The Korean Institute of
Communications and
Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 7 (2021) 210–214

www.elsevier.com/locate/icte

A deep learning based HTTP slow DoS classification approach using flow data

Muraleedharan N.[*], Janet B.

Centre for Development of Advanced Computing (C-DAC), Bangalore, India
Computer Applications Department, NIT Tiruchirappalli, India

Received 28 February 2020; received in revised form 25 July 2020; accepted 23 August 2020
Available online 26 September 2020

Abstract

The popularity of the Internet introduces many network-enabled services that can be accessed by the user. But the adversaries are trying to deny these critical services to the user through Denial of Service (DoS) attacks. Presently, dealing with DoS attack which targets the application layer using slow traffic rate is one of the key challenges faced by the service providers. In this paper, a deep classification model using flow data is proposed to detect slow DoS attack on HTTP. The classifier is evaluated using CICIDS2017 dataset. The results obtained show that the classifier can obtain 99.61% accuracy.

Keywords: Slow DoS; Deep learning; Flow data; Denial of Service

1. Introduction

Nowadays, the Internet has evolved as a powerful platform for communication and collaboration. Hence, many of the services such as e-commerce, cloud computing, finance, health, transport, and citizen services are Internet enabled. These services can be accessed by a user without any geographical boundaries. As these critical services are accessed from the server through the network, the availability of these services to a genuine user needs to be ensured. But the adversaries disrupt or deny the services to a genuine user using Denial of Service (DoS) attacks [1]. The DoS attack is an attack on the availability where the attacker sends unwanted requests to the server to overwhelm the resource. A complex version of DoS known as Distributed Denial of Service (DDoS) attack sends requests from multiple attackers to the same server to deny the services to genuine users.

Traditionally, the DoS/DDoS attack is derived to consume the network bandwidth between the targeted service and clients. To achieve this attack, the adversaries inject a huge volume of traffic using compromised hosts or bot-nets to the targeted machine. Nowadays, the DoS/ DDoS attacks are targeting different environments like cloud infrastructure, [2–4] mobile and wireless networks [5–7].

Another category of DoS attack, known as slow DoS, targets the application and server resources by injecting low volume legitimate traffic at a very slow rate. Since the traffic volume of slow DoS is very low, this attack can be conducted using a lesser number of machines. Moreover, as the slow DoS traffic appears to be legitimate, the traditional mitigation devices may fail to detect these attacks. Fig. 1 illustrates the difference between normal traffic, volumetric DoS/DDoS and slow DoS attacks. As shown in the figure, by considering the traffic volume and transmission speed, the normal traffic and slow DoS traffic regions overlap. This makes it difficult to distinguish slow DoS attacks from normal traffic and further to prevent it.

The major challenges in slow DoS classification from normal traffic are

- It uses legitimate connection during the attack.
- Lesser number of connections are required to launch the attack.
- Bandwidth usage and traffic volume of the slow DoS attack are low. Hence, traditional defencive systems are not able to detect it.

Various approaches have been proposed to detect the slow DoS attacks [8,9]. Recently, Software Defined Network (SDN)

[*] Corresponding author at: Centre for Development of Advanced Computing (C-DAC), Bangalore, India.
E-mail addresses: murali@cdac.in (Muraleedharan N.), janet@nitt.edu (Janet B.).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2020.08.005

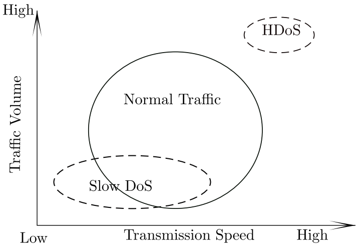

Fig. 1. Illustration of Normal traffic, slow DoS and Volumetric DoS.

[10,11] and machine learning based approaches [12] were used to detect the slow DoS attacks. In this paper a deep learning based slow DoS classifier using flow data is proposed. The novelty of this work is the usage of deep neural network classification technique on the flow data for slow HTTP DoS detection.

The proposed approach has following advantages compared to the host based slow DoS attack detection.

- As the flow data can be collected and analysed from the network gateway, a preventive system for slow DoS attack can be implemented before the attack traffic reaches the victim machine.
- As the flow data is host and operating system independent, the slow DoS classifier can be used in any web server without any configuration changes at server level.

Nowadays, web-based services are one of the common methods used by the service providers to provide services including citizen services, cloud-based services, banking, and financial services. Hence, the slow DoS attack on the web servers in these environments may have a catastrophic effect. As the proposed system addresses the slow DoS on web servers, it can be used for the detection and prevention of slow HTTP DoS in these environments. The remaining sections of the paper are organized as follows. Section 2 describes the low and slow DoS attack targeted on HTTP. The model used for classification is explained in Section 3. Results obtained and analysis are shown in Section 4.

2. Low and slow DoS on HTTP

The slow DoS targets the application layer by sending legitimate traffic at a very low rate. The common property of slow DoS is that the servers would appear to have a large number of connected clients but the actual processing load would be very low. Since HTTP is a prominent application layer protocol used in the Internet, it has become one of the common targets for slow DoS attack.

In their evaluation, Tripathi et al. [13] observed that most of the modern web servers are vulnerable to slow DoS attack. Another study on the impact of application layer DoS on popular web server is presented in [14]. Recent studies show that HTTP/2, the updated version of HTTP protocol, is also vulnerable to many slow DoS attacks [15]. Hence, the detection and prevention of slow DoS attack has paramount importance in the present day Internet.

The different types of slow DoS attacks targeting HTTP are explained below.

2.1. Slowloris

Web servers that are vulnerable to the slowloris attack start to process the request only after they receive the entire request from the client. By knowing this, the attacker sends partial HTTP requests to open connections to the vulnerable web server. Once the connection is opened, the attacker then tries to keep those connections live as long as possible by sending the next portion of the request just before the connection timeout, thus overwhelming and slowing down the victim server [16].

2.2. Slow POST

In slow POST attack, the attacker uses a legitimate HTTP POST method by setting a very high number for the ‘content-length’ value in the request. Upon receiving this request, the server allocates the necessary resources to process the specified content length data. Later, the client sends the data at an extremely slow rate which results in a prolonged open connection in the server [8].

2.3. Slow read

In slow read, the client sends legitimate HTTP requests to the server and reads the response at a very slow rate. The attacker prevents the server from resetting the connection by setting the zero window size in the packet. Upon receiving the packet with zero window size, the server thinks that the client is actually reading the data and therefore keeps the connection open [9].

3. Slow DoS classifier model using deep learning

The workflow of the proposed slow DoS classification model is depicted in Fig. 2. The model takes network traffic as the input and provides the classified results based on the deep learning process. The network traffic is aggregated into flow data based on the flow probe configuration in the network devices. Further, the preprocessing of the data is done to clean the data and fed into the deep learning model. One of the advantages of deep learning over machine learning is that it can learn the important features automatically without manual inputs. As the number of features and the volume of data is more, we have used a deep neural network model for our classification. Details of the model and its components are described below.

3.1. Network flow data

Network flow can be defined as a unidirectional sequence of packets of a given protocol travelling between a source and destination IP address and ports within a time period. The flows are generated from packets by grouping them using key fields. The fields used to derive the flow data from network packets are shown below.

$$Key = \{SrcIP, DstIP, SrcPort, DstPort, Proto\} \qquad (1)$$

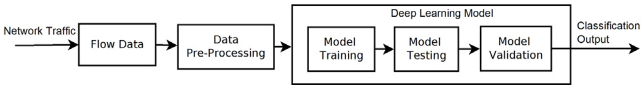

Fig. 2. Workflow of the slow DoS classification model.

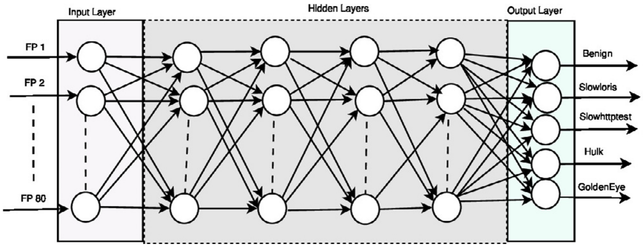

Fig. 3. Deep learning model used for slow DoS classification.

In addition to these fields, the flow data consists of information derived from the connection such as duration, bytes, packets transferred, and TCP flag values. The flow data can be further extended by including the granular level information derived from intra and inter packets details [17].

There are several reasons for selecting flow level data as the input for our classification. Firstly, the volume of data is less compared to the packet level information. Secondly, as the data volume is less, long term data can be collected and analysed. This will help to detect slow attacks which last more time. Thirdly, as it is derived from the packet header, this dataset can be generated from the traffic even when it is encrypted. Details of the dataset used and pre-processing carried out are explained below.

3.1.1. *Dataset used*

The Intrusion Detection Evaluation Dataset available in the Canadian Institute of Cybersecurity (CICIDS2017) is used for the model training and evaluation [18]. It contains multiple attack data including brute force, DoS, web attacks, infiltration, botnet and DDoS. We have selected the DoS dataset for our classification model.

3.1.2. *Data pre-processing*

The CICIDS2017 dataset consists of labelled bidirectional flows in comma-separated (.CSV) format, with 84 parameters in each flow record. The list of parameters and descriptions of each parameter is available in [18]. As it may bias the training, the fields flowID, timestamp, source and destination IP addresses are removed from the flow records used for our classification. Hence, the final dataset selected for classification consists of 80 parameters. Other than the benign traffic, as per the tools used, the flow records are labelled as ‘Slowloris’, ‘Slowhttptest’, ‘Hulk’, and ‘GoldenEye’. These labels are converted into integer values starting from one and ending by five which represent ‘Benign’, ‘Slowloris’, ‘SlowHTTP’, ‘Hulk’, ‘GoldenEye’ flows respectively.

Table 1
Summary of the data used for training, testing and validation.

Total records	Training	Testing	Validation
32,190	19,314	6,438	6438

3.2. *Deep learning based classifier*

As shown in Fig. 3, the deep learning model consists of 3 different layers i.e. ‘input layer’, ‘hidden layers’, and ‘output layer’. The input layer gives information to the network. The hidden layer handles the non-linearly separable relations and passes information from the input layer to the output layer. The output layer performs classification of the traffic into benign or slow DoS.

A fully connected feed forward deep network is used in our classification model. In this model, the input layer takes flow level parameters as inputs. As we have selected 80 features in the flow data, the number of neurons in the input layer is fixed as 80. The flow level features used in the input layer are represented from ‘FP1’ to ‘FP80’. The output layer contains the same number of neurons in the number of classes in the dataset. Hence, the neurons in output layer are fixed as five.

3.2.1. *Model training*

To classify the slow DoS flow records from the benign flow, we trained the system using training data. As the training data consists of the label in it, a supervised learning technique is applied. The data used for training, testing, and validation of the classifier is summarized in Table 1. As it is computationally inexpensive, Rectified Linear Unit (ReLU) is used as the activation function in the hidden layer. Our model is used to classify four different slow DoS from the benign traffic. Hence, ‘softmax’ activation function is used in the output layer. The ‘adam’ (adaptive moment estimation) optimization algorithm was used to optimize the cost function. This optimization technique is used in our model to support large data with more number of parameters. As our model is a multi-class classification, the ‘categorical cross entropy’ function was used as the loss function. The model was implemented using ‘Keras’ [19] API and ‘SciKit’ [20] model selection library. We have configured the ‘early stopping’ option in the epoch with five as the patience value. Based on experiments, we have finalized four hidden layers in the model.

4. Results and analysis

The results obtained from the slow DoS classification shows that the model is able to achieve 99.61% overall accuracy. The confusion matrix of the model which obtained this accuracy is shown in Fig. 4. The ‘x’ axis of the confusion matrix shows the predicted label and the ‘y’ axis shows the true label. By comparing the number of records in each class, we can observe that the records in each class are not distributed uniformly. Hence, we calculated the F1 score as the performance measure in addition to precision and recall. The summary of the obtained classification precision, recall, and F1 score is shown in Table 2.

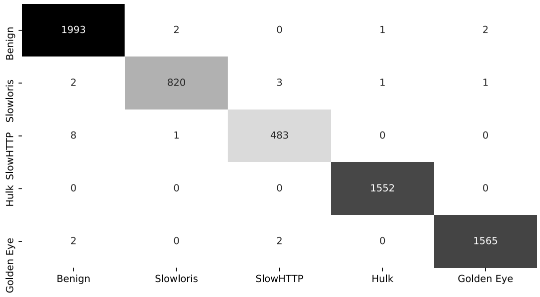

Fig. 4. Confusion matrix of the classifier output.

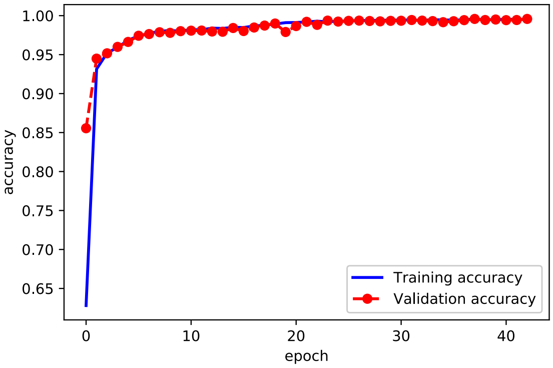

Fig. 5. Accuracy on the training and validation datasets.

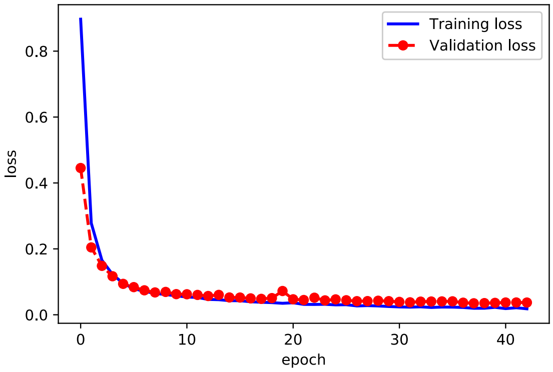

Fig. 6. Loss on the training and validation datasets.

The plots of accuracy and loss on the training and validation datasets over training epochs are shown in Fig. 5 and Fig. 6 respectively. The 'x' axis of the plot indicates the epoch values and the 'y' axis shows the accuracy or loss value. From these plots, we can observe that the model has used 43 epochs and achieved validation accuracy similar to the training accuracy. The system is able to achieve 100% precision, recall, and F1 score for 'Hulk' and 'GoldenEye' DoS traffic. However, compared to other traffic types, the classification performance of slowhttptest data is less. It may be due to the lesser number of slowhttptest records (only 7%) available in the dataset.

By analysing the false positive generated by the model, we can observe that 'benign' class generated 12 false positives and 'hulk' generated two false positives during the classification. However, by considering the number of records, the 'slowhttptest' class generated more false positive entries. The analysis of false negative reveals that 'slowhttptest' and 'slowloris' have nine and seven false negatives respectively. From the accuracy, precision, recall and F1 score obtained, we can conclude that the flow based deep classifier model is effective for classifying the slow DoS traffic.

Table 2
Summary of classification result.

Traffic type	Precision	Recall	F1 score
Benign	0.99	1.00	1.00
Slowloris	1.00	0.99	0.99
Slowhttptest	0.99	0.98	0.99
Hulk	1.00	1.00	1.00
GoldenEye	1.00	1.00	1.00

4.1. Conclusion

A deep learning based classifier for slow HTTP DoS detection using flow data is presented. The classifier is evaluated using CICIDS2017 dataset. The results obtained show that the classifier can classify the attack with accuracy of 99.61%. Although the classifier achieved higher accuracy, it is important to evaluate and benchmark the classifier with real traffic. As a future activity, we are planning to extend this work to provide a comprehensive model for detecting and preventing slow DoS attack on HTTP.

CRediT authorship contribution statement

Muraleedharan N.: Conceptualization, Methodology, Software, Data curation, Visualization, Writing - original draft preparation. **Janet B.:** Supervision, Validation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.T. Zargar, J. Joshi, D. Tipper, A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks, IEEE Commun. Surv. Tutor. 15 (4) (2013) 2046–2069.

[2] A. Almomani, M. Alauthman, F. Albalas, O. Dorgham, A. Obeidat, An online intrusion detection system to cloud computing based on neucube algorithms, Int. J. Cloud Appl. Comput. 8 (2) (2018) 96–112.

[3] A. Bhardwaj, S. Goundar, Comparing single tier and three tier infrastructure designs against DDoS attacks, Int. J. Cloud Appl. Comput. 7 (3) (2017) 59–75.

[4] K. Bhushan, B.B. Gupta, Distributed denial of service (ddos) attack mitigation in software defined network (SDN)-based cloud computing environment, J. Ambient Intell. Humaniz. Comput. 10 (5) (2019) 1985–1997.

[5] B.B. Gupta, D.P. Agrawal, H. Wang, Computer and Cyber Security Principles, Algorithm, Applications, and Perspectives, CRC Press, 2019.

[6] M. Chhabra, B. Gupta, A. Almomani, A novel solution to handle DDOS attack in MANET, J. Inf. Assur. Secur. 04 (03) (2013) 165–179.

[7] S.M. Kasongo, Y. Sun, A deep long short-term memory based classifier for wireless intrusion detection system, ICT Express 6 (2) (2020) 98–103.

[8] M.M. Najafabadi, T.M. Khoshgoftaar, A. Napolitano, C. Wheelus, RUDY Attack: Detection at the Network Level and Its Important Features., in: FLAIRS Conference, 2016, pp. 288– 293.

[9] J. Park, K. Iwai, H. Tanak, T. Kurokawa, Analysis of Slow Read DoS Attack and Countermeasures, in: The International Conference on Cyber-Crime Investigation and Cyber Security (ICCICS2014), The Society of Digital Information and Wireless Communication, 2014, pp. 37–49.

[10] K. Hong, Y. Kim, H. Choi, J. Park, SDN-assisted slow HTTP DDoS attack defense method, IEEE Commun. Lett. 22 (4) (2017) 688–691.

[11] M. Latah, L. Toker, Minimizing false positive rate for DoS attack detection: A hybrid SDN-based approach, ICT Express 6 (2) (2020) 125–127.

[12] F.S. d. Lima Filho, F.A. Silveira, A. de Medeiros Brito Junior, G. Vargas-Solar, L.F. Silveira, Smart detection: An online approach for DoS/DDoS attack detection using machine learning, Secur. Commun. Netw. (2019).

[13] N. Tripathi, N. Hubballi, Y. Singh, How secure are web servers? An empirical study of slow HTTP DoS attacks and detection, in: 2016 11th International Conference on, in: Availability, Reliability and Security (ARES), IEEE, 2016, pp. 454–463.

[14] H.H. Jazi, H. Gonzalez, N. Stakhanova, A.A. Ghorbani, Detecting HTTP-based application layer DoS attacks on web servers in the presence of sampling, computer, Networks 121 (2017) 25–36.

[15] N. Tripathi, N. Hubballi, Slow rate denial of service attacks against HTTP/2 and detection, Comput. Secur. 72 (2018) 255–272.

[16] E. Cambiaso, G. Papaleo, M. Aiello, Taxonomy of slow DoS attacks to web applications, in: International Conference on Security in Computer Networks and Distributed Systems, Springer, Berlin, Heidelberg, 2012, pp. 195–204.

[17] R. Hofstede, P. Celeda, B. Trammell, I. Drago, R. Sadre, A. Sperotto, A. Pras, Flow monitoring explained: From packet capture to data analysis with NetFlow and IPFIX, IEEE Commun. Surv. Tutor. 16 (4) (2014) 2037–2064.

[18] CIC DoS dataset(2017), [Online]. Available: http://www.unb.ca/cic/datasets/dos-dataset.html.

[19] Keras: The Python Deep Learning library, [Online]. Available: https://keras.io/.

[20] scikit-learn: Machine Learning in Python, [Online]. Available: https://scikit-learn.org/stable/index.html.