

A novel RF-SMOTE model to enhance the definite apprehensions for IoT security attacks

V. Dankan Gowda *

Department of Electronics and Communication Engineering

B. M. S Institute of Technology and Management

Bangalore

Karnataka

India

K.D.V. Prasad [†]

Symbiosis Institute of Business Management

Hyderabad

and

Symbiosis International (Deemed University)

Pune

India

Pratik Gite [§]

Department of Computer Engineering

St. John College of Engineering and Management, Palghar (East)

Mumbai

Maharashtra

India

S. Premkumar [‡]

Department of Computer Science and Engineering

Galgotias University

Greater Noida

Uttar Pradesh

India

* E-mail: dankan.v@bmsit.in (Corresponding Author)

[†] E-mail: kdv.prasad@sibmhyd.edu.in

[§] E-mail: pratikgite135@gmail.com

[‡] E-mail: premambal@gmail.com

Naziya Hussain [@]
School of Computers
IPS Academy
Indore
Madhya Pradesh
India

Veera Sivakumar Chinamuttevi [#]
Department of Engineering
Univeristy of Technology and Applied Sciences-IBRI
Sultanate of Oman IBRI

Abstract

The Internet of Things (IoT) environment must prioritise security because of the IoT's significant attack susceptibility for a variety of reasons. The IoT attack detection technique or mitigation procedure is the extent of the currently available solutions. However, there are fewer autonomous security provider approaches available, and they are inappropriate for the IoT environment's evolving threats. Because of this, there has to be a security system in place that can detect and counteract both known and unknown threats for the increasing number of IoT devices. Deep Learning (DL) based intrusion detection systems does not consider attack signatures and normal behavior to obtain detection rules, it requires large data sets for training and takes a longer time to train the data. Many a time, insufficient dataset configuration prompts the minimization of a learning calculation, bringing about over fitting and helpless grouping rates. The goal of the proposed study is to create and implement a machine learning-based self-protection system to safeguard the IoT environment.

Subject Classification: Primary 93A30, Secondary 49K15, 65Y10 Numerical algorithms for specific classes of architectures, 68M25 Computer security, 68M07 Mathematical problems of computer architecture, 68M18 Wireless sensor networks as related to computer science.

Keywords: Attacks, Security, Random forest, Machine learning, IoT and Intrusion.

1. Introduction

Objects in the real world are connected in the IoT, like smartphones, cars, household appliances, buildings, and other things that have sensors, electronic software, and connection built into them. This network enhances the ability of the devices to exchange and gather data. By using an embedding computer system, each item is given a unique identification that allows it to communicate with other internet infrastructures. The IoT

[@] E-mail: naziyahussain@gmail.com

[#] E-mail: veera.siva@ibriict.edu.om

is a cutting-edge paradigm that comprises of billions of autonomously communicating objects. The IoT has recently gained popularity across a range of application fields. Many individuals now use mobile phones, PCs, and laptops to access the internet thanks to developing technology. IoT encompasses a variety of research areas that result in the creation and development of a practical application for society [1]. The IoT is a collection of items that have many online identities. The interconnected things in the environment of IoT are called smart things due to their potential applications in the real-world environment. IoT includes heterogeneous devices due to the combination of virtual and physical components [2]. The IoT devices will be connected wirelessly by using communication technology like Bluetooth, Wi-Fi, etc. The terms ‘Internet’ and ‘Things’ are referred to as the interconnected network worldwide based on communication technology, sensors that become a new version of communication and information technologies. IoT is also called a prototype that includes simple units which will be recognised and conveyed among the gadgets via the internet. The gadgets in the environment of IoT are smaller in size, lower force, and assets the compelled gadgets [3]. The traditional algorithms are costly because of their complexity and require more rounds for encryption. The aim of IoT is to utilise intelligent actuators and sensors in order to deliver a ubiquitous experience and a pervasive environment. IoT is emerging rapidly to deliver innovative dimensions of service by enhancing economic and societal growth [4]. IoT has the capability of connecting with anyone, anything, at any place, at any time with the network for any type of services as shown in Figure 1. The improvements in the industry of electronics, the growth in different communication protocols, the miniaturisation of transceivers and devices, providing the opportunity to convert the isolated objects into interacting devices.

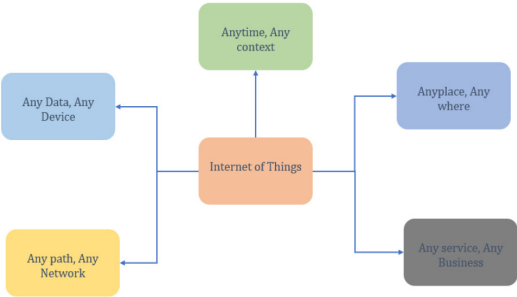


Figure 1
Services and connectivity of IoT

2. Related Work

IoT is widely utilized to provide interaction between physical objects and devices using wireless technologies. This IoT technology became popular because of its wireless communication technologies, small low-cost sensors, and modern Internet techniques [5]. IoT includes many interconnected devices that include interconnected vehicles, uncrewed aerial vehicles, household appliances, public facilities, medical equipment, and wearable equipment. These applications need networking [6]. In IoT devices, security is considered an important issue. Generally, the IoT network brings everything closer and connects the entire world [7]. Moreover, it is affected by various types of attacks. Typically, an IoT system comprises a whole world along with its services and smart technologies. Therefore, IoT links the virtual world and humans using smart devices and their communication protocols. Nowadays, people cannot imagine this world without IoT devices. A review indicates that 50 billion things were linked with the internet until 2021, and it's increasing exponentially with time. Figure 2. shows the estimated percentage of IoT devices used in 2020. Moreover, the evaluation of IoT will gain around 3.9–11.1 trillion USD economic market by the year 2025. Figure 2. shows the growth of IoT devices until the year 2025.

IoT devices are generally utilized for different requirements over the open network [8]. This creates higher accessibility between the users and devices. Moreover, IoT system provides various technological advances

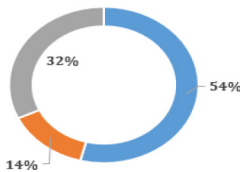


Figure 2
IoT device user estimation in 2021

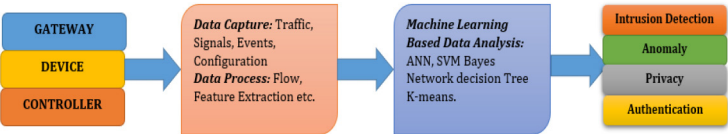


Figure 3
Security model of IoT

to human life like conformability and ease of access. However, multiple threats/attacks occurred in the IoT devices that affected the privacy of users. The security model of the IoT is depicted in Figure 3. This model is used to analyze configuration files, wireless signals, device events, and traffic. Various machine learning methods are utilized to classify data by employing the extracted features from the data sources. Accordingly, the results are used for authentication and privacy or classifying whether the event is related to the anomaly or intrusion event. Next, a detected threat/attack is needed to be mitigated for improving communication performance [9, 10]. This machine learning method helps the devices by using the learning process based on prior knowledge. Here, the learning process is used instead of doing programming for the devices. Generally, machine learning does not require complex mathematical equations and human support. In [11] author's presented attacker and defender games for analyzing their performance. The request was noted when it became similar to MRT. It was dropped without affecting the 'forwarding state table.' Therefore, eDLAM has adopted the optimal threshold update method for MRT that was based on the packet request matching process. This adoption was used to obtain the high defender utility. Based on the entries of expired states, eDLAM identified the attacks in the IoT system and created a higher latency which affected the performance of the IoT system. In [12] author's applied the Naive Bayes categorization method for the Intrusion Detection Systems (IDSs). The IDSs behaved as multi-agents for detecting irregular or misbehaving traffic and node actions in the network. Then, the information of the malicious node was reported to the IoT objects or the administrator for preventing the attacks. Moreover, an entire load was disseminated among users of IoT because of the existence of the agents. In [13] author's presented deep autoencoders for each IoT device to identify the botnet attacks. In [14] author's, analyzed various machine learning algorithms for the detection of Botnet DDoS attacks. The ML methods used for detecting attacks were: Unsupervised Learning (USML). The popular USML methods used for Botnet identification were K-means and X-means. Here, the UNBS-NB-15 dataset was used for examining algorithm performances. From the raw packets, general tools like Bro and Argus were used for generating the features. The features from the packets were determined using IP address and protocol [15]. Moreover, the filtering method was used to choose the features based on the scores, which was the result of certain statistical tests. The Chi-square and information gain were considered as filtering methods. Subsequently, selected features were given to machine learning algorithms to detect

Botnet attacks. Therefore, USML algorithm provided better performance when compared with DT, SVM, NB and ANN. This work treated only Botnet DDoS attacks.

3. RF-SMOTE

As of now, security is a hot exploration point because of the effect in everyday data foundation. Artificial Intelligence (AI) classifications have been improving over traditional studies, even though detection tasks utilize unpredictable metrics of information as the number of cases addresses few examples which can essentially vary. For uneven information, characterization models should have high accuracy for all the classes, while minority classes are viewed as clamour because of the absence of data that they give [16]. For instance, the quantities of strange and consistent associations created by gatecrashers inside an organization are extensively more modest than those from considerate applications. Much of the time, insufficient dataset configuration may prompt the minimization of a learning calculation, bringing about overfitting and helpless grouping rates. To address these issues, propose a resampling strategy called RF (Random Forest) and the SM (Synthetic Minority) Oversampling Technique with a network search calculation improvement method. This work shows grouping result enhancements for N-BaIoT and NSL-KDD datasets by blending artificially produced adjusted information and tuning diverse directed learning calculations [17]. In the result segment, the suggested RF-SMOTE is investigated through precision, accuracy, recall, AUC (Area Under the Curve), f-measure, and FAR (False Alarm Rate). A novel RF with SMOTE procedure is projected to enhance the definite apprehensions for IoT attacks classification.

4. Proposed Method

Contrasted with a solitary choice tree calculation, RF runs productively on huge datasets with superior exactness. RF can deal with ostensible information and doesn't over-fit. The ultimate choice for characterization of test information is finished by a greater part, casts a ballot from expectations of the outfit of trees. A dataset is imbalanced if the order classifications are not around consistently conveyed. To overcome the imbalanced dataset issues, in this work, SMOTE is utilized as a pre-processor. The measure of oversampling is impacted by the quantity of haphazardly chose neighbours from the k closest neighbours. It creates

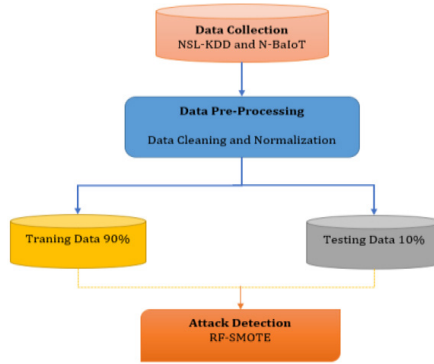


Figure 4

Proposed RF-SMOTE Model

synthetic tests by taking the distinction between include vector of the case viable and its closest neighbour, afterwards duplicating this distinction by a random number somewhere in the range of 0 and 1. Figure 4. illustrates the overview of the proposed RF-SMOTE model.

5. Classification of IoT Attacks by RF-SMOTE Model

RF-SMOTE is introduced to classify the traffic kinds in both of the given datasets once the data splitting procedure is complete. For big datasets, RF is regarded as an effective classifier. The rationale is because it takes advantage of better data standards to classify different assaults, generates a large number of decision trees (DTs), then syndicates the trees in order to get toward an effective classifier [18]. The decision tree method picks the most important characteristics for branching based on information gain and entropy values. Equations (1), (2), and (3) show the entropy value and information gain, which are mentioned below:

$$Gain(t, x) = E(t) - E(t, x) \quad (1)$$

$$E(t) = \sum_{i=1}^c -P_i \log_2 P_i \quad (2)$$

$$E(t, x) = \sum_{c \in H_i} P(c) E(c) \quad (3)$$

When a class is given the letter c , its entropy characteristics are denoted as $E(t, x)$, and its class entropy is given the letter $E(t)$. Here, the

relevant tree t is used to enhance the decision-making process as well as being considered as a unique classifier. The Random Forest classifier relies on two elements to reduce error rates. Thus, the normalised data vectors n_i were properly balanced using the SMOTE approach, enabling the Random Forest classifier to provide better traffic categorization results. The SMOTE, which uses data replacement as an alternative, is an efficient oversampling approach. The minority classes are oversampled by producing “synthetic examples.” The SMOTE strategy powers the selected area of alternative classes to turn into more extensive. The test assessment of the projected RF-SMOTE is specified in the result segment. The new tree classifier is formed by combining the DT with ensemble learning by using RF:

$$\{f(x, \delta_k), k = 1, \dots\} \quad (4)$$

Where, $f(x, \delta_k)$, is a meta classifier; the key in vector is x , and the random vector is δ_k . The two vectors are not dependent on each other. However, both the vectors share a similar distribution and determine the development of a single decision tree.

6. Performance Evaluation

The Proposed project work is replicated by python 3.8 platform and Anaconda navigator with windows 10 OS, 22 GB 2070 Ti GPU, the memory of 2 TB, RAM as 256 GB, and i9 processor. The performance of projected RF-SMOTE is associated with a small number of benchmark models, for example, shallow models [4], to authenticate the presentation of RF-SMOTE. Here, both N-BaIoT, NSL-KDD datasets are assumed for validating the effectiveness of projected RF-SMOTE in terms of accuracy FAR, f-measure, AUC, recall, and precision. The commenced performance metrics are accurately characterized in the Equations (5-10). FAR characterizes the possibility in that the top score is imperfectly categorized. Furthermore, Receiver Operating Characteristic (ROC) curvature designates the connection amongst FPR and TPR by synchronizes (0, 1). Where, TN is represented as True Negative; TP is signified as True Positive; FN is mentioned as False Negative, and FP is characterized as False Positive; t is designated as $1 - specificity$, and sensitivity is quantified as $ROC(t)$.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (5)$$

$$\text{Recall} = \frac{TN}{TP + FP} \times 100 \quad (6)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (7)$$

$$F\text{-measure} = \frac{2TP}{2TP + FP + FN} \times 100 \quad (8)$$

$$FAR = \frac{FP}{TN + FN} \times 100 \quad (9)$$

$$AUC = \int_0^1 ROC(t) dt \quad (10)$$

7. Results and Discussion

Here, the proposed RF-SMOTE presentation is being authenticated using the dataset NSL-KDD. Figures 5. and 6., respectively, show the representation of the classification model in this case. The predicted RF-SMOTE method's performance is assessed using the figures and the f-measure accuracy and recall. The supporting methodologies that are given in the figure are used to study and compare the RF-performance SMOTE's in terms of AUC, accuracy, and FAR. By looking at Figure 6, it can be shown that suggested RF-SMOTE standard achieved accuracy of 98.82%, FAR of 0.23%, and AUC of 99.91%, which is much better when compared to traditional approaches. As a consequence, when compared to other methodologies, the study showed that the suggested RF-SMOTE model produced better outcomes.

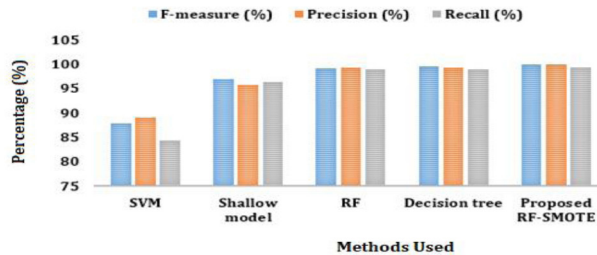


Figure 5

Analysis of Recall, Precision, and F-measure in the dataset NSL-KDD

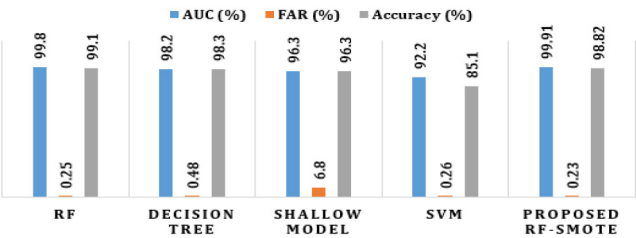


Figure 6

Comparison of NSL-KDD dataset in terms of FAR, Accuracy, AUC

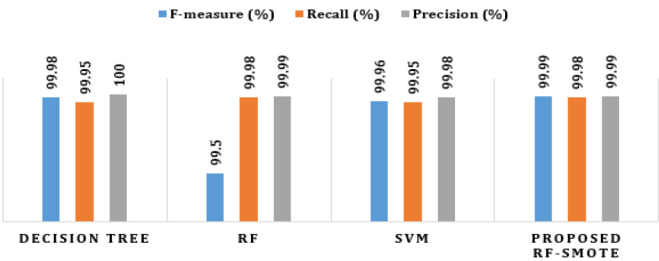


Figure 7

Comparison of binary class on N-BaIoT dataset

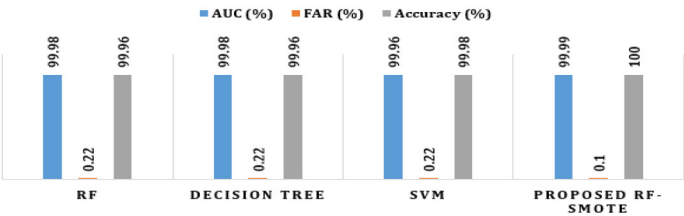


Figure 8

Performance analysis of AUC and Accuracy in N-BaIoT dataset

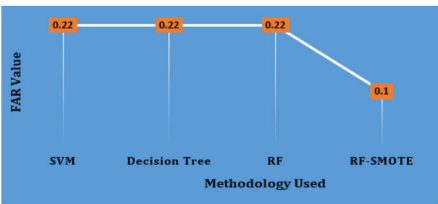


Figure 9

Performance analysis of FAR

N-BaIoT is assumed to analyze the presentation of RF-SMOTE and existing proportional classifiers. In N-BaIoT dataset, the training set consumes 90%, whereas the testing set consumes 10% of records. From Figure 7, RF-SMOTE is examined in terms of f-measure, precision, and recall. Via analyzing Figure 7, the RF-SMOTE accomplished improved outcomes when associated with conventional methods in the process of binary classification. Figure 8, the accuracy performance of RF-SMOTE is considerably superior than the current decision Tree, SVM, and RF. It also depicts RF-SMOTE as finest in classifying both Botnets traffics and normal through accuracy of 99.98%, 0.08% of FAR, and 100% of AUC. The FAR performance analysis in the N-BaIoT dataset is shown in Figure 9. From the comparative analysis, the proposed RF-SMOTE exhibited enhancement on both NSL-KDD and N-BaIoT datasets. Additionally, it showed better improvement in all the performance metrics when compared with existing methods.

8. Conclusion

The RF classifier is used to develop a functional and well-organized IDS. SMOTE is added, which is also picked for every importance given of the minority class by use of the attack technique, to increase the detection accuracy of all assaults in the unbalanced dataset. Results of the research's beginning show that the SMOTE approach significantly raised the detection rate. The SMOTE method is used with Random Forest. The evaluation of the N-BaIoT and NSL-KDD databases for the identification of threats is acknowledged. The evaluation's findings show that the suggested RF-SMOTE effectively separates the regular traffic from IoT threats using f-measure, recall, accuracy, AUC, accuracy, and FAR. When compared to traditional classifiers like deep learning model, logistic regression, SVM, shallow model, and RF, the recommended RF-SMOTE showed the least possible value and highest value of 0.02% and 0.06% improvement on N-BaIoT. The findings shown that the proposed RF-SMOTE provides greater distinguishability against IoT attacks.

References

- [1] TahsienSyedaManjia, Petros, "Machine learning based solutions for security of Internet of Things (IoT): A survey", *Journal of Network and Computer Applications*, Vol. 161, pp.102630 (2020).

- [2] Abebe AbeshuDiro, Naveen Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things", *Future Generation Computer Systems*, Vol. 82, pp. 761-768 (2018).
- [3] Jiabin Li, Ming Liu, ZhiXue, Xiangjian He, "Rtvd: A real-time volumetric detection scheme for ddos in the internet of things", *IEEE Access*, Vol. 8, pp. 36191-36201 (2020).
- [4] Tuan, T.A., Long, H.V., Son, L.H., "Performance evaluation of Botnet DDoS attack detection using machine learning", *Evolutionary Intelligence*, Vol. 13, pp. 283-294 (2020).
- [5] Neha Singh & Deepali Virmani, "Competence computation of attacks in wireless sensor network", *Journal of Statistics and Management Systems*, 23:7, 1227-1239 (2020).
- [6] Rahman, M.A., Asyhari, A.T. and Zolkipli, M.F., "Scalable machine learning-based intrusion detection system for iot-enabled smart cities", *Sustainable Cities and Society*, Vol. 61, pp. 102324 (2020).
- [7] Latif, S., Zou, Z., Idrees, Z. and Ahmad, J., "A novel attack detection scheme for the industrial internet of things using a lightweight random neural network", *IEEE Access*, Vol. 8, pp.89337-89350 (2020).
- [8] Vu, L., Nguyen, Q.U., Nguyen, and Dutkiewicz, E., "Deep Transfer Learning for IoT Attack Detection", *IEEE Access*, Vol. 8, pp.107335-107344 (2020).
- [9] Bhawna Narwal, Amar Kumar Mohapatra & Kaleem Ahmed Usmani, "Towards a taxonomy of cyber threats against target applications", *Journal of Statistics and Management Systems*, 22:2, 301-325 (2019).
- [10] Otoum, Yazan, Dandan Liu, and Amiya Nayak, "DL-IDS: a deep learning-based intrusion detection framework for securing IoT", *Transactions on Emerging Telecommunications Technologies*, Nov 2019.
- [11] Makkar, Aaisha, and Neeraj Kumar, "An efficient deep learning-based scheme for web spam detection in IoT environment", *Future Generation Computer Systems*, Vol. 108, pp. 467-487 (2020).
- [12] Shailendra Rathore, Jong Hyuk Park, "Semi-supervised learning based distributed attack detection framework for IoT", *Applied Soft Computing Journal*, Vol. 72, pp. 79-89 (2018).
- [13] H. G. Govardhana Reddy & K. Raghavendra, "Vector space modelling-based intelligent binary image encryption for secure

- communication", *Journal of Discrete Mathematical Sciences and Cryptography*, 25:4, 1157-1171 (2022).
- [14] 14. Mariam Elnour, Nader Meskin, Khaled Khan, Raj Jain, "A Dual-Isolation- Forests-Based Attack Detection Framework for Industrial Control Systems", *IEEE Access*, Vol. 8, pp. 36639-36651 (2020).
- [15] 15. Ratan Singh Gaharwar & Roopam Gupta, "Vulnerability assessment of android instant messaging application and network intrusion detection prevention systems", *Journal of Statistics and Management Systems*, 23:2, 399-406 (2020).
- [16] 16. Khushal Singh & Nanhay Singh, "An ensemble hyper-tuned model for IoT sensors attacks and anomaly detection", *Journal of Information and Optimization Sciences*, 41:7, 1715-1739 (2020).
- [17] 17. Pushkar Kishore, Gayatri Nayak & Swadhin Kumar Barisal, "Applying unsupervised system-call based software security techniques for anomaly detection", *Journal of Information and Optimization Sciences*, 43:5, 915-922 (2022).
- [18] 18. Neha Singh & Deepali Virmani, "Competence computation of attacks in wireless sensor network", *Journal of Statistics and Management Systems*, 23:7, 1227-1239 (2020).