International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 8, No. 1, 2021, pp. 53-57.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Machine Learning-enabled Intrusion Identification Model for IoT Environment

Aakansha Soy [a], Ashu Nayak [b]

[a] *Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.*
E-mail: ku.aakanshasoy@kalingauniversity.ac.in
[b] *Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.*
E-mail: ku.ashunayak@kalingauniversity.ac.in

Abstract

The Internet of Things (IoT) encompasses millions of autonomous and diverse intelligent devices that interact with one another without people's involvement. IoT-based solutions have enhanced experiences across many industries, including medical care, farming, supply chain management, learning and traffic monitoring, and utilities. Node heterogeneity has produced security concerns, constituting one of the most complex IoT issues. Employing security mechanisms such as data encryption, access management, and identification for IoT gadgets is unsuccessful in ensuring safety. This research identifies different types of IoT assaults. It discusses both superficial learning techniques and Machine Learning (ML) methods utilized in systems for Intrusion Detecting Systems (IDS) within the context of IoT. The efficacy of these models has been assessed utilizing five benchmark databases. Performance criteria were employed to analyze the effectiveness of shallow and ML-oriented IDS. ML-IDS surpasses shallow computer learning in identifying IoT assaults.

Keywords: Machine Learning, Intrusion Identification, IoT, Security.

Received: 11 January 2021/Revised: 21 January 2021/Accepted: 17 February 2021/Published: 31 March 2021

Introduction

The Internet of Things (IoT) (Madakam et al., 2015) is increasing and integral to daily life. IoT devices can utilize an internet protocol address to connect to the World Wide Web. These smart devices are advancing numerous smart applications, like automation of tasks, controlling homes, intelligent vehicles, choice analytics, clever grids, healthcare systems, learning, and industrial advancement. Analysts envision a world where the amount of linked gadgets will outdo the populace. The research projected that by 2024, 42.4 billion internet-connected gadgets would generate 78.3 zettabytes (ZB) of information, in contrast to a predicted worldwide populace of 7.3 billion.

In IoT structures, diverse gadgets are linked to a sophisticated structure, raising safety issues. The primary problem is to guarantee safety in resource-constrained IoT gadgets. These IoT gadgets are susceptible to various forms of threats. Intrusion Detection Systems (IDSs) (Elrawy et al., 2018) are fundamental but potent security mechanisms for ensuring adequate network safety in any IoT-integrated environment. They excel in tracking, evaluating, and identifying practical information via data gathering, regardless of whether the sources are invaders. IDS are conventionally categorized into networking-based IDS (NIDS) and host-oriented IDS (HIDS) according to their detecting locations (Deshpande et al., 2018).

The primary problem is guaranteeing security in assets-constrained IoT gadgets (Iftikhar et al., 2021). These IoT gadgets are susceptible to various forms of threats. IDSs are a fundamental and potent safety technique for ensuring adequate network protection in any IoT-embedded context. They excel at tracking, evaluating, and identifying real-time data packages through passive data collecting, regardless of whether the sources are intruders. IDSs are conventionally categorized into NIDS and HIDS according to their detecting locations.

IDS seeks to track traffic and promptly identify various malicious behaviors. As the frequency of anomalies increases, the enhancement and advancement of IDSs is critically vital, as the primary issue in ID is identifying covert attacks inside standard traffic patterns. Machine Learning (ML) (Saranya et al., 2020) has garnered significant attention from researchers aiming to enhance the effectiveness of IDSs due to its dynamic capability to distinguish between viruses and benign information with a high detection rate. Various supervised and unsupervised ML tools have been effectively implemented.

Background

This area encompasses studying ML techniques and classifiers implemented into intrusion detection systems to identify intrusions in IoT systems. The study presents a Bi-Long Short-Term Memory (Bi-LSTM) Recurrent Neural Network (RNN) (Bi-LSTM RNN) methodology for detecting intrusions, focusing on binary categorizing normal and suspicious attacks (Li et al., 2018) .The developed system taught with an accuracy exceeding 94% in detecting IoT attacks.

The study employed Broad Learning Systems (BLS) methods on the database to categorize known intrusions. The results indicate that the BLS decreases training duration, achieving accuracies of 81.3% and 71.3% for the datasets. The paper empirically investigated IDS utilizing various ML and Deep Learning (DL) algorithms (Ahmad et al., 2021) Two databases have been used to evaluate the efficacy of these methods in identifying and categorizing threats across 22 assessment measures. The experimental results indicate that DL models leverage ML models, particularly that Deep Belief Networks (DBN) improve the detection rate of errors by 4% to 12% compared to others.

The study examined various paradigmatic ML algorithms for identifying breaches in IoT networks that fail the system. Verification was conducted on the database utilizing every evaluated ML method (Hasan et al., 2019) Among these, RF achieves an accuracy of approximately 91.2% in detecting IoT assaults; however, this level of accuracy is not assured in scenarios involving real-time or extensive unknown data. The study presented a model for detecting abnormalities in IoT traffic. The study utilized two forms of the starting set of DS2OS: the information collected over 12 days and the under-sampled version of this data. The HS-TCN classifier is better than LSTM and Support Vector Machines (SVMs) across the changed datasets. A novel ML-based approach to the detection of anomalies was developed using Logical Regression (LR) and Artificial Neural Network (ANN) classification methods. LRs and ANNs achieve approximately 91.3% precision with the complete database, whereas 91.2% is attained by excluding approximately 110M data points. The database is separated into 74% and 22% proportions in both instances.

Proposed ML-based IDS

The strategy framework for detecting assaults on an IoT network comprises fundamental stages. Fig. 1 illustrates the comprehensive structure, wherein the initial stage involves database collection, meticulous observation, and analysis

of features and their respective data kinds. The subsequent stage consists of preprocessing the database for training models. The generated database is separated into learning and verifying groups in a 75:25 ratio. The training set is utilized with the specified ML methods for the learning method. Finally, the ultimate predictions are evaluated using the testing set based on the selected result markers.

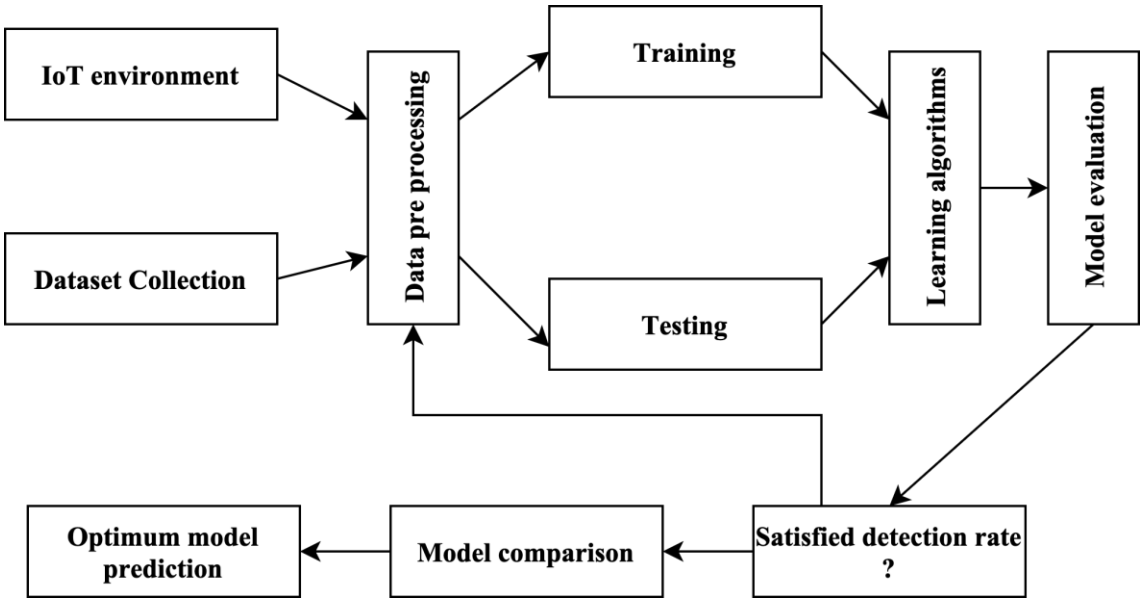


Fig. 1: Workflow of the Model

All ML studies necessitate comprehensive data evaluation to align with the training methods for achieving precise efficiency. Therefore, executing and transforming information from group parameters to algorithms is essential. Two essential pre-processing techniques are Numericalization and Standardization. Before execution, several procedures must be undertaken, including dataset cleansing, identifying missing data, and substituting 'NaN' (Not a Number) data.

Data Sanitization

The IoT Botnet Database 2020 and the IoTIDS20 database exhibit comparable types and quantities of characteristics, albeit with slightly differing attack classes. Vacant spaces are eliminated before standardizing these two records, irrelevant indices are discarded, and data kinds are transformed to the appropriate model (float) to prevent mistakes. These databases have eliminated 16 characteristics with missing data to ensure precise results.

Quantification

Before numericalization, the data kinds comprising each dataset's characteristics must be ascertained. In NSL-KDD, the 'Services,' 'Flag,' and 'Protocol_type' columns comprise three category nominal variables similar to DS2OS; all columns, save 'Value' and 'Timestamp,' consist of categorical nominal parameters. The IoTDevNet database lacks categorical values; however, the databases contain item values in two rows. Category parameters are transformed into numerical values by encoding. Label encryption maintains the original dataset's dimensionality while assigning identical values to repeated designations, resulting in minimal memory usage. The encoded label is utilized in all two databases to transform category characteristics into numerical values.

Standardization

Following numericalization, the Standard-Scaler technique is employed to normalize the numerical information across the relevant databases, namely the higher-range values. Standard-Scaler adheres to a typical distribution, normalizing the values by removing the average zero and dividing by the mean error of 1.

After this, ML algorithms are used to test the classification. The results are analyzed in the next section.

Results and Discussions

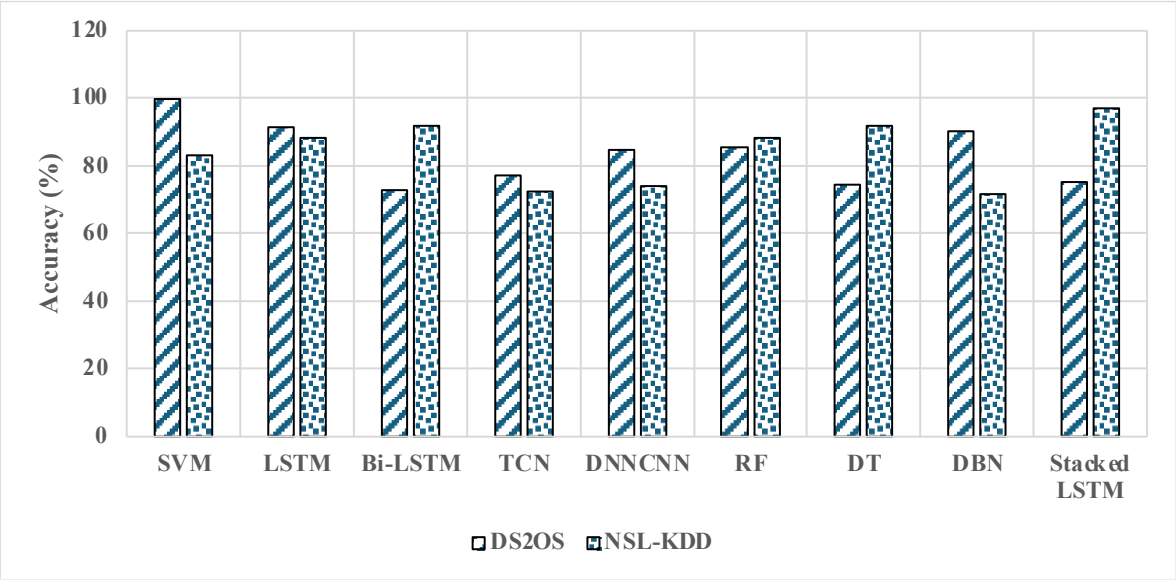


Fig. 2: Accuracy Analysis

The study uses an individual HP notebook with a 2.4 GHz Intel i5 chip. The simulations are implemented on the freely available Google Colab Notebooks. Shallow ML and DL methods have been built utilizing the Keras library with TensorFlow. Pandas and NumPy libraries are used to import and clean up data.

Fig. 2 facilitates an intuitive comparison of the evaluation metrics of the provided models against the previously established IDS systems, utilizing the different datasets. These pictures offer a comparative analysis of various learning models derived from significant studies represented by distinct color sets. The utilization of the DS2OS dataset in the examination of IDSs is very recent. This dataset's number of directed studies about network safety is limited. The study merits acknowledgment. This study evaluates various ML methods, revealing that the SVM achieved an accuracy of 94.1%, which is inferior to the trained SVM model's efficiency of 92.5%. The study evaluated SVM and LSTM, reaching 81.5% and 85.4%, respectively, whereas the analogous design of LSTM and Bi-LSTM attained 91.2% and 94.7%. Yet, a Convolutional Neural Networks (CNN) model has a greater accuracy of 93.4%, but the selected Deep Neural Network (DNN) achieves an anomaly detection accuracy of 88.4%.

The NSL-KDD database yielded an accuracy of 74.6% with decision forests and 75.7% with choice jungle, but the more straightforward decision tree and random Forest (RF) achieved 91.2% and 90.6%, respectively. This study evaluates DNN and LSTM-RNN, achieving accuracies of 81.4%, 91.2%, and 90.1%, respectively. Moreover, the study introduced a transmitted RNN-based method with an accuracy of 90.3%, similar to the suggested stacked LSTM design with an efficiency of 91.4%. The accuracies of 72.1%, 80.3%, 82.3%, and 76.2% were achieved using RFs, CNNs, and Bi-LSTMs, respectively, while the reported accuracies for RFs, LSTMs, and Bi-LSTMs are 90.3%, 91.6%, and 89.4%, respectively. The investigation achieved 95.2% accuracy with its suggested IDS, which is highly equivalent to the Bi-LSTMs.

Conclusion

The IoT is employed to facilitate communication between various application gadgets and clients. The connected architecture of IoT and the ability of gadgets to communicate with each other have heightened security concerns in these networks. A legal security mechanism for IoT gadgets and networks must be established. This study presents an approach to data analysis for identifying breaches in the IoT context. The research commences with an overview of the latest advancements in various systems for intrusion detection and a broad introduction to potential dangers in the IoT. The study delineates the intricacies of five data sets, including two established ones—NSL-KDD and DS2OS—and three relatively novel datasets. This paper examines three and five ML strategies for differentiating IoT threats in known and obscure environments. The framework addresses the challenges of deploying complex ML algorithms on low-capacity IoT devices, accurately identifies several threats with elevated precision and detection costs, and sustains a system for detection through regular updates for enhanced attack recognition. Based on the examination, Bi-LSTM

exhibits superior performance compared to the other examined DL algorithms for this specific study involving many datasets. It does not assure that, due to the extensive information and other ambiguous factors, the system will perform in this manner. Additional analysis is necessary on the issue, utilizing practical information and energy-time maximization. Micro services exhibit unique behaviors during various events, leading to deviations from standard operations in IoT services and thus resulting in inconsistencies. Additional study is necessary to comprehensively explain these concerns, potentially leading to developing a hybrid approach that includes various methodologies.

References

Madakam, S., Ramaswamy, R., & Tripathi, S. (2015). Internet of Things (IoT): A literature review. *Journal of computer and communications*, *3*(5), 164-173. https://doi.org/10.4236/jcc.2015.35021

Elrawy, M. F., Awad, A. I., & Hamed, H. F. (2018). Intrusion detection systems for IoT-based smart environments: a survey. *Journal of Cloud Computing*, *7*(1), 1-20. https://doi.org/10.1186/s13677-018-0123-6

Deshpande, P., Sharma, S. C., Peddoju, S. K., & Junaid, S. (2018). HIDS: A host based intrusion detection system for cloud computing environment. *International Journal of System Assurance Engineering and Management*, *9*, 567-576. https://doi.org/10.1007/s13198-014-0277-7

Iftikhar, Z., Javed, Y., Zaidi, S. Y. A., Shah, M. A., Iqbal Khan, Z., Mussadiq, S., & Abbasi, K. (2021). Privacy preservation in resource-constrained IoT devices using blockchain—A survey. *Electronics*, *10*(14), 1732. https://doi.org/10.3390/electronics10141732.

Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*, *171*, 1251-1260. https://doi.org/10.1016/j.procs.2020.04.133

Li, F., Zhang, M., Tian, B., Chen, B., Fu, G., & Ji, D. (2018). Recognizing irregular entities in biomedical text via deep neural networks. *Pattern Recognition Letters, 105*, 105-113. https://doi.org/10.1016/j.patrec.2017.06.009

Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1), e4150. https://doi.org/10.1002/ett.4150

Hasan, M., Islam, M. M., Zarif, M. I. I., & Hashem, M. M. A. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things, 7*, 100059. https://doi.org/10.1016/j.iot.2019.100059