Indonesian Journal of Electrical Engineering and Computer Science
Vol. 20, No. 3, December 2020, pp. 1413~1418
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v20.i3.pp1413-1418

Performance analysis of flow-based attacks detection on CSE-CIC-IDS2018 dataset using deep learning

Rawaa Ismael Farhan[1], Abeer Tariq Maolood[2], NidaaFlaih Hassan[3]
[1]Department of Computer Science, University of Technology, Wasit University, Iraq
[2,3]Department of Computer Science, University of Technology, Iraq

Article Info

Article history:

Received Mar 9, 2020
Revised May 7, 2020
Accepted Jun 20, 2020

Keywords:

Anomaly-based intrusion
Detection system
CSE-CIC-IDS2018
Deep learning
Flow-based intrusion detection
Internet of thing (IOT)

ABSTRACT

The emergence of the internet of things (IOT) as a result of the development of the communications system has made the study of cyber security more important. Day after day, attacks evolve and new attacks are emerged. Hence, network anomaly-based intrusion detection system is become very important, which plays an important role in protecting the network through early detection of attacks. Because of the development in machine learning and the emergence of deep learning field, and its ability to extract high-level features with high accuracy, these systems have been included to work with real network traffic CSE-CIC-IDS2018 for a wide range of intrusions and normal behavior as an ideal method of testing and evaluation. In this paper, we test and evaluate our deep model (DNN) which has achieved a good detection accuracy of about 90%.

Corresponding Author:

Rawaa Ismael Farhan,
Department of Computer Science,
Technology University, Industry Street, Iraq, Baghdad.
Email: ralrikabi@uowasit.edu.iq

1. INTRODUCTION

Due to the wide growth of the Internet of Things applications at the present time and its dynamic heterogeneous nature, difficulty in measuring it, there are serious challenges facing it, such as IP addressing, privcy, data analyzing and management and security [1]. The network intrusion detection system (NIDS) has a fundamental role in solving security challenges. NIDS monitors the traffic of network to detect any suspicious activity, its analyze information from network traffic to detect breaches of security, that comprise intrusions, misuse and anomaly [2]. The basic target of this system is to propose an intelligent, secure and reliable Internet-based infrastructure that can detect its weaknesses and a secured firewall facing cyberattacks and automatically restore itself [3].

There are two types of attack detection approachs in NIDS. In signature-based detection the NIDS detects intrusions based on matching signatures for known attack patterns while an anomaly-based detection learns the normal activity of the network that monitors its and makes alert when the monitored activity deviates from the normal profile [4]. As [5] A NIDS monitors the network flow of both incoming and outgoing traffic. So, NIDS differs from a firewall where arise alerts if an attack is detected while firewall represent safeguard only allow to traffic from trusted network predefined in ruleset.

Most of research on NIDS have been based on the simulated datasets which mean there is no real datasets, only simulated dataset such as KDDcup99 or NSL-KDD, but simulated database couldn't reflect the scenario of real network intrusion. It is important to use real flow-based datasets to ensure precise evaluation of methods [6]. Thus, in this paper the realistic CSE_CIC_IDS2018 dataset on AWS platform have been used as a shift from static data sets to dynamically generated data sets that not only reflect network traffic and log file at that time but can be modified, replicated, and expanded.

According to [7], it is so important to use effective deep learning algorithms for hierarchical features extracting that represent data better than using manual extracted features in machine learning methods, where deep learning techniques depicted great excellence in reduced test time and high accuracy especially in the big data field. This paper [8] proposed that deep learning techniques are RNN and CNN detect attacks without human-defined rules or signatures then Compared Snort IDS and the deep learning model in detecting time.

Mingyi Zhu [9] proposed a new deep learning model for abnormal traffic detection called AMF-LSTM where features have been used from multi-flows as input to obtain temporal correlation between flows instead of a single flow or the extracted features from log then added an attention mechanism to the LSTM which detect traffic flow with more contributions in final results. This paper [10] implemented preprocess on packets by using (E-GRU) encoders with variant-gated recurrent units for payload-aware intrusion detection where GRU automatically learn payload and header features OF network packet to improve the detection rate of the IDS.

2. RELATED WORKS

In the following section, an importanted works has been studied to investigate (NIDS), the following are the most important works: In [11] a proposed system was detecte a botnet attack classification that represent a famous attacks in banking services and financial transactions. The proposed system applied neural network on a realistic dataset of cyber defence (CSE-CIC-IDS2018). Qianru Zhou in [12] implemented six machine learning approachs on flow-based data to classify and detect Zero-Day attacks wheretraining dataset CIC-AWS-2018 contains fourteen types of intrusions, but testing datasets include different eight attacks types. In [13] the proposed system implemented a dynamic system for network anomaly detection using one of deep learning approachs which was Long Short Term Memory (LSTM). Also its added an Attention Mechanism (AM) to improve the model performance while the SMOTE algorithm improved loss function to solve the problem of class-imbalance in the CSE-CIC-IDS2018 dataset. In paper [14] the deep learning algorithms capabilities in the network intrusion detection is explored. It compared varity of deep learning frameworks (e.g., TensorFlow,Keras, PyTorch, Theanoand fast.ai) for network traffic intrusion detecting and network attack types classifying .In [15]CIC- 2018 dataset and proposed convolutional neural network (CNN) model is implemented, which converted dataset to images, each one in size 13× 6 due to 78 features for each data without the feature Label which used for image classification.

3. METHODOLOGY

In this section, it is explained the results of research and at the same time is given the this part to design NIDS by using Deep Learning.At first we preprocess the CSE-CIC-IDS2018 dataset by deleting all unnecessary features like timestamp,numbered all classes from 0-7 by mapping each class of the 8 classes into numeric values. Then, normalize all data into [-1,1]. finally that, we implement an intrusion detection system by using deep learning.

3.1. CSE-CIC-IDS2018 Real Dataset

In this section an implementing the NIDS on a real network traffic like CSE-CIC-IDS2018 dataset is illustrated, which contains a detailes of intrusions with detailes of protocols. The applications and lowest level entities of network are represent best approach of testing and evaluation, also it refers to shifting from Static data to Dynamic data which is real-time traffic on the Amazon platform (AWS). To download this dataset the following description is applied :

" Resource type
S3 Bucket
Amazon Resource Name (ARN)
arn:aws:s3:::cse-cic-ids2018
AWS Region
ca-central-1" [16].

Seven types of attacks are available in this dataset: DOS, DDoS, Brute-force, Heartbleed, Botnet, infiltration and Web attacks. Attacking infrastructure comprised 50 terminals and the infected organizations comprised 30 servers and 420 computers. This dataset represented the captured traffic of AWS network and machines log files with 80 extracted features by using CICFlowMeter-V3 [17]. Table 1 shows A sub set of extracted traffic features.

Table 1. A sub set of extracted traffic features [17]

fl_dur	Flow duration
tot_fw_pk	Total packets in the forward direction
tot_bw_pk	Total packets in the backward direction
bw_win_byt	Number of bytes sent in initial window in the backward direction
Fw_act_pkt	Number of packets with at least 1 byte of TCP data payload in the forward direction
fw_hdr_len	Total bytes used for headers in the forward direction
bw_hdr_len	Total bytes used for headers in the forward direction
down_up_ratio	Download and upload ratio
fw_win_byt	Number of bytes sent in initial window in the forward direction

3.2. Deep Neural Network (DNN)

Deep learning (DL) has arised as a powerful technique to make precise predictions from complex and bigdata such as image, text, or video [18, 19]. The big progress and capabilities of Deep Learning, which contain multi processing layers allows to learn hidden representations of data. Pathways to deep learning technology [20] have been conducted in cybersecurity. Particularly in classification and malware detection, they save significant time cost and enhance accuracy for total malware detection system pipeline.

In [21] a survey of recent DL techniques that have been presented in cybersecurity scope especially intrusion detection. It described types of DL algorithms and DL framework for cybersecurity. DL outperformed the traditional machine learning (ML) methods because its ability of automatic feature extraction in hirarchical manner instead of feature engineering in ML. Supervised Deep Learning Methods RNN and CNN have been applied [22] to categorize five common types of attacks based solely on information of packet header and do not require any payload using Keras at the top of TensorFlow. [23] a hybrid deep learning method has been suggested to detect suspicious behavior of image-based systems for effective malware classification.

3.3. Experemental Environment

Our DNN model implemented on Windows10 using Visual Studio 2019 contain python 3.7 and installed Keras [24] on top of Tensorflow using (Numpy, Scikit-learn, Panda) libraries, 8GB Memory, CPU core i7,512GB Hard disk .

4. RESULTS AND DISCUSSION

In this part we discuss our DNN model results. Then evaluate this results with main performance metrics. After that we achieved comparative study between DNN model with other models

4.1. Results

In this model, three fully connected layers are used, they described as following:
dense1 layer with 78 neurons use ReLu Activation function.
dens2 layer with 64 neurons use ReLu Activation function.
dense3 layer with 8 neurons use Softmax Activation function.
Regularization method with two dropouts ratio (0.2) are used to avoid overfitting.

Table 2 shows that good tunning of hyperparameter values is important to avoid overfitting.

Table 2. Experimental hyperparameter of proposed DNN model

Parameters	Value
Epoch	100
Batch size	500
Activation function	ReLu,Soft max
Loss function	categorical_crossentropy
Optimizer	Adam

In this paper used two types of non-linear Activation function are ReLu and softmax. [25] ReLu is faster than other non-linear Activation function which maximize the deep learning efficient while Softmax used for multi classification as output layer in DNN model because it outputs the probability of each class then choose biggest value for accurate result.
Loss function represent the difference between the predicted and actual output.
The Optimizer Adam used to minimize Loss function by calculate gradients of a loss after that apply gradients to update values and therefore enhance the DNN results.

4.2. Evaluation Metrics

In this paper, five performance metrics are used for model evaluation.

1) Confusion Matrix

It is a graphical tool to depict the true positive and false positive values of each attack type, its used in this model to make prediction about number of normal and attacked packets in network traffic as shown in Table 3.

Table 3. Confusion matrix of DNN

	Normal	DDOS	BOT	Inf	Brute-force	Dos	Web attack
Normal	161507	0	3	13	0	0	0
DDOS	2	57	0	0	0	0	0
BOT	0	0	37	0	0	0	0
Inf	40	0	0	80	0	0	0
Brute-force	29	0	0	0	91	0	0
Dos	78	0	0	0	0	102	0
Web attack	239	0	0	0	0	0	236

2) Accuracy: A percentage of positive detection of all data cases.
3) Precision: How many attacks are properly returned.
4) Recall: How many attacks the system returns.
5) F1-score: Rate of Precision and Recall in our model.

These metrics described in Table 4 from which we determine the detection rate (DR) and false alarm report (FAR).

Table 4. DNN performance

	Precision	Recall	F1-score
0	0.89	1.00	0.94
1	1.00	0.97	0.98
2	0.99	1.00	0.99
3	0.82	0.67	o.73
4	0.00	0.00	0.00
5	0.53	0.57	0.55
6	0.00	0.00	0.00
7	1.00	0.50	0.66

4.3. Comparitive Analysis

A comparative study in Table 5 is deducted for showing the differences between our implemented Deep Neural Network (DNN) and other methods. As shown in table because all 78 features of CSE-CIC-IDS2018 datasetare used.

Table 5. Comparison between proposed DNN and other methods

	MLP [11]	DT [12]	LSTM [13]	The proposed DNN
Precision	1.0	1.0	0.96	0.65
Recall	1.0	0.40	0.96	0.59
Accuracy	99.97%	100%	96.2%	90.25%

In paper [11] the proposed system designed for a classification to detect botnet attack only in CSE-CIC-IDS2018 that represent a serious threat in banking services. This work used six types of machine learning algorithm on CSE-CIC-IDS2018 with fourteen types of attacks for training and different eight types from zero day attacks for testing [12]. Handeled imbalanced classification problem in CSE-CIC-IDS2018 Dataset by using SMOTE algorithim with LSTM [13]. The accuracy of proposed model is 90.25% which consider good, but we aspire to increase it using one of the available feature selection methods that reduce the dataset dimensionaliy and select the most relevant features only to increase system accuracy, decrease false alarm report (FAR) and reduce the computation time.

5. CONCLUSION

In this paper fully connected dense deep neural network (DNN) is proposed, its used for Flow-based intrusion detection on real-world dataset CSE-CIC-IDS2018 available at AWS platform, few papers addressed this dataset since its emergance yet.In this paper accuracy of detection is 0.90%, its consider good, but there are still challenges to be overcomed, these challenges are focused on: Large data size, Higher dimensionality and Data preprocessing. As future works , some feature selection method are recommended to increase accuracy, detection rate, decrease False Alarm Report (FAR) and minimize computing time. Also, hyperparameter tunning is recommende to be used for further efficiency.

ACKNOWLEDGEMENTS

Special thanks to my teachers in the Techology University. Sincere gratitude to Wasit University.

REFERENCES

[1] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," *Future Generation Computer Systems*, 2017.
[2] Rao U. H. and Nayak U., "Intrusion Detection and Prevention Systems," in The InfoSec Handbook, Apress, Berkeley, CA, 2014.
[3] M. Hasan, et al., "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," Elsevier, Internet of Things, vol. 7, 2019.
[4] S. M. Mehibs and S. H. Hashim, "Proposed Network Intrusion Detection System In Cloud Environment Based on Back Propagation Neural Network," *Journal of Babylon University/Pure and Applied Sciences*, vol. 26, no. 1, 2018.
[5] B. Ram B, et al., "Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks," *Journal of Internet Services and Information Security (JISIS)*, vol. 9, no. 4, pp. 1-17, 2019.
[6] M. Parashar, et al., "Packet and Flow Based Network Intrusion Dataset," Springer, CCIS 306, pp. 322-334, 2012.
[7] G. Karatas, et al., "Deep Learning in Intrusion Detection Systems," *IEEE International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism Ankara*, Turkey, 2018.
[8] N. Chockwanich and V. Visoottiviseth, "Intrusion Detection by Deep Learning with TensorFlow," *International Conference on Advanced Communications Technology (ICACT)*, 2019.
[9] M. Zhu, et al., "A Deep Learning Approach for Network Anomaly Detection Based on AMF-LSTM," Springer, NPC 2018, LNCS 11276, pp. 137-141, 2018.
[10] Y. Hao, et al., "Variant-Gated Recurrent Units With Encoders to Preprocess Packets for Payload-Aware Intrusion Detection," *IEEE*, vol. 7, 2019.
[11] V. Kanimozhi and T. P. Jacob, "Artificial Intelligence based Network Intrusion Detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing," *ICT Express*, vol. 5, pp. 211-214, 2019.
[12] Q. Zhou and D. Pezaros, "Evaluation of Machine Learning Classifiers for Zero-Day Intrusion Detection - An Analysis on CIC-AWS-2018 dataset," ArXiv, abs/1905.03685, 2019.
[13] Lin P., et al., "Dynamic Network Anomaly Detection System by Using Deep Learning Techniques," Springer, Cham, in D. Da Silva, et al. (eds), *CLOUD 2019*, vol. 11513, pp. 161-167, 2019.
[14] R. B. Basneand, et al., "Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks," *Journal of Internet Services and Information Security (JISIS)*, vol. 9, no. 4, pp. 1-17, 2019.
[15] J. Kim, et al., "An Intrusion Detection based on a Convolutional Neural Network," *Journal of Multimedia Information System*, vol. 6, no. 4, pp. 165-172, 2019.
[16] https://registry.opendata.aws/cse-cic-ids2018/
[17] https://www.unb.ca/cic/datasets/ids-2018.html
[18] Tang, et al., "Deep Learning Approach for Network Intrusion Detection in Software Defined Networking," *International Conference on Wireless Networks and Mobile Communications (WINCOM'16)*, Morocco, 2016.
[19] M. K. Ibraheem, et al., "Network Intrusion Detection Using Deep Learning Based On Dimensionality Reduction," 2019. Available: www.ausrevista.com/.
[20] S. Lu, et al., "New Era of Deeplearning-Based Malware Intrusion Detection: The Malware Detection and Prediction Based on Deep Learning," ArXiv, abs/1907.08356v1, 2019.
[21] S. Mahdavifar and A. A. Ghorbani, "Application of deep learning to cybersecurity: A survey," *Neurocomputing*, vol. 347, pp. 149-176, 2019.
[22] N. Chockwanich, et al., "Intrusion Detection by Deep Learning with Tensor Flow," *International Conference on Advanced Communications Technology(ICACT)*, 2019.
[23] S. Venkatraman, et al., "A hybrid deep learning image-based analysis for effective malware detection," *Journal of Information Security and Applications*, vol. 47, pp. 377-389, 2019.
[24] Keras.io., "Keras Documentation," 2019. Available at: https://keras.io/.
[25] Vinayakumar R., et al., "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE*, 2019.

BIOGRAPHIES OF AUTHORS

Rawaa Ismael Farhan is a Ph.D. student in the Computer Science Department, University of Technology, Baghdad, Iraq since Septemper 2017. She was received Msc. in Computer Science from Osmania University, Heyderabad, India in 2014. She is an Lecturer in the College of Computer Science, Wasit University, Kut . She has around 16 years of teaching experience.

Assist. Prof. Dr. Abeer Tariq Maolood received theMSc. and PhD. in Computer Science from University of Technology, Iraq, 2005 and 2010 respectively. She has around 15 years of teaching experience. Her areas of interest's computer and network security, neural networks and web applications security

Assist. Prof .Dr. Nidaa F. Hassan received theMSc. and PhD. in Computer Science from University of Technology, Iraq, 1996 and 2005 respectively. She has around 21 years of teaching experience. Her areas of interest's computer security and image processing.