International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Secure IoT Communication Through Machine Learning-Based Anomaly Detection

Y.Nagamalleswarao[1],K.Pavani[2],K.Keerthi[3]

#1 Assistant Professor in the Department of MCA,SRK Institute of Technology, Vijayawada

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: Cybersecurity threats and network vulnerabilities have expanded dramatically due to the Internet of Things' (IoT) fast expansion in smart homes, healthcare, transportation, and industrial automation. In IoT contexts, traditional security measures and signature-based intrusion detection systems frequently fail to detect novel and changing cyberthreats. This study proposes a machine learning-based anomaly detection and attack categorization framework for secure IoT networks in order to tackle these issues.

In order to identify anomalous behavior and categorize malicious activity, the suggested system continuously monitors IoT network traffic, preprocesses data, extracts features, and uses machine learning algorithms like Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Denial of Service (DoS), Distributed Denial of Service (DDoS), virus assaults, reconnaissance attacks, botnet operations, and unauthorized access attempts are just a few of the cyberattacks that the framework can detect.

Through clever categorization techniques, the system increases real-time threat analysis, reduces false positives, and improves detection accuracy. The suggested method offers dependable performance in differentiating between legitimate and malicious communication patterns in IoT settings, according to experimental study. The created framework provides a scalable, economical, and effective way to improve IoT security and shield linked devices from contemporary cyberthreats.

Index terms *- — IoT Security, Machine Learning, Anomaly Detection, Intrusion Detection System, Attack Classification, Random Forest, Support Vector Machine, K-Nearest Neighbors, Network Traffic Analysis, Cyber Threat Detection, Botnet Detection, DDoS Detection, Smart Devices Security.*

1. INTRODUCTION

IoT technology has connected billions of smart gadgets to the internet, revolutionizing communication and automation. IoT devices are utilized in healthcare, smart homes, transportation, agriculture, industrial automation, smart cities, and environmental monitoring. These gadgets communicate massive quantities of data and deliver intelligent services that boost productivity, efficiency, and user comfort. Although IoT devices are becoming more widespread, limited computing resources, inadequate authentication processes, unsecured communication protocols, and insufficient security setups have created cybersecurity issues.

Traditional firewalls, signature-based intrusion detection systems, and antivirus software use predetermined attack signatures to identify known threats. These solutions cannot stop current cyberattacks, zero-day vulnerabilities, and fast emerging IoT network threats. IoT vulnerabilities allow attackers to do DoS, DDoS, malware injection, reconnaissance attacks, botnet operations, unauthorized access, and data theft. IoT networks are varied and dynamic, making traditional threat detection challenging.

Machine Learning (ML) approaches can analyze large-scale network traffic data, find hidden patterns, and automatically recognize anomalous behaviors, making them strong cyber threat detection tools. Historical network data can help ML-based intrusion detection systems categorize network activity as legitimate or malicious with more accuracy and flexibility. Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) algorithms are frequently employed in cybersecurity because to their efficient classification performance and capacity to handle complicated datasets.

A machine learning-based anomaly detection and attack categorization framework for IoT devices is provided in this study to improve network security and threat detection. The system analyzes IoT network data, preprocesses and extracts features, and uses machine learning techniques to detect anomalies and categorize cyberattacks. The framework improves detection accuracy, reduces false positives, supports real-time threat analysis, and protects IoT infrastructures.

The suggested system provides automatic attack detection, intelligent traffic analysis, increased scalability, cost-effective deployment, and efficient processing of huge IoT datasets. Experimental results show that the suggested technique can detect normal and malicious traffic patterns and classify attacks reliably. The framework helps design safe, intelligent, and scalable IoT security systems that can protect against contemporary cyberthreats.

2. LITERATURE SURVEY

a) Anomaly detection: A survey:

Anomaly detection is a significant issue that has been studied in a variety of application domains and academic fields. While some anomaly detection methods are more general, several have been created especially for certain application areas. The goal of this survey is to present an organized and thorough summary of anomaly detection research. Based on the fundamental methodology used by each technique, we have classified current methods into several groups. We have determined the fundamental

presumptions for each category that the methods employ to distinguish between typical and abnormal behavior. These presumptions can serve as a basis for evaluating a technique's efficacy in a certain domain. We present a fundamental anomaly detection method for each category and then demonstrate how the many methods now in use in that category are variations of the fundamental method. This template makes it simpler and more concise to comprehend the methods that fall under each category. Additionally, we list the benefits and drawbacks of each category's strategies. Since the computational complexity of the methods is a significant problem in practical application fields, we also explore it. With the help of this survey, we seek to gain a better knowledge of the many approaches taken in this field of study and the ways in which methods created in one field may be used in other fields for which they were originally designed.

b) A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection:

This survey paper describes a focused literature survey of machine learning (ML) and data mining (DM) methods for cyber analytics in support of intrusion detection. Short tutorial descriptions of each ML/DM method are provided. Based on the number of citations or the relevance of an emerging method, papers representing each method were identified, read, and summarized. Because data are so important in ML/DM approaches, some well-known cyber data sets used in ML/DM are described. The complexity of ML/DM algorithms is addressed, discussion of challenges for using ML/DM for cyber security is presented, and some recommendations on when to use a given method are provided.

c) Intrusion detection system: A comprehensive review:

The research of intrusion detection systems (IDSs) has drawn a lot of interest from the computer science community due to the growing network traffic and security concern. Current intrusion detection systems (IDSs) present issues due to their enormous processing capacity as well as arbitrary incursion categories. We try to provide a more detailed picture for a thorough evaluation, even if there is a lot of previous literature on IDS concerns. We propose the taxonomy to describe contemporary IDSs by the comprehensive survey and sophisticated arrangement. Additionally, the content's summaries of tables and figures make it easier to understand IDSs as a whole.

d) Distributed attack detection scheme using deep learning approach for Internet of Things:

Because security breaches are becoming more frequent, cybersecurity is a major concern for all sectors of the internet. Due to the advent of several protocols, mostly from the Internet of Things (IoT), hundreds of zero-day attacks are believed to be constantly appearing. The majority of these assaults are minor variations on cyberattacks that have already been identified. This suggests that even sophisticated processes, like conventional machine learning systems, have trouble identifying these little attack mutations over time. However, the success of deep learning (DL) in a number of large data domains has sparked interest in cybersecurity. The advancements in CPU and neural network algorithms have made the implementation of DL feasible. Because of its high-level feature extraction capacity, the application of DL for cyberattack detection might be a robust

approach against tiny alterations or novel assaults. Deep learning architectures' self-taught and compression capabilities are crucial tools for identifying hidden patterns in training data and separating malicious traffic from legitimate traffic. The goal of this project is to use deep learning, a novel method to cybersecurity, to identify assaults in the social internet of things. Distributed attack detection is assessed against the centralized detection system, and the deep model's performance is contrasted with the conventional machine learning method. Our distributed attack detection system outperforms centralized detection methods that use deep learning models, according to the studies. Additionally, it has been shown that the deep model outperforms its shallow counterparts in terms of attack detection.

e) N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders:

IoT-based botnet assaults have increased due to the expansion of IoT devices, which are more readily infiltrated than desktop PCs. New techniques that identify attacks launched from hacked IoT devices and distinguish between IoT-based assaults lasting hours and milliseconds are required to lessen this danger. In this paper, we present N-BaIoT, a novel network-based anomaly detection technique for the Internet of Things that employs deep autoencoders to identify abnormal network traffic from hijacked IoT devices by extracting network behavior snapshots. We used two well-known IoT-based botnets, Mirai and BASHLITE, to infect nine commercial IoT devices in our lab in order to test our approach. The assessment findings showed that our suggested approach could precisely and quickly identify the assaults as they were being launched from the infected IoT devices that were a member of a botnet.

3. METHODOLOGY

i) Proposed Work:

The proposed system introduces an intelligent machine learning-based framework for anomaly detection and attack classification in IoT devices to enhance cybersecurity in modern IoT environments. The system continuously monitors network traffic generated by IoT devices and analyzes communication behavior to identify suspicious activities in real time. Initially, the collected network traffic data undergoes preprocessing operations such as data cleaning, normalization, feature extraction, and feature selection to improve data quality and increase machine learning performance. The processed dataset is then used to train machine learning algorithms including Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) for accurate anomaly detection and attack classification.

The proposed framework is designed to classify multiple cyber-attacks such as Denial of Service (DoS), Distributed Denial of Service (DDoS), malware attacks, reconnaissance attacks, botnet activities, and unauthorized access attempts. The trained models analyze incoming traffic patterns and automatically distinguish between normal and malicious network behavior with improved accuracy and reduced false positive rates. The system also generates attack alerts, prediction reports, and analytical results for efficient network monitoring and threat management. By integrating machine learning techniques with IoT security mechanisms, the proposed work provides a scalable, cost-effective,

intelligent, and real-time solution for protecting IoT infrastructures against evolving cyber threats.

ii) System Architecture:

The proposed system architecture is designed to provide intelligent anomaly detection and cyber-attack classification in IoT networks using machine learning techniques. Initially, IoT devices such as smart sensors, healthcare devices, industrial systems, and smart home appliances continuously generate network traffic data. The data collection module captures both real-time and stored traffic information from IoT environments. The collected data is then forwarded to the preprocessing module, where operations such as data cleaning, normalization, missing value handling, feature extraction, and feature selection are performed to improve dataset quality and enhance machine learning performance.

After preprocessing, the processed data is provided to the machine learning module, which uses algorithms such as Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) for anomaly detection and attack classification. The trained models analyze incoming network traffic and identify whether the traffic behavior is normal or malicious. If abnormal activity is detected, the attack classification module categorizes the attack into classes such as DoS, DDoS, malware attacks, reconnaissance attacks, botnet activities, and unauthorized access attempts. Finally, the prediction results, attack alerts, and analytical reports are stored in the database and displayed through a user-friendly web interface for efficient monitoring and IoT network security management.

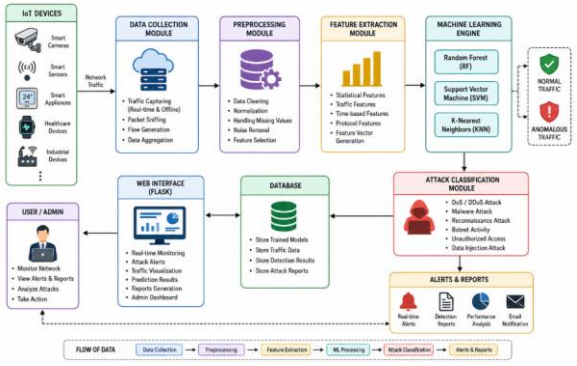

Fig1 proposed architecture

iii) Modules:

1. Data Collection Module

This module collects IoT network traffic data from smart devices, sensors, healthcare systems, industrial equipment, and other connected devices. It captures both real-time and stored traffic information required for anomaly detection and attack analysis.

2. Data Preprocessing Module

The preprocessing module cleans and prepares the collected network traffic data for machine learning analysis. It performs operations such as data cleaning, normalization, missing value handling, noise removal, and feature selection to improve data quality and model performance.

3. Feature Extraction Module

This module extracts important network traffic attributes and behavioral features from the processed dataset. The extracted features help machine learning algorithms identify patterns related to normal and malicious activities in IoT environments.

4. Machine Learning Module

The machine learning module trains and tests algorithms such as Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These models analyze IoT traffic patterns and classify network activities as normal or malicious with improved accuracy.

5. Anomaly Detection Module

This module continuously monitors incoming IoT traffic and identifies abnormal behavior or suspicious activities. It compares current network patterns with trained machine learning models to detect anomalies in real time.

6. Attack Classification Module

Once an anomaly is detected, this module classifies the attack into specific categories such as DoS, DDoS, malware attacks, reconnaissance attacks, botnet activity, and unauthorized access attempts. It helps administrators understand the type and severity of threats.

7. Database Management Module

The database module stores datasets, trained machine learning models, prediction results, attack logs, user details, and analytical reports. It supports efficient data management and retrieval for future analysis.

8. Alert and Report Generation Module

This module generates real-time alerts, attack notifications, prediction summaries, graphical analysis, and security reports. The generated reports help users and administrators monitor IoT network security effectively.

9. User Interface Module

The user interface module provides a web-based platform developed using Flask for monitoring network activities and viewing prediction results. It allows users to upload datasets, analyze attacks, monitor alerts, and generate reports through an interactive dashboard.

iv) Algorithms:

1. Random Forest Algorithm

Random Forest is a supervised ensemble machine learning algorithm used for classification and prediction tasks. It works by creating multiple decision trees using different subsets of training data and combines their outputs using majority voting. In the proposed IoT anomaly detection system, Random Forest analyzes network traffic features and predicts whether the traffic behavior is normal or malicious. The algorithm provides high accuracy, better handling of large datasets, reduced overfitting, and efficient attack classification performance in IoT environments. The Random Forest algorithm processes preprocessed IoT traffic data, generates multiple decision trees, and combines the prediction results from all trees to determine the final attack category. It effectively identifies attacks such as DoS, DDoS, malware attacks, reconnaissance attacks, and botnet activities by analyzing network behavior patterns. Due to its robustness and reliability, Random Forest is highly suitable for cyber security and intrusion detection applications.

2. K-Nearest Neighbors (KNN) Algorithm

K-Nearest Neighbors is a supervised machine learning algorithm used for classification based on

similarity and distance measurement. In the proposed IoT security system, KNN compares incoming network traffic data with previously trained traffic records and predicts the attack category based on the nearest neighboring data points. The KNN algorithm calculates the distance between new traffic records and stored training samples using distance measurement techniques such as Euclidean distance. The majority class among the nearest neighbors is selected as the final prediction result. KNN is simple, efficient, and suitable for identifying anomalous behavior in IoT traffic environments with improved classification accuracy.

3. Support Vector Machine (SVM)

Support Vector Machine is a supervised machine learning algorithm mainly used for classification and anomaly detection. SVM works by creating an optimal hyperplane that separates normal traffic data from malicious traffic data. In the proposed system, SVM analyzes IoT network traffic patterns and classifies whether the incoming traffic belongs to a normal category or an attack category with high precision and accuracy.The SVM algorithm identifies important traffic features from the dataset and maximizes the distance between different classes to improve classification performance. It is highly effective in handling complex and high-dimensional network traffic datasets. The algorithm helps detect cyber threats such as unauthorized access, malware attacks, and abnormal communication behavior in IoT networks.

4. EXPERIMENTAL RESULTS

The experimental evaluation of the proposed system was carried out using IoT network traffic datasets containing both normal and malicious traffic records. The system was implemented using Python, Flask framework, and machine learning libraries such as Scikit-Learn, NumPy, and Pandas. Machine learning algorithms including Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were trained using preprocessed IoT traffic data to perform anomaly detection and attack classification. The experimental results demonstrate that the proposed system successfully identifies abnormal traffic behavior and classifies various cyber-attacks such as DoS, DDoS, malware attacks, reconnaissance attacks, botnet activities, and data theft attacks with improved accuracy and reduced false positive rates.

The developed web-based application provides real-time anomaly detection, live input analysis, traffic monitoring, attack confidence evaluation, and risk reference analysis through an interactive user interface. Experimental outputs show that the system effectively analyzes network traffic samples, detects suspicious activities, and generates attack alerts along with confidence scores and traffic feature analysis. The system also supports manual live input analysis and reference-based risk prediction for identifying high-risk and low-risk network traffic conditions. The obtained results confirm that the proposed machine learning framework provides efficient, scalable, and intelligent security protection for IoT environments against modern cyber threats.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

$$\text{Accuracy} = TP + TN / (TP + TN + FP + FN)$$

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

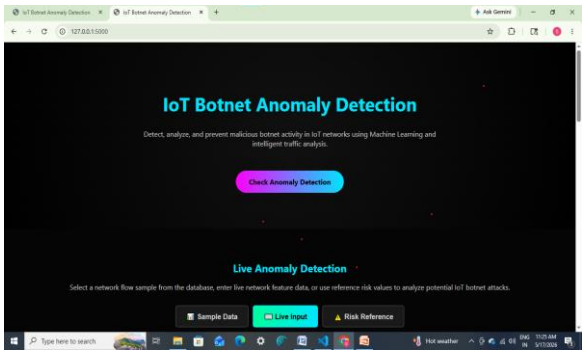

Fig2 Home Page of IoT Botnet Anomaly Detection System

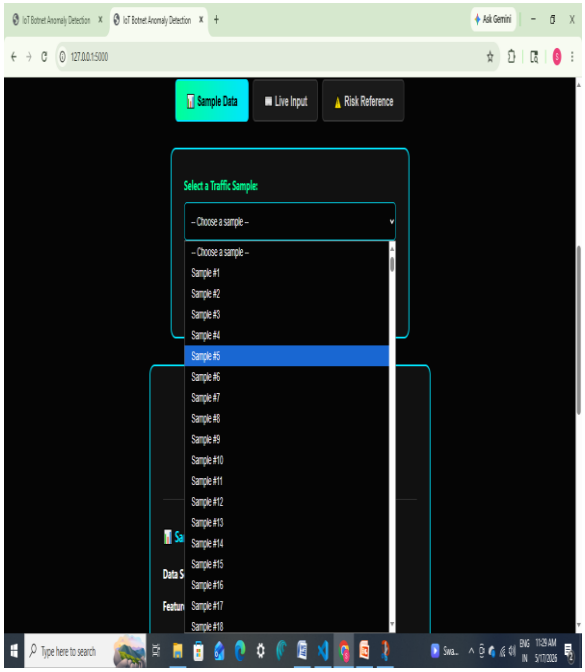

Fig3 Sample Data Selection Interface

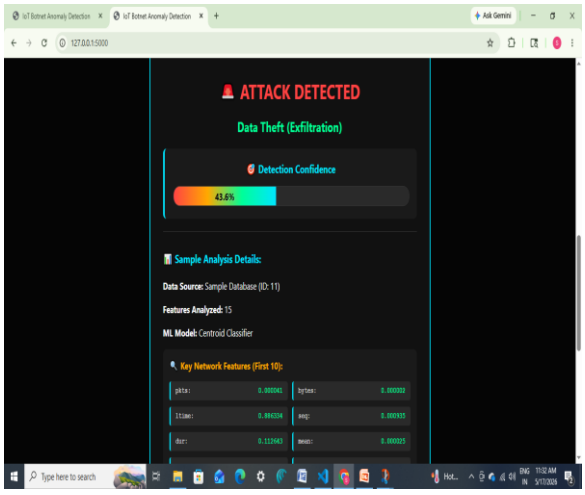

Fig4 Attack Detection Result

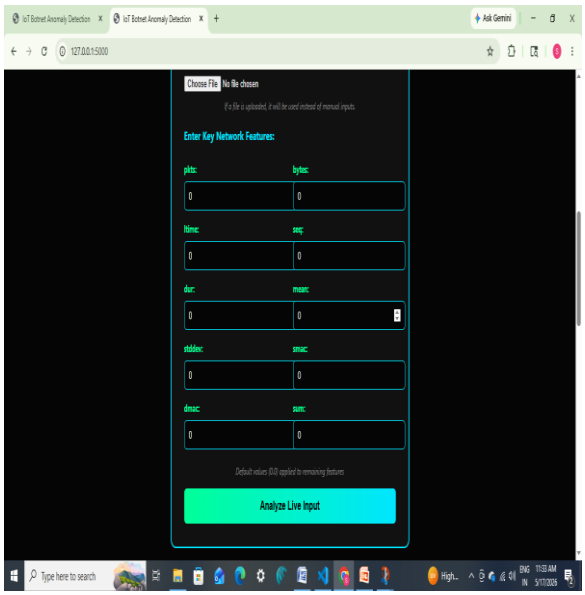

Fig5 Live Input Analysis Module

Fig6 High Risk Reference Analysis

5. CONCLUSION

The proposed system "Machine Learning Approach to Anomaly Detection and Attack Classification in IoT Devices" successfully provides an intelligent and efficient security framework for protecting IoT environments against modern cyber threats. The system utilizes machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) to analyze IoT network traffic, detect anomalous behavior, and classify various attack categories including DoS, DDoS, malware attacks, reconnaissance attacks, botnet activities, and unauthorized access attempts. Through effective preprocessing, feature extraction, and real-time traffic monitoring, the system improves attack detection accuracy while reducing false positive rates.

The experimental results demonstrate that the proposed framework can efficiently identify malicious network activities and generate reliable prediction results with confidence analysis. The developed web-based application supports real-time monitoring, live input analysis, alert generation, and attack reporting, making the system practical for IoT security management. The proposed machine learning-based approach offers scalability, flexibility, cost-effectiveness, and improved performance for securing IoT infrastructures. Therefore, the developed system contributes significantly toward building intelligent, automated, and reliable cybersecurity solutions for modern IoT networks.

6. FUTURE SCOPE

The proposed machine learning-based anomaly detection and attack classification system can be further enhanced by integrating advanced artificial intelligence and deep learning techniques for improving IoT security performance. Future improvements may include the implementation of Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and hybrid learning models for achieving higher detection accuracy and better identification of complex cyber-attacks. The system can also be extended to support real-time intrusion prevention mechanisms that automatically block suspicious traffic and prevent attacks before they affect IoT devices and networks.

In future developments, the proposed framework can be integrated with cloud computing, edge computing, and blockchain technologies to provide scalable, distributed, and secure IoT infrastructures. The system may also be enhanced to detect zero-day attacks, advanced persistent threats (APT), and encrypted malicious traffic using intelligent behavioral analysis techniques. Additionally, support for large-scale smart city environments, industrial IoT systems, healthcare monitoring networks, and autonomous devices can further improve the applicability of the proposed solution. Future versions of the system can include automated alert systems, mobile monitoring applications, and real-time visualization dashboards for efficient IoT security management and cyber threat analysis.

REFERENCES

1. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

2. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

3. Liao, H. J., Lin, C. H. R., Lin, Y. C., & Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 36(1), 16–24.

4. Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761–768. https://doi.org/10.1016/j.future.2017.08.043

5. Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. (2018). N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing, 17(3), 12–22. https://doi.org/10.1109/MPRV.2018.03367731

6. 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

7. Anthi, E., Williams, L., Słowińska, M., Theodorakopoulos, G., & Burnap, P. (2019). A supervised intrusion detection system for smart home IoT devices. IEEE Internet of Things Journal, 6(5), 9042–9053.

8. Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems, 100, 779–796.

9. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS), 1–6.

10. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, 1–6.

11. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. Proceedings of the Network and Distributed Systems Security (NDSS) Symposium. https://doi.org/10.14722/ndss.2018.23234

Author Profiles

Mr.Y.Nagamalleswarao Completed his Masters of Technology from JNTUK,MSC(IS) from ANU,BCA from ANU. He has System Administrator ,Networking Administrator and Oracle Administrator. He also a web developer and python developer, Currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest include Artificial Intelligence and Machine Learning.

Mrs. K. Pavani is working as an Assistant and Head of Department of MCA, in SRK Institute of technology in Vijayawada. She completed her MCA and M.Tech in Computer Science. She has 10 years of teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include AI and ML, etc.

Ms. K. Keerthi is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc.(computers) fromVagdevi Womens Degree College Jaggayapheta. Her area of interest are in Machine Learning with Python and Java.