International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 12th Jan 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-01

DOI: 10.48047/IJIEMR/V10/I01/13

Title: **Optimised Feature Selection with CNN for NIDS of Cloud Environment**
Volume 10, Issue 01, Pages: 74-78
Paper Authors

S.Asha Varma, K.Swathi, S.Nahida

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Optimised Feature Selection with CNN for NIDS of Cloud Environment

[1]S.Asha Varma, [2]K.Swathi, [3]S.Nahida
[1]Assistant Professor NRI Institute of Technology, Pothavarapadu, Andhra Pradesh, India
asha.varma@nriit.edu.in
[2]Professor NRI Institute of Technology, Pothavarapadu, Andhra Pradesh, India
kswathi@nriit.edu.in
[3]Associate Professor NRI Institute of Technology, Pothavarapadu, Andhra Pradesh, India
nahida@nriit.edu.in

Abstract: In recent years predicting the security threats become more complex for the system administrators, especially in the cloud environments.This is because of the distribution of attack samples and benign samples available in the network traffic datasets are much skewed.Using Machine learning techniques, the Network Intrusion Detection System(NIDS) are yielding good results in this type of problems.However even there are more challenges that how far the predicted data is more efficient and accurate as the false alarm rates are high. This paper proposed a framework to design an efficient NIDS by adopting correlation coefficient as a feature selection method to get an Optimized feature subset and a Deep Convolution Neural Network Technique for attack prediction. A benchmark dataset, CICIDS-2017 is used to implement the proposed model. Two performance metrics, Accuracy and computational complexity are considered and the proposed model is compared with traditional classification techniques. Finally, results were presented and conclusions were drawn.

Keywords-Intrusion Detection Systems;Correlation Coefficient;Feature selection; Convolutional Neural Networks.

1. INTRODUCTION

In recent years cloud computing has taken a big stretch in the field of computing because of its pay-on demand feature. For delivering the enterprise applications and for extending infrastructure or launching new additions,the one stop solutions is “Cloud Computing”. It provides “Software as a service”, “Platform as a service” and “Infrastructure as a service”. All these pool of resources are available to user on demand. It provide access to computing over Internet through virtual environment through various cloud models.The network security should receive more and more attention because of the fast development of Internet. Research on identifying abnormal behaviour over a network has occupied prominent role in the field of network security.Intrusion Detection System (IDS) is a part of Network Security design that monitors network traffic for malicious activity and alerts the system or Network Administrator to take necessary actions.To monitor all the network traffic, a large NIDS server can be set up as backbone network. Intrusion Detection builds a predictive model to identify attack instances[1].In order to build a model, the network traffic data should be analysed.As the data consists of too many features that contains false correlation,classification of abnormal behaviour is a complex task.These features or instances may be irrelevant or redundant because of large scale high dimensional data.In order to

reduce dimensionality, many machine learning techniques have evolved.The feature selection algorithm plays a major role in removing irrelevant and redundant attributes without decreasing performance and improve the accuracy in less time. In Machine learning there are different Feature Selection algorithms available presently. These algorithms worked on existing datasets but not proven for latest attacks over network. In view of above necessities, an efficient feature selection algorithm Correlation coefficient approach with CNN is proposed on latest dataset.This proposed methodology obtains the optimized feature set and predicts the abnormal behaviour(attack) by minimizing the computational complexity and maximizes the accuracy.

2. LITERATURE REVIEW

Feature selection methodologies removes the features which are unrelated to the task from the dataset [2].The three types of feature selection algorithms are filter, wrapper and embedded.Filter methods are preprocessing methods.They evaluate the feature based on the scores [3]. The wrapper models select the feature by optimising the predictor[4].In embedded method the model is trained first to get the weights of eachfeature and selects the feature based on the weights. It is said that to detect the unknown attacks and to have effective Intrusion Detection System , the deep CNN is the much helpful technique[5].Now a days as the data is high dimensional data, traditional techniques cannot identify the latest attacks over a networks. To face the challenges of NIDS Deeplearning techniques improve the performance of NIDS.[6].

3. METHODOLOGY

In order to predict the malicious attacks, the traditional machine learning techniques tend to identify with less accuracy and more computational complexity. So, in order to enhance the performance,this paper aims at developing an efficient methodology that predicts and classify the malicious behaviour over a network with more accuracy and less computational complexity. This methodology is implemented using feature selection method Correlation coefficient with Deep CNN for NIDS of Cloud environment on latest bench mark dataset CICIDS 2017 which has latest attacks.The classification and prediction is done by using CNN for the Optimised Feature Subset which is obtained by applying Correlation coefficient filter approach for the dataset.The Proposed methodology is represented diagrammatically in the following Figure.1 respectively.

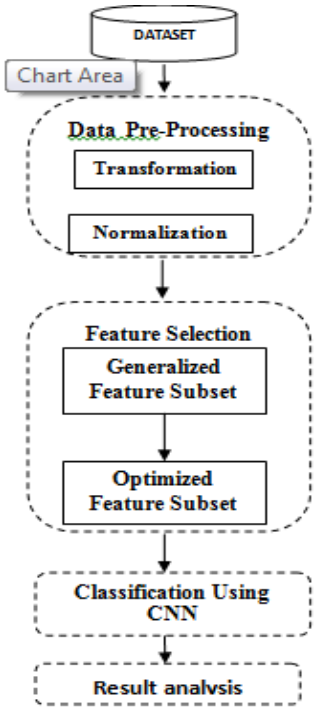

Figure.1: Proposed Methodology for Finding Optimal Feature Subset Selection with CNN.

3.1. Dataset used:

We use the latest benchmark dataset CICIDS 2017 which has the latest type of attacks on cloud environment.The dataset contains benign data which look like real world data.By using this dataset the anomaly based intrusion detection approaches may get correct and accurate performance evolutions.

3.2. Data Preprocessing CICIDS-2017:

Data preprocessing is necessary to make it as input for filtering.This process is done by performing Data Transformation and Data Normalization.

3.2.1. Data Transformation:

The features in the dataset should be transformed from nominal to numeric values through Data Transformation which is suitable for CNN.

3.2.2. Data Normalization:

As the datasets are too large we need to normalize the dataset to improve the performance of the IDS [7].

3.3. Feature Selection:

As we are bounded with huge amount of high dimensional data the features may redundant or irrelevant which may lead to curse of dimensionality. This may even lead to high false alarm rates. So, in view of this the efficient feature selection algorithm is selected to reduce the features. In this paper ,the Correlation coefficient algorithm is used for this feature reduction process.

3.3.1. Correlation Coefficient:

Correlation defines the linear relationship between the two variables and specify how similar they are.Features having high correlation value are more linear dependent and have same effect on dependent value.Using this value we can able to drop one of the feature [8].Hence by using this approach we get Optimised Feature Subset which serves as input to CNN.

3.4. Convolution Neural Networks CNN:

CNN is a Deep learning technique that can recognize and classify the features in images for computer vision [9]. The CNN has four different layers:

i. Convolution tool:

It extracts various features from input for analysis. In this convolution, it has many layers of kernels that helps us to extract a feature map from the input. The feature map tells us the positions of feature in the input.

ii. Pooling Layer:

This layer is sandwiched between the two convolution layers obtains the feature map after applying the pooling operation on the data received from above layer.This operation reduces the size of the input data which helps in improving the efficiency of the network.

iii. ReLU Correlation layer:

It substitutes the negative values received as input with zeros.This works as an activation function.

iv. Fully-connected layer:

This layer receives an input vector to which it applies linear combination and activation function then produces new output vector. Each element specifies the

probability of the input data that predicts the best representation of input.Depending on the size of the vector the activation function Logistic or softmax is selected.

The above drawn feature subset is passed as input to CNN for classification of normal and abnormal attacks.

4. RESULTS:

This methodology is developed on windows platform with 2.8GHz Intel core i5 processor and 4GB RAM. The WEKA tool is used for comparing the traditional techniques with the proposed methodology.The CNN is executed on colabs using Python. By using Correlation coefficient the features of the original dataset CICIDS 2017 got reduced to 16 features. The list of the features is displayed on the Table 1 given below respectively. This Optimized feature subset is supplied as input to CNN for classification of attacks. The CNN has produced 98.7 accuracy with less computational complexity ,when compared with other algorithms. The comparison results were tabulated in the following table Table.2. respectively. Time taken to build the model is 0.28 seconds.

Table.1: List of features drawn after correlation filter approach

S.No	Correlation Rank	Feature Name
1	0.627	AvgBwd Segment Size
2	0.627	Bwd Packet Length Mean
3	0.61971	Bwd Packet Length Std
4	0.61719	Bwd Packet Length Max
5	0.61554	Fwd IAT Std
6	0.61209	Packet Length Std
7	0.60872	Idle Max
8	0.60689	Idle Mean
9	0.60612	Fwd IAT Max
10	0.60537	Flow IAT Max
11	0.60152	Idle Min
12	0.5969	Max Packet Length
13	0.58052	Packet Length Mean
14	0.57516	Average Packet Size
15	0.57286	Packet Length Variance
16	0.56089	Flow IAT Std

Table.2: Classification algorithms and its parameters.

Algorithm	Accuracy	TP RATE	FP RATE
Naïve Bayesian	57.54%	57.5	1.3
Adaboost	85.89%	85.9	2.33
Decision Tree	95.91%	95.9	3.6
J48	95.65%	95.6	2.2

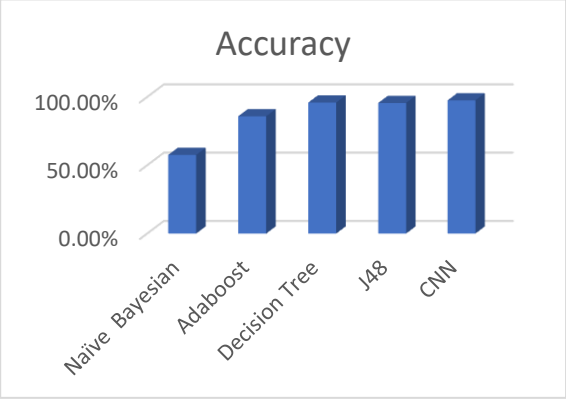

Figure.3: The effect of various classifiers on CICIDS-2017 dataset in terms of Accuracy

It is observed from above figure.3 that there is a steep increase in the accuracy of CNN with 98.7% while keeping the FP rate to minimum. It is also displayed in the table format. Hence Convolution Neural Networks are selected for attack classification.

5. CONCLUSION AND FUTURE WORK

This paper proposed a methodology for NIDS in identifying the attacks over a cloud environment. This is done by drawing the optimised feature subset using Correlation coefficient and applying CNN for attack prediction. In the evaluation process the proposed methodology is compared with the traditional machine learning techniques. The results showcased that CNN achieved high accuracy with less time in predicting the attacks. In future we want to use Multi class classification model for identifying what type of attack it is to be.

REFERENCES

[1] S.N.Dhage, B.B.Meshram, R.Rawat, S.Padawe, M.Paingaokar, A.Mishra, "Intrusion Detection System in Cloud computing environment" - ICWET'11, February 25-26,2011.

[2] Ms.ShwetaSrivastava, Ms.NikitaJoshi, Ms.Madhvi gaur, " A review paper on feature selection Methodologies and their applications ", www.ijerd.comVolume7, Issue 6(June 2013), PP. 57-61

[3] https://www.analyticsvidhya.com

[4] Noelia Sanchez-Marona, Amparo Alonso- Betanzos, Rosa M. Calvo-Estévez, "A wrapper method for feature selection in multiple classes datasets" , Bio-Inspired Systems: Computational and Ambient Intelligence. IWANN 2009.Springer,Berlin, Heidelberg

[5] Leila Mohammad pour, Teck Chaw Ling, Chee Sun Liew and Chun Yong Chong, " A Convolutional Neural Network for Network Intrusion Detection System"- Proceedings of the APAN – Research Workshop 2018, ISBN 978-4-9905448-81

[6] Dr.K.Praana, SVP.Sruthi, K.V.Kalyani, A.Sai Tejaswi, www.joics.org, vol 10 issue 3 2020 ISSN-1548-7741.

[7] M.S. Irfan Ahmed, Riyad A.M, Mohamed Jamshad K ," Information gain based feature selection for intrusion detection systems" International Journal of Scientific & Engineering Research Volume 8, Issue 7, July-2017 ISSN 2229-5518.

[8] https://towardsdatascience.com.

[9] https://missinglink.ai/.