Classification and Detection of Network Attacks in Encrypted Traffic Model

Hrushikesh R Pujar[1], Dr TNR Kumar[2]
[1]Student, Department of Computer Science, Ramaiah Institute of Technology, Bangalore, India
[2]Associate Professor, Department of Computer Science, Ramaiah Institute of Technology, Bangalore, India

Abstract— **It is common for wireless networks to use encryption protocols like WPA or WPA2 in order to provide secured communication. However, while encryption offers protection for the users' data, it reduces the capability of the traditional IDS to perform inspection of packets' payload.**
In the present research, the Wi-Fi traffic, which is encrypted, is analyzed by considering the statistical and flow features alone. The trilayer neural network is developed with the help of MATLAB function fitcnet, and the traffic is classified into various attack classes. The neural network is trained by considering 13 statistical features and employing fivefold cross-validation methodology. As the basis, the logistic regression and ensemble learning models are taken into account.
It can be seen that the neural network model provides an accuracy of 99.9%, whereas the logistic regression and ensemble learning models provide accuracies of 88.46% and 99.60%, respectively.

Index Terms— **Encrypted Traffic, Wi-Fi Security, Neural Network, Machine Learning, Intrusion Detection.**

I. INTRODUCTION

Wireless communication forms an integral part of the current communication process, and hence, the need arises to make sure that the wireless communication is kept secure. Some of the popular techniques of encryption involve WPA and WPA2 that help ensure the security of data transmission. Nevertheless, while it is clear that encryption is an important factor, the use of encryption gives rise to several issues because it hinders intrusion detection.
In view of the above problem, I think that there might be another method which could help us stop using traffic payloads to solve the problem. As such, the characteristics of statistics and flows may be the best options.
The neural network classifier used in the present work is designed to classify packets when these are encrypted. Next comes the evaluation of the efficiency of this classifier compared to others.

II. DATASET AND FEATURE DESCRIPTION

This particular piece of data will be used in the data analysis process. This data consists of the traffic generated using the Wi-Fi method by using the MATLAB software. Several classes can be found in the data sets, such as the normal class and attack class. Generally, several data points can be found in each of the classes, where the rows are entry points and columns are features.13 features are employed in this research, each characterizing certain behaviors within a network.
The characteristics shall assist in the comprehension of several aspects such as the movement of packets, protocols and signal employed. Characteristics like statistical characteristics are highly important in this study since they are based on the lack of consideration of the content in the packets, thereby making them significant in the network encryption.
Pre-processing of the data is highly critical before the training process. The reason behind this is that it ensures that all the characteristics present in the data have the same importance.

III. PROPOSED METHODOLOGY

The of purpose of this technique is used for the categorization of encrypted wireless network traffic based on particular statistical attributes. As the

encrypted traffic payload cannot be examined, the classification analysis should take place based on certain characteristics, such as packet production rate and interpacket period of time.

Overall, the process involves several steps, including feature extraction, training of the model, testing, and predicting the traffic class. In this case, the general structure of the system used consists of three layers of an artificial neural network.

A. Feature Representation and Preprocessing

For any given network, the network has characteristics that are explained using statistical language. These characteristics can include various types of network characteristics such as packet characteristics, time-based characteristics, and protocol ratios.

The input feature vector is defined as:

$X= [x_1, x_2, x_3,,x_{13}]$ (1)

where each xix_ixi represents a statistical feature such as packet rate, byte rate, or flow duration.

Before training, feature standardization is applied so that all input features are on a comparable scale. This step ensures that features with larger numerical values do not dominate the learning process. The data is also organized in tabular form, where each row corresponds to a network instance and each column represents a feature

B. Neural Network Model and Training

The main model used in this work is a multilayer feedforward neural network implemented using MATLAB's fitcnet function. The network consists of three hidden layers, each containing 10 neurons. This trilayer architecture is chosen to capture complex patterns present in encrypted traffic.

For each layer, the transformation is computed as:

$Z^{(l)} = W^{(l)}A^{(l-1)} + b^{(l)}$ (2)

$A^{(l)} = f(Z^{(l)})$ (3)

where $W^{(l)}$ represents the weight matrix, $b^{(l)}$ is the bias vector, and $A^{(l)}$ is the output of layer l.

The activation function used is the Rectified Linear Unit (ReLU), defined as:

$f(x) = \max(0,x)$ (4)

ReLU is selected because it introduces non-linearity and helps the model learn complex feature relationships efficiently.

The final prediction is obtained by selecting the class with the highest output value:

$\hat{y} = \arg\max(A^{(L)})$ (5)

The model is trained using supervised learning with labeled data. To ensure reliable evaluation, 5-fold cross-validation is applied. This method divides the dataset into five subsets, where the model is trained on four subsets and tested on the remaining one.

The classification accuracy is calculated as:

Accuracy = Number of Correct Predictions / Total Number of samples (6)

C. Baseline Models for Comparison

In order to test the effectiveness of the suggested neural network model, another two models will be suggested using the same dataset and features.

The first model used to compare the proposed neural network model is the Logistic Regression classifier with its extension into the case of multi-class classification based on the use of the Error Correcting Output Codes (ECOC). One-vs-One encoding is used in order to convert a multi-class problem into a set of binary problems, each one predicting the class label.

The second classifier to compare with the proposed one is based on ensemble learning techniques. In particular, the Subspace model is used where learners are trained using a random selection of attributes. As the base algorithm for training, the K-nearest neighbor classifier (KNN) is used

D. Prediction and System Workflow

The prediction phase is the final stage of the system, where the trained model is used to classify new input data. Upon training and validation, the model is implemented using a GUI interface created in MATLAB.

Input data can either be fed to the system from the dataset or through manual input into the GUI. The input data contains the statistical features as well. Preprocessing of the input data is done in the same way as was performed during training before making predictions.

The trained model predicts the class label assigned to the input data.The predicted class is obtained as:

$\hat{y} = \text{argmax}(P(y|X))$ (7)

where P(y|X) denotes the probability of classy given the input feature vector X.

The system also provides a confidence score, defined as:

Confidence=max(P(y|X)) (8)

The prediction results, including the attack class and confidence score, are displayed through the user interface. The overall workflow of the system is illustrated in Fig. 1.

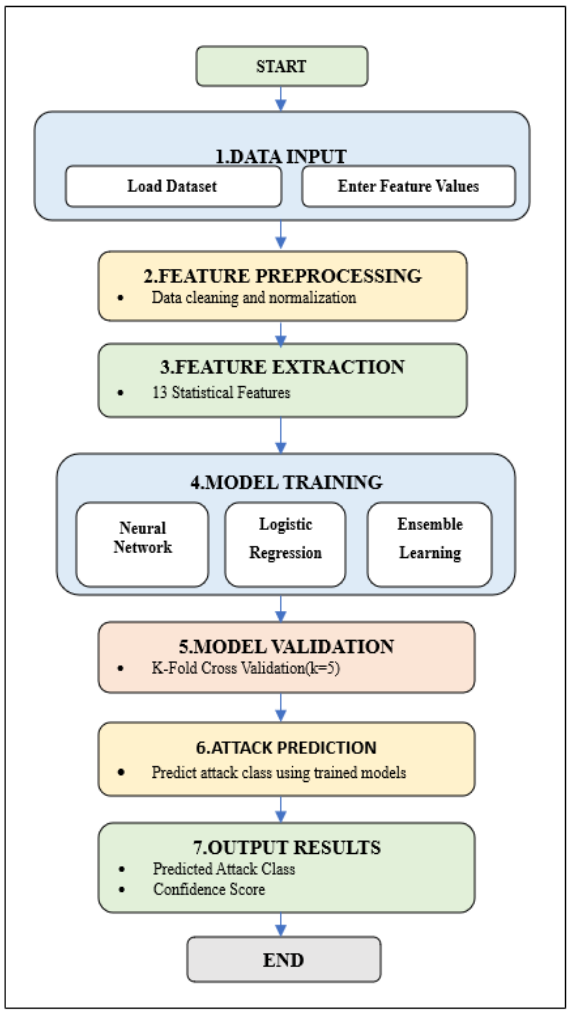

Fig. 1. Flow diagram of the proposed encrypted Wi-Fi attack detection system

Fig. 1 shows the overall workflow of the proposed system. The process begins from the phase of data input, where the input of the data set occurs, or is fed manually into the system. The next phase is preprocessing which is performed for the input data set, where features extraction is performed through statistical means. Training of machine learning models is done using these features, followed by the validation phase, which employs k-fold cross-validation technique. Prediction of attack class is performed through the output generated by the trained model.

IV. IMPLEMENTATION

Implementation of the algorithm is carried out to achieve the purpose of implementing the attacking detecting system through the use of Wi-Fi encryption algorithm. The steps involved in implementing an algorithm include data set preparation, modeling of the algorithm, testing the validity of the results received and finally embedding the algorithm in the prediction interface.

Currently, when carrying out our research activities, the dataset of the attacks, which include statistics, has been used. However, due to the encryption of the payload, we could not employ the above-mentioned data in training; consequently, all stages have been executed solely using traffic attributes. The dataset was produced by MATLAB tables in which each row represented traffic data, and the last column was the class label.

The procedure of training begins with the initial step of loading the data set and making the data ready for subsequent processing by means of machine learning methods. The initial step of implementation will be the loading of the data set and extraction of predictors from class labels.

A. Neural Network Implementation

The main architectural model used in this research work is through the use of the multilayer feed forward artificial neural network, which is executed using the fitcnet command within the MATLAB programming software. The architecture of this network consists of three hidden layers, and in each layer, there are ten neurons.

The neural network takes in 13 statistical data as input and identifies the correlation between those inputs and the different types of attacks. When building the neural network, we standardized the inputs so that the inputs with larger numerical ranges would not overwhelm the others during the training phase.

The training stage of the artificial neural network model will involve tagging the data related to traffic as well as cross-validation. Cross-validation involves splitting the data set into five parts in order to prevent overfitting.

B. Logistic Regression Implementation

Having considered that in this situation, the logistic regression classifier serves as the base classifier to evaluate the comparative results of different

classifiers, it is appropriate to use the Error-Correcting Output Coding technique.

In my opinion, this is the perfect way to solve the problem of classification, especially considering the number of classes in the logistic regression classifier.

This technique of error correction will make use of the one-on-one approach. It means that there will be multiple classifiers in use.

Use of neural networks classifier will be chosen due to the level of complexity that comes with it as compared to other classifiers.

It should be noted that the use of logistic regression will be preferred in cases where there are linear classifier decision boundaries.

C. Ensemble Learning Implementation

Fig. 2. Graphical user interface for encrypted Wi-Fi attack prediction

Fig. 2 shows the developed graphical user interface used for encrypted Wi-Fi traffic analysis. The interface accepts statistical feature values as input and displays the predicted attack class along with the corresponding confidence score.

V. RESULTS AND DISCUSSION

The performance of the proposed encrypted Wi-Fi attack detection system was evaluated using multiple classification measures. Experiments were carried out using the same dataset and validation strategy for all implemented models to ensure a fair comparison. The analysis focused on classification accuracy, confusion matrices, ROC characteristics, prediction confidence, and comparative model behavior.

The proposed trilayer neural network was evaluated against logistic regression and ensemble learning approaches. The generated results provide insight into how different learning methods perform on encrypted Wi-Fi traffic represented through statistical features.

A. Proposed Neural Network Results

The proposed trilayer neural network achieved a validation accuracy of 99.90%, indicating strong classification capability across different attack categories. The results suggest that the model successfully learned the statistical behavior of encrypted traffic and produced highly accurate predictions.

The confusion matrix was used to examine classification performance across all output classes. Most samples were correctly classified, with only a small number of errors observed.

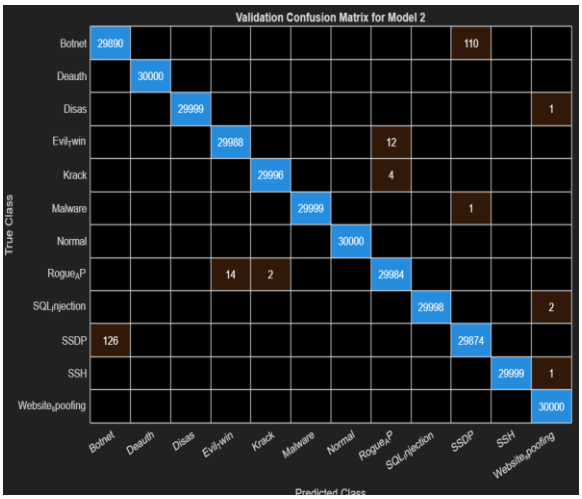

Fig. 3. Confusion matrix of the proposed trilayer neural network

Fig. 3 shows the classification performance of the proposed neural network. Most samples are concentrated along the diagonal region, indicating correct prediction across multiple traffic classes.

To further analyze model performance, ROC curves were generated using a one-versus-all strategy. The generated curves remained close to the ideal region, indicating effective separation among attack categories.

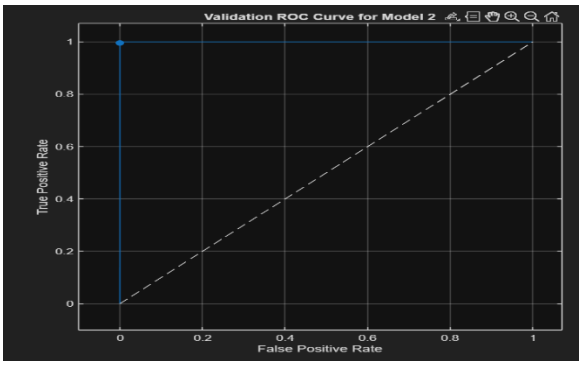

Fig. 4. ROC curve of the proposed neural network

Fig. 4 illustrates the ROC characteristics of the trained model. The obtained curves demonstrate strong discrimination capability among different classes.

B. Logistic Regression Results

The logistic regression model achieved a validation accuracy of 88.46%. Compared with the proposed approach, a larger number of classification errors were observed. Since logistic regression uses a linear decision boundary, its performance becomes limited when dealing with complex encrypted traffic patterns.

C. Ensemble Learning Results

The ensemble learning model achieved a validation accuracy of 99.60%. The use of multiple learners improved prediction performance and produced stronger results than logistic regression. However, the proposed neural network achieved slightly better accuracy.

D. Comparative Analysis

The ensemble learning model achieved a validation accuracy of 99.60%. The use of multiple learners improved prediction performance and produced stronger results than logistic regression. However, the proposed neural network achieved slightly better accuracy.

Table I Performance Comparison of Classification Models

Model	Method	Validation Accuracy
Logistic Regression	ECOC	88.46%
Ensemble Learning	Subspace KNN	99.60%
Proposed Neural Network	fitcnet (Trilayer)	99.90%

This point is illustrated by Table I where the efficiency of all examined methods is shown. It is evident that the proposed neural network performed better in terms of accuracy compared to other approaches.

In this way, the multilayer structure was useful in discovering nonlinear dependencies in data.

The results show that machine learning based on statistical features can successfully recognize encrypted Wi-Fi traffic regardless of the packet payload content. Table I summarizes the performance of all evaluated models.

VI. CONCLUSION AND FUTURE WORK

In this research, a machine learning method is suggested for the detection of the encrypted Wi-Fi attack based on statistical features of the network traffic. Considering that encrypted communications limit the accessibility of the payload information of the packet, the traffic statistics were used as the key for the classification task. The model based on a trilayer neural network using MATLAB fitcnet was built.

The proposed neural network achieved a validation accuracy of 99.90%, outperforming logistic regression and showing slightly better performance than the ensemble model. The results obtained from the experiment prove that the use of statistical features enables effective classification of encrypted traffic without any need to inspect packet payloads. Another advantage is that the graphical user interface created successfully proves the feasibility of the model to perform predictions.

In future, studies may be aimed at testing the system with real-time encrypted traffic, as well as examining other deep learning techniques to enhance the model's effectiveness and practical applicability.

REFERENCES

[1] Ye *et al*., "Detection of spoofing attacks in WLAN-based positioning systems using WiFi hotspot tags," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2025.

[2] L. Wang *et al*., "ACG: Attack classification on encrypted network traffic using graph convolution attention networks," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 1–13, 2025.

[3] Y. Zeng *et al*., "Monitoring and analysis of encrypted attack traffic based on machine learning," *IEEE Access*, vol. 13, pp. 1–10, 2025.

[4] R. Bahlali *et al*., "Malicious encrypted network traffic detection using deep auto-encoder with a custom reconstruction loss," *IEEE Transactions on Network Science and Engineering*, vol. 12, pp. 1–14, 2025.

[5] Q. Meng *et al*., "Detection of unknown attacks through encrypted traffic: A Gaussian prototype-aided variational autoencoder framework," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 1–15, 2025.

[6] J. Shan and H. Ma, "Optimization of network intrusion detection model based on big data analysis," *IEEE Access*, vol. 13, pp. 1–15, 2025.

[7] T. Senevirathna *et al*., "Enhancing accountability, resilience, and privacy of intelligent networks with explainable AI," *IEEE Transactions on Network and Service Management*, vol. 22, no. 1, pp. 1–20, 2025.

[8] Z. Wang *et al*., "Robust classification of encrypted network services using convolutional neural networks optimized by information bottleneck method," *IEEE Transactions on Network Science and Engineering*, vol. 12, pp. 1–10, 2025.

[9] G. Han *et al*., "AI-based malicious encrypted traffic detection in 5G data collection and secure sharing," *Electronics*, vol. 14, no. 51, pp. 1–24, 2025.

[10] Z. Li *et al*., "Unlocking few-shot encrypted traffic classification: A contrastive-driven meta-learning approach," *Electronics*, vol. 14, no. 4245, pp. 1–16, 2025.

[11] Z. Liu *et al*., "Fine-grained encrypted traffic classification using dual embedding and graph neural networks," *Electronics*, vol. 14, no. 778, pp. 1–19, 2025.

[12] S. Alhazbi *et al*., "LLMs have rhythm: Fingerprinting large language models using inter-token times and network traffic analysis," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 5050–5065, 2025.

[13] M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[14] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.

[15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2017.

[16] H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2016.

[17] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in *Proc. Military Communications and Information Systems Conf.*, Canberra, Australia, 2015, pp. 1–6.

[18] H. Hindy *et al*., "A taxonomy of network threats and intrusion detection systems," *Future Internet*, vol. 12, no. 10, pp. 1–33, 2020.

[19] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. Oh, "Semi-supervised deep reinforcement learning in support of IoT and smart city services," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 624–635, Apr. 2018.

[20] S. Dua and X. Du, *Data Mining and Machine Learning in Cybersecurity*. Boca Raton, FL, USA: CRC Press, 2016.