , & Alonso- Betanzos, A. (2011). Feature selection and classification in multiple class dataset.An application to KDD cup 99 dataset. Expert System with Applications, 38 , 5947–5957.

[16] Grochocki and W.Sanders (2014), *A framework for evaluating intrusion detection architectures in advanced metering infrastructures, IEEE Transactions on Smart Grid,* vol.5(2), pp.906–915.

[17] [19] E. Hernndez-Pereira , J. Surez-Romero , O. Fontenla-Romero and A. AlonsoBetanzos (2009), Conversion methods for symbolic features: a comparison applied to an intrusion detection problem, Expert Syst. Appl. 36 (7), 10612– 10617

[18] Maulik,D.Chakraborty,A novel semi supervised SVM for pixel classification of remote sensing imagery, Int J Mach Learn Cybern 3 (3) (2011) 247-258.*IEEE Transcation dependable and secure computing vol*(13)(2) March\April(2016).