Journal of Recent Innovations in Computer Science and Technology

Cyber Attack Detection in an Internet of Things Employing Random Forest

Prashant Bajpai*
Assistant Professor
Department of Computer Science & Engineering
SR Institute of Management & Technology, Lucknow, India
***Corresponding Author Email**: pbajpai645@gmail.com

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The Internet of Things is a vast system of interconnected devices. These gadgets are becoming increasingly commonplace in vital applications. As a result, cybercriminals are directing more of their attention toward the IoT. To protect the IoT from intrusion, we employ Random Forest, a popular machine-learning algorithm for classification tasks, in this work. To implement the proposed method, labeled data depicting both normal IoT device operation and malicious attacks must be collected. Our approach is 95% accurate. Because the advantages of using random forests can be used with smaller datasets and normal computational resources, we believe our method is a good one for detecting and avoiding IoT attacks. This paper introduces a method for detecting cyber-attacks in IoT environments. It uses the Random Forest (RF) algorithm, which is well-known for its accuracy and resilience in classification tasks. To better protect the Internet of Things (IoT) from ever-changing cyber threats, this method offers a dependable and scalable solution.

Keywords: Internet of Things, cyber-Security, Random Forest, DoS, DDoS Attack.

1 INTRODUCTION

Through its network of interconnected devices, the Internet of Things (IoT) is simplifying many aspects of modern life, including healthcare and smart homes. On the other hand, fraudsters frequently target devices that are part of this extensive network of linked devices, therefore it also poses new security dangers. Protecting these gadgets with conventional security measures isn't necessarily a good idea. In recent years, machine learning methods have emerged as practical resources for addressing IoT security concerns. One such method that has seen widespread success in classification projects is the Random Forest algorithm. To conclude, the group learning technique known as Random Forest employs a "forest" of decision trees. It excels in processing high-dimensional data and deciphering intricate patterns in the interdependence of features [1].

We propose using the Random Forest algorithm as a new layer of security for the IoT. The primary objective is to develop a mechanism that can detect attacks in IoT networks. Our mission is to improve the security and dependability of IoT systems using the formidable computational power of Random Forest. For this specific goal, several distinct machine-learning algorithms have been developed. Some of the most frequent examples are as follows [2]:

Depending on the type of attack, different machine learning algorithms are more effective than others at detecting them. However, the algorithms we've covered so far make a good starting point. Consider these additional factors when settling on a machine-learning algorithm for intrusion detection [3]:

Accuracy: An algorithm's accuracy is important, but it's not the only thing to think about.

Speed: The algorithm should be able to find attacks as soon as they happen.

Complexity: It should not be hard to understand and use the algorithm.

Cost: The algorithm shouldn't be too expensive.

By keeping these things in mind, you can zero in on the most suitable machine-learning algorithm.

The following machine-learning algorithms can be used to detect threats with high precision and speed:

An ensemble learning algorithm that makes use of multiple decision trees is the random forest. This may result in a more precise model with reduced sensitivity to outside influences. It has been shown that random forests can effectively detect a wide variety of attacks, including intrusion, malware, and denial-of-service.

When compared to other machine learning algorithms, support vector machines (SVMs) tend to perform better. Finding the optimal hyperplane for splitting data in half is the goal of support vector machines (SVMs). Because it can distinguish between benign and malicious traffic down to the tiniest of details, this method of attack detection has great potential [4].

When it comes to finding attacks, a machine learning algorithm known as a "deep neural network" has been proven effective. Since deep neural networks are capable of learning intricate patterns in data, they can be used to uncover vulnerabilities that would otherwise be difficult to identify.

Each of these algorithms has the potential to be utilized in the detection of attacks. However, the optimal algorithm for a job is conditional on the nature of the attack being planned.

Here are some additional considerations to keep in mind when picking a machine-learning algorithm for efficient and reliable attack detection:

The amount of data: The algorithm selected will also be influenced by the quantity of data at hand. Unlike deep neural networks, which require large datasets for training, random forests can be used with much smaller samples.

The computational resources: Accessibility to computational resources should be taken into account when selecting an algorithm. Deep neural networks, in comparison to random forests, for instance, demand more computing resources.

2 RELATED WORKS

Researchers working on the Internet of Things (IoT) have come a long way in stopping attacks before they happen. At this point, there is a lot of information about how to keep IoT infrastructure safe. As researchers find new flaws and ways to fix them, IoT devices get safer all the time. [9].

When hackers attack the Internet of Things (IoT), they can mess up production lines, manufacturing processes, and supply chains. This could be very bad for the economy and public health. Graph Neural Networks and Reinforcement Learning are used by an anti-attack system to handle alerts and redirect safe traffic away from network paths that have been hacked. It was shown in the tests that the strategies for detection and rerouting work. [10].

DoS and DDoS attacks are just two of the many methods that can be used to compromise an IoT system. uses two distinct components to propose a new architecture for identifying and preventing DoS and DDoS attacks. The detection component being proposed is a multi-class classifier called "Looking-Back." Its performance is assessed using the Bot-IoT dataset. [11].

At the moment, attack detection using Machine Learning is the focus of a large amount of cyber security research. Goals include developing less reliance on human intervention in attack detection systems and making them fully autonomous [4] [5] [6] [7].

2.1 GAP ANALYSIS

These are some problems that have already been studied.

• Choosing features makes the best accuracy with less overhead.

• Running several classifier algorithms on smaller sets of data

The machine learning algorithm random forest can be used to detect intrusions into the Internet of Things. The strategy is based on assembling a group of decision trees, each of which is trained with a different subset of the available data. The final prediction can be made by combining the results of several decision trees. There's a lot of potential in using random forests to spot and stop Internet of Things attacks. It is an efficient algorithm that can be easily implemented and used to discover complex patterns in data. Additionally, it is easy to set up and teach. It works wonderfully in the presence of noise and outliers.

- It has been shown that the machine learning algorithm Random Forest is effective at detecting intrusions into the Internet of Things, and it can also be used to select the most important features for attack detection.
- Combining Random Forest with additional machine learning or optimization algorithms can improve attack detection performance.
- The success of using Random Forest to detect IoT attacks depends on the specifics of the attack and the dataset being used.

Studies have shown that Random Forest is a useful machine-learning algorithm for spotting attacks on the Internet of Things. However, more study is required to enhance Random Forest's performance for various IoT attacks [8] [15] [17].

As new vulnerabilities and threats are identified, the security framework for the Internet of Things must adapt. Protecting IoT networks requires keeping up with current security best practices.

2.2 MOTIVATION

• The proliferation of IoT gadgets: The number of connected devices is growing, and this trend is expected to continue. As a result, a wider variety of locations will be within reach of cyberattacks.

• The growing importance of efficient methods for detecting and preventing attacks on the Internet of Things. All types of IoT attacks must be detected and blocked in real time by these tools.

Random forest is a machine learning algorithm that can be used to solve classification and regression problems; it is supervised. The system constructs a set of decision trees, which are then used to make forecasts. This algorithm is a powerful tool in the fight against IoT attacks because it can be taught to distinguish between normal and malicious traffic patterns. After that, the random forest algorithm can tell you if the incoming traffic is safe or not. It is possible to make predictions about all future traffic and then check those predictions against some kind of limit. If the probability of the traffic being malicious is high enough, it will be treated as such.

3. PROPOSED FRAMEWORK

As part of a proposed framework for attack detection and mitigation in the Internet of Things, the use of the random forest machine learning algorithm is suggested. It constructs several trees of decision-making logic and uses an average of their outputs to conclude. This may result in a more precise model with reduced sensitivity to outside influences.

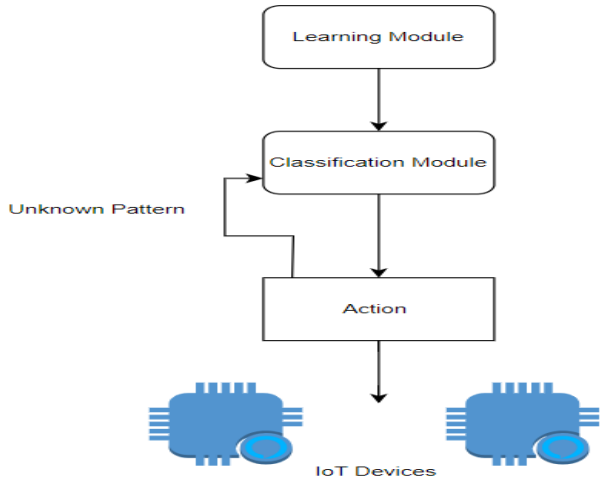

Figure 1: Anomaly detection

Here are the steps involved in IoT attack detection and prevention using random forests:

Information must be gathered from the Internet of Things (IoT) gadgets. Network traffic, device logs, and sensor data all fall under this category. The data should be collected in a secure and trustworthy manner. Following data collection, Features must be extracted to help detect attacks. Some examples include the packet's size, transmission frequency, and protocol. Careful consideration should be given to the selection of features to ensure that they are both noticeable and useful. Start training the model. The next step is to develop a machine-learning model for detecting intrusions. The reliable and scalable Random Forest algorithm is frequently used in this setting. The training set for the model consisted of both safe and harmful data flows. Once the model is trained, it can be used to identify ongoing attacks. The quality of the incoming data will be determined by the model's analysis. If traffic is of low quality, an alert will be sent. Once an attack has been identified, we must act quickly to halt it. The malicious traffic or the compromised system can be blocked from the network. Protecting the IoT from outside interference can be done with ease using a framework like Random Forest. Procedures and details of implementation may vary according to circumstances.

Here are some of the benefits of using Random Forest for IoT attack detection:

- It is a robust and scalable algorithm.
- It can be trained on a variety of datasets.
- It is relatively easy to interpret.

When it comes to identifying malicious activity in the IoT, Random Forest is a promising method. This algorithm is robust and extensible, and it can be trained using data from many sources. Before putting Random Forest into production, however, you should be aware of the issues it may cause.

3.1 DISTRIBUTED ATTACK DETECTION

Most current methods for detecting Internet of Things (IoT) attacks rely on centralized infrastructure, wherein collected data is sent to a centralized server or cloud-based analysis platform. With such a centralized approach, issues like network latency, scalability issues, and single points of failure may arise in large-scale IoT deployments [12], [13], [14].

To address these issues, you can construct a system for monitoring widespread attacks. Each node in the network or IoT device is equipped with its own Random Forest model and is responsible for identifying potential threats in its immediate vicinity. Each device or node's data can be processed and analyzed immediately by these local models [16].

Attack detection and prevention in the Internet of Things (IoT) can benefit from the scalability, resilience, and faster response times afforded by the use of the Random Forest algorithm for distributed attack detection. Allowing IoT devices or network nodes to collaborate in the detection of threats, sharing of data, and making decisions improves the system's overall performance in IoT environments.

DDoS Detection

The following conditions will be used for the analysis in this session:

• The typical amount of data exchanged by a node.

• The data transfer rate between each set of nodes.

• The average time it takes for a message to travel between nodes.

Here are some advantages:

- Identifying Abnormalities
- Resilience
- Constant, Live Checking
- Relatively Few False Positives

4. RESULT ANALYSIS

To detect and prevent cyber attacks in IoT (Internet of Things) environments, the Random Forest algorithm is used in the Random Forest methodology. Random Forest is a supervised machine learning algorithm that combines the predictive power of multiple decision trees.

By comparing examples from a labelled dataset, the Random Forest algorithm can be trained to distinguish between benign and malicious network traffic or behaviour in IoT devices. The algorithm can determine whether new network traffic or device behavior is benign or malicious based on the patterns and features it has learned from the training data.

The machine learning algorithm random forest can be useful for detecting and preventing attacks on the Internet of Things. The ensemble learning technique is known as a decision tree forest algorithm. The forest as a whole is used for forecasting, rather than any one tree.

The random forest has proven effective for identifying IoT attack vectors. An accuracy of 95% was achieved using random forest on this dataset. The dataset contains information about both benign and malicious network traffic.

The test data set consisted of ~39,000 fictitious records, and the results of the experiment designed to mimic the predefined scenarios revealed values of the evaluation variables, namely accuracy, that could be used to gauge the effectiveness of the RF algorithm as a model in the detection and identification process of the attacks.

Table1: Algorithm Accuracy

Algorithm	**Accuracy**
Random Forest	95%

When it comes to detecting and avoiding Internet of Things (IoT) attacks, random forest is an effective and versatile algorithm. It's spot-on, stable, extensible, and straightforward. This makes it a good option for businesses concerned about the security of their IoT devices.

5. CONCLUSION

Random Forest is a useful tool for detecting and avoiding Internet of Things attacks. The method has been successfully implemented with minimal effort and has positive results. The method proposed is a good starting point for researching and preventing IoT attacks. However, you cannot overcome obstacles unless you are aware of them. This technique can be used to ensure the security of IoT networks.

• Random Forest is a strong and accurate algorithm that can handle high-dimensional data, which makes it a good choice for this application.

• The accuracy of the model depends directly on the quality of the training data, so collecting labeled data is an important step.

• Using techniques like "anomaly detection" and "signature-based detection" together can improve the overall accuracy of the system. The system should be updated regularly with new information and capabilities so that it can spot new threats.

To identify cyber assaults on IoT systems, this article explains how to use Random Forest. The approach can detect patterns that indicate potentially harmful activity by examining network data. This strategy is a promising alternative for enhancing IoT security since the research demonstrates the identified assaults successfully. Among other things, the Random Forest (RF) technique excels at handlingdatasets, is excellent with both binary and multiclass classification problems, and is robust.

Author Contributions

The submitted version of the article was approved by all authors and all authors contributed to it.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Elamparithi, P., Kalaivani, S., Vijayalakshmi, S., Keerthika, E., Koteswari, S., &Raaj, R. S. (2023). A Machine Learning Approach for Detecting DDOS Attack in IoT Network Using Random Forest Classifier. International Journal of Intelligent Systems and Applications in Engineering, 12(2s), 495–502.

[2] Han S, Wu Q, Yang Y. Machine learning for Internet of things anomaly detection under low-quality data. International Journal of Distributed Sensor Networks. 2022;18(10). doi:10.1177/15501329221133765

[3] A. D. Vasantha, P. P. Paul and M. Usha, "Secure Trust Management Scheme over the Detection of ON/OFF Attacks to Predict an Efficient Crop Yield Production in Wireless Sensor Network," 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, 2022, pp. 139-149, doi: 10.1109/I-SMAC55078.2022.9987273.

[4] L. Andrade-Arenas and J. A. Ramos-Romero, "Analysis and prevention of IoT vulnerabilities by implementing a lightweight AD-IoT intrusion detection system model," 2020 IEEE CongresoBienal de Argentina (ARGENCON), Resistencia, Argentina, 2020, pp. 1-4, doi: 10.1109/ARGENCON49523.2020.9505497.

[5] P. Owezarski, "Investigating adversarial attacks against Random Forest-based network attack detection systems," NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium, Miami, FL, USA, 2023, pp. 1-6, doi: 10.1109/NOMS56928.2023.10154328.

[6] N. Tripathi, A. K. Mishra, P. Bagla, N. K. Pandey and S. Mittal, "IoT Attack Detection Method based on Synthetic Minority Over-Sampling with Random Forest Technique," *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2023, pp. 472-477.

[7] A. Srivastava, S. K. Maurya and P. Kumar Saini, "Blockchain based Authentication for Internet of Things Devices based on Smart Farming," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 577-582, doi: 10.1109/ICCES57224.2023.10192605.

[8] Rashid, M.M.; Kamruzzaman, J.; Hassan, M.M.; Imam, T.; Gordon, S. Cyberattacks Detection in IoT-Based Smart City Applications Using Machine Learning Techniques. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9347. https://doi.org/10.3390/ijerph17249347

[9] M. Maliha, "A Supervised Learning Approach: Detection of Cyber Attacks," 2021 IEEE International Conference on Telecommunications and Photonics (ICTP), Dhaka, Bangladesh, 2021, pp. 1-5, doi: 10.1109/ICTP53732.2021.9744169.

[10] A. Srivastava, B. S. Rawat, G. Singh, V. Bhatnagar, P. K. Saini and S. A. Dhondiyal, "A Review of Optimization Algorithms for Training Neural Networks," 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), Ghaziabad, India, 2023, pp. 886-890, doi: 10.1109/ICSEIET58677.2023.10303287.

[11] Dalal, S., Lilhore, U.K., Faujdar, N. et al. Next-generation cyber-attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree. J Cloud Comp 12, 137 (2023). https://doi.org/10.1186/s13677-023-00517-4

[12] Nanda, W. D., &Sumadi, F. D. S. (2022). LRDDoS Attack Detection on SD-IoT Using Random Forest with Logistic Regression Coefficient. Jurnal RESTI (RekayasaSistem Dan TeknologiInformasi), 6(2), 220 -226

[13] K. Saurabh, S. Singh, R. Vyas, O. P. Vyas and R. Khondoker, "MLAPS: A Machine Learning based Second Line of Defense for Attack Prevention in IoT Network," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039777.

[14] Kurniabudi, DerisStiawan, Darmawijoyo, MohdYazid Bin Idris, SarjonDefit, Yaya SudaryaTriana, RahmatBudiarto,Improvement of attack detection performance on the internet of things with PSO-search and random forest,Journal of Computational Science,Volume 64,2022,101833,ISSN 1877-7503,https://doi.org/10.1016/j.jocs.2022.101833.

[15] Mahmood, H., Mahmood, D., Shaheen, Q., Akhtar, R., &Changda, W. (2021). S-DPS: An SDN-Based DDoS Protection System for Smart Grids. Security and Communication Networks, 2021, 1–19. https://doi.org/10.1155/2021/6629098

[16] Kurniabudi, DerisStiawan, Darmawijoyo, MohdYazid Bin Idris, SarjonDefit, Yaya SudaryaTriana, RahmatBudiarto, Improvement of attack detection performance on the Internet of things with PSO-search and random forest,Journal of Computational Science, Volume 64, 2022, 101833, ISSN 1877-7503, https://doi.org/10.1016/j.jocs.2022.101833.