Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd

ISSN,e-ISSN: 2548-1304
Volume:IDAP-2022, pp:219-224, 2022
https://doi.org/ 10.53070/bbd.1172706
Research Paper

Novel Machine Learning (ML) Algorithms to Classify IPv6 Network Traffic in Resource-Limited Systems

Yıldıran Yılmaz*[1], Selim Buyrukoğlu[2], Muzaffer Alım[3]

[1]Computer Engineering Department, Recep Tayyip Erdoğan University, Rize, Turkey

[2]Computer Engineering Department, Çankırı Karatekin University, Çankırı, Turkey

[3]Beşiri Vocational School, Batman University, Batman, Turkey

*Corresponding Author

(yildiran.yilmaz@erdogan.edu.tr, sbuyrukoglu@karatekin.edu.tr, muzaffer.alim@batman.edu.tr)

Received:Sep.08,2022 Accepted:Sep.16,2022 Published:Oct.10,2022

Abstract— Providing machine learning (ML) based security in heterogeneous IoT networks including resource-constrained devices is a challenge because of the fact that conventional ML algorithms require heavy computations. Therefore, in this paper, lightweight ProtoNN, CMSIS-NN, and Bonsai tree ML algorithms were evaluated by using performance metrics such as testing accuracy, precision, F1 score and recall to test their classification ability on the IPv6 network dataset generated on resource-scarce embedded devices. The Bonsai tree algorithm provided the best performance results in all metrics (98.8 in accuracy, 98.9% in F1 score, 99.2% in precision, and 98.8% in recall) compared to the ProtoNN, and CMSIS-NN algorithms.

Keywords : *Embedded systems, machine learning, lightweight ML algorithms, IPv6 Network, cyber attack*

1.Introduction

IoT networks process and store a lot of valuable user data, so they become valuable targets of malicious attackers. Due to the constant presence of attackers, machine learning-based attack detection systems are needed to detect abnormal network activity. In the development of these systems, an IoT network traffic dataset termed as the KDD Cup99 (Alieksieiev and Andrii, 2019) containing normal and malicious activities is used. This data set, which contains the data of the IoT network, is given as training data to the machine learning algorithm for identifying and classifying activities in the IPv6 network. Machine learning algorithms can monitor network activities and detect network attacks and traces effectively and efficiently.

Various ML approaches have been put forward to classify the IoT network traffic in order to detect malicious activities such as denial of service attacks (Ge et al., 2019). It has been proposed by Ge (2019) that a new intrusion detection scheme based on deep learning can classify traffic flow in IoT networks. Khraisat (2019) in another study suggested using ensemble hybrid attack detection system for IoT networks. Elsewhere, Ferrag et al. (2020) presented a new detection model which adopted deep learning-based ML algorithm and provided a classification of 35 well-known datasets using deep learning.

The aforementioned studies have provided a high classification accuracy in their experimental analysis. However, deep learning and ensemble hybrid-based ML models are computationally heavy algorithms as they use more processing power and memory compared to single-based ML algorithms (Yang et al. 2017). IoT networks include not only computationally rich devices but also resource-scarce end devices such as sensor nodes. Smooth implementation of computationally heavy ML models in such heterogeneous IoT networks could be a challenge because of the existence of resource-limited devices with constrained processing power and memory capacity (Tuor et al.,2018). Therefore, this paper evaluates new lightweight ML algorithms in resource-scarce embedded devices to test their classification ability.

In this study, ProtoNN, CMSIS-NN, and Bonsai tree algorithms were evaluated on the heterogeneous IoT network dataset. The aforementioned novel lightweight ML algorithms are implemented and evaluated on the classification of network traffic as DOS attacks and innocuous in computing resource-scarce embedded systems.

2. Related Work

In the literature, many attempts using machine learning algorithms have been made with the purpose of attack detection in IoT networks.

Lonea et al. (2013) proposed an attack detection approach using Dempster-Shafer Theory (DST) processes and Fault Tree Analysis (FTA) for virtual machine (VM) intrusion detection system (IDS) based attacks to detect and analyze distributed denial of service attacks in cloud computing services. The approach quantitatively represents uncertainty and is used efficiently in IDSs to reduce false alarm rates. The rules of communication for IoT-based sensor networks are being developed over time with new standards. However, denial of service attacks threatens the availability of resources by threatening sensor nodes with massive network attacks. To prevent this, Beloglazov et al. (2019) demonstrated an ADE (Averaged Dependence Estimator) based intrusion detection scheme for IoT sensors.

Tertytchny et al. (2020) analyzed the anomalies in the network and stated that this situation could be caused by malfunctions or network attacks. Their analysis results show that supervised machine learning methods achieve high accuracy rates in classifying failures or attacks.

Network traffic classification is an efficient method of detecting and analyzing communication network anomalies. KNN, SVM and RF algorithms have been commonly applied to detect DDoS attacks (Aamir, 2021). In another study, KNN, SVM and RF models with optimized parameters provided 95%, 92% and 96.66% accuracy rates respectively, on the dataset created by using the Riverbed Modeller network traffic generator (Aamir and Zaidi, 2019).

Volkov and Kurochkin (2020) tested the applicability of artificial neural networks for the purpose of defining network attacks, they classified the data set containing 7 different classes with the LSTM model. It has been observed that the LSTM model in the 2-class structure selected from the data set is more successful than the MLP model in classifying network attacks.

An OCSA and RNN-based intrusion detection system in cloud computing services has been proposed by SaiSindhuTheja and Shyam (2021). A metaheuristic OCSA algorithm was used for feature selection. The KDDcup99 dataset was classified with an accuracy rate of 94.12% in the model created with RNN.

Tekerek (2021) proposed a web attack detection architecture by using CNN deep learning algorithm to detect anomalies in HTTP web traffic using the CSIC2010v2 HTTP dataset. Using the specified data set, HTTP data on the test data in the CNN model were classified as normal and abnormal with an accuracy rate of 97.07%.

3. Application of novel ML Methods for Embedded Systems

This section presents novel algorithms and tools used in the following sections to understand optimizations of machine learning algorithms on end devices in resource-scarce embedded systems.

3.1. Dataset Description

In this paper, KDD Cup99 (Alieksieiev and Andrii, 2019) dataset was used to evaluate the novel lightweight ML models. The dataset includes normal and attacks samples of IPv6 networks related to denial-of-service attacks. The dataset was divided into two parts training and testing datasets. The training dataset was 80% of the total dataset and the testing dataset was 20% of the total dataset.

The identifying feature or information for the attack can be obtained by observing the network. In general, monitoring the network can be seen as an expensive and time-consuming task. However, in order for IoT networks or computer systems to work, data collection from the network is a straightforward process. In practice, network communication traffic can be monitored using sniffer tools (Lamping and Warnicke, 2004). However, simply observing network packets may not give an overall view of network traffic (Lamping and Warnicke, 2004). Despite these drawbacks, several datasets are generated for the purpose of testing intrusion detection systems. The KDD Cup'99 dataset is developed specifically for the testing of intrusion detection systems and made available in a competition organized within the framework of the 5th International Knowledge Discovery and Data Mining conference (Alieksieiev and Andrii, 2019). Therefore, in this paper, we employed the KDD Cup'99 dataset for training and testing purposes.

3.2. Lightweight Machine Learning Algorithms and Tools for Resource-scarce IoT devices

This section overviews the ML tools and algorithms currently available in scarce-resource IoT devices for understanding the optimizations and properties of ML. This section focuses on the main optimizations when discussing these algorithms and tools, not the foundation of implementing these new algorithms and libraries.

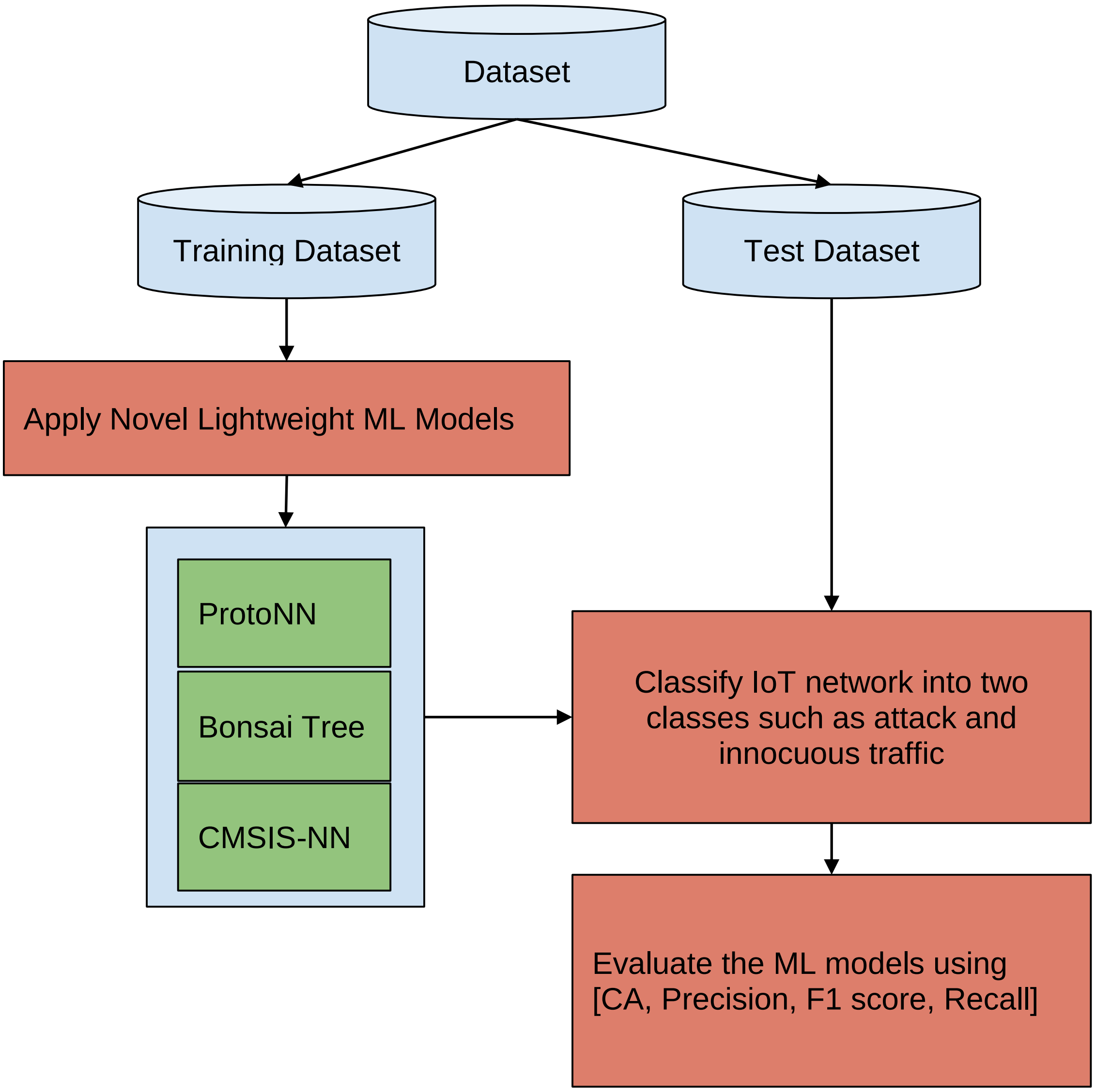
Dataset
Training Dataset
Test Dataset
Apply Novel Lightweight ML Models
ProtoNN
Bonsai Tree
CMSIS-NN
Classify IoT network into two classes such as attack and innocuous traffic
Evaluate the ML models using [CA, Precision, F1 score, Recall]

Figure 1. Flow diagram of the lightweight ML model to classify IoT network traffic.

Figure 1 illustrates the flow diagram of the ML model to classify IoT network traffic into two classes such as attack and innocuous traffic.

ProtoNN algorithm (Gupta et al., 2017) is a new ML model as an alternative to the KNN algorithm in resource-limited devices including MCUs. The KNN algorithm is expressed as a supervised/supervised machine learning method in which the class (learning cluster) and the nearest neighbour (element) of the sample data point to be classified are determined according to the k value (similarity). However, the KNN algorithm is not preferred to be implemented on devices with limited memory space, because this algorithm stores the entire training dataset that it will use in the prediction phase. On the other hand, the ProtoNN algorithm can model and learn the entire training set with a small dataset prototype.

Bonsai is a novel decision tree-based ML algorithm (Khandagale, 2020). Decision trees are used in a classification problem. It creates a learning model based on a tree structure containing leaf nodes and decision nodes by feature and target. The decision tree learning method is carried out by dividing the data set into smaller pieces. A decision node may contain one or more branches. The first node is called the root node. A decision tree can be used for both categorical and numerical dataset. Bonsai tree ML algorithm reduces the model size to minimize resource usage and still maintains predictive accuracy. To achieve this, it chooses a low-dimensional data space from which the entire tree can be learned and sparsely reflects all data.

CMSIS-NN is a lightweight library to embed Neural Network based ML algorithms in resource-scarce MCUs. This library is proposed by Lai et al. (2018). Neural networks emerged as a result of the mathematical modelling of the learning process (Lai et al. 2018). The low-cost version of the neural network named CMSIS-NN can be implemented in Cortex-M processor cores by achieving 4 times performance improvement and energy efficiency when compared to conventional neural network implementation (Sakr et al., 2021). CMSIS library performs data optimizations such as fixed-point uniform quantization and converting 8-bit to 16-bit data types so that neural networks can be implemented smoothly in resource-constrained IoT devices.

3.3. Evaluation Metrics

This study has been analyzed based on four metrics. These are Accuracy, Precision, F1-Score and Recall. The formula for each metric is presented as follows:

- Accuracy = (TP+TN)/(TP+FN+TN+FP)
- F1-Score:(2*Precision*Recall)/(Precision+Recall)
- Precision: TP/(TP+FP)
- Recall: TP/(TP+FN)

where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.

4. Results and Discussion

This section provides the performance results of the novel machine learning models evaluated on the IPv6 network dataset generated on resource-scarce embedded devices. The performance metrics such as testing accuracy, precision, F1 score and recall are used to discuss the classification ability of the ProtoNN, CMSIS-NN, and Bonsai tree ML algorithms.

Table 1 presents the DOS detection performance results of lightweight ML algorithms. The best statical score is provided through the Bonsai model while the worst performance is obtained by the ProtoNN model. Even if the CMSIS-NN model is the second-best model in terms of each metric, the accuracy rate between them was 0.003%. In other words, the CMSIS-NN model can be considered an effective model as much as Bonsai in the DOS detection of lightweight.

Table 1. DOS detection performance results of lightweight ML algorithms

ML Model	ACC	F1 Score	Precision	Recall
ProtoNN	96.2	94.3	93.5	96.2
CMSIS-NN	98.5	98.5	98.5	98.5
Bonsai	98.8	98.9	99.2	98.8

As can be seen from Table 1, the Bonsai model has a better performance when compared to the CMSIS-NN and ProtoNN models in the DOS detection performance results of lightweight ML algorithms based on F1-score, Precision, and Recall. The Bonsai and CMSIS-NN can almost present similar statistical scores (around 98.5% and 99.2%).

5. Conclusion

In this study, three novel machine learning algorithms (ProtoNN, CMSIS-NN, and Bonsai tree) were employed to classify communication traffic on heterogeneous IoT networks including resource-scarce embedded devices. Those algorithms were evaluated by using performance metrics such as testing accuracy, precision, F1 score and recall. The highest accuracy score (98.8%) was achieved by the Bonsai tree ML algorithm. In addition, the Bonsai tree algorithm provided the best performance in other metrics such as 98.9% in F1 score, 99.2% in precision, and 98.8% in the recall.

References

Aamir, M., & Zaidi, S. M. A. (2021). Clustering-based semi-supervised machine learning for DDoS attack classification. Journal of King Saud University-Computer and Information Sciences, 33(4), 436-446.

Alieksieiev, V., & Andrii, B. (2019, September). Information analysis and knowledge gain within graph data model. In 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 3, pp. 268-271). IEEE.

Beloglazov, A., Abawajy, J., & Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Future generation computer systems, 28(5), 755-768.

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419.

Ge, M., Fu, X., Syed, N., Baig, Z., Teo, G., & Robles-Kelly, A. (2019, December). Deep learning-based intrusion detection for IoT networks. In 2019 IEEE 24th pacific rim international symposium on dependable computing (PRDC) (pp. 256-25609). IEEE.

Gupta, C., Suggala, A. S., Goyal, A., Simhadri, H. V., Paranjape, B., Kumar, A., ... & Jain, P. (2017, July). Protonn: Compressed and accurate knn for resource-scarce devices. In International conference on machine learning (pp. 1331-1340). PMLR.

Khandagale, S., Xiao, H., & Babbar, R. (2020). Bonsai: diverse and shallow trees for extreme multi-label classification. Machine Learning, 109(11), 2099-2119.

Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J., & Alazab, A. (2019). A novel ensemble of hybrid intrusion detection system for detecting internet of things attacks. Electronics, 8(11), 1210.

Lai, L., Suda, N., & Chandra, V. (2018). Cmsis-nn: Efficient neural network kernels for arm cortex-m cpus. arXiv preprint arXiv:1801.06601.

Lamping, U., & Warnicke, E. (2004). Wireshark user's guide. Interface, 4(6), 1.

Lonea, A. M., Popescu, D. E., & Tianfield, H. (2012). Detecting DDoS attacks in cloud computing environment. International Journal of Computers Communications & Control, 8(1), 70-78.

SaiSindhuTheja, R., & Shyam, G. K. (2021). An efficient metaheuristic algorithm-based feature selection and recurrent neural network for DoS attack detection in cloud computing environment. Applied Soft Computing, 100, 106997.

Sakr, F., Bellotti, F., Berta, R., De Gloria, A., & Doyle, J. (2021). Memory-Efficient CMSIS-NN with Replacement Strategy. In 2021 8th International Conference on Future Internet of Things and Cloud (FiCloud) (pp. 299-303). IEEE.

Tekerek, A. (2021). A novel architecture for web-based attack detection using convolutional neural network. Computers & Security, 100, 102096.

Tertytchny, G., Nicolaou, N., & Michael, M. K. (2020). Classifying network abnormalities into faults and attacks in IoT-based cyber physical systems using machine learning. Microprocessors and Microsystems, 77, 103121.

Tuor, T., Wang, S., Salonidis, T., Ko, B. J., & Leung, K. K. (2018, April). Demo abstract: Distributed machine learning at resource-limited edge nodes. In IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 1-2). IEEE.

Volkov, S. S., & Kurochkin, I. I. (2020). Network attacks classification using Long Short-term memory based neural networks in Software-Defined Networks. Procedia Computer Science, 178, 394-403.

Yang, T. J., Chen, Y. H., Emer, J., & Sze, V. (2017, October). A method to estimate the energy consumption of deep neural networks. In 2017 51st asilomar conference on signals, systems, and computers (pp. 1916-1920). IEEE.