Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 241 (2024) 421–426

Procedia Computer Science

www.elsevier.com/locate/procedia

The 1st International Workshop on Intelligent Mobile Systems based on Internet of Things
August 5-7, 2024, Marshall University, Huntington, WV, USA

Enhancing IoT Security: Effective Botnet Attack Detection Through Machine Learning

Tamara Zhukabayeva[a,b], Lazzat Zholshiyeva[b,c,*], Khu Ven-Tsen[b], Aigul Adamova[a,b], Yerik Mardenov[b,d], Nurdaulet Karabayev[a,b]

[a] *Astana IT University , Mangilik El avenue 55/11, Astana and 010000, Kazakhstan*
[b] *International Science Complex "ASTANA", Kabanbay Batyr 8, Astana and 010000, Kazakhstan*
[c] *Astana IT College, Mangilik El avenue 55/11, Astana and 010000, Kazakhstan*
[d] *Astana International University, Kabanbay Batyr 8, Astana 010000, Kazakhstan*

Abstract

One of the most dangerous threats in WSNs is botnet attacks, in which attackers use mutual communications between IoT devices to launch large-scale malicious activities. In this regard, developments in the field of effective and reliable means of defence against this type of threat, in particular, reliable methods for detecting, identifying, and countering botnet attacks, are becoming increasingly important and relevant. This paper presents a comprehensive study that applies machine learning techniques, namely Random Forest and XGBoost, to identify botnet attacks on IoT effectively. These algorithms are analyzed, compared, and shown to be highly effective in detecting complex patterns indicative of botnet activity, thus achieving a significant improvement in IoT security. The conducted research aims to make a useful contribution to the problem of securing WSNs and IoT in general. The results of the study demonstrated high accuracy in detecting attacks with an accuracy of 99.18% for XGBoost and Random Forest showed an accuracy of 99.21%. Thus, it was shown that the significance of applying machine learning techniques such as Random Forest and XGBoost can be one of the key approaches in combating botnet attacks and securing the IoT. The results of the work emphasize the promising application of machine learning techniques for effective defense against cyber threats and highlight the importance of further.

© 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Conference Program Chair

Keywords: IoT; Identification of attack; Random Forest; XGBoost; Botnet attack

* Corresponding author.
E-mail address: Lazzat.Zholshiyeva@astanait.edu.kz

1877-0509 © 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Conference Program Chair
10.1016/j.procs.2024.08.058

1. Introduction

Internet of Things (IoT) devices can be used to launch various attacks that are potentially dangerous, but the most common attacks are IoT-based botnet attacks. The reason is that an IoT botnet spreads faster and has more consequences than other attacks on computer networks [1].

The modern world cannot be imagined without IoT, which has become an integral part of our daily lives. WSN is a subset of IoT and it plays an important role in collecting data for the broader IoT system. From smart homes to industrial systems, IoT devices provide us with connectivity, data collection, and process automation. However, along with the proliferation of these devices, comes an increase in the level of threats to their security. One of the most serious and widespread is botnet attacks [2]. This poses a great threat to IoT systems as they can spread rapidly and have serious consequences. Traditional security measures are often ineffective against botnet attacks due to their complex and dynamic nature.

The relevance of this research is due to the increasing threats associated with the rapid proliferation of IoT devices. It is important to note that the security and privacy of IoT devices play a key role in their successful operation, and therefore, developing effective attack detection techniques is an important task. The main objective of this research is to study the application of ML algorithms for detecting botnet attacks in IoT environments to improve the effectiveness of security measures.

This work is organized as follows:

- Section 2 describes the materials and search strategy of the research work.
- Section 3 presents the performance of the proposed method and compares the results of machine learning algorithms for attack detection.
- Section 4 presents the research results and future works.

2. Material and Search Strategy

This section identifies the main research questions to identify the detection gaps of botnet attacks in IoT. A strategy for finding research papers on relevant issues is described. To evaluate the effectiveness of existing security protocols and identify the gaps that can be filled by machine learning algorithms and their applicability for detecting botnet attacks on IoT. Comparative advantages of using machine learning techniques for implementation and performance in detecting botnet activity, the following research questions are initially identified:

- What are the strengths and limitations of traditional security measures in detecting botnet attacks in IoT networks?
- What machine learning algorithms are used to detect botnet attacks in IoT environments?
- What are the comparative advantages of using Random Forest over XGBoost (or vice versa) in the context of IoT security against botnet attacks?

Considering the research questions of this paper, the search query starts by defining keywords such as IoT, Identification of attack, Random Forest, XGBoost, and Botnet attack. To identify the appropriate literature that addresses the research question in a structured manner, the Prisma [3] framework is used which utilizes different databases. According to this framework, three constituent steps are proposed to narrow down the search. The search strategy for research papers has been formulated both based on the research questions and according to the recommendations.

The search query was applied to three different databases Scopus, Google Scholar, and Crossref, and all databases were queried from 2020 to 2024. The base search keywords from the databases are set as follows: ((IoT) and ("Random Forest") and (XGBoost) and (botnet) or (attack) and (Identification) or (of) or (attack)).

Machine learning algorithms are the most commonly used methods for information retrieval and also due to high detection accuracy another reason can be efficient memory and contribute to a higher level of performance [4]. For purposefulness, the above keywords were selected for the study and then left articles, reviews and sections of book or book 2020-2024 for further screening.

The inclusion criteria were articles and reviews in scientific journals and conferences published in English with a high h-index, books and sections from books were excluded. The search was conducted, using keywords, from databases and this procedure resulted in an initial sample of n=469 articles, specifically from Scopus n=69, Google Scholar n=200, and Crossref n=200. After selecting journal articles and conference articles with a high h-index, there were n=63. The exclusion criteria applied resulted in the detection of n=46 duplicates and their removal, after which n=22 were excluded and n=41 articles were included for further screening. Full-text screening helped to narrow this

sample. More articles that did not address machine learning methods with accuracy and f1-score were removed, after which became n=19. The screening resulted in 19 articles being included for further investigation, where the algorithm is shown in Figure 1.

All training methods were compared in terms of types of attacks, methods used to detect attacks, and datasets to select the best methods or techniques to detect attacks [5]. Several ensemble and individual classifiers including machine learning algorithms were used in the classification phase.

The paper [6] describes an attack detection system in an IoT network that adopts a novel hybrid approach to reduce the number of features. XGBoost compared to other machine learning algorithms consistently outperforms DT and RF in terms of accuracy in both binary and multilevel classification tasks in the paper [7]. This suggests that machine learning is an important algorithm to achieve high accuracy in detecting attacks in IoT networks. In addition, the choice of dataset plays an important role in influencing the model's performance.

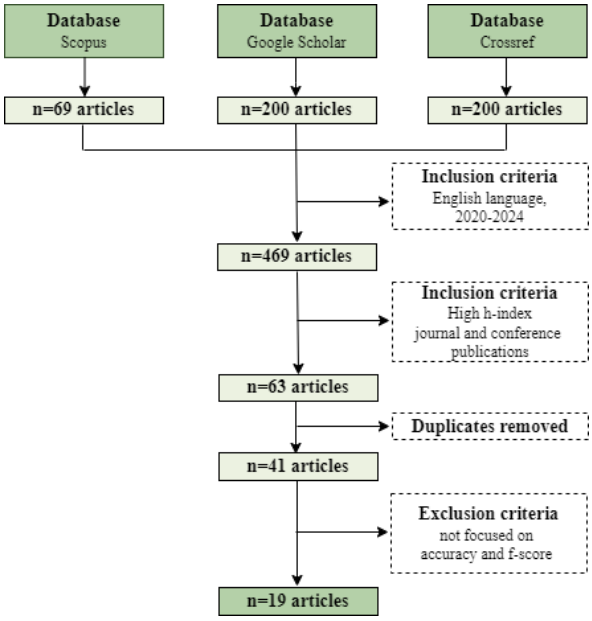


Fig. 1. PRISMA article selection algorithm.

Machine learning algorithms are effective tools for detecting botnet attacks in IoT systems. Their ability to analyse large amounts of data and detect imperceptible patterns enables organisations to improve the security of IoT devices and networks.

Selected studies from the PRISMA algorithm are summarised below in Table 1, which shows the machine-learning techniques and their attack detection results. The papers in this table utilise different IoT datasets.

Table 1. A comparative analysis of various studies of machine learning techniques in attack detection.

Authors	Year	Used Methods	Results
Benamor et al. [8]	2023	Random Forest, XGBoost, Decision Tree	XGBoost achieved accuracy rates of 93.15% RF model had an accuracy of 94.51%
Abdullahi et al. [9]	2022	XGBoost, SVM, Fuzzy C-Means	XGBoost accuracy 97.986%
Swapna Siddamsetti et al. [10]	2022	Random Forest, XGBoost, ML models	XGBoost accuracy 96% RF accuracy 94%
Mustafa S. Ibrahim Alsumaidaie [11]	2023	Random Forest, XGBoost, Decision Tree	XGBoost - 94% accuracy, 95% precision, IDDOSAD Approach showed 92% to 100%
Almomani et al. [12]	2024	Random Forest, XGBoost, Decision Tree, Gaussian NB, Logistic Regression, SVM	RF classifier , XGBoost - 98.92% , SVM - 98.29% Gaussian NB had a sensitivity rate of 91.87% in detecting DoS attacks

Gunupusala Satyanarayan [13]	2023	Random Forest, XGBoost, and Extra Trees for classification	Random forests - 98.06%, XGBoost - 97.99%, Extra-Trees - 97.80%
Mayes A., Anwar, A. [14]	2022	XGBoost, Decision Tree	Classification of cyberattack
Thi-Thu-Huong Le et al. [15]	2022	XGBoost	XGBoost model achieved F1 scores of 99.9% and 99.87% on X-IIoTDS and TON_IoT
Faysal et al [16]	2022	XGB-RF	XGB-RF model detects 99.94% of IoT attacks effectively
Alduailij et al. [17]	2022	Random Forest, GB	RF achieved better performance
Adel Assiri et al. [18]	2020	GA-based RF	Detection rates of 97.20% for KDD99 test set and 86.70% for UNSW-NB15 test se
Kumar et al. [19]	2021	XGBoost, KNN	XGBoost model achieved F1 scores of 99%
Zuech, Richard et al. [20]	2021	Random Forest, XGBoost, DT, CatBoost, LightGBM	LightGBM performs best
Waqas et al. [21]	2022	RF, XGBoost, DT, CatBoost, LightGBM, Fuzzy Classifier, KNN, GB	Tree based algorithm achieved more than 99% accuracy
Ikram et al. [22]	2021	XGBoost, DNN	XGboost was 92.93%
Ismail et al. [23]	2022	Random Forest, XGBoost	XGBoost: 90%, Random Forest: 89%

Thus, the effectiveness of botnet attack detection methods in IoT networks based on intelligent machine learning is confirmed by research results demonstrating high accuracy and reliability in detecting various types of attacks.

3. Results and Discussion

This section shows the results of using XGBoost and Random Forest machine learning algorithms for multiclass classification of botnet attacks with dataset detection. This work utilises the N-BaIoT dataset collected from real network traffic of IoT devices, including benign and attack traffic for identifying Mirai and Bashlite attacks [7]. The network traffic classes and their corresponding percentages of the total observed network traffic are used in this work: g_tcp- 28.6%, g_udp- 29.2, g_combo- 14.9%, benign- 11.0%, g_junk- 8.4%, g_scan- 7.9%.

XGBoost and Random Forest machine learning algorithms are used in the experimental work.

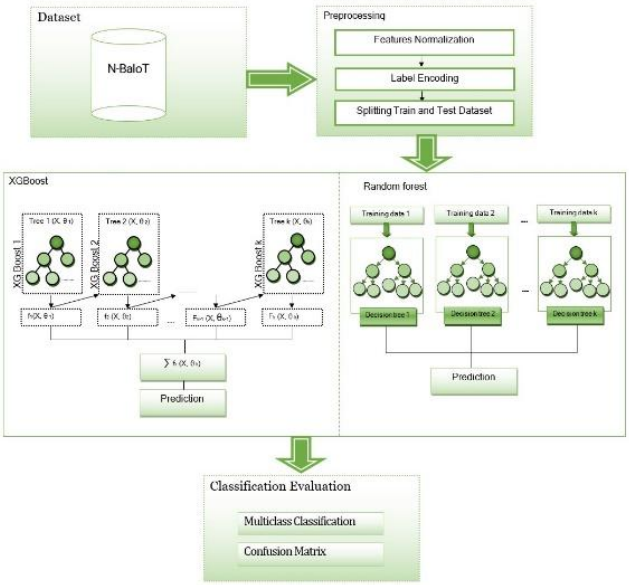


Fig. 2. XGBoost and Random Forest algorithm for multiclass classification in N-BaIoT datasets.

XGBoost algorithm is an efficient machine learning method used for data processing to achieve high performance in various machine learning tasks while requiring fewer resources compared to other systems [8].

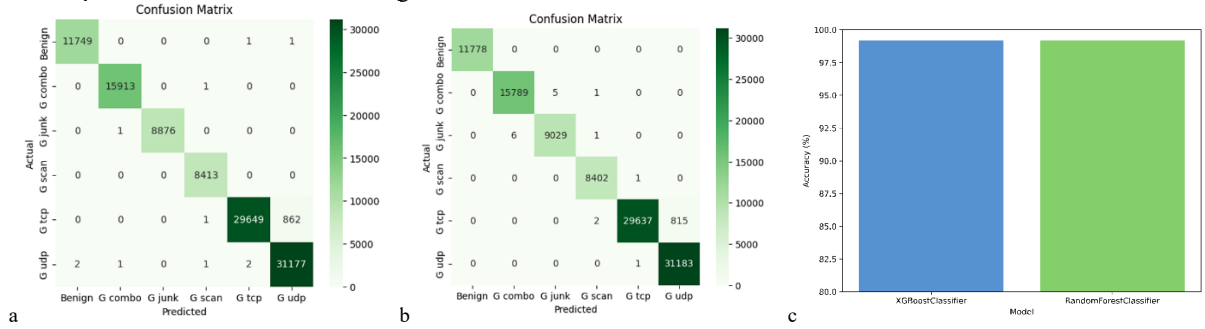
The Random Forest algorithm is an ensemble of decision trees that was proposed by Leo Breiman and Adele Cutler. Unlike individual trees, Random Forest combines the results of multiple trees to reduce variance in noisy datasets. This method is to utilize multiple solving trees to reduce the overfitting problem and improve prediction accuracy [24].

Figure 2 shows the XGBoost and Random Forest algorithm for multiclass classification of attacks in N-BaIoT datasets. The dataset is distributed. Performance evaluations of classification models. The suitability or selection of a particular classification method is based on the performance evaluation. Figures 3(a,b) show the descriptions of the evaluations using XGBoost and Random Forest error matrix. Figure 3(c) shows the accuracy percentage of the algorithms in XGBoost and Random Forest of the proposed model.

Table 1 lists the studies from 2020 to 2024 on machine learning models such as Random Forest, XGBoost, Decision Tree, Fuzzy C-Means, SVM, Gaussian NB, LSTM, XGB-RF, Catboost, Logistic Regression, LGBM, CNN, KNN, Extra Trees with different datasets. XGBoost consistently shows high accuracy in various studies with other datasets: from 93.15 to 97.986%.

Random Forest also performs well, with accuracies ranging from 94% to 98.92%. Other models such as SVM, Gaussian NB and LSTM also show results with accuracy ranging from 91.87% to 98.29%. Ensemble methods such as XGB-RF and combinations of tree-based algorithms (Random Forest, XGBoost, CatBoost, etc.) tend to achieve the highest accuracy, often exceeding 99%.

The choice of dataset can affect the performance of the model, as evidenced by different results in different studies. Our experimental result showed a high result with the N-BaIoT dataset.



a b c

Fig. 3. (a) XGBoost confusion matrix; (b) RF confusion matrix; (c) Comparison of Model Accuracies.

As can be seen from the obtained training results: for wireless IoT sensor networks, all machine learning methods showed high classification accuracy. In this paper, experimental evaluations of attack detection performance using machine learning models are obtained. The effectiveness of XGBoost and Random Forest algorithms lies in their superior performance and key role in achieving accurate intrusion detection of attacks in IoT networks.

4. Conclusion

In conclusion, this study represents a significant contribution to the field of IoT security. The proposed XGBoost and Random Forest models, tested on the N-BaIoT dataset, showed high performance on the task of classification and attack detection, achieving an accuracy of 99.18% and 99.21%, respectively. These results emphasize the potential of applying these models in real-world applications, especially in the context of IoT applications.

Based on the results obtained, it can be concluded that the proposed XGBoost and Random Forest models exhibit similar high performance. This indicates that both methods are effective in solving the problems of classifying and detecting attacks on IoT devices.

Since the security and privacy of IoT devices are crucial for their successful operation, the proposed approach can significantly contribute to improving the security of IoT systems.

In the future work is planned to test the performed developments and other machine learning algorithms in real-world IoT applications and develop practical recommendations for their use in WSN and IoT security systems. This will ensure further development and improvement of the proposed methods, as well as expand their application in the field of IoT security.

Acknowledgments

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. No AP19680345).

References

[1] Ali I., Ahmed A. I. A., Almogren A., Raza M. A., Shah S. A., Khan A., and Gani A. (2020) "Systematic Literature Review on IoT-Based Botnet Attack." *IEEE Access,* **8**: 212220–212232.

[2] Ahsan Nazir, Jingsha He, Nafei Zhu, Ahsan Wajahat, Xiangjun Ma, Faheem Ullah, Sirajuddin Qureshi, and Muhammad Salman Pathan. (2023) "Advancing IoT security: A systematic review of machine learning approaches for the detection of IoT botnets." *Journal of King Saud University - Computer and Information Sciences,* **35:** 101820.

[3] Moher D., Liberati A., Tetzlaff J., and Altman D.G. (2009) "Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *Ann. Intern. Med*., **151(64)**: 264-269.

[4] Alduailij M, Khan QW, Tahir M, Sardaraz M, Alduailij M, and Malik F. (2022) "Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method." *Symmetry,* **14(6)**:1095.

[5] Inayat U., Zia M.F., Mahmood S., Khalid H.M., and Benbouzid M. (2022) "Learning-Based Methods for Cyber Attacks Detection in IoT Systems: A Survey on Methods, Analysis, and Future Prospects." *Electronics,* **11:** 1502.

[6] Chen, T., and Guestrin, C. (2016) "XGBoost: A scalable tree boosting system", *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA*, *13–17 August* pp. 785–794.

[7] MeidanYair, Bohadana Michael, Mathov Yael, Mirsky Yisroel, Breitenbacher Dominik, Asaf, and Shabtai Asaf (2018) "N-BaIoT-Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders." *IEEE Pervasive Computing,* **17(3)**: 12-22.

[8] Benamor Z., Seghir Z.A., Djezzar M., Hemam M. (2023) "A comparative study of machine learning algorithms for intrusion detection in IoT networks." *Revue d'Intelligence Artificielle*, **37(3)**: 567-576. https://doi.org/10.18280/ria.370305

[9] Alkhudaydi O.A., Krichen M., and Alghamdi A.D. (2023) "A Deep Learning Methodology for Predicting Cybersecurity Attacks on the Internet of Things." *Information*, **14**: 550.

[10] Siddamsetti, S., and Srivenkatesh, M. (2022) "Implementation of blockchain with machine learning intrusion detection system for defending IoT botnet and cloud networks." *Ingénierie des Systèmes d'Information*, **27(6)**: 1029-1038.

[11] Mustafa S. and Ibrahim Alsumaidaie. (2023) "Intelligent Detection of Distributed Denial of Service Attacks: A Supervised Machine Learning and Ensemble Approach." *Iraqi Journal for Computer Science and Mathematics*, **4(3)**: 12-24.

[12] Omar Almomani, Adeeb Alsaaidah, Ahmad Adel Abu Shareha, Abdullah Alzaqebah, and Malek Almomani. (2024) "Performance Evaluation of Machine Learning Classifiers for Predicting Denial-of-Service Attack in Internet of Things." *(IJACSA) International Journal of Advanced Computer Science and Applications*, **15(1)**: 263-271.

[13] Satyanarayana G, Chatrapathi KS. Improving Intrusion Detection Performance with Genetic Algorithm-Based Feature Extraction and Ensemble Machine Learning Methods. *International Journal of Intelligent Systems and Applications in Engineering* 2023;**11**(4):100-112.

[14] Mayes A., and Anwar A. (2022) "Machine Learning Based IDS for Cyberattack Classification." *In: Ahmed, M., Islam, S.R., Anwar, A., Moustafa, N., Pathan, AS.K. (eds) Explainable Artificial Intelligence for Cyber Security. Studies in Computational Intelligence*, *Springer, Cham*, **1025:** 93-111.

[15] Le T.-T.-H., Oktian Y.E., and Kim H. (2022) "XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems." *Sustainability,* **14**: 8707.

[16] Faysal, J.A., Mostafa, S.T., Tamanna, J.S., Mumenin, K.M., Arifin, M.M., Awal, M.A., Shome A., and Mostafa, S.S. (2022) "XGB-RF: A Hybrid Machine Learning Approach for IoT Intrusion Detection." *Telecom*, **3**: 52–69.

[17] Adel Assiri. (2020) "Anomaly Classification Using Genetic Algorithm-Based Random Forest Model for Network Attack Detection." *Computers, Materials & Continua,* **66(1).**

[19] Kumar, P., Gupta, G.P. and Tripathi, R. (2021) "Toward Design of an Intelligent Cyber Attack Detection System using Hybrid Feature Reduced Approach for IoT Networks." *Arab J Sci Eng* **46:** 3749–3778.

[20] Zuech, Richard et al. (2021) "Detecting web attacks using random undersampling and ensemble learners." *Journal of Big Data,* **8.**

[21] Waqas M., Kumar K., Laghari A., Saeed U., Rind M.M., Shaikh A.A., Hussain F., Rai A., and Qaz A.Q. (2021) "Botnet attack detection in Internet of Things devices over cloud environment via machine learning." *Concurrency and Computation: Practice and Experience,* **34.**

[22] Sumaiya Thaseen Ikram, Aswani Kumar Cherukuri, Babu Poorva,Pamidi Sai Ushasree, Yishuo Zhang, Xiao Liu, and Gang Li. (2021) "Anomaly Detection Using XGBoost Ensemble of Deep NeuralNetwork Models." *Cybernetics and Information Technologies*, **21(3)**: 175-188.

[23] Ismail, Muhammad Ismail Mohmand, Hameed Hussain, Ayaz Ali Khan, Ubaid Ullah, Muhammadzakarya, Aftab Ahmed, Mushtaq Raza, Izaz Ur Rahman, and Muhammadhaleem. (2022) "A Machine Learning based Classification and Prediction Technique for DDoS Attacks." *IEEE Access,* **10***:* 21443-21454.

[24] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L.F., and Abdulkadir, S.J. (2022) "Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review", *Electronics* **11**: 198.