Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 1 PAGE: 232-246 YEAR: 2024
DOI:10.17798/bitlisfen.1380547

New Hybrid Distributed Attack Detection System for IoT

Çiğdem Bakır[1*]

[1]*Software Engineering Department, Engineering of Faculty, Dumlupinar University, Kütahya, Türkiye*

(ORCID:0000-0001-8482-2412)

Keywords: Random Forest, Artificial Neural Network, Security, IoT, Distributed System

Abstract

Internet of Things (IoT) is expressed as a network of physical objects with applications and various technologies that provide data connection and sharing with various devices and systems over the Internet. Security vulnerabilities in IoT devices are one of the biggest security issues in connecting devices to the internet and collecting and processing user data. These vulnerabilities can lead to increased attacks on IoT devices and malicious use of user data. In this article, we discuss these security problems that arise in IoT systems in detail in distributed systems technology. Distributed systems are increasingly used in the modern computing world. These systems are a structure where multiple independent computers communicate with each other for a common purpose. Distributed system technologies have become more common with the development of internet and cloud computing systems. However, the use of distributed systems has brought with it important security challenges such as security vulnerabilities, access controls and data integrity issues. Therefore, the security of distributed system technologies has been an important focus of work in this area. In this study, information about distributed system technologies and security for IoT is given. The all attack types were classified using Artificial Neural Network (ANN), developed Random Forest (RF) and hybrid model. In RF, all feature vectors created from all datasets (bank and two financial datasets) were also analyzed separately and the classification performance was examined. In addition, a new RF algorithm based on weight values using the Gini algorithm has been proposed. With this algorithm, the traditional RF algorithm has been developed and the success rates have been increased. In addition, a hybrid method was created by classifying the datasets obtained by RF with ANN. With the hybrid method ANN and the enhanced RF method, its accuracy in detecting normal behaviors and attack types was calculated and the success of the methods was presented comparatively. In addition, the working times of the methods were determined.

1. Introduction

The concept of the IoT is a popular technological trend that has gained increasing momentum in recent years. IoT refers to the network of devices that are connected to the internet and can communicate with each other. These devices can include all kinds of devices such as smartphones, computers, home appliances, industrial devices, vehicles, wearables, and even healthcare devices [1]. IoT devices are designed to make the world around us and people's lives easier. For example, smart home devices can enable you to remotely control your home, save energy, increase your security, and improve your quality of life at home. Industrial IoT devices, on the other hand, can increase efficiency in areas such as production and logistics, and offer new opportunities in areas such as agriculture and health [2].

* Corresponding author: cigdem.bakir@dpu.edu.tr
Received: 24.10.2023, Accepted: 29.01.2024

Despite all these advantages, there are some security threats to IoT devices due to heterogeneous structures that communicate over different networks over the internet. In particular, it is possible to damage IoT devices and hack the system by various cyber attacks. These damages cause some irreversible problems in the future. For example, an attacker could attack an IoT device in the home, gaining access to all the devices in the home and even the personal information of the host. Another threat to IoT devices is malware infection of devices. These software can take control of devices and turn them into botnets. To deal with security threats to IoT devices, it is important to provide regular security updates to detect and fix vulnerabilities. In addition, security measures such as encryption of communication between devices and authentication must be taken in order for IoT devices to work securely. However, when these studies are evaluated as a whole at the point of detecting different types of attacks, they are often insufficient. In particular, it has shown that a stronger security mechanism is needed to detect and prevent attacks such as Denial-of-Service Attack (DoS) and Distributed Denial of Service Attack (DDoS) [3,4].

Research on data security has revealed that vulnerabilities and leaks in distributed systems in IoT devices are major problems [5,6]. For this reason, studies on solutions and measures for data security problems in distributed systems are very important. In particular, institutions and organizations need to create a strong security mechanism against important security vulnerabilities such as security vulnerabilities, access controls and data integrity problems.

In our study, security problems and precautions that arise in distributed systems related to IoT devices and networks are discussed. A method has been developed to detect and prevent possible attacks from different sources that may arise as a result of various violations such as unauthorized access, weak encryption methods, neglect of device updates, increasing attacks and threats, and management difficulties that will disrupt data confidentiality and data integrity. The developed model proposed in the study is applied on a real-life dataset to detect and prevent different attacks. In addition, short-term attacks that went undetected in the database were also observed. The working times and performance analyzes of the methods were also made. The success of the proposed new intrusion detection and prevention model is shown comparatively. When the results obtained are compared, it has been observed that the proposed model gives very successful results in preventing security problems in distributed systems IoT devices.

With the increase in the number and use of IoT devices in many areas, security problems have occurred in IoT devices. Within the scope of this study, first of all, attacks on IoT devices were mentioned and studies in the literature were mentioned. Secondly, a hybrid (RF+ANN) model was developed to detect possible attacks and the results of the proposed model were applied on three different real datasets. The success of the proposed method on all three datasets was presented by comparing it with metrics such as Accuracy, Precision, Recall and F1Score.

2. Related Works

Existing security mechanisms for distributed system technologies are insufficient due to the complexity and diversity of IoT devices [7]. Therefore, stronger and customized security mechanisms are needed for the security of IoT devices. Data privacy and security in these systems has become a major problem due to problems such as collecting, storing, processing and sharing user data. These issues include security protocols, user privacy, detection and prevention of attacks against IoT devices, scalability of IoT devices, and new security threats to IoT [8]. Therefore, stronger data privacy and security measures must be developed for IoT devices. There are many vulnerabilities in the security of IoT devices in distributed system technology and these devices are vulnerable to cyber attacks [9]. Security issues in IoT devices are due to device constraints such as low power consumption, limited processing capacity, and limited memory Many international organizations and standards organizations are developing security standards for IoT devices. Current challenges regarding the security of IoT include detection and prevention of cyber attacks, encryption of communication between devices, secure software updates. Although studies are carried out in the literature to ensure security in IoT, these studies are limited in order to solve the problems that arise. Important works done in recent years are presented below:

Table 1. Studies in the literature

Study	Technique	Attack Types	Performance	Disadvantages
Jaber et al. [10]	Autoencoder Network Model Genetic algorithm	DDoS	90.26%	the study brings reliability and computational complexity for very large data. For these reasons, the proposed model needs to be further optimized in order to increase its performance and reduce computational complexity.
Moudoud et al. [11]	Hidden Markov Model	False Data Injection (FDI)	97%	Although the performance is good, the model needs to be improved as it takes too much time to integrate into the learning process and new technologies.
Labiod et al [12]	Multi-layer sensor	DDoS	99.99% other attacks 86%	More improvements are needed to detect all types of attacks.
Habiba et.al [13]	Deep learning-based algorithm	DDos	99.99%	In this study, the efficiency of Edge AI for IoT platforms was demonstrated. Only DDos attacks are covered.
Alotaibi and Ilyas [14]	Ensemble method	TON-IoT dataset	98.63%	A multi-classification approach is required that can detect anomalies and intrusions in IoT network traffic.

Mahmoud et.al [15]	Intrusion Detection Systems based on Machine and Deep Learning (IDS_MDL)	Botnet attack dataset	99.7 %	Short-term observed attacks need to be detected
Sun et.al [16]	SVM	SCT dataset	98.71%	More features can be added to determine the physical behavior of more systems to improve the SVM method.
Elsayed et.al [17]	Secured Automatic Two-level Intrusion Detection System (SATIDS)	ToN-IoT and InSDN datasets.	96.35 %	Short-term attacks that occur during the day need to be detected and applied to the real-world IoT.
Sasikala et.al [18]	Logistic Regression	NB15 data set	97.8 %	Inaccurate missing data can be improved by using various approaches.
Jasim et.al [19]	Convolution Neural Network (CNN)	Unspecified	99.18 %	In terms of time, the proposed model works slowly
Almiani et.al [20]	Deep recurrent neural network	NSL-KDD	98.27 % (for only DoS attacks)	Probe is more susceptible to DoS attacks than detection of Remote to Local (R2L) and User to Root (U2R) attacks.
Kareem et.al [21]	Metaheuristic Algorithms	NSL-KDD, CICIDS-2017, UNSW-NB15 and BoT-IoT	95.5%, 98.7%, 81.5%, and 81.5% in the NSL-KDD,	Optimization of hypermaterials is required in solving multi-

			CICID2017, UNSW-NB, and BoT-IoT datasets, respectively	objective problems.
Pehlivanoğlu et.al [22]	Gradient Boost classifier	Bot_IoT and ToN_IoT datasets	Port scanning attacks about 99%	The model proposed in the study should be tested on different data sets, especially large data sets.
Kozik et.al [23]	Deep learning	IoT-23 about 90%	(precision)	By adjusting the hyperparameter, the success and performance of the study can be increased.
Gökdemir and Çalhan [24]	Long Short-Term Memory (LSTM)	IoT dataset 99.17 %		There are some disadvantages in terms of working time.

Gökdemir and Çalhan [24]	Long Short-Term Memory (LSTM)	IoT dataset 99.17 %		There are some disadvantages in terms of working time.
Ölmez ve İnce [25]	Support Vector Machines (SVM)	Bot-IoT, CICIDS-2017, IoT-23 and N-BaIoT	99.94% for Bot-IoT dataset, 99.95% for CICIDS-2017 dataset, 99.96% for IoT-23 dataset and 99.92% for N-BaIoT dataset	Data preprocessing and converting the data into the appropriate format is a separate workload.
Yaman and Tekin [26]	XGBoost	"Brute force ftp", "brute force ssh", "dos http flood", "dos icmp flood", "dos syn flood", "syn scan" and "udp scan"	92.55%	The detection success of some attacks can be increased.

In our study, it is aimed to detect and prevent all kinds of attacks that may occur in order to ensure IoT security in distributed systems. By using artificial neural networks, proposed random forest tree and hybrid method, an attack detection system has been implemented on banks and two different financial data. Our work includes intrusion detection comprehensive and a significant contribution to

detecting different similar attacks and integrated in IoT environments. In addition, performance measurements of all models used are presented comparatively using different evaluation metrics (precision, recall, F1 Score).

3. Material and Method

Currently, intelligent environments are spreading with the IoT in all areas where computing resources are applied. However, as the techniques to exploit computer infrastructure security vulnerabilities are constantly evolving, applications and systems may inevitably be exposed to some attacks. With attacks, accessing systems, obtaining confidential information and improper use, rendering resources unusable can become unavoidable. For this reason, security becomes the most important issue for the IoT environment. Unresolved challenges regarding the security of IoT devices can be summarized as follows [34]:

Poor security practices: Unlike traditional computers, IoT devices can often have low resources and limited processing power. This can result in poor security practices being used to secure IoT devices.

Weak authentication methods: Authentication methods used in communication between IoT devices can often be insufficient or weak. This may require the development of new and more secure authentication methods for the security of IoT devices.

Hard-to-protect physical locations of IoT devices: IoT devices can be used in a variety of physical environments, and these environments can threaten the physical security of devices. Therefore, for the security of IoT devices, security issues arising from the physical location of devices that are difficult to protect may need to be addressed. Data privacy issues: IoT devices can often handle sensitive data, and it's important to protect the privacy of this data. Therefore, it may be necessary to develop new and more secure solutions to data privacy problems for the security of IoT. Some security measures taken are not sufficient to detect and prevent attacks. In order to detect attacks in IoT environments, a model that can be implemented on real platforms and in real time is needed.

Traditional pain detection systems have difficulties to cope with and implement the above-mentioned security challenges. IoT components' limited computing power, large number of interconnected devices and objects, data sharing between users, and vulnerability to security and privacy risks. Difficulty in detecting attacks such as DoS, DDoS, Man in the Middle, unsafe connections, malicious code injection livable. In our study, a model tested on real data is presented that takes into account the context of intrusion detection and prevention in IoT-based environments. The performance of the proposed RF model and the proposed hybrid model ANN model were calculated based on different evaluation criteria

3.1. ANN

ANN have been developed by taking advantage of the working principle of the human brain. An artificial neural network consists of neurons (nodes) that make connections between input and output data, and layers that make connections between these neurons [35]. A neural network consists of a large number of neurons and layers. There are many neurons in each layer, and each neuron receives information by weighted connections that connect nerve cells. This information is passed through different activation functions and transmitted to other neurons. In addition, each neuron receives input from many nodes and passes these inputs through a function and transmits its output to other neurons. While the neurons of all layers that will form in the neural network are in series with the neurons of the previous and next layer, they are parallel among themselves. Since information will be transmitted between neurons to each layer, it contains connections between them. These links represent the weights between the layers and end at the output layer. Determining the correct weights during the training of the network is crucial for accurate and reliable output. Initially, weight values are randomly assigned. Weight values are updated in each iteration according to the neural network and learning rule to be formed. When different samples are trained according to the resulting network, the weights are changed again and the most accurate weights are tried to be determined. These processes are repeated until the optimum weights are found.

Figure 1 shows the structure of the artificial neural network. In our study, the input size was determined by the number of samples and the number of features for each data set. The N value expressed here includes the number of samples. There are 3 hidden layers for each data set. There are 150 neurons in the first layer, 120 neurons in the second layer, and 100 neurons in the last hidden layer. The learning coefficient was taken as 0.01. In each dataset, the number of output layers is determined by the general classes (5 attack types). There are also similar attacks within these classes. Since the number of similar attacks, which are subclasses, is too high, their names are not included.

Classes of test samples are tried to be determined according to the training data determined in the ANN model [36]. If the classes of test data are correctly

determined, the network is considered to be properly trained. The correct determination of the ANN model depends on the activation function, the aggregation function in which the artificial nerve cells transmit information, the learning rule and the topology of the network. The aggregation function calculates the net input of the ANN model using different functions. The activation function, on the other hand, converts the information obtained as a result of the addition function to the output value by using different functions. In some cases, the created ANN model can be memorized. If this happens, training should be stopped. If there is no memorization and the network is learning, the training is not continued. The classes of test samples are determined according to the last iteration. ANN model is created by considering the following criteria [37]:

• Identifying layers
• Inputs and outputs to be used in education
• Number of neurons
• Activation functions
•The objective function used to determine whether the network is successful.

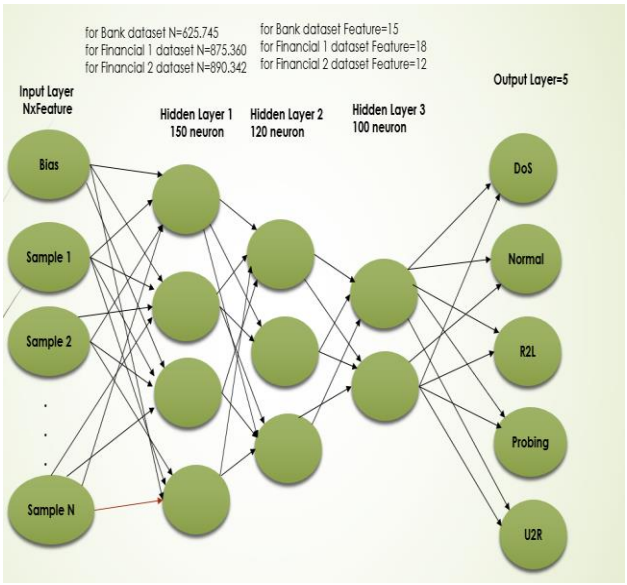

Figure 1. Proposed of ANN model

In ANN models, the structure of a pre-trained network can be used for other datasets. The purpose of transfer learning methods is to increase classification performance and training speed by using previously trained models on a new dataset [38]. If the dataset to be trained is small and insufficient, the attributes of previously trained large datasets can be directly applied to this dataset. The output layers of these trained large datasets directly form the network structure of the dataset to be trained. In addition, it can be learned in a short time and with fewer parameters by fine-tuning it to make it more suitable for the dataset to be trained. During the training phase, the weights of the pre-trained layers are precisely adjusted and updated, and the classification layers are learned in accordance with the training. Fine-tuning the model's layers and newly added classifier layers by training them simultaneously allows the model to be improved. It is to create the local minimum value of the parameters (parameters used in activation, training and performance functions, etc.) that best suit the ANN network structure. Thus, classification success is increased by fine-tuning. This fine-tuning method can be done by backpropagation

3.2. RF

RF method is an ensemble learning method built on Decision Trees (DT) [39]. Community learning methods are popular in the field of machine learning. It can be used for classification and regression operations. As a basic principle, ensemble learning methods decide on the solution of a classification problem depending on the decisions of more than one classifier. In decision making, the highest vote (majority) or the lowest error (minimized error) approaches can be preferred.

In the RF method, there is a forest structure consisting of individual decision trees for the same problem [38]. The decision of each tree in the forest is independent of the decisions of other trees. The randomness in the method is due to the selection of variables in the creation of a decision tree. The traditional decision tree consists of root, twig (brunch) and leaves (leaf). While the root and branch structures are determined according to the variables that will determine the classification result, the leaves show the class decisions. The information gain of the variable is calculated in the selection of the variable for the roots and branches in the traditional decision tree. One of the most used calculation methods is the GINI index. Information gain is the effect of the variable on the class decision. The higher this effect, the higher the effect of the variable on the outcome. The root of the decision tree is established with the variable with the highest information gain and the branches are completed according to the gain order of the variables. In the RF method, variable selections are made randomly while creating each tree. This randomness is useful in preventing overfitting. At the same time, the samples to be selected for each tree created are selected as a subset of the whole sample with the bootstrap technique with a random approach. The RF method was developed by Breiman [40, 41]. Breiman, who previously developed the Classification and Regression Trees (CART) method,

developed the RF technique by first developing Bagging- Bootstrap Aggregating and then Random Subspace methods and combining these methods because the CART method, which has a very good learning ability, is prone to excessive learning. Our proposed RF model proposes a cloud-random forest (IoT - RF) model that combines the IoT environment and random forest for intrusion detection (Figure 2). In this model, based on traditional CART, a weight determination algorithm based on the cloud model and the decision-making trial and evaluation lab is applied to obtain the evaluation weights. The feature weight and the gain value of the smallest Gini coefficient corresponding to the same feature are weighted and summed. The weighted sum is then used to replace the original gain value. This value rule is used as a new CART node split criterion to create a new decision tree, thus creating a new random forest, i.e. IoT - RF. In the proposed RF model according to the IoT environment, the weight is determined by the following steps:

Step 1: Attack types are determined as A={a1,a2,....,aN}.

Step 2 Draw the correlation diagram between any two attack types.

Step 3 The first matrix that directly affects the attack type is created.

Step 4 The property values are determined.

Step 5 The IoT matrix of all decision makers is collected.

Step 6 The matrix of the relationship between the features is found.

Step 8 The standardized comprehensive impact matrix is measured.

Step 10 All attack types have weights.

Classification success in RF methods varies depending on hyperparameters such as the number of trees, tree depth, number of features and training rate [42]. The accuracy of the model can be increased by fine-tuning these hyperparameters manually or with optimization methods such as grid search and random search. Too many features may reduce classification performance. For this reason, the success of the model can be increased by removing features that are not related to feature reduction methods such as information gain or gain ratio. Increasing the number of trees, which affects the success of RF methods, may have a positive effect on the success of the model. Although the success of the model increases as the depth of the tree increases, the success of the model does not change when the depth of the tree reaches a certain point. Additionally, choosing the tree depth too small may cause model incompatibility. For this reason, it is very important to fine-tune these hyperparameters that affect the classification success of RF methods.

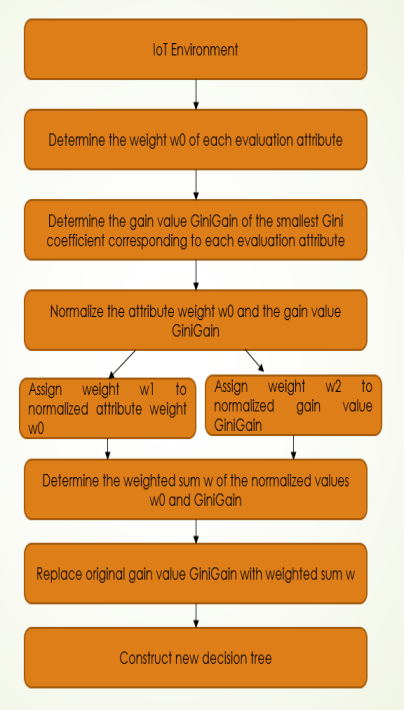

Figure 2. Proposed RF flow chart

3.3. Hybrid Model

The data sets used in the study are not open access and the study was carried out in IoT environments. Due to the confidentiality of the data, fields and features of the data set cannot be mentioned. All datasets used in the study are divided into 70% train and 30% test. Different analyzes
were also performed (60% train, 40% test or 80% training and 20% test). However, it gave the best and most optimum solution when we divided it into 70% training and 30% testing.

IoT has attracted a lot of attention in recent years in many fields such as automation, industry and smart environments. There are many different and similar types of attacks [43]. However, since IoT is a heterogeneous environment, it is also exposed to many attacks. For this reason, a hybrid model has been developed in our study to ensure IoT network security. This hybrid model is formed by combining the proposed RF and ANN model, and the flowchart is shown in Figure 3. In the hybrid model we recommend, first data with the classification model. We have classified. In studies on trained data, observed noise removal, feature extraction, and selection methods increased classification success and. These methods make the available data more useful. Therefore, the results of the studies vary

according to the situation. normalization techniques, noise removal, feature extraction and feature selection. We are at every stage determined the accuracy rate of normal and attack types, and calculated the average. Clustering of datasets is done on the basis of the variables and criteria of the proposed RF model. Then, the ANN classifier is applied sequentially to each clustered dataset. This hybrid model we propose is formed by combining the developed RF and ANN classification models. We propose a hybrid machine learning system based on RF and ANN for detecting all types of attack on bank and financial datasets. The performance evaluation and error rates of the existing models and the proposed hybrid model in the diagnosis of normal and all attack types were calculated and compared with the proposed models in terms of accuracy and time.

In the hybrid model we proposed, we first grouped the data close to each other with the proposed RF method. We performed classification process on the data we grouped with various methods. At each stage, we determined the accuracy rate of normal and preventable attack types with the model we suggested and calculated the mean. It is calculated by averaging the correct and normal behaviors found here and the types of attacks that can be prevented with the model we propose. In hybrid model we recommend, it provides data protection by better detecting attack types compared to single classification or single clustering methods. It gave more successful results. The hybrid model made more successful results than the ANN and RF method. In addition, the reason for the decrease in the success rate as the amount of data increases is due to the multidimensional and complex structure of the data. When the results are compared in all models, the hybrid model in ANN is more successful than the first hybrid model. The reason for this is that when clustering is performed first, normal behaviors and different attack types are clustered in a group as far away from each other. This ensures correct detection and prevention of attack types

Figure 3. Proposed hybrid model flow chart

4. Results and Discussion

In our study, bank and financial data were used to ensure IoT security for distributed systems. The features of the dataset we used are given in Table 2. In addition, normal and attack types and numbers are shown for all three data sets. These attacks are grouped into 4 main groups as DoS, Probing, U2R, R2L and these attacks also consist of similar attack types. Within each attack cluster, there are subclasses containing between 10 and 30 similar attack types. Similar attacks were gathered in the same cluster. Table 3 shows the number of samples in each data set according to attack types. As can be seen, the number of samples with DoS attacks is higher for all three data sets than for other types of attacks.

In data preprocessing, it was first checked whether there were null values in all three datasets. Empty values are filled by taking the average of the column. Additionally, outliers in the datasets were normalized between [0,1]. Thus, data imbalance in the datasets was eliminated. The normalization method was applied to increase the accuracy of the classical model and the proposed models. By examining the correlation matrix of the features in three datasets, the connections between the features were tried to be determined. The most important feature was chosen as the feature with the highest correlation value. These features were used in the proposed classification methods to increase the success. In addition, dimensionality reduction was performed by removing features with low correlation from the data set. The success and performance of the models used were increased by applying the min-max normalization technique to some features. Thus, the classification success of multi-class datasets has been increased by feature selection and dimensionality reduction.

Each attack occurring in these datasets is categorical. However, these categorical classes have been converted to integer data type, starting from one. In addition, the success of the models used in the study was increased by adding an extra new feature to include similar attacks in the same data set. Attacks were also investigated according to their protocol types.

In Figure 4, the accuracy results obtained in the detection of all attack types for all datasets of the artificial neural networks developed are given. In our study, unlike other studies, important results were obtained in detecting all types of attacks. In Figure 5, the values of the detection of all attack types in terms of time are given for all datasets of the artificial neural networks developed. All results are calculated based on the accuracy of the test data and the running time of the test data.

Table 2. Used Datasets

	Number of Samples	
Dataset	**Normal**	**Attack**
Bank	625.745	178.034
Financial1	875.360	279.693
Financial2	890.342	300.476

Table 3. Number of samples by attack types

Dataset	**DoS**	**Probing**	**U2R**	**R2L**
Bank	87.341	12.634	24.677	53.382
Financial1	184.302	48.072	34.974	12.345
Financial2	108.342	85.974	81.366	24.794

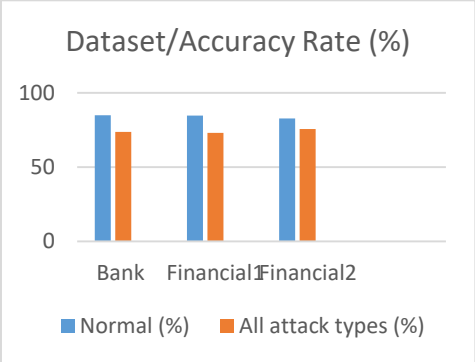

Figure 4. ANN accuracy results for all datasets

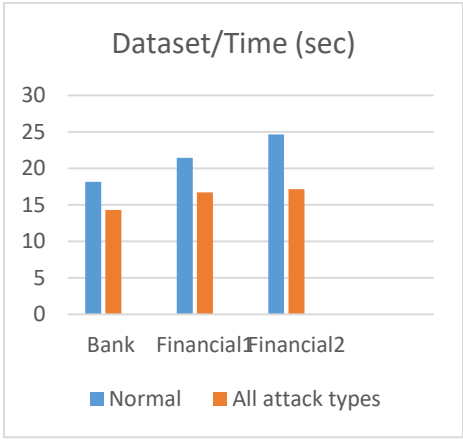

Figure 5. ANN time results for all dataset

Table 4. ANN results by different evaluation metrics

Dataset/ Performance Criteria (%)	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Bank	73.64	73.50	74.42	74.10
Financial 1	73.05	72.82	72.14	72.45
Financial 2	75.64	73.64	72.42	73.02

In Table 4, the results of the ANN according to different evaluation metrics for all three data sets are given. It gave more successful results for bank data compared to all datasets. The reason for this is that the examples of DoS attacks are less than other datasets. Because the subclasses of DoS attacks are more than other attacks. In short, since there are many similar attack types, it is more difficult to detect DoS attacks when compared to other types of attacks when we look at the studies in the literature in general

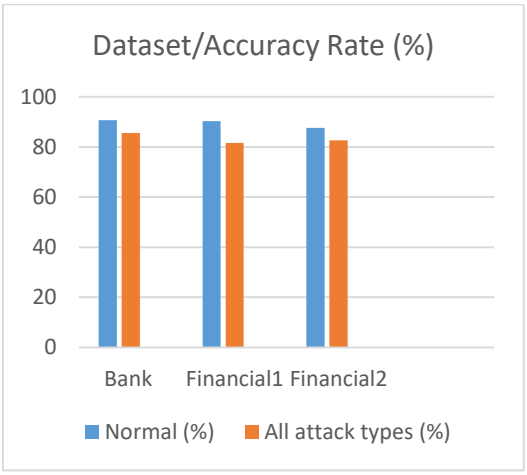

Figure 6. Proposed RF accuracy results for all datasets

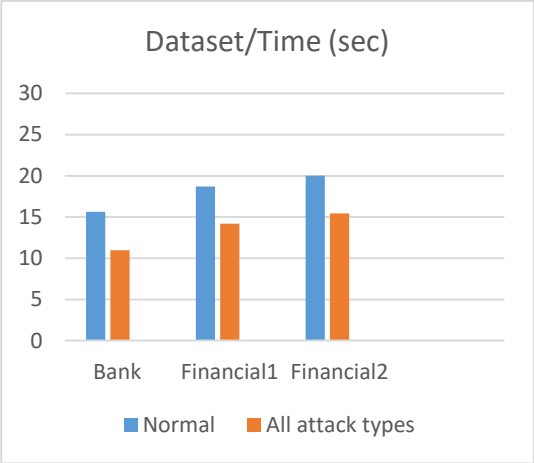

Figure 7. Proposed RF time results for all dataset

Table 5. Proposed RF results by different evaluation metrics

Dataset/ Performance Criteria (%)	Accuracy	Precision	Recall	F1-Score
Bank	85.60	86.73	84.72	84.99
Financial 1	81.67	84.60	83.15	83.01
Financial 2	82.64	82.62	85.03	85.07

In Table 5, the results of the proposed RF according to different evaluation metrics for all three data sets are given. In Figure 6, the accuracy results obtained in the detection of all attack types for all datasets of the proposed RF developed are given. In our study, unlike other studies, important results were obtained in detecting all types of attacks. In Figure 7, the values of the detection of all attack types in terms of time are given for all datasets of proposed RF developed.

In Figure 8, the accuracy results obtained in the detection of all attack types for all datasets of the Hybrid model developed are given. In our study, unlike other studies, important results were obtained in detecting all types of attacks. In our future work, we aim to increase the accuracy by using different hybrid methods and to use it in different real datasets. In Figure 9, the values of the detection of all attack types in terms of time are given for all datasets of hybrid model developed. When we compare it with other studies in terms of time, we see that it detects the attack in a short time.

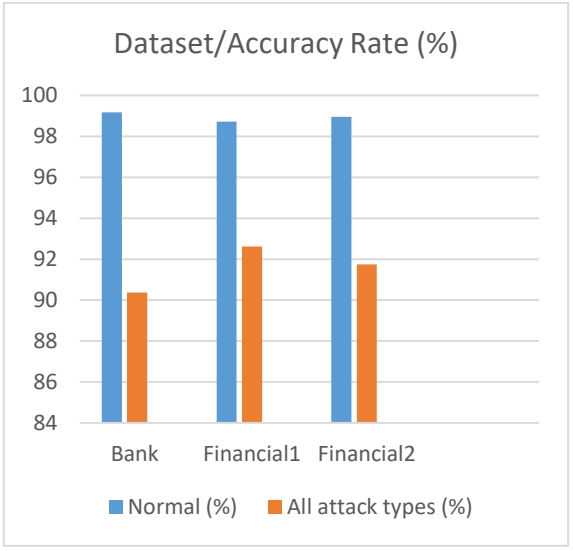

Figure 8. Hybrid model accuracy results for all datasets

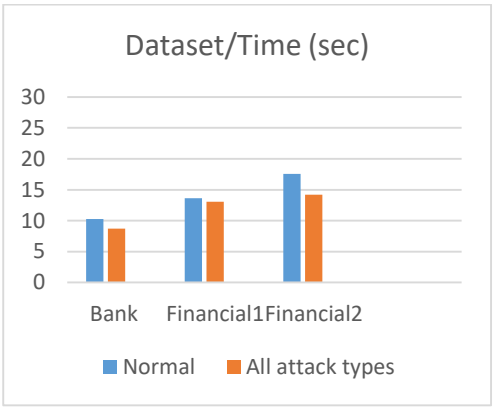

Figure 9. Hybrid model time results for all dataset

Table 6. Hybrid results by different evaluation metrics

Dataset/ Performance Criteria (%)	Accuracy	Precision	Recall	F1-Score
Bank	90.37	91.36	91.36	91.07
Financial 1	92.62	92.62	92.43	91.45
Financial 2	91.74	91.43	93.78	92.72

Table 6 presents the results of the proposed hybrid model according to different evaluation metrics for all three data sets. In the first stage of our study, the data trained with ANN was classified with RF. In addition, data pre-processing steps such as feature extraction and noise removal increase the success in classification. For this reason, the results of the studies vary depending on the data processing steps. We determined the accuracy rate of normal and attack types at each stage and calculated their average.

In our study, k-fold cross validation was used to ensure that the methods used, other than voting, were

resistant to the training and testing sections. 5 and 10 were chosen for the K value. While the solution is produced according to the voting technique, the ANN results for k - fold cross validation for k=5 and k=10 have been added in the accompanying Table 7. When

the results were compared, using k-fold cross validation gave more successful results than the voting technique. Additionally, in general, k=10 is more successful than k=5 in all methods used in the study.

Table 7. ANN results according to k-fold craoss validation method

Dataset/ k value	k=5	k=10
Bank	75.45	77.74
Financial 1	75.97	76.37
Financial 2	78.95	77.34

While the solution is produced according to the voting technique, the RF results for k-fold cross validation for k=5 and k=10 have been added in the accompanying Table 8.

Table 8. RF results according to k-fold craoss validation method

Dataset/ k value	k=5	k=10
Bank	88.04	88.67
Financial 1	85.34	86.72
Financial 2	85.17	87.95

While the solution is produced according to the voting technique, the hybrid model results for k-fold cross validation for k=5 and k=10 have been added in the accompanying Table 9.

Table 9. Hybrid model results according to k-fold craoss validation method

Dataset/ k value	k=5	k=10
Bank	92.32	93.01
Financial 1	93.71	92.37
Financial 2	92.82	93.54

5. Conclusion and Suggestions

Connecting IoT devices to the internet can pose a significant threat to the security of the devices. IoT devices have many different security vulnerabilities, such as malware infections, phishing attacks, and attacks through physical access. These threats can affect both consumer and industrial IoT devices and have serious consequences.

Distributed system technologies are a structure that is widely used in many areas today and allows services to be offered in a more efficient and scalable way. However, these systems have a more complex structure than central systems. In this study, we evaluated the distributed system technology for IoT and showed the distributed system architecture in detail. The security of IoT devices is very important for distributed systems technology. IoT devices, their usage areas and numbers are increasing day by day and the security of these devices is of great importance in terms of personal privacy and data security. In our study, different security methods and mechanisms that can be used in distributed system technology for the security of IoT devices are discussed. However, it was also emphasized that more research is needed to find solutions to security problems in IoT devices. Therefore, those working on the security of IoT devices need to identify different security vulnerabilities and risks and ensure the security of users by taking appropriate security measures.

The security of distributed systems is a more challenging issue than centralized systems because communication between different components has difficulty in ensuring security. Therefore, various precautions should be taken for the security of distributed systems. Among them; topics such as authentication, authorization, encryption, data integrity, confidentiality and reliability. Today, many research and development studies are carried out on the security of distributed systems. These studies provide important steps for reducing the risks of cyber security vulnerability, preventing attacks and making systems more secure. As a result, distributed system technologies and security are an increasingly important topic for IoT today. The development and security of these technologies will increase the efficiency of enterprises and enable them to provide better service. For this reason, it is very important to raise awareness about the security of distributed systems and to ensure the continuation of research studies.

In this study, a system is designed to detect and prevent different and same types of attacks that occur as a result of security vulnerabilities in distributed systems. The aim of our study is to ensure attack security in the IoT environment. The three different data sets we used were evaluated using ANN, RF and our proposed hybrid model. The results of the proposed hybrid model and the proposed RF model and the classical ANN model are shown comparatively. In addition, the operating times of all

models used in the study were calculated and compared. ANN for bank, financial 1 and financial 1 data sts are 0.7364, 0.7305, 075.64, respectively; RF 0.85, 0.8167, 0.8264; hybrid model showed success of 0.9037, 0.9262 and 0.9174. In addition, the results for all three models were analyzed with different evaluation criteria. When the results are compared, the proposed hybrid model showed approximately 10% more success than the ANN model in all three datasets. When the results are compared, it is observed that more accurate attack prediction is made in hybrid models. The effective use of hybrid models in attack detection and prevention is of great importance for future studies. Using artificial intelligence techniques at the beginning of these studies provides a greater advantage. Our study achieved very successful results in detecting different and new attack types occurring in IoT environments. In addition, the proposed hybrid model aims to address security and privacy issues that may arise in different areas in the future. Unlike previous studies, data security is ensured in all operations for IoT.

In the future, it is aimed to design attack detection and prevention systems on different and larger complex data sets with deep learning methods. It will be aimed to contribute to the studies in the literature by conducting more studies in this field. In addition, it is aimed to increase the success of the models by automatically fine-tuning the hyperparameters for the proposed models in the future using optimization algorithms.

Contributions of the Authors

All authors contributed equally to the study.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics

References

[1] M. Wang, and Q. Zhang, "Optimized data storage algorithm of IoT based on cloud computing in distributed system," *Computer Communications*, vol. 157, pp.124-131, 2020.

[2] G. Eleftherakis, D. Pappas, T. Lagkas, and K. Rousis, "Architecting the IoT paradigm: a middleware for autonomous distributed sensor networks," *International Journal of Distributed Sensor Networks*, vol. 11, no.12, pp.139735-139735, 2015.

[3] X. Yu, J. Chu, and K. Yu, "Energy-efficiency optimization for IoT-distributed antenna systems with SWIPT over composite fading channels," *IEEE Internet of Things Journal*, vol. 7, no.1, pp. 197-207, 2019 .

[4] A. R. Sfar, E. Natalizio, Y. Challal, and Z. Chtourou, "A roadmap for security challenges in the Internet of Things," *Digital Communications and Networks*, vol. 4, no.2, pp. 118-137, 2018.

[5] S. Keoh, S. Kumar, and H. Tschofenig, "Securing the internet of things: A standardization perspective," *IEEE Internet of things Journal*, vol. 1, no.3, pp. 265-275, 2014.

[6] K. Jaswal, T. Choudhury, and R. Chhokar, "Securing the Internet of Things: A proposed framework," *In 2017 International Conference on Computing, Communication and Automation (ICCCA),* 2017, pp. 1277-1281.

[7] P. Sivaraman, C. Sharmeela, P. Sanjeevikumar, "Health Monitoring of a Transformer in a Smart Distribution System using IoT," *In IoT, Machine Learning and Blockchain Technologies for Renewable Energy and Modern Hybrid Power Systems River Publishers*, pp. 79-91, 2023.

[8] G. Bhandari, A. Lyth, and A. Shalaginov, "Distributed Deep Neural-Network-Based Middleware for Cyber-Attacks Detection in Smart IoT Ecosystem: A Novel Framework and Performance Evaluation Approach," *Electronics*, vol. 12, no. 2, pp. 298-298, 2023.

[9] A. Ukil, J. Sen, and S. Koilakonda, "Embedded security for Internet of Things," *In 2011 2nd National Conference on Emerging Trends and Applications in Computer Science,* 2011, pp. 1-6.

[10] M. H. Ali, M. Jaber, and S. Abd, "Threat analysis and distributed denial of service (DDoS) attack recognition in the internet of things (IoT)," *Electronics*, vol. 11, no.3, pp. 494-494, 2022.

[11] H. Moudoud, Z. Mlika, L. Khoukhi, and S. Cherkaoui, "Detection and prediction of fdi attacks in iot systems via hidden markov model," *IEEE Transactions on Network Science and Engineering*, vol. 9, no.5, pp. 2978-2990, 2022.

[12] Y. Labiod, Y, A. Korba, and N. Ghoualmi, "Fog computing-based intrusion detection architecture to protect iot networks," *Wireless Personal Communications*, vol. 125, no.1, pp. 231-259, 2022.
[13] M. Habiba, M.R. Islam, S.M. Muyeen, and A.S. Ali, "Edge intelligence for network intrusion prevention in IoT ecosystem," *Computers and Electrical Engineering*, vol.108, pp.108727-108727, 2023.
[14] Y. Alotaibi, and M. Ilyas, "Ensemble-Learning Framework for Intrusion Detection to Enhance Internet of Things' Devices Security," *Sensors*, vol. 23, no. 12, pp. 5568-5568, 2023.
[15] W. A. Mahmoud, M. Fathi, H. El-Badawy, and R. Sadek, R, "Performance Analysis of IDS_MDL Algorithm to Predict Intrusion Detection for IoT Applications", *In 2023 40th National Radio Science Conference (NRSC)*, vol. 1, 2023, pp. 139-149.
[16] C. Sun, D. J. Cardenas, A. Hahn, and C. Liu, "Intrusion detection for cybersecurity of smart meters", *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 612-622, 2020.
[17] R. A. Elsayed, and R.A.Hamada, "Securing IoT and SDN systems using deep-learning based automatic intrusion detection", *Ain Shams Engineering Journal*, vol. 14, no. 10, pp.102211- 102211, 2023.
[18] K. Sasikala, and S. Vasuhi, "Anomaly Based Intrusion Detection on IOT Devices using Logistic Regression", *In 2023 International Conference on Networking and Communications (ICNWC)*, 2023, pp. 1-5.
[19] A. F. J. Jasim, and S. Kurnaz, "New automatic (IDS) in IoTs with artificial intelligence technique", *Optik*, vol. 273, pp.170417-170417, 2023.
[20] M. Almiani, A. AbuGhazleh, and A. Razaque, "Deep recurrent neural network for IoT intrusion detection system", *Simulation Modelling Practice and Theory*, 101, 102031, 2020.
[21] S. S. Kareem, R. R Mostafa, F.A. Hashim, and H. M. El-Bakry, "An effective feature selection model using hybrid metaheuristic algorithms for iot intrusion detection", *Sensors*, vol. 22, no. 4, pp. 1396-1396, 2022.
[22] M. K. Pehlivanoğlu, A.Kuyucu, K.A. Recep, "IoT Veri Kümelerinde Makine Öğrenmesi Tekniklerine Dayalı Saldırı Tespiti", *Avrupa Bilim ve Teknoloji Dergisi*, vol. 52, pp. 19-26, 2023.
[23] R. Kozik, M. Pawlicki, and M. Choraś, "A new method of hybrid time window embedding with transformer-based traffic data classification in IoT-networked environment", *Pattern Analysis and Applications*, vol. 24, no. 4, pp. 1441-1449, 2021.
[24] A. Gökdemir, and A. Calhan, "Deep learning and machine learning based anomaly detection in internet of things environments", *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 37, no. 4, pp. 1945-1956, 2022.
[25] E.G. Ölmez, and İ. Kenan, "IoT Botnet Verisetlerinin Karşılaştırmalı Analizi", *Computer Science*, 151-164, 2022.
[26] O. Yaman, and R. Tekin, "Akıllı Ev Sistemleri için XGBoost Tabanlı Saldırı Tespit Yöntemi", *Journal of Intelligent Systems: Theory & Applications*, vol. 6, no. 2, 2023.
[27] Ş. Okul, and M. A. Aydın, "Security attacks on IoT", *In 2017 International Conference on Computer Science and Engineering (UBMK)*, 2017, pp. 1-5.
[28] A. A. Ismael, and A.Varol, "IoT Sistemini Güvenliği: Yeni Bir Model", *5th National Informatics Congress*, 2018.
[29] K. İlhan, and Ş. Abdülkadir, Ş. "IoT Ağ Güvenliği için 802.1 x, DMZ ve SSL-VPN Birleştirme Tabanlı Etkili bir Güvenlik Yöntemi". *Acta Infologica*, vol. 4, no. 2, pp. 65-76, 2020.
[30] J. Azimjonov, and T. Kim, "Designing accurate lightweight intrusion detection systems for IoT networks using fine-tuned linear SVM and feature selectors", *Computers & Security*, vol. 137, no. 103598, 2024.
[31] P. Vijayan, and S. Sundar, "Original Research Article IoT intrusion detection system using ensemble classifier and hyperparameter optimization using tuna search algorithm", *Journal of Autonomous Intelligence*, vol. 7, no.2, pp. 1-10, 2024.
[32] A. Biju, and S.W. Franklin, "Evaluated bird swarm optimization based on deep belief network (EBSO-DBN) classification technique for IOT network intrusion detection", *Automatika*, vol. 65, no. 1, pp. 108-116, 2024.
[33] S. Shen, C. Cai, and S. Yu, "Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks", *Applied Soft Computing*, vol. 150, no. 111080, 2024.

[34] M. Abomhara, and G. M. Køien, "Security and privacy in the Internet of Things: Current status and open issues," In 2014 international conference on privacy and security in mobile systems (PRISMS), 2014, pp. 1-8.

[35] J. Park, and Y. S. Jeong, "Dynamic analysis for IoT malware detection with convolution neural network model," *IEEE Access*, vol. 8, pp. 96899-96911, 2020.

[36] S. Smys, H. Wang, and A. Basar, "5G network simulation in smart cities using neural network algorithm," *Journal of Artificial Intelligence*, vol. 3, no. 1, pp. 43-52, 2021.

[37] D. K. Reddy, H. S. Behera, and J. Nayak, "Deep neural network based anomaly detection in Internet of Things network traffic tracking for the applications of future smart cities," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 7, 2021.

[38] W. Pannakkong, K. Thiwa-Anont, K. Singthong, K., and J. Buddhakulsomsiri, "Hyperparameter tuning of machine learning algorithms using response surface methodology: a case study of ANN, SVM, and DBN", *Mathematical problems in engineering*, 2022, 1-17, 2022.

[39] S. S. Roy, S. Dey, and S. Chatterjee, "Autocorrelation aided random forest classifier-based bearing fault detection framework", *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10792-10800, 2020.

[40] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[41] J. Wang, C. Rao, and X. Xiao, "Risk assessment of coronary heart disease based on cloud-random forest", *Artificial Intelligence Review*, vol. 56, no. 1, pp. 203-232, 2023.

[42] H. Parmar, S. Bhanderi, S., and G. Shah, "Sentiment mining of movie reviews using Random Forest with Tuned Hyperparameters", *In International Conference on Information Science*, Kerela, 2014, pp. 1-6.

[43] F. James, "IoT cybersecurity based smart home intrusion prevention system", *In 2019 3rd Cyber Security in Networking Conference (CSNet)*, 2019, pp. 107-113.