DDoS Risk Assesment using Machine Learning Techniques

Mutyala Sridevi , Nandeesh V

Abstract *- The result of exploiting holes in safeguarding Internet of Things (IoT) devices, the amount of cyber-attacks and data breaches has skyrocketed across various corporations, companies, and sectors. Because its capacity to extract and learn deep characteristics of known assaults and detect novel attacks without the need for manual feature engineering, machine learning is used in cyber-attack detection. Despite the use of improved Machine Learning (ML) techniques for intrusion detection, the assault remains a huge danger to the Internet. The primary target of this research is to identify and detect assaults on the network. The expansion of social networks is now increasing on a daily basis.*

However, detecting the assaults is a difficult task. By examining the information in the KDDCUP Dataset, this project will dynamically detect the attack. The method of feature scaling is employed to standardize a range of independent variables or data constituents. The feature reduction PCA technique is used to locate the directions of highest variance in high-dimensional data and project it onto a new subspace with the same or less dimensions than the original one. Finally, the ML classification technique is used to categorise the data. Based on the assaults and the typical event, the final report is created..

Key Words: DDOS, Machine learning, Cyber Attacks,PCA Algorithm.

1. INTRODUCTION

The Internet of Things (IoT) is regarded as a fast evolving paradigm in computer history. IoT has expanded tremendously in several technology domains during the last few years. It has resulted from the convergence of hundreds of billions of devices from various systems (such as smart automobiles, smart health care, smart grid, smart home, and so on) with the internet. However, because IoT merges the digital and physical worlds, this convergence has resulted in numerous cyber-attacks against IoT equipment. Because of the heterogeneity, enormous scale, limited hardware resources, and worldwide accessibility of IoT systems, IoT security has become a challenge. With the advancement of information technology, IoT technology has swiftly grown and is now extensively employed in industries such as agriculture, military, and so on.

Although IoT is highly utilised and technologically diversified, numerous devices are continually being integrated with the IoT, either as IoT terminals or as IoT branches. As an open environment based on the Internet, the Internet of Things (IoT) has complex and diversified security concerns in its gadgets, which are continually attacked and destroyed by the outside world. As a result, it is important to improve the identification of security concerns in IoT.Among the present security technologies are security gateways, firewalls, code signatures, encryption technology, and so on, but they are all passive security defensemeasures that cannot perform active detection and reaction. The goal of IoT intrusion security detection is to assess whether the IoT is in a harmful environment by collecting data and analysing attack behaviour. To accurately categorise and predict data.

To address the sparsity issue.

Enhancing the overall performance of forecast results.

2. LITERATURE SURVEY

R.S. Miani explains that the Internet of Things (IoT) represents a novel paradigm that merges the internet with physical devices across diverse domains such as home automation, industrial processes, human health, and environmental monitoring. It increases the prevalence of Internet-connected gadgets in our everyday lives, bringing with it, in addition to numerous benefits, security difficulties. Intrusion Detection Systems (IDS) have been an essential tool for network and information system security for more than two decades. However, because of features such as constrained-resource devices, unique protocol stacks, and standards, implementing classic IDS approaches to IoT is problematic. We give an overview of IDS research activities for IoT in this article. Our goal is to uncover emerging trends, unresolved challenges, and future research opportunities. The IDSs proposed in the literature were categorised based on the following characteristics: detection technique, IDS deployment strategy, security threat and validation approach. We also reviewed the many options for each characteristic, delving into parts of works that either offer unique IDS schemes for IoT or provide attack detection methodologies for IoT threats that may be included in IDSs.

Advantages

Placing the IDS in the border router detects Internet intrusion threats against items in the physical domain.

Disadvantages

Identify obstacles to making Cloud Computing a reliable platform for providing IoT services.

3. EXISTING SYSTEM

Existing systems are used to identify intrusion attacks on IoT devices. It use the deep learning auto encoder approach to detect intrusion attacks in IoT devices, whether they are normal or malicious.

When utilising a deep learning encoder, essential variables will be misunderstood, and the data will be trained using the incorrect use case.
Many academics are utilising deep learning and machine learning approaches to detect DDoS attacks that have the most impact on social networking. Few of many current works in this field is covered below.
Encryption and accessibility control are similar in that they both refer to privacy and prevention. The main distinction is that encryption generally concerns with data secrecy. Data can be accessed by either a trustworthy or an untrusted party. Encryption guarantees that only authorised and trustworthy parties have access to the data. Access control, conversely, attempts to limit data access. Data constraints frequently occur among trusted parties. As result, encryption solutions should be more powerful than access control mechanisms. Encryption places severe constraints on data secrecy.

4. VARIOUS TYPES OF DDOS ATTACK

ICMP FLOOD

The hacker utilises ICMP echo request packets to deliver a service request to a genuine user. The attack can eat incoming and outgoing bandwidth and cause servers to reply to packets, resulting in overall system delay.

SYN FLOOD

It may target any device that is linked to the system via the internet. The sender of the SYN flood makes several SYN requests but does not respond to the host's SYN-ACK and sends SYN queries with a faked IP address. The host waits for the acknowledgement, which results in a denial of service.

PING OF DEATH

It transmits a larger-than-allowable-size packet into the system to the intended source. The maximum length of an IP packet is 65535 bytes. This exploit has the potential to overflow a packet-allocated memory buffer.

SLOWLORIS

It makes a request between a single computer and a server. It establishes a connection to the target server but only delivers a portion of the request, and it always sends HTTP headers in an incomplete manner. Because the target server has all failed connections open, the maximum concurrent connection results to connection denial.

NTP AMPLIFICATION

The attacker concentrates on assaulting the publicly available source via UDP packets in this case. The attacker acquires a list of open NTP servers from which he may quickly construct a high-volume, high-bandwidth assault.

HTTP FLOOD

This attack is carried through via HTTP request. To carry out this attack, the attacker sends a request to the end user over http, after which the attack is carried out.

5.METHODOLOGY

The model that is suggested is introduced to address all of the shortcomings of the current system. In the proposed system, feature scaling is a mechanism for normalising the range of independent variables or data features. It will assist in selecting the best feature by utilising the PCA method. As result, system's performance will improve.
It speeds up training and allows for more easy weight decay and Bayes optimisation.

6. IMPLEMENTATION

1) Data Selection and Loading

Identifying the appropriate data type and source, along with selecting suitable devices for data collection, constitutes the practice referred to as data selection.
Data selection comes before data collection and is the process through which data relevant to the analysis is determined and obtained from data gathering.
After data is received and integrated from many sources, cleaned and formatted, and then loaded into a storage system, such as a cloud data warehouse, data loading refers to the "load" component.
The KDD CUB dataset is being utilised in this study to detect intrusion attacks.

2) Data Preprocessing

The process of deleting undesirable data from a dataset is known as data pre-processing. Missing data removal: In this step, null values such as missing values are eliminated using the imputer library.

3) Splitting dataset into training dataset and testing dataset

The act of separating accessible data into segments is known as data splitting. Two sections, commonly for cross-validators.
One component of the data is used to create a predictive model, while the other is utilised to assess the model's performance. It is critical to separate data into training and testing sets when analysing data mining methods.

Upon dividing data into a training set and a testing set, a significant portion of the data is allocated for training, while a smaller segment is employed for testing purposes.

4) Classification

Random forests, referred to as random decision forests, represent an ensemble learning technique used for tasks such as classification and regression. This method involves the creation of a substantial quantity of decision trees during the training process. The final output is determined by either selecting the mode of classes (for classification) or calculating the mean/average prediction (for regression) across the individual trees.

5) Prediction

The method of anticipating DDoS attacks based on dataset.This project will successfully forecast data from a dataset by improving the overall prediction outcomes.

6) Result Generation

The total classification and forecast will be used to create the Final Result. The effectiveness of the suggested technique is assessed using metrics such as,

- Accuracy
- Precision
- Recall
- F1 score

7. CONCLUSION

The Machine learning classifier is used in this work to assess intrusion attacks using IoT device data. The algorithm Principle Component Analysis (PCA) is used to extract the feature from the dataset. In classification, a machine learning technique such as the Support Vector Machine (SVM) is used to predict the outcome based on accuracy, precision, recall, and f1 score.

This study's dataset was derived from a single network. This research can be carried out using a bigger or smaller network region.In the near future, systems will rely on analytics and data networks to make predictions using machine learning or deep learning.

8. FUTURE ENHANCEMENTS

The dataset analysed in this study was from a single network. This study can be conducted with a larger and smaller network area. In near future systems are moving towards Analytics and Data network to find the Prediction through machine learning or deep learning.

REFERENCES

[1] S. Sarkar, S. Chatterjee, S. Misra, “Assessment of the suitability of fog computing in the context of internet of things,” IEEE Transactions on CloudComputing, vol. 6, no. 1, pp. 46-59, 2018.

[2] G. Yang, Q. Zhang, Y. C. Liang, “Cooperative ambient backscattercommunications for green internet-of-things,” IEEE Internet of Things Journal, vol. 5, no. 2, pp. 1116-1130, 2018.

[3] H. Subir, G. Amrita, C. Mauro, “Limca: an optimal clustering algorithm for lifetime maximization of internet of things,” Wireless Networks, vol. 4, pp.1-19, 2018.

[4] S. K. Choi, C. H. Yang, J. Kwak, “System hardening and security monitoring for iot devices to mitigate iot security vulnerabilities and threats,” Ksii Transactions on Internet & Information Systems, vol. 12, no. 2, pp.906-918, 2018.

[5] S. U. Haq, Y. Singh, “On iot security modelstraditional and block chain,” International Journal of Computer Sciences & Engineering, vol. 6, no. 3, pp.26-31, 2018.

[6] P. Bajpai, A. K. Sood, R. J. Enbody, “The art of mapping iot devices innetworks,” Network Security, vol. 4, pp. 8-15, 2018.

[7] Z. A. Al-Odat, S. K. Srinivasan, E. M. Al-Qtiemat, S. Shuja, “A reliableiot-based embedded health care system for diabetic patients,” InternationalJournal on Advances in Internet Technology, vol. 12, pp. 50-60, 2019.

[8] S. Y. Hashemi, F. S. Aliee, “Why dynamic security for the internet ofthings?,” Journal of Computing Science & Engineering, vol. 12, no. 1, pp.12-23, 2018.

[9] M. Meddeb, A. Dhraief, A. Belghith, et al., “Named data networking: apromising architecture for the internet of things (iot),” International journal onSemantic Web and information systems, vol. 14, no. 2, pp. 86-112, 2018.

[10] A. Kamilaris, X. P. Francesc, “Deep learning in agriculture: a survey,”Computers and Electronics in Agriculture, vol. 147, no. 1, pp. 70-90, 2018.

[11] Y. B. Wang, Z. H. You, L. P. Li, et al., “Improving prediction ofself-interacting proteins using stacked sparse auto-encoder with PSSMprofiles,” International journal of biological sciences, vol. 14, no. 8, pp. 983,2018.

[12] V. Badrinarayanan, A. Kendall, R. Cipolla, “Segnet: A deepconvolutional encoder-decoder architecture for image segmentation,” IEEEtransactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp.2481-2495, 2017.

[13] G. X. Cui, D. K. Li, “Overview on Deep Learning Based on AutomaticEncoder Algorithms,” Computer Systems & Applications, vol. 9, pp. 7, 2018.

[14] J. He, L. Zhao, H. Yang, et al., “HSI-BERT: Hyperspectral ImageClassification Using the Bidirectional Encoder Representation FromTransformers,” IEEE Transactions on Geoscience and Remote Sensing, vol. 58,no. 1, pp.165-178, 2019

[15] M. Peter, K. Gergana, M. Radoslav, P. Nevena. “Curve fitting problem:torque – velocity relationship with polynomials and boltzmann sigmoidfunctions,” acta of bioengineering & biomechanics, vol. 20, no. 1, pp. 169-184,2018.

[16] Y. Liu, S. Liu, Y. Wang, et al., “A stochastic computational multi-layerperceptron with backward propagation,” IEEE Transactions on Computers, vol.67, no. 9, pp. 1273-1286, 2018.

[17] D. Rathore, A. Jain, “Design hybrid method for intrusion detection usingensemble cluster classification and som network,” International Journal ofAdvanced Computer Research, vol. 2, no. 3, pp. 181-186, 2019.

[18] M. E. Boujnouni, M. Jedra, “New intrusion detection system based onsupport vector domain description with information gain metric,” InternationalJournal of Network Security, vol. 20, no. 1, pp. 25-34, 2018.

[19] R. F. Molanes, K. Amarasinghe, J. Rodriguez-Andina, et al., “Deeplearning and reconfigurable platforms in the Internet of Things: Challenges andopportunities in algorithms and hardware,” IEEE industrial electronicsmagazine, vol. 12, no. 2, pp. 36-49, 2018.

Y. J. Ai, P. Liang, Y. X. Wu, et al., “Rapid qualitative and quantitativedetermination of food colorants by both Raman spectra and Surface-enhancedRaman Scattering (SERS),” Food chemistry, vol. 241, pp. 427-433, 2018., no. 1, pp. 70–76, Jan. 2007.