Proceedings of the 7th International Conference on Inventive Material Science and Applications (ICIMA-2025)
DVD Part Number: CFP254B2-DVD; ISBN: 979-8-3315-2150-9

IDENTIFICATION OF UNAUTHORIZED ACCESS POINT IN WIRELESS NETWORK USING SUPERVISED MACHINE LEARNING TECHNIQUES

Dr.C.Arul stephen
Associate Professor/Department of Electronics and Communication Engineering
Vels Institute Of Science, Technology, &Advanced Studies, (VISTAS)
Chennai, India
arulstephenc@gmail.com

Dhatsanamurthy S
UG Scholar/Department of Electronics and Communication Engineering
Vels Institute Of Science, Technology, &Advanced Studies, (VISTAS)
Chennai, India
dhasnamoorthysk@gmail.com

Livingston Rahul Raj D
UG Scholar/ Department of Electronics and Communication Engineering
Vels Institute Of Science, Technology, &Advanced Studies, (VISTAS)
Chennai, India
freakylivngstone@gmail.com

Dr.A. Vijayalakshmi
Professor, *Dept. of ECE,*
VelsInstitute of science Technology and Advanced
Chennai, India
vijayalakshmi.se@vistas.ac.in

Dr.R.Kumudham
Associate Professor, *Dept. of ECE,*
VelsInstitute of science Technology and Advanced
Chennai, India
kumudham.se@vistas.ac.in

Dr.R.Chandrasekaran
Assistant Professor, *Dept. of Biomedical,*
VelsInstitute of science Technology and Advanced
Chennai, India
chandrasekar.se@vistas.ac.in

Abstract-- A growing widerange use of wi-fi or mobile hotspots in public are prone to various risks in wireless environment.There is a significant risk of contracting different attacks resulting in falling victim to unauthorized attackers, particularly when utilizing Access points in different government sector office etc. Information protection requires the detection of unauthorized Access points.The main objective of the proposed work is to use machine learning methods to assess the round trip time data set values in order to identify authorized and illegitimate Access points in wireless environments.Three different machine learning algorithms employed in the proposed work are Support vector machine, Random forests and K-Nearest Neighbors and their performance were analysed.The empirical results shows that Random forest algorithm achieves maximum accuracy of 99% compared to other machine learning models.

Keywords:Machine learning algorithms, round trip time, and access point.

I. INTRODUCTION

In today's world, wireless WiFi, also known as "Wireless Fidelity," is technology that has revolutionized internet connectivity and mobility.WiFi essentially uses radio waves to transfer data in wireless medium between devices, doing away with the requirement for conventional connected connections

WiFi is compatible with a wide range of devices and runs in the 2.4 and 5 GHz radio frequency band it widely adopted. Ideal for both residential and commercial situations, it allows users to connect several devices to a network without being constrained by physical wires.The development of wireless routers, access points, and range extenders—which form the foundation of WiFi networks has resulted from the widespread use of WiFi technology. These gadgets enable wireless communication between gadgets, including tablets, laptops, cellphones, and smart home appliances.WiFi is a necessary component for supporting ,online gaming and smart home automation.

II.LITERATURE SURVEY

When developing, maintaining, and assessing wireless networks,protocols and their applications.Latency is a crucial factor to consider in wireless communication .The point-to-point roundtrip time, which calculates the interval between data transmission and the receiption of a favorable acknowledgement .one measure that is frequently used to quantify network is latency[1] . The RTT needs to be measured often because it is used by many applications and protocols to predict network load. The best-known example is Transmission Control Protocol, or TCP[2].

The paper illuststrates that network applications and protocols that use the RTT do not have to worry if they use our prediction technique[3]. In this work, the authors have presented a unique RTT estimate method based on the Experts Framework, a machine-learning-based methodology [4] and [5]. The Expert's Framework uses "online" learning, where the learning process takes place in "trials," as is covered in detail [6]. An aggregate prediction for each trial is provided by several"experts, which is then compared to the actual value (obtained, for example, through measurement) . The next three algorithm iterations employ the revised weights, which are determined by the algorithm using the prediction error to fine-tune each expert's

contribution to the fore cast. The suggested RTT estimation technique to estimate the RTT required by TCP's error and congestion control as an example application is used in proposed work [7].The means of comprehensive simulations and real-time experiments, demonstrate that the machine-learning method is able to adjust more fast to variations in the RTT and, as a result, forecast its value more precisely than the existing EWMA strategy used by the majority of TCP versions[8] and [9] . TCP computes its Retransmission Time-Out timer using the RTT estimate, as explained in methodology .After the RTO expires, the TCP sender retransmits therelated packet because it thinks it was lost. TCP's RTO is dependent on RTT estimates and measurements in order to calculate its value effectively[10] . The return transit time (RTT) is the amount of time that passes between a packet's departure from the sender and the sender's receipt of a positive acknowledgment for that packet[11]. If the RTO is set too long, there could be lengthy idle waits until the sender responds to the supposedly lost message[12] and [13]. On the other hand, if the RTO is set too aggressively—that is, too quickly—it can expire too frequently and force unnecessary retransmissions. Naturally, RTO setting has an impact on TCP performance[14] .
Several methods are utilized to categorize rogue access points, even in unforeseen circumstances [15].The difference, mean, variance and standard deviation of the delay times of each authorized and unauthorized AP are used to choose the feature points for RTT values [16].

There have been algorithmic research on the detection of unauthorized Aps [17]. The aim of project is to form a different algorithm that can be used to detect rogue wireless networks[18].

The following are the machine-learning algorithms and characteristics that were employed for categorization in this work.

1.An ensemble learning technique called Random Forest predictions by combining several decision trees . To arrive at the final forecast, the predictions of all the decision trees are averaged, with each tree being taught on a distinct division of the data.

2. Machine learning techniques like the KNN classifier are applied for regression and classification issues . The method determines the K closest points in the training dataset and predicts the class or value of a new data point based on their presence .

3.Robust machine learning methodology Regression, outlier identification, and linear or nonlinear classification are all handled by Support Vector Machines (SVM) [20]. It can handle high-dimensional data and nonlinear interactions, which makes them flexible and strong in a range of contexts.
The main limitations of existing work is based only on the time delay.The Round trip delay alone cannot be considered as a factor for classifying if it is an authorized or legitimate Access point because hackers use sophistacted methods to deceive legitimate users.Hence in our proposed work all features in dataset with correlations are considered to give better results.

III.RELATED WORK

The increasing wireless environment has led to a greater awareness of threat to wireless (AP) security.There is a greater chance of contracting different viruses and becoming victim to various attacks while utilizing unauthorized wireless access points in business, government, and military settings. For information security, it is essential to identify unauthorized APs. In this project, authorized and illegal access points (APs) in wired and wirelessly integrated environments are identified using the round trip Time (RTT) value. The SVM .Decision Tree Classifier, KNN, and MLP machine learning techniques are used in this work to examine themThe study report concludes that the decision tree classifier method has yielded the highest accuracy.

Owing to the quick development of wireless network-using devices, it is difficult to find locations in our daily lives without wifi. Public institutions, businesses, cafes, military installations, and schools all have easy access to Wi-Fi. Numerous anonymous individuals utilize Wi-Fi, making it challenging to verify each one. Additionally, it might be challenging to identify yourself even while using approved Wi-Fi to tether like a hotspot unless you pay special attention to the settings and look directly at the list of APs (Access Points).An access point, also known as a transceiver, is a data-transmitting and data-receiving station in a (WLAN). In addition to enabling user connections within the network, an access point can strong in the point of connection between a frame wire network and a wireless local area network. RTT (round trip time) values were used in this project to construct a dataset. The best algorithm is determined by comparing the outcomes after the generated data set is used to run the machine learning algorithm. Examine the methods being utilized for unauthorized AP categorization and explain how the experimental setup and the assign values applied in the dataset used for experiment analysis relate to each other.

The network attacks are familiar these days and can compromise service, steal, or misrepresent confidential information belonging to network users.In this study, we have established a method for safeguarding wireless local area networks from unwanted access points..Because it doesn't require any specialized technical knowledge, unauthorized access points have become a well-known security problem in the world of wireless local area networks. There is no requirement for an intrusion device in this kind of attacks. Threats of this nature cannot be stopped if they are not promptly identified and countered. The suggested methodology offers a means for dealing with unapproved access points. Illegitimate users create unauthorized accounts in order to steal the credentials of authorized users. All traffic begins to travel through the illicit user's computer when the genuine clients connect to the rough AP, allowing it to lose confidentiality and the legitimate users' sessions and other data.We worked on this

project to identify and categorize unwanted access points using four distinct classification algorithms in order to protect sensitive information. Furthermore, the analysis of performance data reveals that the Decision Tree Classification method exhibits superior accuracy compared to the SVM, KNN, and MLP algorithms.

Users who are connected to an internet network sometimes worry about unlawful access points.Using Round Trip Time gathered data from all networked node, we aim in this study to identify the unauthorized access point entries.To accurately identify the access points as authenticated or unauthenticated, we will use machine learning models. In this work, our primary goal is to investigate the efficacy of identifying different unauthorized access points using a variety of machine learning methods, including genetic algorithms SVM and KNN .We compared the performance of other algorithms and found that, at 98.15%, the Ant colony Optimization algorithm produced the best results. In the suggested work, network simulation has been used to generate the synthetic RTT dataset for predictive analysis.

We may conclude that (ACO) has the best accuracy of 98.15% based on all of the data. The synthetically generated data is almost exactly like real-world network data. Thus, it can be concluded that the Ant colony Optimization technique should be employed in the production system to locate unlawful access points because it offers a lower false rate and greater accuracy than other models. Further research can investigate various genetic algorithms. According to certain research, ensemble learning techniques offer high-accuracy binary classification utilizing discrete features.

IV.METHODOLOGY

The idea behind the RTT dataset is gathering and evaluating information on Round-Trip Time (RTT), a key performance indicator in wireless networks.The time it takes for a data packet to leave its source, travel to its destination, and then return is known as the RTT. This is a brief explanation of RTT dataset's concept.

Data Collection: Measurements of RTT over a network are used to create RTT datasets. Usually, specialized equipment or software is used to transfer data packets and time how long it takes them to travel to their destination and back to the source.The dataset is collected from Kaggle and the software used for the proposed work is Python using Google colab.The dataset used for training and testing are 70 and 30 % respectively. The behavioural pattern of each features are plotted and based on the results .After the RTT data is gathered, it can be examined to learn more about the behaviour and performance of the network. Finding patterns or trends in RTT values over time, determining the cause of network congestion or latency problems, evaluating the effect of packet size or network protocols on RTT, and establishing a connection between RTT measurements and other network metrics like packet loss and throughput are some examples of analysis.

Table 1:RTT DATASET FEATURES

S.NO	FEATURES
1	fps_mean
2	fps_std
3	rtt_mean
4	rtt_std
5	dropped_frames_mean
6	dropped_frames_std
7	dropped_frames_max
8	bitrate_mean
9	bitrate_std
10	Target
11	Label

Table 1 shows the 11 features in RTT data set out of which 10 are input class and label is the output class.Based on the features and their average values the label is categorized into o and 1 representing authorized and unauthorized access points.

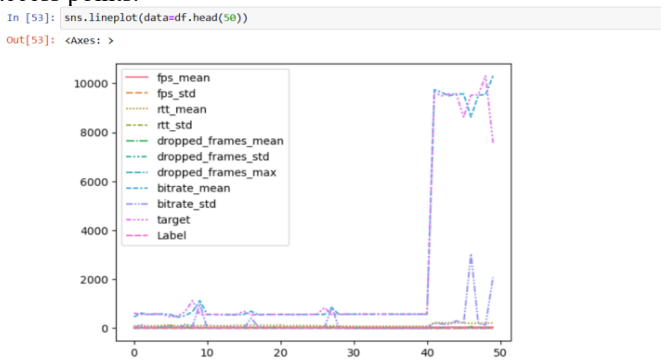

Fig 1 Behavioral pattern of Features

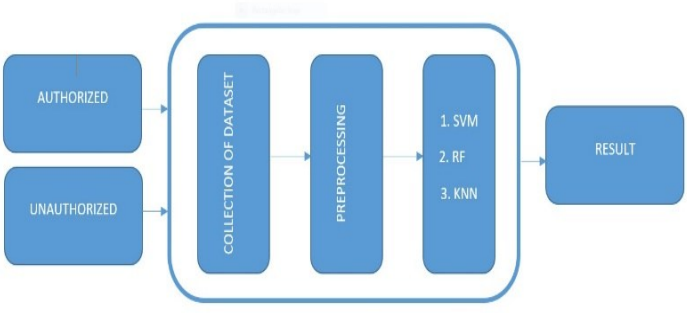

Fig 2:Block diagram of proposed work

The proposed work makes use of an RTT (round trip time) supervised data set with 11 features: bitrate mean, bitrate mean, target, label, dropped frames mean, dropped frames max, fpsmean, fpsstd, rttmean, rttstd, dropped frames mean,

bitrate mean, bitrate mean, and so on. The Fig.1 shows the behavioral pattern of various features used in RTT dataset.The dataset contains 3,79,021 values, which are classified as 0 and 1 by label shown in Table 2.The feature of dropped frames mean value and rttmean value is equal to zero and target value is less than 500; these are determined as 1 (unauthorized) and the others are determined as 0 (authorized).Fig.2 shows the methodology used in the proposed work.

Table 2 Number of authorized and un authorized values in dataset

Value	Type	Total value
1	Unauthorized	377877
0	Authorized	1144

And also using three machine learning random forest algorithm, SVM algorithm and KNN algorithm to test and train the data from the data set.

V.MACHINE LEARNING MODELS

Supervised machine learning algorithms are the backbone of many modern AI systems, allowing computers to learn from labeled data and make predictions or decisions based on that learningThe algorithm gains knowledge from a training dataset made up of input-output pairs in supervised learning. The input is typically a set of features or attributes, while the output is the label or target variable we want to predict or classify [19].

1.Random Forest is a learning method its builds each decision tree in the forest during training using a random subset of features and a random portion of training data.

.In tasks involving classification and regression, it integrates the forecasts from several decision trees to produce predictions that are more trustworthy and accurate.

2.KNN is an algorithm for supervised machine learning that finds the 'k' nearest points in the trained set to a given input data in order to generate predictions.TheKNN , a highest vote determines the projected class for classification tasks. In regression tasks, the average of the k nearest neighbors target values represents the projected value.

3.SVM is a method for supervised machine learning that looks for the optimum hyperplane in a high-dimensional space to divide input points belonging to distinct classes[20]. The hyperplane maximizes the margin between the classes and acts as the decision boundary in classification problems.

VI.PERFOMANCE ANALYSIS

PRECISION

A statistical measure called precision is used to assess how well a classification model performs, especially when dealing with binary classification. It calculates the percentage of all cases that the model properly predicts as positive cases.Precision measures how the model predicts outcomes.

$$\text{Precision} = TP / (TP + FP)$$

RECALL

Recall, which stands for the model's capacity to find all pertinent events, is often referred to as true positive rate or sensitivity. It is compute as TP divided by the sum of TP and FN.The model can efficiently identify the majority of the positive occurrences when it has a high recall.

$$\text{Recall} = TP / (TP + FN)$$

F1-SCORE

The F1-score is a valuable metric for assessing the overall efficacy of a classification model since it offers a single value that accounts for both precision and recall. When there is an imbalance in the dataset between the number of positive and negative cases, it is especially useful.

$$\text{F1-score} = 2\times(P\times R) / (P + R)$$

The performance of a classification model is assessed using a table in machine learning called a confusion matrix. Through a comparison of predicted and actual class labels from the dataset, it makes the model's performance visually evident. A square matrix is commonly used to depict it, with the rows denoting the actual class labels and the columns denoting the anticipated class labels. For every combination of actual and anticipated classes, the number of cases for that combination is contained in each cell of the matrix. Fig 3 shows the confusion matrix of Random forest machine learning model.

Predicted	0	1	All
Actual			
0	234	2	236
1	8	75561	75569
All	242	75563	75805

Fig 3:Confusion matrix-random forest

ACCURACY

The term "accuracy" in machine learning algorithms particularly refers to the model's ability to correctly anticipate outcomes or labels for a given dataset. It measures the proportion of correctly identified cases among all the examined instances. It is imperative to bear in mind that, despite its popularity, accuracy may not always be the optimal machine learning parameter [21], especially when dealing with imbalanced datasets

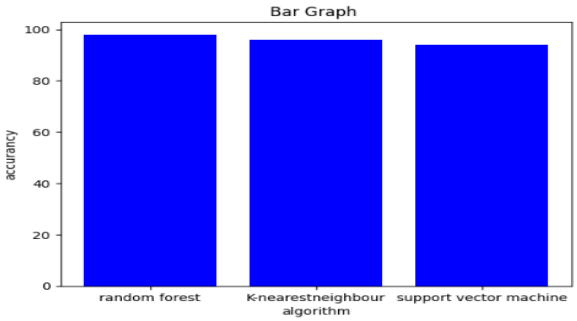

Fig 4 :Accuracy plot

Fig 4 shows the accuracy of various machine learning models used in RTT data set random forest algorithm achieves a maximum accuracy of 99% compared to other two algorithms KNN algorithm and SVM algorithm.

VII .RESULT AND DISCUSSION

Out of all the machine-learning algorithms, the algorithms relevant to categorization were chosen for comparison in this paper are KNN, Random Forest and SVM .Based on the performance analysis of the three different machine learning models used in RTT dataset. Random forests making use of ensemble technique majority votes decide the prediction of attacks and final decision is made taking average of all outcomes from each decision tree. This model achieves a maximum accuracy of 99% compared to other models.

It is imperative to bear in mind that, despite its popularity, accuracy may not always be the optimal machine learning parameter, especially when dealing with imbalanced datasets. A clearer view of a model's performance in specific circumstances may be provided by additional metrics such like the ROC curve (AUC-ROC), F1 score, precision,recall.

Table 3:Comparision of machine learning parameters

Algorithm	Random forest	KNN	SVM
Accuracy	0.99	0.98	0.96
Precision	0.98	0.98	0.96
Recall	0.98	0.95	0.92
F1-score	0.99	0.96	0.94

Table 3 depicts the comparison of all the three machine learning models used in the proposed work.To find a balance between detecting all the real positives and producing precise positive predictions, recall and precision are frequently combined. The threshold or decision boundary of the model can be changed to manage the trade-off between recall and precision.

OUTPUT

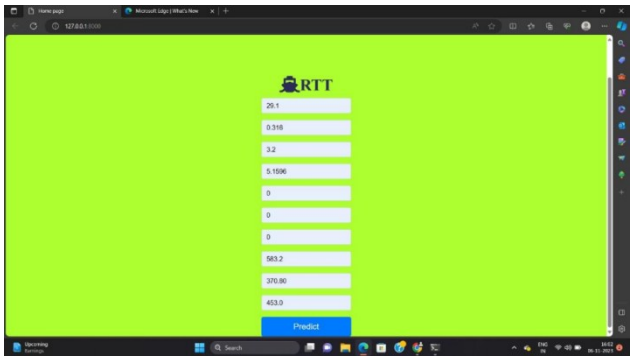

Fig 5(a).RTT input values

Fig 5(a);shows the input data to be entered and based on the values output is predicted if the access point is authorized or not.Fig5(b);shows the output is unauthorized

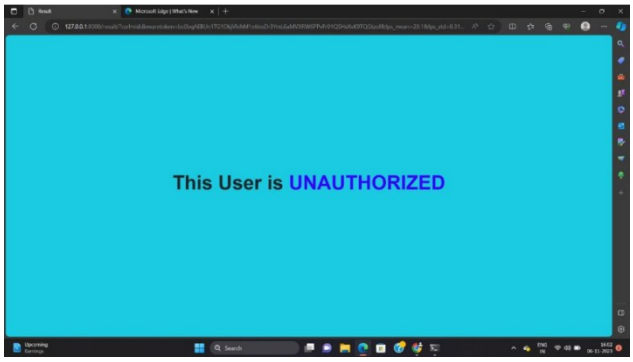

Fig 5(b).Output for unauthorized

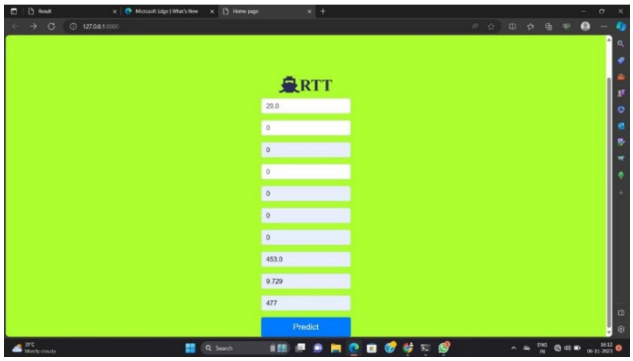

Fig.6(a).RTT input values

Fig 6(a)shows the input data to be entered and based on the values output is predicted if the access point is authorized or not.Fig 6(b);shows the output is authorized

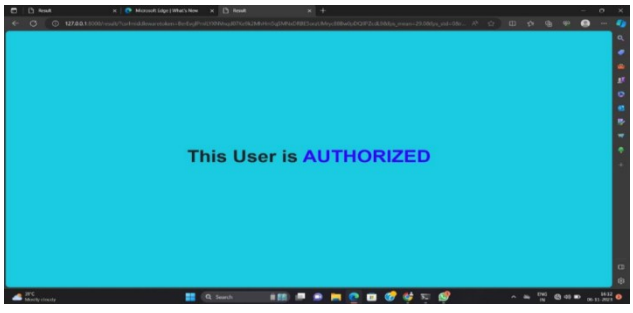

Fig.6(b). Output for authorized

VIII.CONCLUSION

In this proposed work,three machine learning approaches were used to classify APs into authorized and unauthorized.. Thus random forest is much suitable for this dataset. To safeguard the system against attacks machine learning approaches are used to train and test dataset and the best model is chosen for detecting attacks.Since Random forest algorithm working on ensemble technique was able to classify attacks precisely despite the fact it is time consuming for large dataset with large number of trees. Beside the fact of more training time the model shows accurate prediction results thus outperforms other two machine learning models in terms of machine learning parameters.RF is much suitable for the proposed work using RTT dataset.

REFERENCE

[1] Doyeonkim ,Dongil shin, Dongkyoo shin ."Unauthorized access point detection using machine learning algorithm for information protection" .in IEEE international conference on big data science and engineering.

[2] C. Wang, X. Zheng, Y. Chen and J. Yang, "Locating Rogue Access Point Using Fine-Grained Channel Information," in IEEE Transactions on Mobile Computing, vol. 16, no. 9, pp. 2560- 2573, 1 Sept. 2017.

[3] M. S. Gondal, A. J. Malik and F. A. Khan, "Network Intrusion Detection Using Diversity-Based Centroid Mechanism," 2015 12th International Conference on Information Technology - New Generations, Las Vegas, NV, 2015.

[4] H. Han, B. Sheng, C. C. Tan, Q. Li and S. Lu, "A Timing-Based Scheme for Rogue AP Detection," in IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 11, pp. 1912-1925, Nov 2011.

[5] S. Kitisriworapan, A. Jansang and A. Phonphoem, "Evil-Twin Detection on Client-side," 2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Pattaya, Chonburi, Thailand, 2019.

[6] M. S. Gondal, A. J. Malik and F. A. Khan, "Network Intrusion Detection Using Diversity-Based Centroid Mechanism," 2015 12th International Conference on Information Technology - New Generations, Las Vegas, NV, 2015.

[7] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.

[8] S. Mathur and A. Badone, "A methodological study and analysis of machine learning algorithms," International Journal of Advanced Technology and Engineering Exploration, vol. 6, pp. 45–49, 02 2019.

[9] B. Alotaibi and K. Elleithy, "Rogue access point detection: Taxonomy, challenges, and future directions," Wireless Personal Communications, vol. 90, pp. 5021– 5028, 10 2016.

[10] V. Bhusari, "Application of hidden markov model in credit card fraud detection," International Journal of Distributed and Parallel systems, vol. 2, 11 2011.

[11] K. Khan, A. Mehmood, S. Khan, M. A. Khan, Z. Iqbal, and W. K. Mashwani, "A survey on intrusion detection and prevention in wireless ad-hoc networks," Journal of Systems Architecture, vol. 105, p. 101701, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1383762119305089

[12] R. R. Reddy, Y. Ramadevi, and K. V. N. Sunitha, "Effective discriminant function for intrusion detection using svm," in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016, pp. 1148–1153.

[13] Asaju, L. Bolaji, P. B. Shola, N. Franklin, and H. M. Abiola, "Intrusion detection system on a computer network using an ensemble of randomizable filtered classifier, k-nearest neighbor algorithm," 2017.

[14] J. Hounsou, T. Nsabimana, and J. Degila, "Implementation of network intrusion detection system using soft computing algorithms (self-organizing feature map and genetic algorithm)," Journal of Information Security, vol. 10, pp. 1–24, 01 2019.

[15] B. Jahromy, A. Honarvar, M. Saif, and M. Jahromy, "A new method for detecting network intrusion by using a combination of genetic algorithm and support vector machine classifier," vol. 11, pp. 810–815, 01 2016.

[16] IEEE, '1999 edition (r2003) part 11: wireless LAN medium access control (MAC) and specification.

[17] L Jolliffe. Principle component analysisSpringer series in statistics, 2002.

[18] S . Jana and S. Kasera" On fast and accuratedetection of unauthorized access point using clock skews, "proc mobicom, 2008.

[19] J.R. Quinlan c4.5 program for machine learning ,morgankaufmann publishers inc , 1993.

[20] V. Vapniuk "support vector machine" in the nature of statistical learning theory , springer science & business media 2013.

[21] Arul Stephen, Mathesh"Evaluation of various machine learning algorithm for detection of attacks in 5G",Lecture Notes in Electrical Engineering 795, pp.397-495