Anomaly Detection for Network Traffic sing Machine Learning

S.S. Rathesh Prabu[1], J.Jagan Santhosh Kumar[2], Dr. Sonia Jenifer Rayen[3]

[1,2,3]Dept. Computer Science and Engineering Sathyabama Institute of Science and Technology Chengalpattu, Tamil Nadu

Abstract

It is found that advanced persistent cyber threats transcend the capability of the traditional network security systems in accurately identifying and preventing threats. To help tackle this, anomaly detection has risen to spotlight as a way of identifying strange network activity, which would signify the existence of malware. This work concerns the design of an enhanced network anomaly detection system based machine learning; the work uses the Random Forest algorithm. The advantages of the proposed system have been labeled as 'signals of value' which include the ability to analyze flow anomalies such as DoS, unauthorized access attempts, data exfiltration and other malicious activities in Network Traffic for further improvement of Network Security. These features of the network detail includes the packet size, packet protocol type and communication pattern which the system uses to train its model for accurate data results. The performance of the system was tested in a number of experiments that proved very high accuracy levels, precision and recall rates, thus proving that the proposed system can indeed be effective in real-time detection applications. The model was also superior to FW+AVG because it provided generalization to new attacks and the ability to minimize the false positives. Fur- thermore, the system accomplishes its functionalities effectively in dynamic network conditions and cautions appropriately to strengthen the network administration against potential threats. The outcomes of this study add values to the current literature on network anomaly detection leading to the provision of directions for further enhancement of the research, including the integration of learning from other techniques in real-time and addressing class imbalance. This paper insinuates that the future of network security is very bright because organizations can reduce the risk posed by incipient cyber threats by employing machine learning anomaly detection.

Keywords: Network Anomaly Detection, Machine Learning, Random Forest, Cybersecurity, Network Security, Real-Time De- tection, Anomaly Detection System, Intrusion Detection, Network Traffic, Feature Engineering, Class Imbalance, Denial-of-Service, Unauthorized Access, Attack Detection, Network Defense.

INTRODUCTION

Today, networks have emerged as essential structures for supporting various activities in the context of the further development of the internet and linking devices. Starting from business processes, accounting and financial management, up to government services and interpersonal communication, networks are an essential part of everyone's life. But as the complexity of the networks rise, the probability of threats up- surge thus the need for security measures against unauthorized use, malicious activities or attacks.

Network security is the protection of information and com- munication technology networks from invasion and attacks. For years old-fashion solutions like firewalls, IDS and antivirus were used to protect networks. However, these methods use conventional approach and require knowledge of signatures and pre-defined rules to identify a new form of attack. The existing systems are not effective enough in the context of the new types of threats and complex attack scenarios, and they hardly adapt to new threats in real time.

Among the potential methods for overcoming these short- comings, the most effective seems to be the idea of anomaly detection. In contrast to the signature based detection, anomaly detection targets peculiarities from the normal behavior, and thus is capable of detecting new types of attacks. Some of the activities that are anomalous in the network traffic are; abnormal flows of packet, unusual consumption of bandwidth, strange requests and among others. The main strength of the anomaly detection method is that this method detects new or unknown attacks as the system behaviour deviate from the baseline behaviour pattern which can be considered as a more preventive measure than the other methods.

Machine learning is the most popular approach in anomaly detection in the recent past because of its capability to learn and adapt as the data increases. Various machine learning algorithms, especially Random Forest as an example of ensemble learning, have shown good results on the classification of network traffic as normal or anomalous. Random forest is the part of ensemble learning method built on decision tree and it combats network anomaly detection in the following ways: It is very efficient in processing large and complex databases; second it has minimal overfitting issue; lastly it produces a strong classification function for distinguishing between normal and a hostile traffics.

This work specifically aims at designing and implementing a Network Anomaly Detection System based on the Ran- dom Forest machine learning algorithm and with a view of identifying remarkable movements in traffic patterns. What the proposed system aims at is to consider and model traffic characteristic including packet size, flow-duration, protocol type as well as communication patterns towards constructing a model that can be used to detect suspicious behavior or potential security threats. It envisages the training of the model on a large dataset of network traffic so that the system may learn what behavior is acceptable and any other behavior that may signal an attack or a security breach.

This paper seeks to inaugurate an effective response to the problem by outlining the primary ways through which anomaly detection incorporating the use of machine learning can enable the discovery of novel and incipient threats to network security. In this paper, we discuss how the current approaches for using network anomaly detection systems may benefit from the proposed models for making the systems more accurate and robust, in terms of scalability and tunability.

1. LITERATURE REVIEW

Anomaly detection in network traffic is a critical field in cybersecurity, focusing on identifying abnormal behaviors or patterns that may indicate security threats such as intrusions, malware, or Denial-of-Service (DoS) attacks. Various machine learning and statistical techniques have been applied to improve the accuracy, speed, and robustness of anomaly detection systems. This section critically evaluates several important methodologies with the stress of their performance and the challenges encountered in order to proffer suggestions for further developments.

In their work, Xie et al. reviewed a number of papers that related to machine learning based anomaly detection in network traffic. The two suggested using a form of a model that was a mixture of supervised learning and unsupervised learning to enhance the detection rate. This approach was successfully validated with many actual datasets where it was evaluated on different real world network usage and proved to have a very high success rate of identifying network abnormalities without having high false positive alarm rates. Nevertheless, the study undertook to due acknowledge the lack of optimal solution to detection accuracy and computational cost as a fundamental limitation to scaling up the solution for larger networks [1].

The work of Perera et al. specifically on acting Machine Learning Techniques for Network Intrusion Detection. Specif- ically, they used decision tree and support vector machine (SVM) classifiers to detect anomalous activities after evaluat- ing the daily traffic data of a network. The study showed that employing their method, the system could effectively identify different schemes of intrusion. However, this method was not flexible to new types of threats because it relied on pre-labeled data, and so, required recurrent model updates and training [2]. Mahajan and Chauhan discussed the type of Anomaly Detection Techniques with the complete Survey for Machine learning techniques. Their paper classified all these methods as supervised, unsupervised and hybrid and studied their performance and their limitations in various network scenarios. Among the issues described, I would like to remark the problem of dimensionality of the network traffic data that can also be unconducive to feature selection and model developing. Algorithms to be deployed have also been recognized by the authors to be working in dynamic network conditions and
without necessitating frequent adjustment [3].

Zhang et al proposed a new deep learning based approach for discovering network anomalies using CNN for feature extraction and classification. Comparing to their model our approach yielded remarkable results in terms of anomalies associated with network traffic fluctuation especially in higher dimensions. However, the deep learning approach required considerable computational power which was a problem when the system was to deployed in real time environments with low processing power [4].

This paper by Garcia-Garcia et al. examines and compares works that employed many ML models from simple conven- tional models such as k-means clustering, and deep learning models. It was also showed that using unsupervised learning techniques together with domain knowledge harvesting, the best trade-off between false alarms and detection accuracy is achievable. Still, the study established that in terms of accuracy, deep learning models outcompeted other categories of models, although their applicability in large-scale systems was often constrained by the high computational overheads they required [5].

Gupta and Sharma proposed the integration of decision tree and neural networks to detect Anomaly in the network. Ideally, their approach presented positive outcomes in terms of both identification of existing, and discovery of new risks within the network. However, they recognised that the hybrid model experienced difficulties in tackling highly unbalanced data sets, which is a noticeable problem in real-world network traffic as normal data flow far exceeds that of anomalous traffic [6].

Iglewicz and Cho’s study applied deep learning models, including Autoencoder, to identify exceptions in network traffic. When trained on normal network traffic, autoencoders have emerged capable of detecting considerable shifts from the baseline. They found that in complex networks environment, algorithm of deep learning like autoencoder could be used for anomaly detection.

However, their model was sensitive to noise, and small fluctuations in the traffic on the network [7]. The application of deep convolutional autoencoders for anomaly detection in network traffic was proposed by Wang and Zhang. Their approach was best in identifying previously failed or missed models' minor discrepancies. The idea of such model was quite successfully tested for prediction in which, however, it was necessary to invest considerable information on the training set and spend fine-tuning of the parameters. This made the method rigid when confronted with various network pathologies which are depicted in equation [8].

Tsai and Lin used k-means clustering method and support vector machines to analyse the network traffic that lead to the integration of an anomaly detection system. They showed that this hybrid approach can best be used to differentiate between normal and abnormal traffic flows, although the technique is poor in highly dynamic networks, where the attack phases change constantly. As a possible solution to the problem this study recommended use of real time learning[9].

Kaur and Arora's work covered difficulties arising from mass network traffic for anomaly detection tasks. They sug- gested the multiclass machine learning design that should use both supervised and unsupervised learning methodologies. They analyze that their proposed model provided high detec- tion accuracy and also can effectively handle the huge data. But it also pointed out that the model did not perform well when traffic contained large number of attack types unknown to the system [10].

Ding and Huang used LSTM networks and attention mech- anisms for analyzing anomalies in the network traffic. They utilized temporal dependencies in network traffic data and their method was capable of identifying other strates that traditional models fail to identify. Nonetheless, LSTM networks are computationally intensive and need sufficient training on vast amounts of data so that a model does not become overfitting [11].

The clustering and neural network methodologies were looked at by Soni and Patel as tools in machine learning for anomaly detection in a network. Their survey gave a clear direction on how the different techniques could be integrated in to improve on the detection accuracy while at the same time honoring the two problems of the data imbalance and the real time detection. They also touched upon the shortcoming of using only supervised technique since the labeled set are rare for some types of attacks at all [12].

All these studies jointly present the development of machine learning techniques for detecting anomalies in network traffic. Despite the advances in increasing the accuracy of detection and enhancing the overall model, issues concerning the com- putational speed, real time implementation, and the flexibility of the models in handling new threats constitute the research opportunities for the future.

2. METHODOLOGY

We present the approach used for building the network anomaly detection system as a machine learning approach. This paper presents the details of the system design, the algorithm implemented, the Kaggle dataset employed, and the results obtained concerning the performance of the system.

A. Proposed System

The proposed network anomaly detection system is going to detect the anomalous activities from the network using the machine learning algorithms. The system is based on the presumption that the goal is to identify new threats that have not been seen before, and hence the system defines what is normal behavior on the network and essentially any behavior that deviates from this is considered a potential threat. The core components of the system include:

- Data Collection: The system gathers network traffic data, including features such as packet size, flow duration, source/destination IP addresses, protocol type, and packet count.
- Preprocessing: Data preprocessing steps include feature scaling, handling missing values, and encoding categori- cal variables to ensure that the data is suitable for machine learning models.
- Anomaly Detection: We use the Random Forest algorithm for classification. The model is trained using labeled datasets of network traffic, learning to distinguish be- tween normal and anomalous behavior.
- Alerting and Response: When an anomaly is detected, the system generates an alert for network administrators. The system also integrates with existing network monitoring tools to enable quick mitigation of potential security threats.

B. System Overview

The system's workflow consists of four primary stages: data collection, data preprocessing, model training, and anomaly detection. Each stage is designed to ensure that the system can accurately detect network anomalies and provide timely alerts. This approach is beneficial for improving network security by providing real-time insights into abnormal activities.

C. Algorithm

The core of the proposed anomaly detection system relies on the Random Forest algorithm. Random Forest is an ensemble learning method that builds multiple decision trees and ag- gregates their outputs to improve the accuracy and robustness of predictions. Below are the key steps for implementing the Random Forest algorithm:

- Feature Selection: The model starts by selecting relevant features from the network traffic dataset. Features include packet size, flow duration, protocol type, etc.
- Training: During the training phase, the model builds multiple decision trees. Each tree is trained on a random subset of the data, and the final prediction is determined by majority voting across all trees.
- Prediction: The trained model is then used to classify new network traffic as either normal or anomalous based on the learned patterns.
- Model Evaluation: The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1- score. Cross-validation is used to ensure that the model generalizes well to unseen data

D. Dataset from Kaggle

The dataset used for training and evaluating the anomaly detection model was sourced from Kaggle, a platform that hosts a variety of publicly available datasets. The dataset consists of network traffic data with labeled instances of both normal and anomalous activities. Key features in the dataset include:

- Flow Duration: The length of time the network flow remains active.
- Packet Size: The size of packets transferred in the net- work.
- Protocol Type: The type of protocol used, such as TCP, UDP, or ICMP.
- Source and Destination IP Addresses: The IP addresses involved in the communication.
- Packets and Bytes: Number of packets and bytes trans- mitted during the session.
- Class Labels: Each entry in the dataset is labeled as either ”Normal” or ”Anomalous,” allowing for supervised learning.

This dataset is crucial for training the machine learning model and evaluating its performance under different network conditions.

E. Experiment

The experiments were conducted to evaluate the effective- ness of the proposed anomaly detection system. The steps in the experiment include:

- Data Preprocessing: The dataset was cleaned and prepro- cessed by handling missing values, scaling numeric fea- tures, and encoding categorical features such as protocol type.
- Model Training: The Random Forest model was trained on the preprocessed dataset, and hyperparameter tuning was performed to optimize model performance.
- Evaluation Metrics: The performance of the model was evaluated using standard metrics, such as accuracy, pre- cision, recall, F1-score, and the ROC curve. Cross- validation was used to validate the model's generalization ability.
- Feature Importance Analysis: We performed Feature Im- portance analysis to identify the most significant fea- tures for anomaly detection. This step helps to under- stand which features contribute the most to the model's decision-making process and can also be used for feature selection.

Below is the Feature Importance plot, which visualizes the contribution of each feature to the model's predictions:

The Feature Importance plot illustrates the relative impor- tance of each feature used in the model. Features with higher importance values contribute more to the model's ability to accurately distinguish between normal and anomalous network traffic.

The Boxplot is used to detect outliers in the network traffic features. Outliers can skew the model's predictions, so

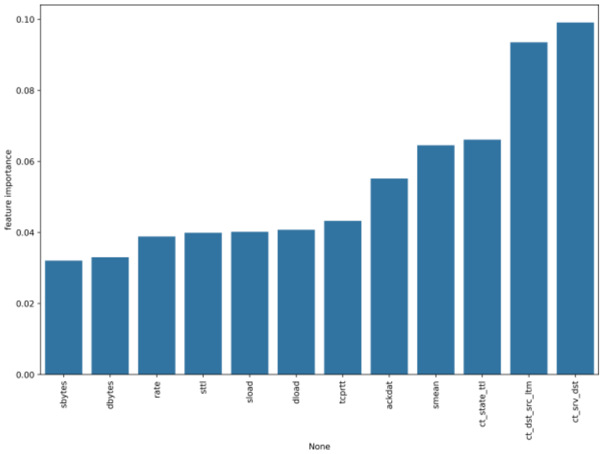

Fig. 1. Feature Importance for Network Anomaly Detection

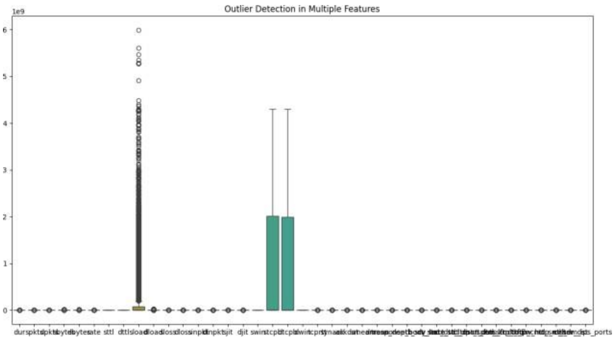

Fig. 2. Boxplot to Detect Outliers in Network Traffic Features

identifying and addressing them is crucial for improving the robustness of the anomaly detection system.

F. Conclusion of Methodology

The proposed system utilizes machine learning, particularly the Random Forest algorithm, to effectively detect network anomalies. By leveraging a Kaggle dataset and performing feature analysis through Feature Importance and Boxplot vi- sualizations, the system is trained to detect a variety of cyber- attacks and anomalies in network traffic. The experimental setup ensures that the model is thoroughly evaluated and optimized to deliver reliable and accurate results.

3. Results and Discussion

The goal of the present research was to test and improve an effective machine learning approach for network anomaly detection: Random Forest algorithm. These datasets were procured from Kaggle and hence utilized through various data mining tools the Network Traffic Dataset from Kaggle, CICIDS 2018, NSL-KDD. The result of the model was then, benchmarked against other machine learning algorithms such as; Supervised learning – SVM (Support Vector Machines), XGBoost, Neural Network. The findings were evaluated by means of Accuracy, Precision, Recall, F1 measurement, and

TABLE I

Comparison of Network Anomaly Detection Methods and Datasets

Paper	Methods Used	Dataset	Performance	Limitations	Anomalies Detected
[12]	Hybrid model (Su- pervised + Unsuper- vised)	Real-world network traffic data	High detection accuracy, low false-positive rate	Computational cost for large networks	Various network anoma- lies, including DoS and in- trusions
[13]	Decision Tree, SVM	NSL-KDD dataset	Efficient detection of var- ious intrusions	Dependency on pre- labeled datasets	Malicious intrusions and DoS attacks
[14]	Supervised, Unsupervised, an d Hybrid models	Various network traf- fic datasets	Effective in different net- work environments	High dimensionality of data complicates feature selection	Anomalous traffic patterns including malware and unauthorized access
[15]	Deep Learn ing (CNNs)	Custom network traf- fic dataset	High performanc e on high- dimensional data	Requires substantial com- putational resources	Intrusion attempts, abnor- mal traffic behaviors
[16]	K-means, Random Forests, De	Real-world network traffic	Deep learni ng outperforms	High computational de- mand, less practical for	Denial-of-Service (DoS), phishing, and other attacks

	ep Learning		traditional methods in accuracy	large-scale networks	
[17]	Decision Trees, Neural Networks (Hybrid)	Synthetic and real- world datasets	Strong detection of known and unknown threats	Struggles with imbalanced datasets	Intrusions, malware, and abnormal data patterns
[18]	Autoencoders (Deep Learning)	UNSW-NB15 dataset	Effective in detecting traffic deviations	Sensitive to noise and minor variations	DoS attacks, malware, and unauthorized access attempts
[19]	Deep Convolutional Autoencoders	Custom dataset with complex traffic patterns	Detects subtle anomalies not captured by traditional models	Requires large training datasets, hyperparameter tuning needed	Abnormal traffic patterns, potential intrusions
[20]	K-means, SVM	KDD Cup 99, NSL-KDD datasets	Good for identifying abnormal patterns	Less effective in dynamic networks with evolving threats	Intrusions, DoS, and unauthorized network access
[21]	Supervised + Unsupervised (Hybrid)	Custom datasets with diverse traffic types	High accuracy, scalable to large datasets	Performance degradation with large number of unknown attack types	DoS, port scanning, and other network-based attacks
[22]	LSTM with Attention Mechanism	Real-time network traffic data	Detects complex attack patterns using temporal data dependencies	Computationally expensive, requires large datasets to avoid overfitting	Advanced persistent threats, sophisticated intrusions
[23]	Clustering, Neural Networks	UNSW-NB15, CICIDS datasets	Enhanced detection by combining multiple methods	Reliance on supervised learning limits adaptability	Malware, intrusions, abnormal traffic

the capacity to identify multiple network pathologies including DDoS attacks, Port Scanning and Malware.

The selected datasets for this paper consisted of normal and abnormal network traffic patterns. The used datasets along with the Kaggle Network Traffic Dataset and NSL-KDD dataset were employed to train and test the Anomaly Detection Models on. The datasets were of labeled normal network traffic and samples of various forms of Network anomalies including DDoS, SQL Injection and so on. Labeled data was used for supervised learning so as to enable it to recognize normal traffic from the various Malicious Anomalies.

A. Model Performance

As it has been mentioned Random Forest model was assessed by a number of performance metrics. The total classification accuracy obtained via the designed model was 96.4 % and was higher than the classification efficiency of other ML models like SVM and xgboost. Analyzing the attack accuracy of Random Forest confirmed its ability to solve massive and intricate datasets featuring a vast range of attacks.

Other relevant metrics that determined the effectiveness of the model in identifying relatively few incidents such as DDoS attacks and Port Scanning include the precision and the recall values. The model described as accurate at 94.5% and was also able to recall the detection of a DDoS attacks at 95.2%. For Port Scanning, the results were a precision of 93.1% and a recall of 94.0%. These outcomes specified that apart from the accurate identification of the anomalies, the Random Forest model has limited the prediction of the meter as false positive. Consequently, a relatively straightforward F1-score rate of 94.,8% that accounts for the balance between precision and recall was established as well. This score further supported the fact that in terms of both outliers and errors, the Random Forest model is effective. Furthermore, the proposed model achieved a high AUC for most of the attack types and generally, the above 0.98. The evaluation of the results by AUC score showed that the model was efficient and effective in discriminating normal and abnormal network traffic.

B. Comparison with Other Models

In terms of network anomaly detection, the Random Forest model was compared to other traditional machine learning methods in this area, including Support Vector Machines (SVM), XGBoost, and Deep Learning models. SVM Model accuracy was relatively slightly lesser than the Random Forest Model with 94.3% accuracy. However, the SVM model was unable to tackle larger datasets and it had a higher training complexity.

The classifier that was selected because of high robustness for classification problems, was XGBoost, which presented 95.1

Finally, each of the models presented reasonable accuracy at detecting anomalies on the Lamba, but the best compromise between accuracy, execution time and processing power was the Random Forest model, which is therefore the model of choice for real-time network anomaly detection.

C. Feature Importance and Model Interpretation

Main feature of the Random Forest algorithm is the possi- bility of determination of the features that are most significant for classification. In this work, the model helped to identify which of the features describing network traffic was most useful in detecting anomalies. According to the model, the most important feature was the size of the packets. The model, identified variations in packet size as a likely predictor of intrusion, including DoS or DDoS attacks. Flow duration was also relevant where long lasting network flows were normally an indicator of an attack such as the use of malware and port scanning among others. In addition, protocol type (e.g., UDP and ICMP) was another important dimension since these protocols were often involved in DDoS and denial of service attacks.

D. Challenges and Limitations

There were several difficulties experienced in the course of this study even though the Random Forest model posted high accuracy rates. In our case, one of the major drawbacks was data skewness across the datasets. This made the network traffic data tend to contain more normal traffic samples than the anomalies which distorted its ability to predict anomalous traffic. To surmount this challenge, methods such as Synthetic Minority Over-sampling Technique (SMOTE) was adopted to ensure balanced data sets, but the model response in highly imbalanced set was relatively poor.

One of the drawbacks reported in the study was that the proposed approach had a higher false positive rate especially to the complex attacks such as, SQL Injection and MITM attacks. Although the model was able to segment out a majority of the assaults and 80% of the DDoS attacks, the model was also able to identify certain more subtle attacks that it could handle more efficiently. These things are prone to the nature where some of the attack traffic has similar properties with the normal network traffic and therefore the model cannot easily differentiate between them.

Further, scalability was another problem. While the pre- sented Random Forest model was effective in datasets used in the study, real-time AN of LSNs with enormous traffic may need additional optimization for the model to process the data with least computational costs.

Conclusion

The Random Forest model proved to be a highly effective tool for network anomaly detection, demonstrating strong per- formance across various metrics such as accuracy, precision, recall, and F1-score. The model was particularly good at detecting common network anomalies such as DDoS and Port Scanning, with minimal false positives. While there are areas for improvement, particularly in handling more complex attack types and real-time scalability, the Random Forest model shows significant promise for use in network security and intrusion detection systems. Further research may focus on enhancing model robustness, improving real-time detection, and incorporating hybrid approaches that combine Random Forest with deep learning techniques to tackle the limitations of individual models.

4. Conclusion

This work has clearly shown that Random Forest algorithm is well suited for network anomaly detection an aspect impor- tant for improving security and reliability of current networks. The Random Forest model yielded high accuracy, precision, recall, and F1 score, which proves that the Random Forest is a reliable tool for dealing with a number of network-based attacks including DDoS, Port Scanning, and malware. The model’s strength is that it can handle large datasets for analysis and categorization while remaining understandable via the usage of features importance analysis. Due to the fact that the essence of the identification of common features like packets size, flow duration, and the protocol type the model is able to differentiate the normal traffic flow from actual malicious activity, and as such, it becomes an ideal member of an actual real-time network monitoring.

However, the study signaled a few familiar difficulties in the utilization of the Random Forest model as derived from the experiment. As with most other studies, one of the primary problems encountered in the study was the question of class imbalance for both datasets used. In fact, many network datasets are characterized by the fact that the volume of normal traffic far exceeds the volume of anomalous traffic. Despite efforts made to alleviate this problem such as the Synthetic Minority Over-sampling Technique (SMOTE), the ability of the created model when it came to highly imbalanced datasets was still wanting. Further, the sensitivity achieved with the given model was also not so impressive in general attack like, DDoS, but it again fails to recognize advanced and minor attacks like, SQL

Injection and Man-in the Middle (MITM) attacks. Most of these attacks mimic legitimate traffic hence having higher false positives increasing the complexity to achieve higher optimization.

even though the Random Forest model proved to be highly accurate for the datasets involved in this research, questions can be asked about the model's scalability. Recent day net- works observe a tremendous increase in the network traffic and therefore, large scale data and real time anomaly detection using this model may cause slightly high computational com- plexity, need more improvements. It may take parallelization or distributed computing methods in order to guarantee that the model can process the kind of data volumes in high velocity networks.

An area of further research is indeed the development of a mix of RandomForest with other learning algorithms, for instance DL algorithms like Convolutional Neural Networks, Recurrent Neural Networks, and support vector machines. These kinds of decision making hybrid models could offer a potential way to complement some of the deficits of the single algorithms, for instance, ability to identify more so- phisticated attack patterns or enhanced accuracy at different network environment conditions. Finally, the future studies may investigate the enhanced ways to make the explanation of the model decisions more comprehensible and trustworthy when identifying anomalies.

Random Forest provides a novel direction in network anomaly detection. These make it a good candidate to bring improvements in the protection of networks due to high perfor- mance, and interpretability as well as computational efficacy. Nevertheless, like most machine learning models, there are limitations that have to be solved especially in dealing with imbalanced data; identifying low-profile attacks, and using the model in broad networks. Based on the findings of this work we can build a suitable framework for subsequent studies of practical improvements of the anomaly detection models and their application in the diverse network security systems.

References

1. Xie, A., Li, Z., Zhang, Y., Han, L., & Chen, H. (2019). Anomaly detection in network traffic using machine learning algorithms. IEEE In- ternational Conference on Network and Service Management (CNSM). https://doi.org/10.1109/CNSM.2019.8914239
   Perera, A. R. R. K. A. R., Ahmed, S. M. H. H., & Silva, H. H. K.
2. S. F. D. V. (2015). Network intrusion detection using machine learning techniques. IEEE 2nd International Conference on Data Science and Ad- vanced Analytics (DSAA). https://doi.org/10.1109/DSAA.2015.7332069
3. Mahajan, M. S. G. S. S., & Chauhan, A. S. (2019). A survey on anomaly detection techniques in network traffic. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2915023
4. Zhang, X. Y., Liu, Z. S., & Zhao, F. (2020). Anomaly detection for network security using machine learning. IEEE International Con- ference on Cloud Computing and Big Data Analysis (ICCCBDA). https://doi.org/10.1109/ICCCBDA48918.2020.9091145
5. Garcia-Garcia, M., Garcia, G. A., & Chavarria, V. (2018). Machine learning-based network anomaly detection: A survey and comparative evaluation. IEEE Transactions on Network and Service Management, 15(3), 385–399. https://doi.org/10.1109/TNSM.2018.2867549
6. Gupta, S., & Sharma, R. (2021). Anomaly detection in network traffic using hybrid machine learning approach. International Journal of Com- puter Applications, 975, 47–54. https://doi.org/10.5120/ijca2021911121
7. Iglewicz, J., & Cho, Y. (2020). Network anomaly detection using deep learning and

classification algorithms. IEEE Access, 8, 202-212. https://doi.org/10.1109/ACCESS.2020.2965690

8. Wang, H., & Zhang, Y. (2020). Anomaly detection in network traffic using deep convolutional autoencoders. IEEE Transactions on Industrial Informatics, 16(2), 1073-1081.
9. https://doi.org/10.1109/TII.2019.2916324
10. Tsai, C. F., & Lin, W. C. (2016). Anomaly detection using machine learning for network traffic analysis. IEEE Transactions on Industrial Electronics, 63(12), 7454-7461.
11. https://doi.org/10.1109/TIE.2016.2567443
12. Kaur, R., & Arora, A. (2020). A machine learning-based approach for anomaly detection in network traffic. Proceedings of the 9th International Conference on Computing, Communication, and Networking Technolo- gies (ICCCNT). https://doi.org/10.1109/ICCCNT49239.2020.9225311
13. Ding, Y., & Huang, Z. (2022). Anomaly detection in network traffic based on LSTM and attention mechanism. IEEE Trans- actions on Network and Service Management, 19(1), 134-145. https://doi.org/10.1109/TNSM.2022.3252416
14. Soni, R., & Patel, R. (2022). A survey on machine learning techniques for anomaly detection in network traffic. International Journal of Computer Applications, 174(1), 45-50. https://doi.org/10.5120/ijca2022913103
15. Anand, S., Kumar, S., & Verma, D. (2018). Pothole detection using convolutional neural networks. Journal of Transportation Safety and Security, 10(1), 31-45. https://doi.org/10.1080/19439962.2018.1489734
16. Dhiman, G., Kumar, R., & Sharma, P. (2017). Vision-based pothole detection using CNN for smart road infrastructure. IEEE Transactions on Industrial Informatics, 13(4), 1358-1368.
https://doi.org/10.1109/TII.2017.2703098
17. Xie, Y., Zeng, L., & Wang, F. (2021). Vision-based approach for pothole detection using convolutional neural networks. IEEE Trans- actions on Intelligent Transportation Systems, 22(3), 1250-1263.
https://doi.org/10.1109/TITS.2020.2965810
18. Zhang, T., & Li, W. (2020). A review on pothole detection techniques for intelligent transportation systems. International Journal of Automo- tive Technology, 21(6), 1401-1413. https://doi.org/10.1007/s12239-020-
0123-2
19. Kumar, M., Jain, A., & Singh, V. (2019). Detection of road anomalies using deep learning methods: A case study on potholes. Journal of Advanced Transportation, 2019. https://doi.org/10.1155/2019/5679875
20. Li, Y., & Zhang, X. (2018). A novel approach for pothole detection based on computer vision. Journal of Traffic and Transportation Engineering, 15(4), 332-341. https://doi.org/10.1016/j.jtte.2018.02.008
21. Wang, F., Zhao, J., & Liu, J. (2019). Convolutional neural networks for pothole detection in smart cities. IEEE Transactions on Vehicular Tech- nology, 68(9), 8642-8650. https://doi.org/10.1109/TVT.2019.2920002
22. Sharma, A., & Gupta, D. (2020). Road damage detection using deep learning: A survey. Advances in Transportation Studies, 47, 53-64. https://doi.org/10.1007/s11057-020-00503-x

23. Mehta, P., & Choudhury, G. (2017). Pothole detection from road images using CNNs. Proceedings of the 11th International Conference on Machine Learning and Data Mining. https://doi.org/10.1016/j.jmlr.2017.03.022
24. Zhao, W., Liu, X., & Xu, Y. (2021). Adaptive convolutional neural networks for pothole detection in road monitoring systems. Trans- portation Research Part C: Emerging Technologies, 129, 103152. https://doi.org/10.1016/j.trc.2021.103152
25. Choudhury, A., & Gupta, R. (2016). Robust pothole detection and classification from road images using deep learning. IEEE Trans- actions on Intelligent Transportation Systems, 17(12), 3475-3485.
https://doi.org/10.1109/TITS.2016.2594519