International Journal of Applied Engineering & Technology

TOWARDS CLASSIFICATION OF CYBER-ATTACKS IN APPLICATION LEVEL FOR INTERNET OF THINGS DEVICES

Amit Kumar Harichandan[1], Biswajit Brahma[2] and Minati Mishra[3]
[1]Research Scholar, Fakir Mohan University, Balasore, Odisha, India
[2]Senior Data & Visualization Engineer, McKesson Corporation
[3]Assistant Professor, Department of Computer Science, Fakir Mohan University, Balasore , Odisha, India
[2]Biswajit.Brahma@gmail.com

ABSTRACT

Malicious attacks on Internet of Things (IoT) devices and network infrastructure have surged dramatically in response to the widespread use of IoT-based platforms. This spike has resulted in information theft and vulnerabilities due to unusual traffic patterns throughout the network, which has partially or totally disrupted regular network operations. As a result, it becomes essential for service providers and network managers to keep an eye on the types of devices connected to their networks as well as the traffic that flows through them. Accordingly, the current work, which focuses on application level security, presents an ensemble-based method to describe traffic patterns unique to IoT systems.To detect anomalous traffic patterns, the suggested system integrates the Naïve Bayes (NB) and Adaptive Boosting (AdaBoost) algorithms. Additionally, this suggested algorithm's performance is contrasted with that of conventional learning models. According to experimental data, the suggested method for detecting abnormal traffic inside the IoT network achieves an impressive prediction accuracy of 96.324%, with a sensitivity value of 0.928, Positive Prediction Value (PPV) of 0.988, and an F-score of 0.949. Furthermore, the suggested approach has an exceptionally high AUC-ROC (Area Under the Receiver Operating Characteristic) value of 0.986. Thus, it is clear that the suggested method has potential for creating a successful Intrusion Detection System (IDS) and controlling traffic flow patterns in large-scale heterogeneous networks with an emphasis on application level security.

Keywords:Cyber security, Machine learning,Attack Identification,Internet of Things, Ensemble learning

1. INTRODUCTION

The seamless connection between people and sensory equipment has been changed by the introduction of the Internet of Things (IoT), especially in application areas including smart homes, cities, healthcare services, and businesses. In order to enable information sharing and archiving between IoT devices and intelligent software applications, these IoT procedures require a high degree of interconnection. IoT devices are typically defined as low-power sensor/actuator-based entities that communicate independently with application servers inside the network as well as with other devices. The purpose of this link is data and information relaying between huge autonomous networks and specialized application servers.

The increasing intricacy of Internet of Things applications has presented difficulties for network traffic control, which might expose networks to hostile attacks. The spread of Internet of Things applications has made it more challenging for network operators and service providers to fully capture device and network characteristics. The necessity of addressing vulnerabilities and unusual traffic patterns resulting from different attackers or compromised devices emphasizes the importance of strong application level security measures even more.

In this situation, network administrators are essential for tracking down devices inside a network and analyzing data flow to differentiate between malicious and legitimate activity. In the context of application level security, in particular, the prompt identification and prevention of such threats necessitates the implementation of a strong intrusion detection system (IDS).

A thorough grasp of application level security, with an emphasis on the nuances of data transfers, communication protocols, and potential vulnerabilities at this layer, is crucial to navigating these issues successfully. In conclusion, protecting IoT environments requires a thorough application level security strategy that emphasizes the necessity of attentive network management and the installation of cutting-edge intrusion detection systems to neutralize possible threats.

Several research works have investigated the use of various machine learning methods to successfully detect cyberattacks as well as compromised devices that initiate these attacks [1-3]. Modern learning techniques have demonstrated effective results in assault detection and model training. Careful tracking of every linked Internet of Things (IoT) device becomes essential for efficient network administration in the context of large-scale IoT systems. Maintaining control over the kind and quantity of traffic produced by these devices is essential to avoiding a decline in network performance.

To improve the evaluation of the underlying network and IoT devices, several frameworks have been developed to solve these issues in IoT platforms [4–9]. But current methods have shortcomings, especially when it comes to managing the dynamic nature of IoT ecosystems. The complex properties of these networks make it difficult for conventional learning algorithms to distinguish between abnormal and typical traffic flow behaviors.

With an emphasis on application level security, this study presents an effective cooperative ensemble learning system designed towards network traffic classification in Internet of Things scenarios. In order to identify network activity from a varied gathering of traffic patterns and the suggested architecture combines the Naïve Bayes algorithm with the dynamic ensemble technique of adaptive boosting (AdaBoost). After conducting experiments, the suggested method characterizes aberrant traffic flow behavior with a notable 96.334% prediction accuracy, 0.928 sensitivity, 0.949 F-score, and 0.988 positive predictive value (PPV). The ensemble-based technique is superior to traditional learning algorithms in identifying network traffic flows, as demonstrated by comparison analyses with Random forest, Naïve Bayes, Bayesian network (BayesNet), and classification and regression tree (CART) methods. This is especially important for Internet of Things environments that have a variety of traffic characteristics.In conclusion, by effectively categorizing network traffic in Internet of Things contexts, the suggested ensemble-based approach shows potential for improving application level security. The results point to its possible use in the creation of automated systems for detecting network intrusions in IoT applications involving monitoring of traffic in real time.

2. RELATED WORK

In this section, wegive a thorough review of important studies in the field of machine learning methodologies with implications for IoT application security in the context of network traffic categorization for Internet of Things systems. A semi-supervised fuzzy clustering technique was presented by Glennan et al. [10] with the goal of labeling various traffic types, with a focus on anonymous application flows. To improve classification performance, their architecture combines fuzzy cluster labeling and feature selection methods. A remarkable accuracy of 95.10% was found in the study, indicating significant advancements in all courses.

Due to the usage of real data in attack packets, Distributed Denial of Service (DDoS) amplification assaults provide a challenge to conventional detection and isolation techniques. Khan et al. [11] suggested a method for detecting these attacks that is based on artificial neural networks (ANNs). The authors evaluated packet flow complexity by computing the Lyapunov exponent, which gave them insights into the characteristics of the traffic. They used a standard ANN algorithm as well as a suggested chaotic neural network method to classify network traffic, with the chaotic model reaching an amazing 98% classification accuracy.

These studies highlight how important sophisticated machine learning techniques are for handling complex problems with network traffic categorization in Internet of Things systems, especially when it comes to application level security. The approaches demonstrated demonstrate how neural network-based methods and fuzzy clustering may effectively categorize various traffic patterns, which supports the continuous endeavors to strengthen the security environment of Internet of Things systems.

Perera et al. [12] carried out a thorough investigation on the automated classification of network traffic through the use of supervised machine learning models. The study was centered on network Quality of Service (QoS) management according to the underlying model's performance in the IoT-based communication network. There were two different feature selection methods used: relief attribute evaluation and correlation-based feature selection. To identify network traffic, sixmachine learning models wereused:Naïve Bayes (NB), Decision Trees (DT), Bayes Network (BayesNet), NB-Tree, Random Forest (RF), and Multilayer Perceptron (MLP). According to the investigation, when it came to classifying network traffic, RF and DT performed well, offering high accuracy and minimal computational complexity.

A framework designed exclusively for threat analysis of privacy concerns in Android-based IoT apps was introduced by Sharma and Gupta [13]. The framework tackled the issues of malware propagation and covert assaults, given that Android applications are vulnerable to attacks because of their extensive availability and subpar software infrastructures. Three sets of data - Drebin, M0-Droid, and AndroZoo—were used for experimental studies in order to distinguish between malicious and safe apps. The assessment parameters were F-score, accuracy, specificity, and sensitivity that the model fulfilled, with 92.98% peak accuracy and 92.25% F-score. The suggested machine learning method showed effectiveness in locating and examining Android-based malware's permissions.

A system for categorizing various IoT security threats by examining the properties of network traffic was given by Rathore and Park [14]. Their method featured a centrally located attack detection mechanism that was able to gather and analyze data in order to detect different types of Internets of Things security assaults. For the purpose of identifying IoT device assaults in fog computing settings, the authors suggested combining supervised machine learning methods with a semi-supervised learning strategy. The suggested plan was created to handle issues with unlabeled datasets and satisfy the majority of Internet of Things (IoT) device Quality of Service (QoS) demands, for example, lower latency, scalability, resource constraints, and ubiquity. The suggested method detected IoT security threats with an accuracy of 86.53% using the NSL-KDD dataset for assessments.

Kozik et al. [15] presented a novel method for automating the identification of cyberattacks on Internet of Things (IoT) devices by utilizing Extreme Learning Machines (ELMs) in a cloud computing setting. The framework leverages the powerful features that come with cloud platforms to analyze network data in-depth and classify threats efficiently. This method differs from edge computing platforms with limited resources, where it might be difficult to create and maintain detection models for training massive amounts of traffic data. In order to overcome this constraint, Kozik et al.'s work recommends using pre-built ELM models that are hosted in the cloud, which makes it possible to handle large traffic datasets effectively. Their results highlight the applicability of this approach in supporting application level security in Internet of Things environments by showing that it can be simply built for edge node-corresponding aggregated traffic analysis and classification and is well-suited for conducting comprehensive network inspections.

3. PROPOSED FRAMEWORK

This study presents an approach that combines a statistical learning model, the Naïve Bayes (NB) algorithm, with an ensemble learning algorithm, the AdaBoost algorithm. The architecture of the suggested framework is shown in Figure 1, whereby specialized network traffic collecting technologies are first used to gather data on IoT network traffic across various IoT networks [16]. These data sets are then used to train the model in the attack detection framework, which is comprised of the ensemble learning module and the AdaBoost and NB algorithms. After being exposed to labeled instances from test data during the training phase, the model is deployed into the system to conduct a thorough study of several performance criteria.

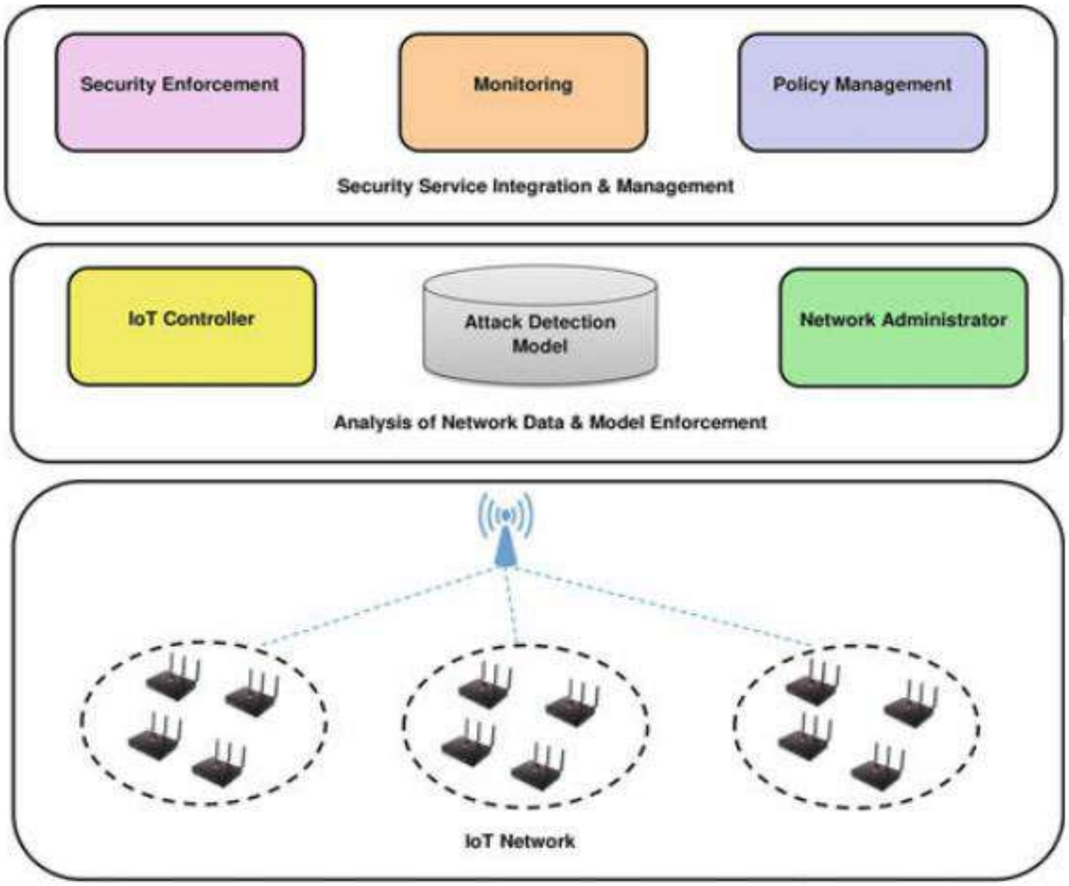

Figure 1: The proposed high-level architecture for attack detection.

The model is the basis for an Intrusion Detection System (IDS) that detects malicious network activity based on the predicted results from the test data. Important responsibilities in this process are played by the IoT controller module and network administrator, who actively detect and identify devices bymanaging the quantity of IoT devices in the network and stop malicious traffic. These interpretations are essential characteristics for putting security enforcement mechanisms in place inside the Internet of Things domain, guaranteeing compliance with network standards.

The next sub-section gives a thorough overview of the history of each method used to create the suggested framework, namely the NB and AdaBoost algorithms. Additionally, the suggested ensemble-based algorithm's performance is methodically compared to traditional machine learning models like Random Forest (RF), Classification and Regression Tree (CART), BayesNet, and Naïve Bayes, highlighting the algorithm's applicability and potency in bolstering application level security in Internet of Things environments.

3.1 AdaBoost Algorithm

In this research, we build a group of learners employing the popular ensemble learning method of adaptive boosting (AdaBoost), to extract insights from datasets [17, 18]. To optimize overall prediction outcomes, this strategy entails training these learners and combining their combined prediction accuracy. One base learner or more base learners can be used with the AdaBoost algorithm. Here, the model is trained using a single base learner, the statistical learning method known as the Naïve Bayes (NB) algorithm. Here, the AdaBoost method works with a group of homogeneous learners that are in line with the NB algorithm, and it improves their performance by producing precise prediction results.

When it comes to application level security, the AdaBoost algorithm plays a crucial role in maximizing the identification of possible security risks in Internet of Things environments. The AdaBoost technique guarantees a targeted and reliable learning process that is adapted to the subtleties of statistical patterns pertinent to security in IoT data by using a single base learner, the NB algorithm. By combining statistical methods and ensemble learning, this framework for application level security becomes more complex and reliable, increasing the precision and dependability of security forecasts in Internet of Things settings.

Looking at a binary categorization issue currently, the AdaBoost method may be represented as follows in a weighted additive composite formulation for ith base learners:

$$F(x) = \text{sign}^{\left(\sum_i \alpha_i k_i(x)\right)} \quad (1)$$

The linear composite coefficient for the ith base learner, ki(x), applied to the sample data x is represented by the symbol αi in this case. αi is larger than 0 and i goes from 1 to n.

3.2 Naïve Bayes Algorithm

Assuming independence among all features in a given dataset, the Naïve Bayes (NB) algorithm is well-suited for classification tasks [19]. By utilizing Bayes' rule, this algorithm establishes class labels connected by a maximal likelihood hypothesis to a dataset [20]. With regard to application level security in IoT networks, the NB algorithm calculates the probability for every attribute in a given dataset as follows:

$$P\,(C_i|a_1, a_2, \ldots, a_j) \quad (2)$$

Here, Ci stands for the ith class's variables in the context of application level security, where i is a number between 1 and n. Furthermore, each class's attributes are indicated by aj, which is a member of attribute set A and represents j values ranging from 1 to m. The Naïve Bayes algorithm's essential functionality is captured in this form, which is especially pertinent for problems involving application level security.

Consequently, the conditional probability for Eq. (2) in the context of application level security is obtained by applying Bayes' rule as follows:

$$P(C_i|A) = \frac{P(C_i)P(A|C_i)}{P(A)} \quad (3)$$

According to the application level security context [19], the joint probability distribution for A characteristics across Ci is obtained as follows:

$$P(A|C_i) = \prod_{j=1}^{m} P(a_j|C_i) \quad (4)$$

By maximizing Equation (4), the Maximum A Posteriori (MAP) function is calculated, which may be interpreted as follows in the context of application level security, to predict the behavior of network traffic inside a specific dataset:

$$P(C|A) = {}^{arg} \max_{i \in \{1,2,\ldots,n\}} P\,(C_i) \prod_{j=1}^{m} P(a_j|C_i) \quad (5)$$

$\prod_{j=1}^{m} P(a_j|C_i)$serves as a weighted multiplicative combination of the attribute set A corresponding to the Ci classes. It is the joint probability distribution as determined in Eq. (4). This concept is especially important when thinking about security at the application level.

4. EXPERIMENTAL OUTCOMES

Here, we outline the performance measures that are used to evaluate the suggested ensemble learning method's effectiveness in network traffic classification, with a focus on application level security. We then provide the experimental findings that come from our study, with a particular emphasis on forecasting typical and atypical traffic patterns in Internet of Things situations. To determine how well the suggested method works to improve security controls at the application level of Internet of Things systems, an evaluation is carried out by comparing it with well-established machine learning models.

4.1 Performance Evaluation

We carried out an extensive performance evaluation to determine the efficacy of our suggested ensemble learning framework in the field of application level security. For every model, we calculated important parameters including sensitivity, prediction accuracy, F-score, positive predictive value (PPV), and created ROC curves. The ratio of true positives (t_p) to the total of true positives and false negatives (f_n) is used to compute sensitivity, sometimes referred to as recall. Regarding application level security, t_p denotes the correct detection of anomalous traffic in the sample data, whereas f_n denotes the cases in which the learning model incorrectly identified aberrant traffic flows. The following is the expression for the sensitivity value for a certain model:

$$\text{Sensitivity} = \frac{t_p}{(t_p+f_n)} \qquad (6)$$

In application level security for IoT devices, Positive Predictive Value (PPV) refers to the precise detection of assaults on traffic patterns throughout the whole sample space that have been marked as abnormal. The ratio of true positives (tp) to the total of true positives plus false positives (fp), or (tp + fp), is the mathematical expression for positive predictive value (PPV). The PPV value provides a sophisticated indicator of the application level security efficiency and is essential for comprehending how well the learning model identified threats within the observed anomalous traffic. It's provided as:

$$\text{PPV} = \frac{t_p}{(t_p+f_P)} \qquad (7)$$

The term "false positive values" (fp) in this context refers to situations in which the model mistakenly classifies regular traffic as abnormal. The harmonic mean of the Positive Predictive Value (PPV) and the values of sensitivity, as indicatedby Equations(6) and (7), is used to calculate the F-score, an important metric for assessing a model's predictive competency in the context of application level security. This all-inclusive measure encompasses recall as well as accuracy, providing a sophisticated evaluation of the model's ability to accurately detect and categorize aberrant activity in the network. As a result, the F-score for a learning model may be found here:

$$\text{F-score} = 2 \times \left(\frac{PPV \times Sensitivity}{PPV + Sensitivity}\right) \qquad (8)$$

When it comes to application level security, a model's accuracy is defined as the percentage of correctly categorized samples (both typical and unusual traffic kinds combined) throughout the full sample space. The following formula is used to determine a model's accuracy, given as a percentage:

$$ACC = \frac{t_p+t_n}{(t_p+t_n+f_p+f_n)} \times 100 \qquad (9)$$

Furthermore, ROC curves are useful measures of how well a learning model predicts the future, particularly when it comes to particular thresholds. A prediction model is shown by ROC curves near 1 that is operating at its best. Plotted against the true positive rate (TPR) and false positive rate (FPR) are these curves that represent probability curves for a specific prediction model [21–23]. To illustrate the effectiveness of our suggested ensemble-based framework in identifying network traffic flows, we show a comparison study of ROC curves versus popular machine learning models, such as RF, CART, NB, and BayesNet. Furthermore, each ROC curve's Area Under the Curve (AUC) gives a succinct overview of threshold values, which helps distinguish between anomalous and typical traffic classes. AUC-ROC values near a model that performs better at the application level security space, from 1, and a high value is normally desired.

4.2 Results and Discussions

Here, we provide the results of our traffic analysis utilizing the NSL-KDD dataset, with an emphasis on the application level security performance indicators that were previously mentioned. The dataset includes four distinct forms of network attacks: probing, Denial of Service (DoS), user to root (U2R), and remote to local (R2L). The occurrences are divided into normal and anomalous categories despite the variety of assault techniques, allowing for a binary classification of the information. With 125,973 instances, 90% of the sample is made up of the training data as a whole, whereas the testing data has 22,544occurrences.Renowned for its ability to detect harmful traffic patterns and record sophisticated Internet traffic behaviors, the NSL-KDD dataset [24].

The effectiveness between the conventional learning methods and the proposed ensemble learning framework is demonstrated in Fig. 2a, with a particular emphasis on the characterisation of typical network traffic patterns. F-score, Positive Predictive Value (PPV), and sensitivity are among the performance measures taken into account. The ensemble-based AdaBoost-NB method performs better than the other algorithms, with notable results including a sensitivity of 0.989, a PPV of 0.939, and an F-score of 0.957. A thorough summary of the performance metrics for each method under consideration is given inTable 1, where bolded findings represent the best outcomes.

The effectiveness of the algorithms in recognizing abnormal traffic flows is seen in Fig. 2b. The suggested ensemble-based method outperforms traditional learning algorithms in identifying harmful traffic patterns, producing an impressive PPV of 0.988, a sensitivity of 0.928, and an F-score of 0.949. Interestingly, a high PPV value of 0.967 is also demonstrated using the Random Forest (RF) method. The importance of PPV in detecting anomalous traffic flows must be emphasized, as it represents the correctly recognized malicious flows in the whole sample that has been classified as harmful in a certain sample region. Attaining a genuine detection rate of abnormal flows over the full sample space is contingent upon the high sensitivity value. Table 2 presents a detailed comparison of the suggested (AdaBoost-NB) ensemble learning's performance metrics with traditional learning techniques, with a focus on application level security characterization of aberrant traffic flow behavior.

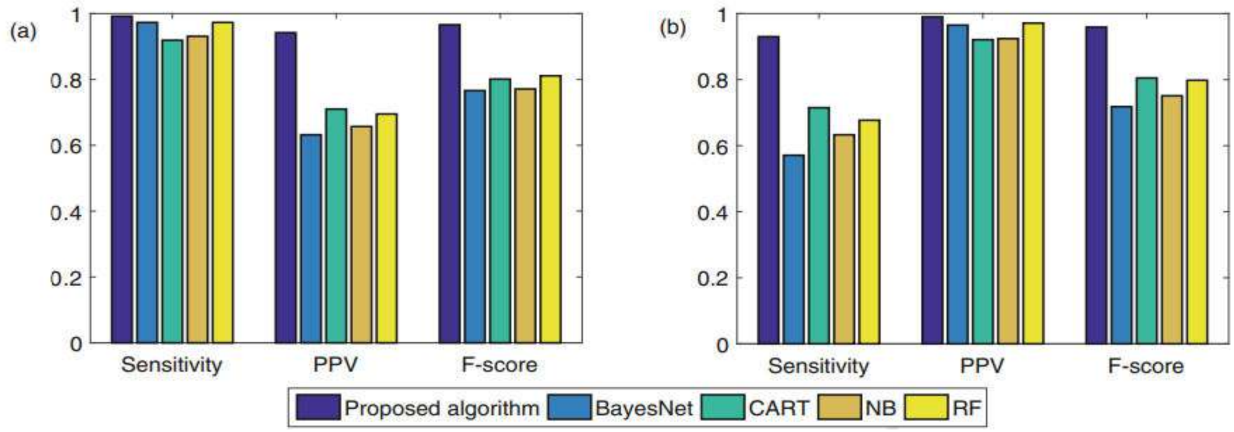

Figure 2: (a) Performance of the proposed algorithm compared with different machine learning algorithms for normal traffic. (b) Performance of the proposed algorithm compared with different machine learning algorithms for abnormal traffic.

Fig. 3 displays the prediction accuracy for each learning strategy, where the remarkable accuracy of 96.324% for the suggested ensemble-based technique is highlighted by the noteworthy findings.Remarkably, the Random Forest (RF) algorithm comesin second with the greatest prediction accuracy at 80.448%; nevertheless, its accuracy is significantly less than that of the suggested method for characterizing network traffic. The ROC curves shown in Fig. 4 with their sharp slope that converges to 1 further highlight the better performance of the suggested approach. The ensemble-based algorithm's excellent predictive potential is demonstrated by this convergence.

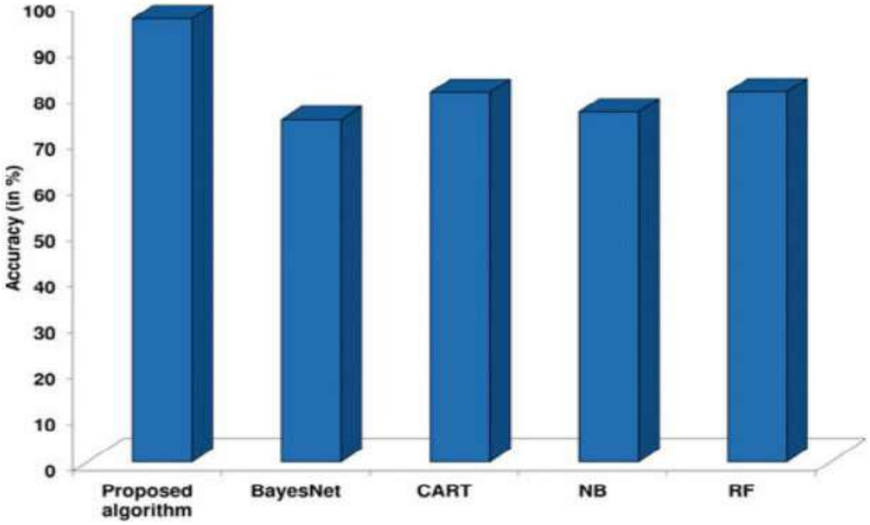

Figure 3: Comparison plot for accuracy pertaining to the proposed algorithm compared with different machine learning algorithms.

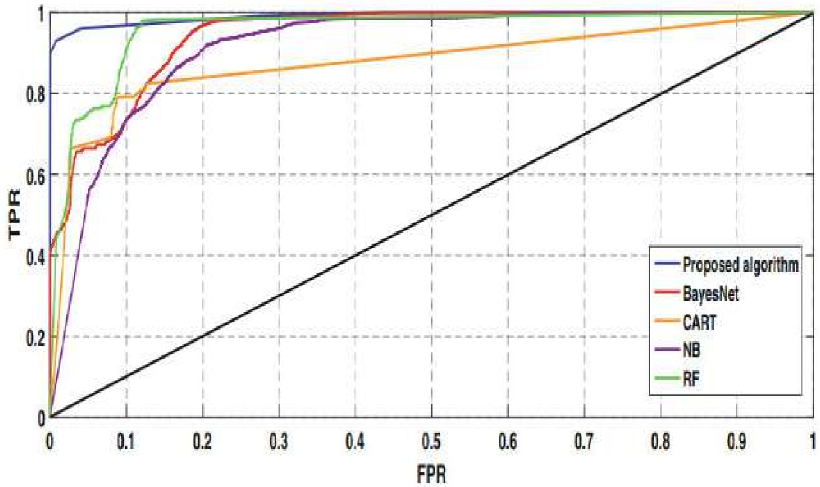

Figure 4: The plot comparing ROC curves for the proposed algorithm with the different machine learning algorithms.

Table 3 gives a comprehensive analysis of each algorithm's prediction accuracy as a percentage and its corresponding Area Under the Receiver Operating Characteristic (AUC-ROC) values. The outcomes clearly demonstrate that the suggested technique is more effective than traditional learning models in characterizing network traffic. The aforementioned results bear noteworthy consequences for strengthening application level security, highlighting the effectiveness of the suggested ensemble-based methodology in precisely identifying network traffic patterns and possible security hazards.

Table 1: Performance measures comparing traditional learning algorithms and the suggested ensemble approach for typical traffic flow

⟶ Algorithms	**Proposed Algorithm**	**BayesNet**	**CART**	**NB**	**RF**
Sensitivity	**0.989**	0.981	0.909	0.926	0.968
PPV	**0.939**	0.627	0.721	0.648	0.688
F-score	**0.957**	0.759	0.811	0.769	0.809

Table 2: Performance measures for traditional learning algorithms and the suggested ensemble approach for anomalous traffic flow

$\longrightarrow$ Algorithms	Proposed Algorithm	BayesNet	CART	NB	RF
Sensitivity	**0.928**	0.568	0.706	0.626	0.668
PPV	**0.988**	0.957	0.917	0.904	0.967
F-score	**0.949**	0.708	0.813	0.741	0.788

Table 3: Accurate prediction and AUC-ROC values for learning algorithms

Algorithms	Accuracy (in %)	AUC-ROC
Proposed Algorithm	**96.324**	**0.986**
BayesNet	74.428	0.953
CART	80.317	0.869
NB	76.123	0.911
RF	80.448	0.949

5. CONCLUSION AND FUTURE CHALLENGES

We investigated the critical functionalstudy of network traffic in dynamic Internet-of-things applications in this study. We suggested a framework for group-based learning that integrates the AdaBoost and Naïve Bayes (NB) algorithms in light of our findings. Our framework's efficacy was methodically assessed by contrasting it with well-known learning algorithms, including Random Forest (RF), BayesNet, Classification and Regression Tree (CART), and Naïve Bayes (NB).The suggested method had exceptional prediction ability, as demonstrated by our testing findings, which showed an accuracy of 96.324%. Furthermore, important measures includingsensitivity, F-score, and Positive Predictive Value (PPV) produced results of 0.928, 0.949, and 0.988, respectively, demonstrating the algorithm's effectiveness in identifying unusual traffic patterns. Additionally, it was found that the suggested algorithm's Area Under the Receiver Operating Characteristic Curve (AUC-ROC) value was 0.986. This demonstrates how well our ensemble-based methodology predicts traffic characteristics, which is particularly important when it comes to improving application level security in Internet of Things scenarios.

It is critical to regularly and quickly assess IoT traffic behavior in order to guarantee real-time network traffic analysis for improved application level security. Proactive threat detection requires the ability to recognize irregularities within the underlying patterns of network activity. A major obstacle in developing Intrusion Detection Systems (IDS) for Internet of Things (IoT) systems is acquiring sufficiently tagged traffic statistics. This is where sophisticated methods, especially learning procedures that are semi-supervised and unsupervised, come into play as a means of getting around labeling constraints and promoting strong application-layer security.

REFERENCES

1. Salman O, Elhajj IH, Chehab A, Kayssi A (2019) A machine learning based framework for IoT device identification and abnormal traffic detection. Transactions on emerging telecommunications technologies, e3743.
2. Panda S, Panda G (2020) Intelligent classification of IoT traffic in healthcare using machine learning techniques. In: 2020 6th international conference on control, automation and robotics (ICCAR). IEEE, pp 581–585.
3. Tripathy SS, Imoize AL, Rath M, Tripathy N, Bebortta S, Lee CC, Chen TY, Ojo S, Isabona J, Pani SK. A novel edge-computing-based framework for an intelligent smart healthcare system in smart cities. Sustainability. 2022 Dec 31;15(1):735.
4. Yao H, Gao P, Wang J, Zhang P, Jiang C, Han Z (2019) Capsule network assisted IoT traffic classification mechanism for smart cities. IEEE Int Things J 6(5):7515–7525.

5. Tahaei H, Afifi F, Asemi A, Zaki F, Anuar NB (2020) The rise of traffic classification in IoT networks: A survey. J Network Comput Appl 154:102538.

6. Bebortta S, Singh SK. An adaptive machine learning-based threat detection framework for industrial communication networks. In2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT) 2021 Jun 18 (pp. 527-532). IEEE.

7. Bebortta S, Panda M, Panda S (2020) Classification of pathological disorders in children using random forest algorithm. In: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE). IEEE, pp 1–6.

8. Tripathy SS, Rath M, Tripathy N, Roy DS, Francis JS, Bebortta S. An Intelligent Health Care System in Fog Platform with Optimized Performance. Sustainability. 2023 Jan 18;15(3):1862.

9. Kumar A, Lim TJ (2019) EDIMA: early detection of IoT malware network activity using machine learning techniques. In: 2019 IEEE 5th world forum on internet of things (WF-IoT). IEEE, pp 289–294.

10. Glennan T, Leckie C, Erfani SM (2016) Improved classification of known and unknown network traffic flows using semi-supervised machine learning. In: Australasian conference on information security and privacy. Springer, Cham, pp 493–501.

11. Khan MS, Ferens K, Kinsner W (2014) A chaotic complexity measure for cognitive machine classification of cyber-attacks on computer networks. Int J Cogn Inf Natural Intell (IJCINI) 322 8(3):45–69.

12. Perera P, Tian YC, Fidge C, Kelly W (2017) A comparison of supervised machine learning algorithms for classification of communications network traffic. In: International conference on neural information processing. Springer, Cham, pp 445–454.

13. Sharma K, Gupta BB (2019) Towards privacy risk analysis in Android applications using 327 machine learning approaches. Int J E-Services Mobile Appl (IJESMA) 11(2):1–21.

14. Rathore S, Park JH (2018) Semi-supervised learning based distributed attack detection framework for IoT. Appl Soft Comput 72:79–89.

15. Kozik R, Chora´s M, Ficco M, Palmieri F (2018) A scalable distributed machine learning approach for attack detection in edge computing environments. J Parallel DistrComput 119:18– 332 26.

16. Lin H, Yan Z, Fu Y (2019) Adaptive security-related data collection with context awareness. J Network Comput Appl 126:88–103.

17. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, Berlin.

18. Zhou ZH (2012) Ensemble methods: foundations and algorithms. CRC Press, Boca Raton.

19. Rish I (2001) An empirical study of the naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence, vol 3, No 22, pp 41–46.

20. Moustafa N, Turnbull B, Choo KKR (2018) An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things. IEEE Int Things J 6(3):4815–4830.

21. Adekitan AI, Abolade J, Shobayo O (2019) Data mining approach for predicting the daily Internet data traffic of a smart university. J Big Data 6(1):11.

22. Rahmat S, Niyaz Q, Mathur A, Sun W, Javaid AY (2019) Network traffic-based hybrid malware detection for smartphone and traditional networked systems. In: 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON). IEEE, pp 0322–0328.

23. Gratian M, Bhansali D, Cukier M, Dykstra J (2019) Identifying infected users via network traffic. ComputSecur 80:306–316.

24. Khonde SR, Ulagamuthalvi V (2019) Ensemble-based semi-supervised learning approach for a distributed intrusion detection system. J Cyber Secur Technol 3(3):163–188.

25. Bebortta S, Singh SK. An opportunistic ensemble learning framework for network traffic classification in iot environments. InProceedings of the Seventh International Conference on Mathematics and Computing: ICMC 2021 2022 Mar 6 (pp. 473-484). Singapore: Springer Singapore.

26. Bebortta S, Das SK, Chakravarty S. Fog-enabled Intelligent Network Intrusion Detection Framework for Internet of Things Applications. In2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2023 Jan 19 (pp. 485-490). IEEE.

27. Bebortta S, Singh SK. An Intelligent Network Intrusion Detection Framework for Reliable UAV-Based Communication. InInternational Conference on Cryptology & Network Security with Machine Learning 2022 Dec 16 (pp. 169-177). Singapore: Springer Nature Singapore.