Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 2, February 2019, pg.1 – 6

A New Distributed Denial-of-Service (DDoS) Attacks Detection System Combining Multistage Auto-Encoders with Radial Basis Function (RBF)

Sefer Kurnaz[1], Ansam Khalid[2]

[1]Computer Engineering & Altınbaş University, Turkey

[2]Information Technology & Altınbaş University, Turkey

[1] sefer.kurnaz@altinbas.edu.tr; [2] ansamkhalidkm@gmail.com

Abstract— Distributed denials of service (DDOS) attack have strong impact on the cyber world. As far as cyber-attack is concerned that it halts the normal functioning of the organization by Internet protocol (IP) spoofing, bandwidth overflow, consuming memory resources and causes a huge loss. There has been a lot of related work which focused on analysing the pattern of the DDOS attacks to protect users from them. A User datagram protocol (UDP) flood is a network flood and still one of the most crucial network floods today. This study presents new method to detect DDOS attacks by using multistage auto-encoders based on Radial Basis Function (RBF). The input data which represented the DDoS features are first analyzed by using auto-encoders and the number of auto-encoders depended on the data nonlinearity and dimension. The output of the first auto-encoder wired to the second auto-encoder etc. The aim of these auto-encoders are to extracted features that have ability to presented the best classification results and to speed up the processing time by reducing the dimension of features. In the last stage, the Radial Basis Function (RBF) trained in supervised method to classify the features into two labels there is attack or not. The obtained results compared with well-known studies presented in this field.

Keywords— DDOS, auto-encoders, radial basis function.

I. Introduction

In 1994, the numeral of internet operators was about 25 million individuals, which was about 0.4 percent of the world's populace in that time. By 2015, 3 billion and 250 million people are using internet, which is about 40 percent of the world's people [1].

In the beginning, internet was designed for functionality, not safety and it was indeed, successful in attainment its purpose. It offers operators debauched, informal and cheap communication. Furthermore, it proposals dependability and sure level of excellence of service. The internet is controlled by spread way, so no common policy can be organized to users. Procedures were open and rules were based on joint admiration. This caring of enterprise has security difficulties and it has approximately subjects which would offer junctures for distributed denial of service attack. Attacks which have a goal to hurdle the availability of computer systems or facilities are generally called DDoS attacks [2].

The earliest time DDoS attacks suited conscious was at year 2000. 15-year-old Canadian teenager ongoing sequence of DDoS attacks to the largest websites of its time; Yahoo!, Fifa.com, Amacon.com, Dell, eBay and CNN. He succeeded to close these websites depressed for a while. Yahoo! was the major search engine in that

period, and it was not nearby for a few hours. Attacks caused 1.2 billion US dollars in worldwide economic compensations [3]. DNS origin servers was below attack at year 2002 and 9 over 13 root servers were pretentious by this attack. In 2007 Estonia and in 2008 Georgia experienced DDoS attack, the basis of both attacks was demanded as Russia. In 2010, after WikiLeaks unconfined some intimate data around world politics, it was success by DDoS attack as well. Then, WikiLeaks groups attacked to MasterCard, Visa and PayPal payment systems, to complaint the obstruction of gift to WikiLeaks by these companies [4].
In 2015, DDoS attacks were amplified around 130 %, paralleling with the same dated of 2014. The lengthiest attack interval was more than 64 days. Furthermore, 20 % of the all attacks lasted more than 5 days [5].

II. MATERIAL & METHODS

A. Dataset

A new dataset was collected in this effort because there are no present data sets that include a contemporary DDoS attack like (SIDDOS, HTTP Flood), and moreover, extra available data sets may contain a countless contract of identical and jobless records, and that may outcome in a final impractical result. The data collected dataset include four kinds of DDoS attack as shadows: (HTTP Flood, SIDDOS, UDP Flood, and Smurf) without jobless and identical records.

B. Auto-encoders

The Sparse Auto-encoder (SAE) is mainly a neural network containing of a numeral of AEs where separately AE denoted a layer and trained in unsupervised manner applying data without labels. The input of every AE is output of the earlier AE. The training of AE is guessing the best variables by applying various techniques which decrease the separation between input $'x'$ and output $'\dot{x}'$. The coding between input $'x'$ and output $'\dot{x}'$. Is illustrated as below in Equations from :

$$\dot{x} = f(x) \tag{1}$$

$$n_1^{(1)} = M_f(w_{11}^{(1)} x_1 + \cdots w_{15}^{(1)} x_{5+} + b_1^{(1)}) \tag{2}$$

$$n_i^{(1)} = M_f(w_{i1}^{(1)} x_1 + \cdots w_{i5}^{(1)} x_{5+} + b_i^{(1)}) \tag{3}$$

where M () is an activation function such as a sigmoid logistic function.
The final mathematical model can be illustrated in equation (4):

$$n_{w,b}(x) = M_f(w_{11}^{(2)} n_1^{(2)} + \cdots w_{15}^{2} n_5 + \cdots + b_1^{(2)}) \tag{4}$$

The input $'x'$ and output $'\dot{x}'$ discrepancy represented by using a cost function. Several algorithms are used to find the optimum parameters of the network; the details of mathematic model are presented in [9][10].
The deep auto-encoder consists of two auto-encoders or more, where the AE attempt to extract the important features from input features X. The purpose of applying several of AEs are to decrease the amount of data step by step which decreasing the amount of data rapidly to loss effective data. The objective function of AE illustrated as Equation (5).

$$E = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{k} (x_{kn} - \hat{x}_{kn})^2 + \lambda * \frac{1}{2} \sum_{l}^{L} \sum_{j}^{n} \sum_{i}^{k} w_{ji}^{(l)^2} + \beta * \sum_{i=1}^{D(1)} \rho \log(\rho || \hat{\rho}_i) + (1-\rho) \log\left(\frac{1-\rho}{1-\hat{\rho}_i}\right) \tag{5}$$

Where, the mistake ratio illustrated with parameter *E*, the data denoted by x , the recreated data denoted by $\hat{\dot{x}}$ which is the output of the AE, λ is the constant for the and β constant too for the and
L is demonstrating hidden layers' number, n is demonstrating the explanations number, and k is demonstrating the hidden layer's amount.
Finally, preferred cost demonstrating by ρ, $\hat{\rho}_i$ denotes the ordinary activation of a neuron i . The simple auto-encoder shown in Figure 1.

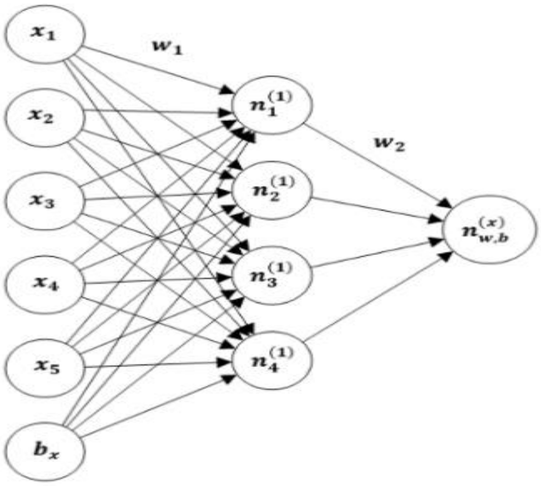

Figure 1: Simple Auto-encoder architecture [10].

C. Radial Basis Function Network (RBFN)

A RBFN is a specific kind of NN. In this paper, I'll be explaining it applies as a non-linear classifier [8].
The mathematic model RBF neurons shown in Eq (6).

$$fa(x) = \int_{-\infty}^{\infty} e^{\frac{||x-xa||}{2!}} \qquad (6)$$

where r is a receptor and σ value of shaping parameter.

h(x) is the Gaussian system with the variable r (radius) and c (center) distinct distinctly at apiece RBF component. The training procedure is depending on regulating the limits of the net to repeat a traditional of input-output designs. There are 3 kinds of variables; the weight w among the concealed bulges and the output bulges, the center c of individually neuron of the hidden layer and the component width r. The manufacture of the net covers of a collection of nodes, unique each collection that we are maddening for categorizing. Unconnectedly produce node computes a kind of entire for the connected collection. Characteristically, a group excellent is comprehensive by broadcast the input to the collection with the highest notch.

III.PROPOSED METHOD

In this paper, deep learning used to solve (DDOS attack detection) by using number of auto-encoder train each one of them alone as unsupervised learning for feature extraction after that train a RBF. multistage Auto-encoders combined with Radial Basis Function to detect DDOS attack automatically. The features from DDOS attack features extracted by using first Auto-encoder. Then, the output of first Auto-encoder wired to the second and three or N number of Auto-encoders according to the input data size and properties. Furthermore, the extracted features from Auto-encoders classified by using RBF function. The RBF presented mark able results in many complex supervised problems. The structure of proposed method presented in Figure 1.

Finally, MATLAB 2018 will be used as a tool to develop this system. The steps of the proposed method listed below:

1. In the first rectangle, Read the data [6] by using MATLAB.

2. In the second rectangle, the auto-encoder1 have five parameters identified by the user, hidden Size, Max Epoch, L2 Weight Regularization, Sparsity Regularization, Sparsity Proportion. The sparsity regularizer effect is controlled by a Sparsity Regularization parameter, dealing to force a chain on the sparsity of the output from the hidden layers. sparsity regularizer parameter is controlled by Sparsity Proportion (SP) parameter. The sparsity of the output from each hidden layer is controlled by the Proportion parameter. A low value for SP normally leads all neurons in the hidden layer specialized by

only producing a high output value for a small amount of training examples. Hidden size represented the number of features that selected in hidden layer. The hidden size different from problem to another depended on the problem type. Furthermore, Max Epoch represented the number of iterations that used to find the optimum results for the network.

3. In the third rectangle, the first auto-encoder trained in unsupervised fashion to extracted important features
4. In the fourth rectangle, the second auto-encoder parameters identified by the user, hidden Size= 27, Max Epoch, L2 Weight Regularization, Sparsity Regularization, Sparsity Proportion.
5. In the fifth rectangle, the second auto-encoder trained in unsupervised fashion to extracted important features.
6. In the sixth rectangle, the Radial Basis Function (RBF) classifier trained in supervised fashion to classify the features of the DDOS to the classes there is attack or not.
7. Train complete mean if the training data not completed return to the read data or continue with test part.
8. In the test part, after the training part complete the system tested by using another data to calculate the accuracy of the system using testing instance. The new data used to test the system that trained by using the same Behaviour data.
9. Test complete mean if the testing data not completed return to the test data or continue with calculate accuracy.
10. The accuracy will be calculated after the testing part will be complete and the confusion matrix occur in the end of the procedure.

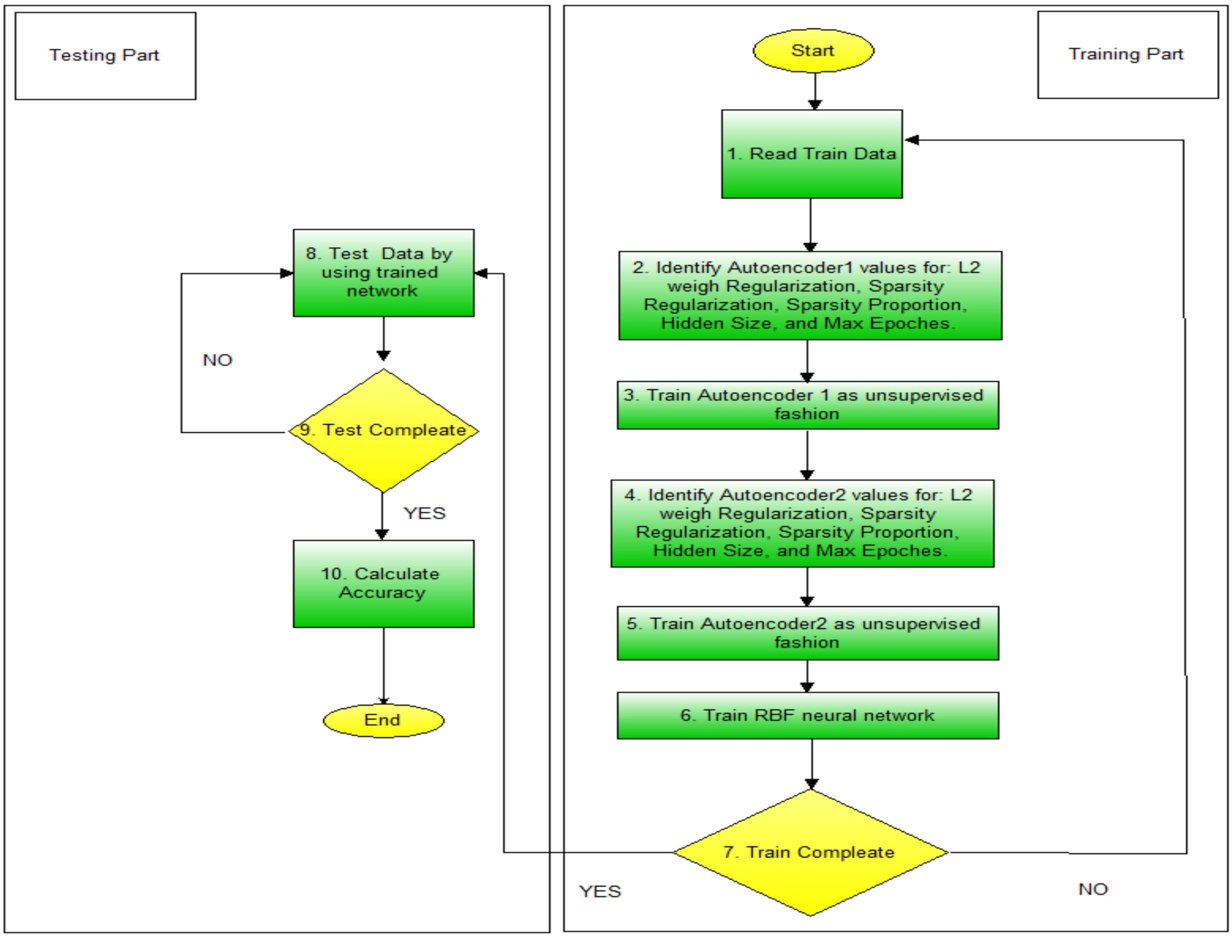
Testing Part
Training Part
Start
1. Read Train Data
2. Identify Autoencoder1 values for: L2 weigh Regularization, Sparsity Regularization, Sparsity Proportion, Hidden Size, and Max Epoches.
3. Train Autoencoder 1 as unsupervised fashion
4. Identify Autoencoder2 values for: L2 weigh Regularization, Sparsity Regularization, Sparsity Proportion, Hidden Size, and Max Epoches.
5. Train Autoencoder2 as unsupervised fashion
6. Train RBF neural network
7. Train Compleate
YES
NO
8. Test Data by using trained network
NO
9. Test Compleate
YES
10. Calculate Accuracy
End

Fig. 2: Proposed Method

IV. PROPOSED METHOD

Numerous experiments is performed in several settings, the test and the outcome were measured applying several measurements, the performance of several experiments were related and the outcomes were highlighted.
In the first stage, the 27 features which represented the DDOS features are become input to the auto-encoder1 which try to extract only 25 features from them. Then, the 25 features wired to the auto-encoder2 when extracted only 23 features which represented high level features.
Furthermore, radial basis function used to classify the features that extracted from last auto-encoder by training in supervised fashion. Several parameters are calculated to evaluate the proposed method see Table 1.

As shown in Table 2 the proposed method presented high results when 10 statistical parameters are calculated. Then, the obtained results compared with well-known studies presented in this field. According to the comparison in the Table 3 our method presented remarkable results compared to previous studies.
As shown in the Table 3 the proposed method presented best results than methods proposed in [6,7,8] which these studies represented the commonly known researches in this field.

Table 1: Proposed Method Results

Parameters	**Results**
Sensitivity	0.9900
Specificity	0.9860
Precision	0.9861
Negative Predictive Value	0.9900
False Positive Rate	0.0140
False Discovery Rate	0.0139
False Negative Rate	0.0100
Accuracy	**0.9880**
F1 Score	0.9880
Matthews Correlation Coefficient	0.9760

Table 2: Results Comparison

Methods	**Results**
MLP [6]	98.63
Random Forest [6]	98.02
Naïve Bayes [6]	96.91
SVM [7]	97.29
SSAE-SVM [8]	97.65
Proposed Method	**98.80**

V. Conclusions

This paper presented a new multistage auto-encoder based on radial basis function classifier for DDOS detection. The data are analyzed by using several auto-encoders to extract high level features from DDOS dataset and then the extracted features are classified using radial basis function. The new method presented 98.80 accuracy which is remarkable results when compared with previous studies.

The presented method is new idea which combine the power of deep learning with radial basis function. The aim of this study extracts high level features by using number of auto-encoders to obtain high accuracy results. The extracted features combined with radial basis function which trained in supervised learning to classify the extracted features.

Furthermore, in this method there is not to stack auto-encoders with the output layer classifier and trained them again in supervised fashion. Because the proposed method presented remarkable results without doing this level.

Then, auto-encoders can have combined with various classifiers to solve different types of problems because auto-encoders presented remarkable results when compared with convolution techniques.

The proposed method can easily apply to various classification problems such as EEG signal classification, EGC signal classification, face recognition, fingerprint recognition and disease detection by modifying only number of parameters like input features, hidden nodes and output classes.

References

[1] “Internet Users”, internetlivestats.com/internet-users, 2015.

[2] Mirkovic, J., J. Martin and P. Reiher., “A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms”, ACM SIGCOMM Computer Communication Review, 2004.

[3] “Mafia Boy”, wikipedia, n.d., https://en.wikipedia.org/wiki/MafiaBoy.

[4] Liu., X., “Mitigating Denial-of-Service Flooding Attacks with Source Authentication”, PhD Thesis in Department of Computer Science in the Graduate School of Duke University, 2012.

[5] Meek, A., “DDoS attacks are getting much more powerful and the Pentagon is scrambling for solutions”, BGR, 2015.

[6] M. Alkasassbeh, G. Al-Naymat, A. B.A, and M. Almseidin, “Detecting Distributed Denial of Service Attacks Using Data Mining Techniques,” International Journal of Advanced Computer Science and Applications, vol. 7, no. 1, 2016.

[7] Ahmad M. Karim, Mehmet S. Güzel, Mehmet R. Tolun, Hilal Kaya, and Fatih V. Çelebi, “A New Generalized Deep Learning Framework Combining Sparse Autoencoder and Taguchi Method for Novel Data Classification and Processing,” Mathematical Problems in Engineering, vol. 2018, Article ID 3145947, 13 pages, 2018. https://doi.org/10.1155/2018/3145947.

[8] Y. Ju, J. Guo, and S. Liu, “A Deep Learning Method Combined Sparse Autoencoder with SVM,” in Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 257–260, Xi’an, China, September 2015.

[9] Karim, A. M., Güzel, M. S., Tolun, M. R., Kaya, H., & Çelebi, F. V. (2019). A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing. Biocybernetics and Biomedical Engineering, 39(1), 148-159. doi:10.1016/j.bbe.2018.11.004

[10] A. M. Karim, Ö. Karal, and F. V Çelebi, “A New Automatic Epilepsy Serious Detection Method by Using Deep Learning Based on Discrete Wavelet Transform,” no. 4, pp. 15–18, 2018.