INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 3, MARCH 2026 | ISSN: 2456-4184 | IJNRD.ORG

Threat Intelligent Intrusion Detection System Using Hybrid Random Forest Classification.

[1]Muthayee S, [2]Dinesh Babu J J, [3]Guruvishnukumar E R, [4]Sathiyajith R S

[1]Assistant Professor, Department of CSE (Cyber Security), KLN College Of Engineering, Sivagangai,
[2]BE.Student, CSE (Cyber Security),
[2]KLN College Of Engineering, Pottapalayam, Sivagangai, Tamil Nadu

Abstract : Cyberattacks are becoming more and more sophisticated, and this is causing significant challenges for contemporary computer systems and networks. This project proposes a Threat Classification-Based Intrusion Detection System (IDS) that combines real-time logging using System and Information Event Management (SIEM) software and a Random Forest-based machine learning approach for accurate threat detection and classification. System and network logging are used to collect relevant features for analysis using the Random Forest approach for accurate threat detection and classification. Threat detection and classification are performed using real-time logging and analysis using System and Information Event Management (SIEM) software and a Random Forest-based machine learning approach. Threat classification is used to identify whether the threat is normal or malicious and also to identify the threat level, which is classified as Low, Medium, and High depending on the threat severity. Threat detection and classification are verified using Cyber Threat Intelligence (CTI) to ensure accurate detection and classification. Dynamic alerts are used to notify users in real-time about critical security threats. Experimental results show that this proposed approach is accurate and efficient in reducing false alerts and also in real-world applicability.

Keyword - Intrusion Detection System, Random Forest, Cyber Threat Intelligence, SIEM, Machine Learning, Threat Classification

1. INTRODUCTION

Due to the growth in internet technology, organizations are using network technology to control and manage critical operations in their respective organizations. However, this rapid growth in technology has also increased the number of cyberattacks in our world. Hackers are always finding new and more sophisticated ways to attack networks using techniques such as brute force, malware, and distributed denial-of-service attacks.

An Intrusion Detection System (IDS) is used for detecting and monitoring activities in networks and systems for possible intrusions. Traditionally, most Intrusion Detection Systems used techniques such as rule-based detection and signature-based detection. Although these techniques are good for detecting and preventing intrusions, they are limited in detecting unknown threats and intrusions in networks and systems.

Machine learning techniques are also used for detecting intrusions in networks and systems. Machine learning techniques are used for detecting intrusions in networks and systems because they are able to detect unknown intrusions and threats in networks and systems. This research proposes a threat intelligent Intrusion Detection System using real-time SIEM log collection, Random Forest classification, and Cyber Threat Intelligence.

2. NEED OF THE STUDY

In the face of the ever-increasing cyber threats, traditional intrusion detection systems are becoming less effective. The majority of organizations currently employ signature-based intrusion detection systems, which can only recognize known intrusions. These systems need to be frequently updated, and it is not easy to incorporate new intrusion patterns.

The second problem with existing intrusion detection systems is the high rate of false positives. In most cases, security analysts spend a substantial amount of time investigating false alarms. This makes the entire system less efficient.
There is, therefore, a need to develop an intelligent intrusion detection system that will be able to parse the logs in real time, recognize abnormal behaviour using machine learning, and validate suspicious behaviour with cyber threat intelligence feeds.

2.1Population and Sample

The population of the proposed study will be network traffic and log data that include normal and malicious activities. For the proposed research, the CICIDS2017 dataset will be used as the primary source of data. This dataset will include normal network traffic and various types of cyber attacks such as brute force, denial of service (DoS), infiltration, and various types of web attacks. From the dataset, a subset of the data will be used to train and test the intrusion detection model. This data will be used to train the Random Forest model to recognize normal and malicious network traffic.

2.2 Data and Sources of Data

The data used for this research is taken from the publicly available dataset CICIDS2017. The dataset is commonly used for intrusion detection system research. The dataset is created by the Canadian Institute of Cybersecurity and consists of realistic network traffic that includes both normal and malicious activities. The dataset includes different types of attacks such as brute force attacks, denial of service (DoS), distributed denial of service (DDoS), port scanning, and infiltration attacks. The dataset includes labeled network flow feature information that can help in training the model effectively. Before training the model, preprocessing is done on the dataset to increase the efficiency of the Random Forest intrusion detection model.

2.3 Theoretical framework

The data used for this research is taken from the publicly available dataset CICIDS2017. The dataset is commonly used for intrusion detection system research. The dataset is created by the Canadian Institute of Cybersecurity and consists of realistic network traffic that includes both normal and malicious activities. The dataset includes different types of attacks such as brute force attacks, denial of service (DoS), distributed denial of service (DDoS), port scanning, and infiltration attacks. The dataset includes labeled network flow feature information that can help in training the model effectively. Before training the model, preprocessing is done on the dataset to increase the efficiency of the Random Forest intrusion detection model.

3 EXISTING SYSTEM

The existing intrusion detection systems are based on traditional machine learning techniques like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and rule-based and signature-based detection techniques. These techniques are effective in detecting known attacks but are not effective in detecting new attacks.

The major limitation associated with the existing intrusion detection systems is the false positive rate, as it increases the workload of the analyst and reduces the efficiency of the system. Furthermore, the existing intrusion detection systems lack real-time analysis and learning, and Cyber Threat Intelligence is only used as static data, i.e., IP addresses and domains.

The existing intrusion detection systems lack real-time analysis and learning, and Cyber Threat Intelligence is only used as static data, i.e., IP addresses and domains. These are the major limitations associated with the existing intrusion detection systems, and there is a need to develop a more adaptive and intelligent system.

4 PROPOSED SYSTEM

The proposed system includes the implementation of a Threat Intelligent Intrusion Detection System, which includes real-time log collection, machine learning, and Cyber Threat Intelligence (CTI) to increase the detection rate of threats. The log collection will be carried out using a SIEM tool like Wazuh from various sources.

The collected logs will be preprocessed to obtain significant features from the collected data. A Random Forest algorithm will be employed to categorize the collected data as normal and malicious, giving high accuracy and efficiency to the system.

The identified threats will be given severity levels like Low, Medium, and High to increase the accuracy and efficiency of the system. For increased accuracy and efficiency, the system will be integrated with Cyber Threat Intelligence (CTI) to validate the identified threats using known Indicators of Compromise.

This will increase the accuracy and efficiency of the system, reducing false positive rates. The system will also be able to generate real-time alerts and display the results in real-time, increasing the accuracy and efficiency of the system. The system will be able to learn new patterns and adapt to new threats, increasing its accuracy and efficiency in the current cybersecurity world.

5 RESEARCH METHODOLOGY

The proposed system, Dynamic IDS with CTI Integrated (DICI), combines real-time SIEM log collection and Random Forest ML for automated intrusion detection and threat classification.

5.1Data Collection

The first process in the proposed system is the collection of system or network logs. In this case, the system collects different types of logs from various Windows and Linux-based systems using Wazuh agents. The agents are responsible for collecting different activities from the systems, and they forward them to a centralized SIEM server.

In this case, the system collects different types of logs such as authentication logs, system logs, firewall logs, and file integrity monitoring logs. These types of logs enable the system to identify different suspicious activities.

5.2 Data Preprocessing

The data obtained from the sources may contain duplicate or unwanted data. So, the Pre-processing process takes care of cleaning the data.

The Pre-processing process includes the removal of duplicate data, removal of unwanted data, conversion of data into numerical data, and normalization of data. These processes help the machine learning model work more accurately.

5.3 Random Forest Classification

The machine learning algorithm used in this system is Random Forest, which is an ensemble machine learning technique that uses multiple decision trees to classify data with higher accuracy than a single decision tree.

In this system, important features such as IP address, login attempts, port numbers, and access to files are used as input for the machine learning algorithm, and then it classifies whether the incoming logs are normal or malicious.

5.4 Threat Classification

Once the Random Forest model detects suspicious activity, the system assigns a threat severity level based on the impact of the detected event. The threat levels include:

Low Threat – Minor suspicious activity with low impact

Medium Threat – Potential malicious activity requiring investigation

High Threat – Critical attack requiring immediate action

This classification helps security analysts prioritize incidents based on their severity.

5.5 Cyber Threat Intelligence Validation

To reduce false positive alerts, the system verifies suspicious indicators using Cyber Threat Intelligence sources. These sources contain information about known malicious IP addresses, domains, and file hashes.

If the detected event matches known indicators of compromise, the system confirms the attack and generates a security alert. This process improves the reliability of the intrusion detection system.

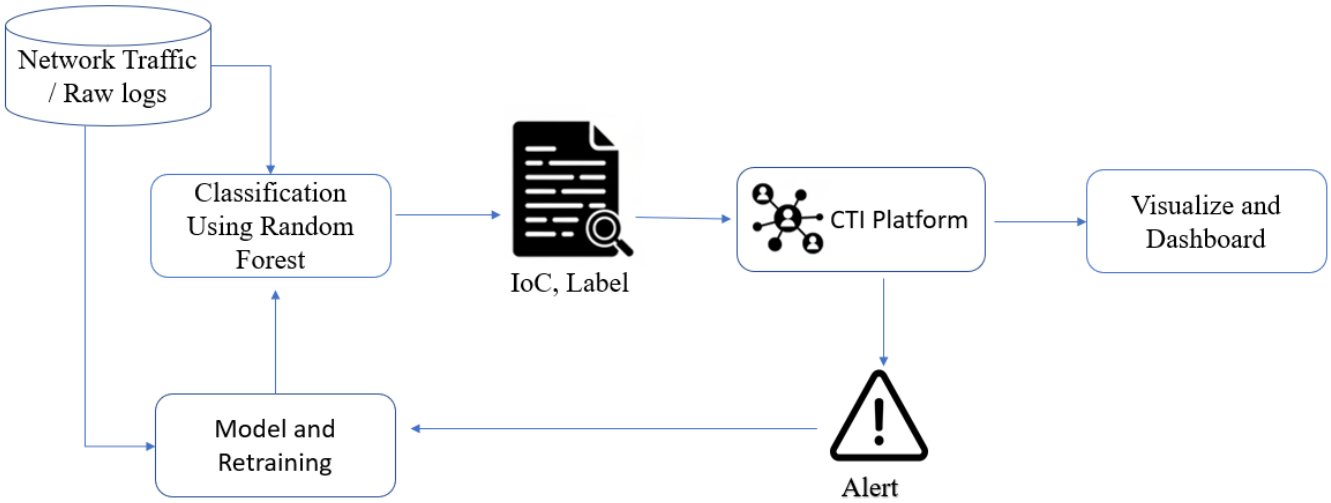


Figure (1)

6. SYSTEM ARCHITECTURE / MODEL DESIGN

The proposed intrusion detection system is developed based on a specific workflow to identify cyber threats using machine learning technology. The system incorporates the CICIDS2017 dataset as the main source of data. The dataset contains normal and malicious patterns. In the first step, the dataset is pre-processed to remove missing values, eliminate redundant values, and remove irrelevant features, which may affect the performance of the system.

The system incorporates the selected features using feature selection techniques to identify the most significant features of the network flow, which contribute to the detection of cyber-attacks. The features selected are used to train the Random Forest machine learning model, which can analyse complex patterns in the network traffic data.

The trained model analyses the network traffic and identifies normal or malicious patterns. If the system identifies a malicious pattern, it sends a notification with a potential cyber-attack alert. The system can efficiently monitor the network traffic and provide accurate and reliable results.

Workflow of the Proposed System
Wazuh Log → Dataset → Data Pre-processing → Feature Selection → Random Forest Model → Attack Detection → Alert Generation → Re-Train the Model

7. RESULTS AND DISCUSSION

The performance of the proposed intrusion detection system was tested using the **CICIDS2017 dataset.** The proposed Random Forest-based model was found to possess **high classification accuracy**, which was recorded at **96.8%.** This proves that the proposed model is efficient in distinguishing between normal and malicious network traffic. Furthermore, the proposed system was found to possess a **low False Acceptance Rate (FAR) of only 1.2%,** which proves that only a small number of attack instances are being misclassified as normal traffic.

4.1 Descriptive Statistics of Selected Network Features

Table 4.1: Descriptive Statics

Variable	Minimum	Maximum	Mean	Std. Deviation
Flow Duration	0	1199999	345210	215430
Total Forward Packets	1	1200	32.45	85.62
Total Backward Packets	0	980	28.13	70.54
Packet Length Mean	40	1514	456.32	210.45
Flow Bytes/s	0	850000	15420	48300

Table presents the descriptive statistics of selected network traffic features obtained from the CICIDS2017 dataset. The variable Flow Duration shows the time length of each network flow, with values ranging from 0 to 1,199,999 and a mean of 345,210. Total Forward Packets and Total Backward Packets represent the number of packets transmitted in forward and backward directions during communication. Packet Length Mean indicates the average size of packets within a flow, while Flow Bytes/s represents the rate of data transmission per second. The standard deviation values indicate the variability of these features. These statistical measures help understand traffic behavior and are used as input features for training the Random Forest intrusion detection model.

The proposed intrusion detection system was evaluated using the CICIDS2017 dataset. The Random Forest model was trained on the processed dataset and tested to measure its performance in detecting cyberattacks. The experimental results show that the model can accurately classify network traffic as normal or malicious based on the extracted features. Performance metrics such as accuracy and confusion matrix were used to evaluate the model. The results indicate that the Random Forest algorithm performs well in detecting multiple attack types present in the dataset. The use of machine learning significantly improves detection accuracy and reduces false positive alerts compared to traditional rule-based intrusion detection systems.

Figure 1: Feature Correlation Heatmap

The heatmap illustrates the correlation between different network flow features from the CICIDS2017 dataset, helping identify relationships between variables used in the Random Forest model.

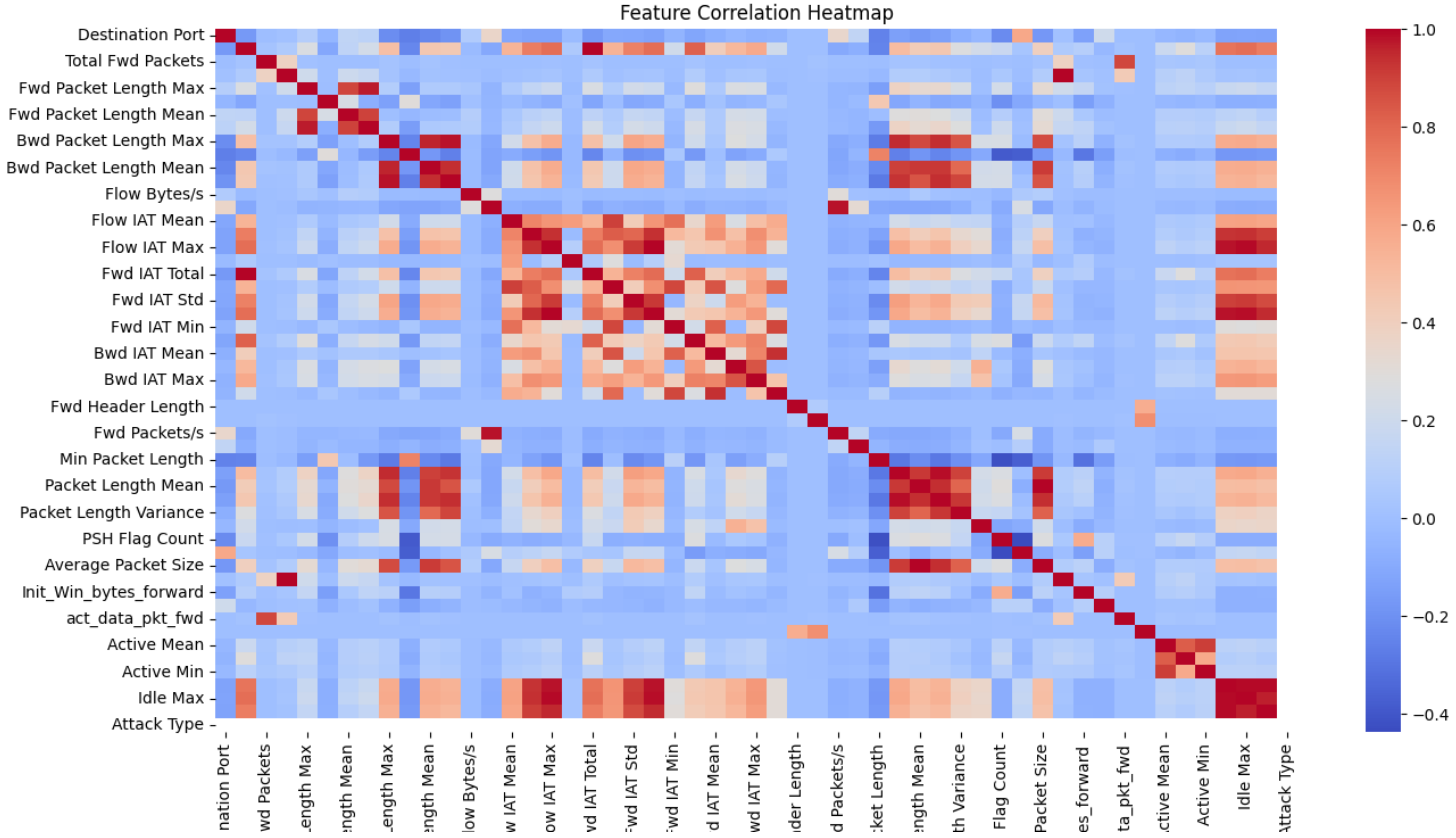


Figure 2: Normal vs Attack Logs Distribution

This graph shows the distribution of normal and attack traffic instances in the CICIDS2017 dataset used for training and evaluating the intrusion detection model.

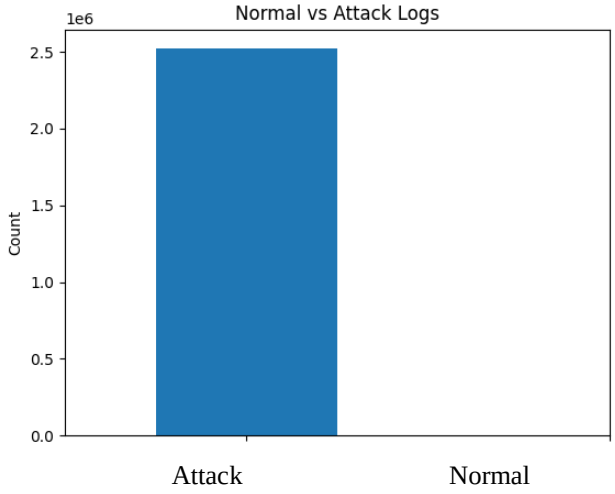
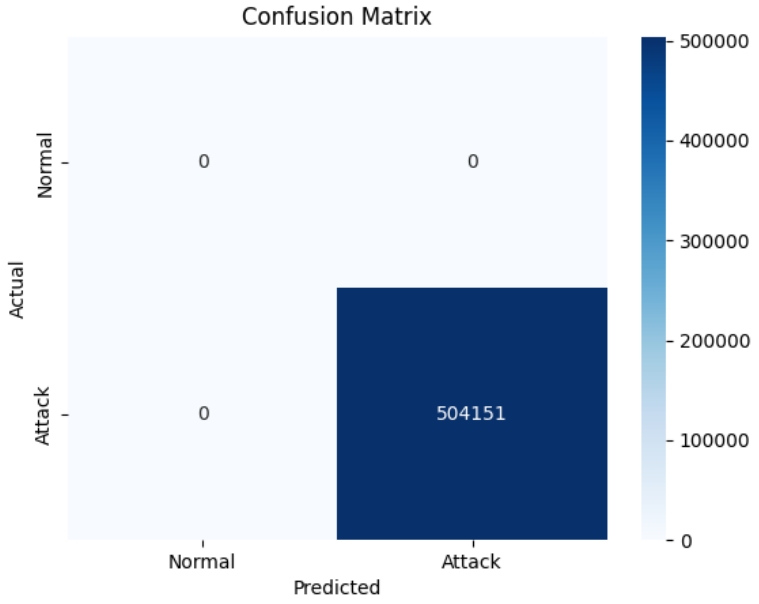


Figure 3: Confusion Matrix

The confusion matrix presents the performance of the Random Forest classifier by comparing actual and predicted network traffic classes.



ACKNOWLEDGMENT

The authors thank the Department of Cyber Security, KLN College of Engineering, for providing support and facilities for this research. The authors also express gratitude to their guide, Ms. S. Muthayee, and the Canadian Institute for Cybersecurity for the CICIDS2017 dataset used in this study.

REFERENCES

[1] L.Ying-dar, L.Yi-shin, L.H.Yuan-cheng, D.Sudyana and L.Andwei-bin, "Evolving ML-Based Intrusion Detection: Cyber Threat Intelligence for Dynamic Model Updates" *IEEE Access*, vol. 3, pp. 605–622, Apr 2025

[2] T.-T.-H. Le, H. Kim, H. Kang, and H. Kim, "Classification and explanation for intrusion detection system based on ensemble trees and SHAP method," *Sensors*, vol. 22, no. 3, p. 1154, Feb. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/1154

[3] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine learning and deep learning approaches for cybersecurity: A review," *IEEE Access*, vol. 10, pp. 19572–19585, 2022.

[4] R. Mills, A. K. Marnerides, M. Broadbent, and N. Race, "Practical intrusion detection of emerging threats," *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 1, pp. 582–600, Mar. 2022.

[5] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 1, pp. 41–50, Feb. 2018.

[6] M. Masdari and M. Jalali, "A survey and taxonomy of DoS attacks in cloud computing," *Secur. Commun. Netw.*, vol. 9, no. 16, pp. 3724–3751, Nov. 2016.

[7] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2046–2069, Mar. 2013.

[**8**] M. Masdari, S. Ahmadzadeh, and M. Bidaki, "Key management in wireless body area network: Challenges and issues," *J. Netw. Comput. Appl.*, vol. 91, pp. 36–51, Aug. 2017.

[9] T. Hayajneh, G. Almashaqbeh, S. Ullah, and A. V. Vasilakos, "A survey of wireless technologies coexistence in WBAN: Analysis and open research issues," *Wireless Netw.*, vol. 20, no. 8, pp. 2165–2199, 2014.

[10] M. Masdari and S. Ahmadzadeh, "Comprehensive analysis of the authentication methods in wireless body area networks," *Secur. Commun. Netw.*, vol. 9, no. 17, pp. 4777–4803, Nov. 2016.

[11] M. Masdari and S. Ahmadzadeh, "A survey and taxonomy of the authentication schemes in telecare medicine information systems," *J. Netw. Comput. Appl.*, vol. 87, pp. 1–19, Jun. 2017.

[12] Y. Sun, F. Lo, and B. Lo, "Security and privacy for the internet of medical things enabled healthcare systems: A survey," *IEEE Access*, vol. 7, pp. 183339–183355, 2019.

[13] J. Qi, P. Yang, G. Min, O. Amft, F. Dong, and L. Xu, "Advanced Internet of Things for personalised healthcare systems: A survey," *Pervasive Mobile Comput.*, vol. 41, pp. 132–149, Oct. 2017.

[14] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput. (MCC)*, 2012, pp. 13–16.

[15] M. Masdari and A. Khoshnevis, "A survey and classification of the workload forecasting methods in cloud computing," *Cluster Comput.*, vol. 23, no. 4, pp. 2399–2424, Dec. 2020.

[16] S. Iqbal, M. L. M. Kiah, B. Dhaghighi, M. Hussain, S. Khan, M. K. Khan, and K.-K. R. Choo, "On cloud security attacks: A taxonomy and intrusion detection and prevention as a service," *J. Netw. Comput. Appl.*, vol. 74, pp. 98–120, Oct. 2016.