JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF DDOS ATTACK USING LSTM TECHNIQUE

Beulah A
Department of Electronics Engineering,
Pondicherry University,
Pondicherry, India.

Samundiswary. P
Department of Electronics Engineering,
Pondicherry University,
Pondicherry, India.

ABSTRACT: The Internet of Things (IoT) has increased significance in the modern period as a result of rapid development in the technology in several ways. The popularity of IoT applications has increased in comparison to earlier times due to the availability of several devices that act as IoT enablers, such as smartwatches, smartphones, security cameras, and smart sensors. However, a number of issues, such as Distributed Denial-of-Service (DDoS) assaults, have been brought on by the absence of security in IoT devices. Many efforts have been made recently to develop intelligent models that can protect IoT networks from DDoS attacks. Making a model that can fight against DDoS attacks and be able to differentiate between legitimate traffic and false alarms is the main topic of research that is still being done. The most popular moniker for distributed network attacks is DDoS attacks. These attacks take advantage of limitations including the layout of the website run by the authorised organisation and arrangement of asset. To determine the current state of DDoS attacks, the University of California Irvine (UCI) machine learning respiratory dataset must be used. Further, Deep learning algorithms are typically used to categorise and forecast the many forms of DDoS attacks. In Existing Method XG Boost is used and performance metrices are analysed. However, the accuracy is not good. Hence in this paper, Long Short-Term Memory (LSTM) method based Recurrent Neural Network (RNN) algorithm is developed for the detection of DDoS attack which is considered as the proposed work. LSTM achieves precision (PR), recall (RE) and F1 score of approximately 99%. Additionally, the above-mentioned technique achieves an Accuracy (AC) of roughly 94%.

Keywords: Distributed Denial-of-Service (DDOS) attacks, Internet of Things, Deep Learning, classification, Recurrent Neural Network, Long Short-Term Memory.

1. INTRODUCTION

The method by which attackers or hackers attack a server by interfering with all the network services used by the users connected to that network is known as a DDoS assault. The attackers employ unwelcome bot traffic to flood the website, preventing regular users from accessing the target website. In order to overwhelm the website's network and prevent legitimate users from using it, the attackers use a variety of IoT devices, network servers, and gadgets [1], [2]. The Internet of Things (IoT) refers to a network of connected, web-connected things that may exchange information automatically between remote organizations [3]. The "Things" could be anything with sensors that can collect and transfer information within an organisation, such as linked clinical equipment, bio-chip transponders, solar panels, and associated cars with sensors that can warn the driver of a number of potential issues. Deep learning algorithms for intrusion detection are suggested [3]. The models used are Convention Neural Network (CNN) and Recurrent Neural Network and XG Boost. The proposal was proven to work well with RNN. The typical accuracy stood at 79%. For intrusion detection, several researchers [5] proposed a hybrid deep learning model. For the categorization of CNN and LSTM from the RNN model, they integrated two deep learning techniques. For all datasets, tree-based meta-estimators outperform others on absolute measures. Almost always, between 5 and 15% of the training dataset results in convergence [6]. It has been demonstrated that the ensemble model [7] significantly increases the detection accuracy when compared to other methods. The proposed adaptive voting algorithm outperforms the alternatives with accuracy, precision, recall, and F1 values of 85.2%, 86.5%, and 84.9%, respectively. Comprehensive experimental findings show that Supervised Adversarial Variational Auto-Encounter Regularization with Deep Neural Network (SAVAER-DNN) has a greater detection rate for low frequency attacks in addition to being able to identify known and unknown attacks. Additionally, comparative experimental results on the UNSW-NB15 dataset show that SAVAER-DNN is capable of detecting advanced network threats. The highest overall accuracy, highest overall Precision, and highest overall F1 score using UNSW-NB15 dataset are reported to be 93.01 percent, 91.94%, and 93.54%, respectively [8]. The recommended model was found to perform satisfactorily with an accuracy of more than 98.76% and a FAR of less than 1.2% using the UNSW-NB15 dataset. The attention-based intrusion detection model surpasses a number of state-of-the-art methods, including auto-encoder, deep

feedforward neural networks, and single-class support vector machines, in terms of detection precision [9]. It focuses on creating a portable Intrusion Detecting System (IDS) for IoT anomaly detection [10]. The intended victim is a common DDoS assault. The two primary issues on which the suggested IDS is focused are the characteristic of the receiving data used to categorise the signal and the classifier that is based on machine learning. This approach outlines the four most common Supervisory Control And Data Acquisition (SCADA) IIoT protocols, their security vulnerabilities, a risk assessment of the most important and pervasive SCADA IIoT system vulnerabilities, and how Machine Learning (ML) -based solutions would be successful to address them. [11]. The LSTM. PQDO method, which is based on the deep neural LSTM network, is proposed to automatically extract the known and potential linguistic statistical features of the Domain Generation Algorithm (DGA) domain name, and dynamically optimise the samples' [12] proportion based on a careful examination of the quantity and makeup of the original samples. The sample ratio is iterated using the prior optimisation method's optimal answer as the initial iteration value. The best option is then found by heuristically searching in the desired direction around the initial solution. The numerous studies on the application of ML algorithms for anomaly-based Network based Intrusion Detecting System (NIDS) are completed and offer new comparisons between the use of different datasets, including UNSW-NB15 and UGR16. The usage of the mean deviation of the variables is made possible by the significance of pre-processing features such fundamental traits and statistical traffic traits using z-score standardisation procedures. An appropriate method is used to choose relevant papers in the field of Artificial Intelligence (AI)-based NIDS [13]. The methodology's benefits and drawbacks in terms of intrusion detection effectiveness and model complexity are discussed. Nearly 80% of the suggested solutions were built using deep learning approaches, with DNN being the most often used algorithms. Even so, deep learning techniques outperform ML-based ones in terms of their ability to independently learn features and to produce models that are more accurately fit [14]. Port scanning and DDoS attacks are classified as a mix of legitimate and malicious traffic using a comparison study on the recently released benchmark dataset CICIDS2017. To test how effectively 22 different machine learning algorithms work against the most recent attack vectors, they have all been built and tested. The classification outcomes indicate that all Support Vector Machine (SVM) and discriminant analysis variants provide testing accuracies of higher than 90%. The SVM variation with the greatest accuracy rating, 99%, is the Fine Gaussian variation [15]. A new payload approach-based video steganography method Divide Embed Component Method (DECM) was developed using the video steganography botnet model based on Simple Notification Service (SNS) messengers by using the free software Virtual Dub and Stegano [16]. To effectively detect DDoS attacks, the suggested methodology employs a number of steps, including pre-processing, feature selection, and categorization. All values are normalised to a predetermined scale range as part of the pre-processing method. The suggested Feature Selection-Whale Optimization Algorithm (FS-WOA) model is applied to the normalised data [17] to determine the optimal collection of characteristics. The goal of this research is to design an LSTM algorithm that can forecast DDOS attacks.

The rest of the article is organized as follows: Section 2 deals with existing work, Section 3 deals with proposed work, Section 4 deals with results and discussions, Section 5 deals with performance comparison and Section 6 deals with conclusion and future scope.

2. EXISTING WORK:

This section describes about the method of XG Boost is used to detect and classify the normal from attacker datasets by separating the testing datasets from training datasets. The block diagram of existing work illustrated in Fig.1.

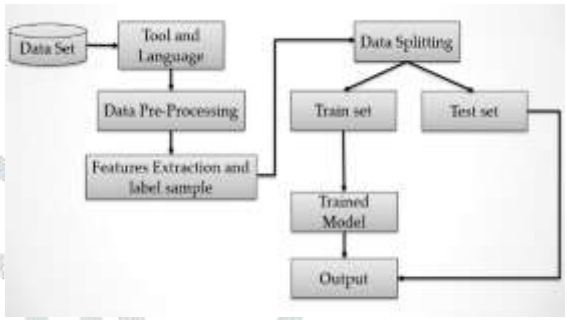

Fig. 1 Block diagram of existing work

Step 1: Datasets: To complete the procedure, the UNSW-nb15 dataset, which includes statistics on DDoS attacks, is obtained from GitHub. This dataset was provided by the Australian Centre for Cyber Security (ACCS), and the anaconda programme uses Python as a simulator language.

Step 2: Data Pre-Processing: Here, the data is purged from any unimportant information and is transformed it into reliable data. Using statistical methods values are removed from the data that are not crucial to the experimental research. Every data analysis for the initial phase evaluation needs to include this. After that, it is required to transform data into trustworthy form.

Step 3: Label Encoding: In order for machines to read labels, they must be transformed into a numeric representation. The effectiveness of the labels can then be assessed using machine learning techniques. It is a crucial supervised learning pre-processing step for the structured dataset.

Step 4: Data Spitting: For evaluating the training and testing dataset, the Scikit learn model selection library is used. This dataset is further categorised into two groups which is:

I. The target class is another name for the dependent class.
II. Classes that are independent of other classes are referred to as independent classes.

Step 5: Algorithm: For regression and classification on large datasets, the well-known supervised learning method XG Boost is used in this case. A training set and a testing set must be made independent of the data. About 75% of the rows are in the training set, while the remaining 25% were all randomly selected without replacement for the testing set.

3. PROPOSED WORK:

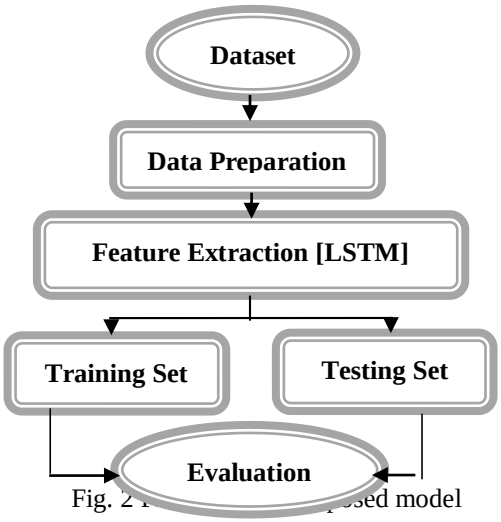

Fig. 2 [illegible]sed model

The flow diagram of the proposed model is shown in Fig.2.

Step 1: Dataset: The machine learning community uses the UCI Machine Learning Repository to conduct empirical research on machine learning algorithms. It is a collection of datasets, domain theories, and data generators. The National Science Foundation's funding assistance is sincerely acknowledged. As a result, it is widely employed to spot network breaches like DDoS assaults.

Step 3: Feature Extraction [LSTM]: Deep learning uses Long Short-Term Memory (LSTMs) which is illustrated in Fig. 3. In particular when conducting tasks involving sequence prediction, a wide number of recurrent neural networks (RNNs) are capable of learning long-term dependencies. Because LSTM has so many feedback links, it can process the entire data sequence. LSTM performs superbly across a broad spectrum of problems.

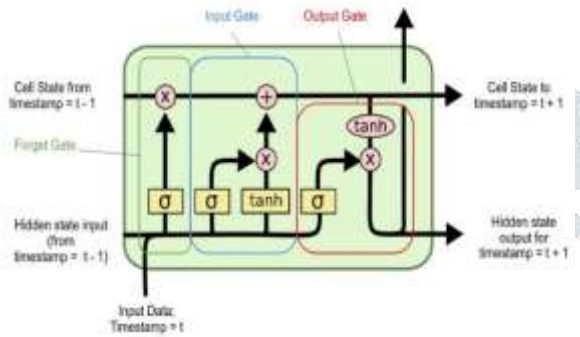

Fig. 3 Long Short-Term Memory

Step 4: Training Set: A machine learning model is trained using a very large dataset referred to as training data. Using training data, prediction models that use machine learning algorithms are taught how to extract characteristics that are important to business goals.

Step 5: Testing Set: The testing dataset evaluates the model's effectiveness and verifies that it successfully generalizes to fresh or untested datasets. The original data from the training dataset is a distinct subset here.

Step 6: Evaluation: Evaluating a model's performance is critical in the early stages of research. Model monitoring and assessment work together. A confusion matrix, an accuracy matrix, and a precision matrix are all present in evaluation classification.

4. RESULTS AND DISCUSSIONS

Existing Work:

The data are loaded in the dataset table shown in Fig. 4

Fig. 4 Uploaded Dataset Table

The structured data set obtained is shown in Fig.5

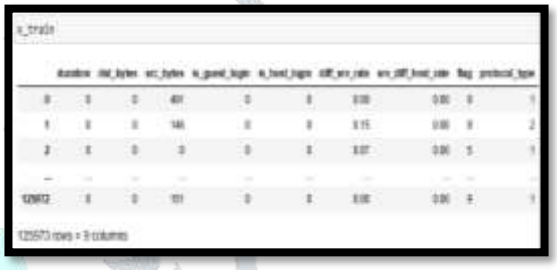

Fig. 5 Structured Datasets table

The intrusion detection system prediction table is shown in Fig. 6

Fig. 6 Prediction table

The table of normal or attacker is shown in Fig. 7

	A	B	C	D	E	F	G	H	I	J
1	duration	protocol_type	src_bytes	dst_bytes	is_host_login	is_guest_login	diff_srv_rate	srv_diff_host_rate	flag	labels
2	0	tcp	491	0	0	0	0	0	9	normal
3	0	udp	146	0	0	0	0.15	0	9	normal
4	0	tcp	0	0	0	0	0.07	0	5	attacker
5	0	tcp	232	8153	0	0	0	0	9	normal
6	0	tcp	0	0	0	0	0.07	0	5	attacker
7	15159	tcp	150	1185	1	0	0	1	9	attacker

Fig. 7 Table of normal or attacker

Proposed work:

The accuracy and loss proposed method is shown in Fig. 8 (a) and Fig. 8 (b)

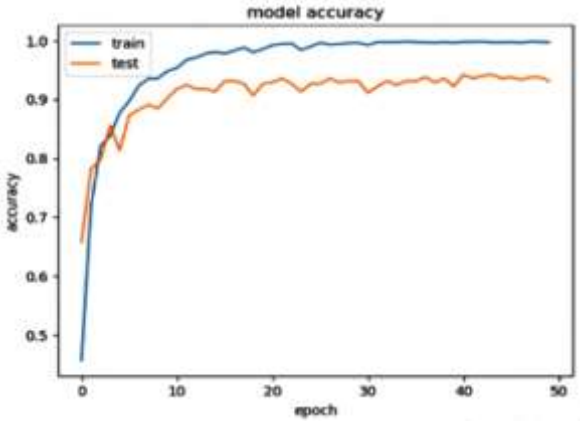

Fig. 8 (a) Graphical Representation of accuracy

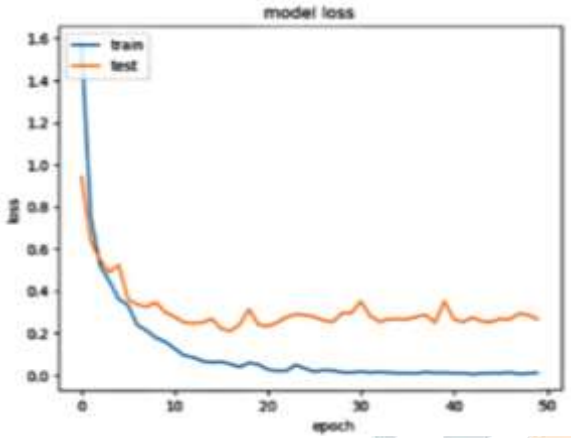

Fig.8 (b) Graphical Representation of loss

The bar graph representing the normal and attacker dataset is shown in Fig. 9

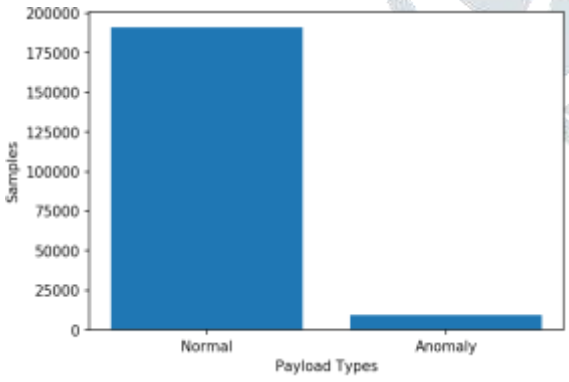

Fig. 9 Bar graph representation of normal & attacker

5. PERFORMANCE COMPARISON:

Table 1. Performance Comparison between Existing and Proposed Method

DATA SETS	ACCURACY	PRECISION	RECALL	F1 SCORE
Proposed (LSTM)	94%	99%	99%	99%
Existing (XG Boost)	89%	92%	88%	92%

Table. 1 Lists the proposed and existing method's performance measures which includes Accuracy, Precision, Recall, and F1-Score.

6. CONCLUSION AND FUTURE WORK:

This article describes a thorough, organised process for spotting DDoS attacks. LSTM is a potential method for detecting distributed denial of service threats. An effective tool for learning and analysing sequential data is the LSTM neural network type, which makes it a good choice for identifying irregularities in network traffic patterns. By training an LSTM model on normal network traffic data, it can learn the typical patterns of traffic and identify any deviations from the normal traffic patterns. This can help in detecting DDoS attacks as they usually involve abnormal traffic patterns that are different from the normal ones. Moreover, LSTM can also be used for mitigating DDoS attacks by predicting the traffic patterns and then blocking any traffic that is not consistent with the predicted patterns. This can effectively prevent the malicious traffic from overwhelming the network and causing a denial of service. The testing accuracy of the suggested model is high at 91.45%, while the training accuracy is high at 94%. However, further research is needed to evaluate the performance of model on larger datasets and in different environmental conditions. The future scope of using LSTM for DDoS attack detection and mitigation looks promising. With the increasing number of connected devices and the growing demand for online services, the risk of DDoS attacks is also increasing. Hence advanced deep leaning techniques can provide an efficient way of addressing this security challenge by enabling networks to quickly identify and respond to DDoS attacks.

REFERENCES:

1. N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, ‘‘Adversarial machine learning applied to intrusion and malware scenarios: A systematic review,’’ *IEEE Access*, vol. 8, pp. 35403–35419, 2020.

2. G. Karatas, O. Demir, and O. K. Sahingoz, ‘‘Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset,’’ *IEEE Access*, vol. 8, pp. 32150–32162, 2020.

3. T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ‘‘BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset,’’ *IEEE Access*, vol. 8, pp. 29575–29585, 2020.

4. H. Jiang, Z. He, G. Ye, and H. Zhang, ‘‘Network intrusion detection based on PSO-xgboost model,’’ *IEEE Access*, vol. 8, pp. 58392–58401, 2020.

5. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty, and V. S. Kiran, ‘‘Similarity based feature transformation for network anomaly detection,’’ *IEEE Access*, vol. 8, pp. 39184–39196, 2020.

6. L. D’hooge, T. Wauters, B. Volckaert, and F. De Turck, ‘‘Classification hardness for supervised learners on 20 years of intrusion detection data,’’ *IEEE Access*, vol. 7, pp. 167455–167469, 2019.

7. X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, ‘‘An adaptive ensemble machine learning model for intrusion detection,’’ *IEEE Access*, vol. 7, pp. 82512–82521, 2019.

8. Y. Yang, K. Zheng, B. Wu, Y. Yang, and X. Wang, ‘‘Network intrusion detection based on supervised adversarial variational auto-encoder with regularization,*’’ IEEE Access*, vol. 8, pp. 42169–42184, 2020.

9. C. Liu, Y. Liu, Y. Yan, and J. Wang, ‘‘An intrusion detection model with hierarchical attention mechanism,’’ *IEEE Access*, vol. 8, pp. 67542–67554, 2020.

10. S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, ‘‘Toward a lightweight intrusion detection system for the Internet of Things,’’ *IEEE Access*, vol. 7, pp. 42450–42471, 2019.

11. M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, ‘‘Machine learning-based network vulnerability analysis of industrial Internet of Things,’’ *IEEE Internet Things* J., vol. 6, no. 4, pp. 6822–6834, Aug. 2019.

12. Y. Chen, B. Pang, G. Shao, G. Wen, and X. Chen, ‘‘DGA-based botnet detection toward imbalanced multiclass learning,’’ Tsinghua Sci. Technol., vol. 26, no. 4, pp. 387–402, Aug. 2021.

13. X. Larriva-Novo, V. A. Villagrá, M. Vega-Barbas, D. Rivera, and M. S. Rodrigo, ‘‘An IoT-focused intrusion detection system approach based on pre-processing characterization for cybersecurity datasets,’’ *Sensors*, vol. 21, no. 2, p. 656, Jan. 2021.

14. Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, ‘‘Network intrusion detection system: A systematic study of machine learning and deep learning approaches,’’ *Trans. Emerg. Telecommun. Technol*., vol. 32, no. 1, p. e4150, Jan. 2021.

15. M. Aamir, S. S. H. Rizvi, M. A. Hashmani, M. Zubair, and J. A. Usman, ‘‘Machine learning classification of port scanning and DDoS attacks: A comparative analysis,’’ *Mehran Univ. Res. J. Eng. Technol.*, vol. 40, no. 1, pp. 215–229, Jan. 2021.

16. M. Kwak and Y. Cho, ‘‘A novel video steganography-based botnet communication model in telegram SNS messenger,’’ *Symmetry*, vol. 13, no. 1, p. 84, Jan. 2021.

17. Agarwal, M. Khari, and R. Singh, ‘‘Detection of DDOS attack using deep learning model in cloud storage application,’’ *Wireless Pers. Commun.*, vol. 2, pp. 1–21, Mar. 2021.