Failure Reasons Identification for the Next Generation WLAN: A Machine Learning Approach

Zhaozhe Jiang, Bo Li, Mao Yang(✉), Zhongjiang Yan, and Qi Yang

Northwestern Polytechnical University, Xi'an 710129, China
jzz@mail.nwpu.edu.cn
{libo.npu,yangmao,zhjyan}@nwpu.edu.cn

Abstract. Artificial Intelligence (AI) is one of the hottest research directions nowadays. Machine learning is an important branch of AI. It allows the machine to make its own decisions without human telling the computer exactly what to do. At the same time, Media Access Control (MAC) is also an important technology for the next generation Wireless Local Area Network (WLAN). However, due to transmission collision, noise, interference, channel fading and other reasons, the transmission between access point (AP) and station (STA) may fail. This is limiting the overall performance. If the node can obtain the real-time failure reasons, it can adjust protocol parameters accordingly such as Modulation and Coding Scheme (MCS) and Contention Window (CW). Then, the overall performance of WLAN is improved. Therefore, a machine learning based failure reason identification approach is proposed for the next generation WLAN. In this paper, access environment is divided into four categories: nice, severe collision, deep fading and both deep fading. Different training models are used to train the data. Through our experiments, the accuracy can reach 83%, while that of Random Forest model can reach 99%.

Keywords: Machine learning · Failure reasons · Access environment state

1 Introduction

In recent years, the global communication business grows rapidly. And WLAN is one of the most important data technologies [1]. In order to meet the increasing users demands, academia and industry are devoted themselves to improving WLAN performance such as throughput, latency, Quality of Service (QoS), etc [2].

Media Access Control (MAC) is one of the key technologies in WLAN [3]. And it is also one of the focuses of communication researchers. Since WLAN is based on distributed access, reasons such as collision, interference and channel fading may affect system performance [4]. Due to the complex and time-variable

B. Li et al. (Eds.): IoTaaS 2019, LNICST 316, pp. 417–426, 2020.
https://doi.org/10.1007/978-3-030-44751-9_35

WLAN environment, there may be a variety of reasons leading to the transmission failures, such as collision, channel fading and so on. If the access point (AP) or STA can obtain the failure reasons, they can adopt corresponding appropriate MAC strategies to improve the access success ratio and the overall WLAN performance. Therefore, whether there exists an intelligent identification mechanism that can accurately determine the reasons for the current user access failure.

There are many studies focusing on the performance improvement of MAC. Chen etc. [5] proposed an Interference Free Full Duplex with power control (IFFD) MAC protocol to avoid Inter-Station Interference Problem (ISIP) for next generation WLAN. Tsurumi etc. [6] proposed the MAC method based on the Synchronization Phenomena of coupled oscillators (SP-MAC) to improve a total throughput of wireless terminals connected to an Access Point (AP). Kim etc. [7] proposed Adaptive Virtual Backoff Algorithm (AVBA) to improve throughput. Qu etc. [8] proposed an OFDMA based Multiple Access for IEEE 802.11ax (OMAX) protocol to increase throughput. The result indicates that OMAX protocol increases the throughput to 160%. However, few of these studies mentioned before take into account the reasons of the access failure.

This paper proposes an identification mechanism based on machine learning for access failure reasons. In this paper, access state can be totally divided into four categories: perfect environment (access success), serious collision (access failure), serious fading (access failure) and both serious channel fading and collision (access failure). The training set and test set under different conditions are obtained through simulation of NS3. After that, train the data by different machine learning models. And test the identification accuracy by the test set. Our simulation experiment shows that the accuracy of Naïve Bayes model can reach 83%. And after feature extraction, the accuracy of Random Forest model can reach 99%.

The sections of this paper are arranged as follows. Section 2 introduces the key idea of this paper. And Sect. 3 introduces the process of machine learning. The fourth section introduces the simulation configuration and performance analysis. Finally, the fifth section introduces the conclusion and future work needed to do.

2 Motivation

In this paper, the method based on machine learning is adopted to determine the reasons of user access failure. We divide access state into four types. We use the information of ACK as our data set. And we use several models to do the machine learning. Finally, we can obtain the accuracy of different models by the test set.

2.1 Access Failure Reasons

We divide STA access environment state into four types. Three of them are failure state. And the other one is successful state (Fig. 1).

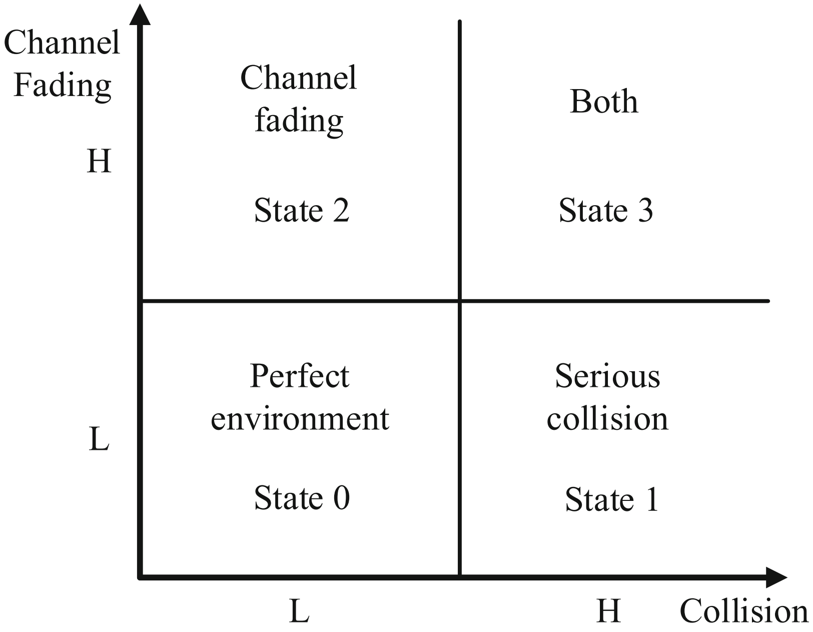

Fig. 1. Access environment state

Perfect environment means that the current state of communication is good. It is with low latency and high throughput. In our simulation configuration, the number of STAs is relatively small and it is close to AP. Therefore, the collision among STAs is relatively small. And there is almost no channel fading.

Serious collision means that the current access fails. In our simulation configuration, STA number is much more than perfect environment and STAs are close to AP. Because many STAs compete for only a channel, access failures must occur.

Channel fading is another reason of access failure. In our simulation configuration, the number of STAs is small and STAs are far away from AP. Due to the small number of STAs, the collision is not serious. It is that the distance becoming longer and the channel becoming worse, resulting in failure of access.

The last reason is both serious channel fading and collision. In our simulation configuration, the number of STAs is very large and the distance from AP to STAs is relatively long. It combines the above two reasons, so it is the worst case of communication state.

2.2 Motivation

If we can obtain the current access environment state, then we can take some steps to improve overall performance of communication. For example, when the access environment state is good, we can increase the MCS to increase the throughput. If the current collision of STAs is serious, we can increase the CW to reduce the collision probability. If the current channel fading is serious, we can reduce the MCS to ensure the successful transmission of information.

3 Machine Learning for Identification

This Section will describe the simulation process in our experiment.

3.1 Machine Learning Process

Machine learning is the use of algorithms to parse data, learn from it, and then make decisions or predictions about something in the world. It poses a deep technical revolution in almost every field [9]. The structure of machine learning is shown below (Fig. 2).

3.2 Data Generation

In this paper, the method of machine learning is used to determine the cause of STAs access failure. First of all, we need to produce enough data for the model to learn. Therefore, we use NS3 simulation software to simulate four different access environment state so as to obtain data sets.

It is well known that the Data/ACK pattern has been used in WLAN to improve the overall performance of the network. This is the foundation that we determine the cause of the current access failure. We use the ACK information in the network as training sets. We consider ACK information from two aspects. The first case is on whether the STA receives an ACK every time a packet is

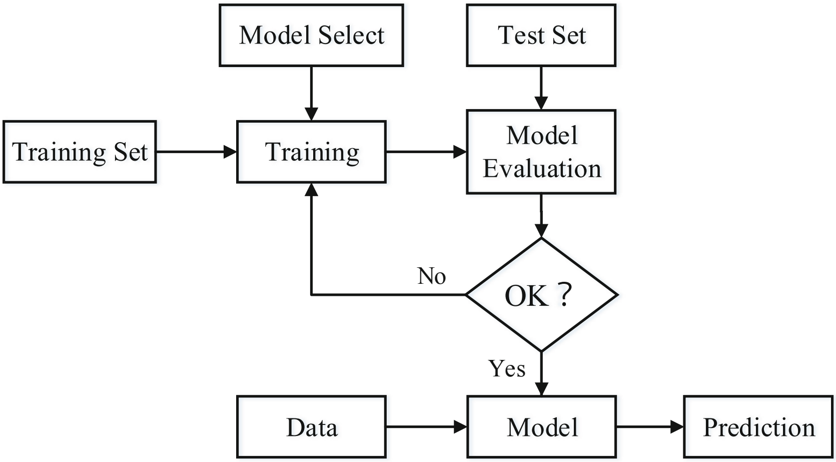

Fig. 2. Process of machine learning

transmitted. The STA will record 1 when it receives an ACK. And it records 0 when it does not receive an ACK. The other case is for each packet. It means that STA records the number of transmission times before the packet has been transmitted successfully. It will record 1 if a packet is transmitted successfully at the first time. Recording 2 means that a packet is transmitted successfully at the second time. The packet is retransmitted once. And record 0 if it exceeds the maximum number of retransmission. It indicates that the packet is abandoned.

In our simulation experiment, we configured four different access environment scenarios. They are named: Nice, Collision, Channel and Both.

Nice: It means that current access environment is perfect environment. The classification label is 0. For our simulation, there are a small number of STAs (Low Collision), low traffic rate, and close to AP (Low Channel Fading) in this scenario.

Collision: It means that current access environment is serious collision. The classification label is 1. For our simulation, there are a large number of STAs (High Collision), high traffic rate, and close to AP (Low Channel Fading) in this scenario.

Channel: It means that current access environment is channel fading. The classification label is 2. For our simulation, there are a small number of STAs (Low Collision), low traffic rate and a long distance from AP (High Channel Fading) in this scenario.

Both: It means that current access environment is channel fading and collision. The classification label is 3. For our simulation, there are as more STAs (High Collision), higher traffic rate, and nodes far away from AP (High Channel Fading) in this scenario.

The above is the overall description of the scene. And the detailed scene configuration is shown in Sect. 4.1.

After the simulation, we got a data set in the form of a matrix. There are hundreds of thousands of rows and 51 columns in the matrix. We use the first 50 transmissions or packets to determine the current access environment. A row of a matrix is called a record. The first 50 elements of a record are data, and the last is the label. There are hundreds of thousands of records in our data set. We use the data set as the training set.

Our test set is also generated by simulation. In order to validate the accuracy of the model effectively, we changed some configuration parameters to generate the test set. For example, in the scenario configuration of the training set, the distance between STA and AP is set to 150 m, 200 m and 250 m. Meanwhile, we set the distance of 180 m in the test set scenario to improve the test effectiveness. This will be closer to the real situation.

In order to further improve the accuracy of the models, we extracted the features of the data set. The mean, variance, maximum, minimum, median and mode of each record were extracted as new records. These records make up the new data set. Finally, there are hundreds of thousands of rows and 7 columns in the data set matrix. And the last column of the matrix is the label.

3.3 Data Cleaning

Due to the configuration of the simulation scenarios, a small part of data sets may be unavailable. We need to delete the unavailable data from our data set. For example, in the Channel scenario, the network performance is poor in the early stage of simulation. And it is a large packet loss ratio. However, Nice state of the network occurred in the final stage of simulation. The reason may be that other STAs no longer produce and send packets. Therefore, the STA competition pressure is reduced and the interference is also reduced. Each packet does not need to be retransmitted, which is inconsistent with the reality. Thus, it is necessary to delete this part of data set to get close to the reality.

3.4 Training Model

In the simulation of this paper, we used five different types of model to train the data respectively. They are the K Nearest Neighbor (KNN) algorithm [10], Random forest algorithm [11], Naive Bayes algorithm [12], Ensemble learning algorithm [13] and Discriminant Analysis (DA) algorithm [14]. The results of the different models are compared in the next section.

4 Simulation and Performance Evaluation

This section focuses on the detailed configuration of four simulation scenarios. And the simulation results are given and analyzed.

4.1 Simulation Scenario

WLAN is configured as a single cell uplink traffic scenario, with Distributed Coordination Function (DCF), DATA/ACK mode, without RTS/CTS interaction. There are four scenario which are Nice, Collision, Channel and Both. Under each scenario, all STAs will send 10,000 packets to AP in total. For each packet case, the data set matrix size is 40,000 *51. The data set matrix will be larger for each transmission case.

Since different parameter configurations may belong to the same scenario category. Therefore, in order to improve the completeness of the data set and the accuracy of the model, four different simulations are carried out under each scenario. The following is 4 detailed configuration tables for the four scenarios (Tables 1, 2, 3 and 4).

4.2 Results and Analysis

We use different training models in our experiment. The accuracy of different models is obtained after calculating with test set. Accuracy is shown in the Table 5.

From Table 5, we can see the experiment results of two different training sets under the same training model. As can be seen from the table, the accuracy of

Table 1. Nice scenario configuration

Parameter	1st	2nd	3rd	4th
STA number	10	20	5	10
Packets/s	100	50	50	50
Traffic rate	0.8 Mbps	0.4 Mbps	0.4 Mbps	0.4 Mbps
Distance	5 m	3 m	3 m	10 m
Packets number	4000	4000	1000	1000

Table 2. Collision scenario configuration

Parameter	1st	2nd	3rd	4th
STA number	200	100	50	200
Packets/s	1000	2000	1000	2000
Traffic rate	8 Mbps	16 Mbps	8 Mbps	16 Mbps
Distance	5 m	5 m	5 m	10 m
Packets number	4000	4000	1000	1000

Table 3. Channel scenario configuration

Parameter	1st	2nd	3rd	4th
STA number	10	10	10	10
Packets/s	100	1000	50	100
Traffic rate	0.8 Mbps	0.8 Mbps	0.4 Mbps	0.8 Mbps
Distance	200 m	150 m	250 m	250 m
Packets number	4000	4000	1000	1000

Table 4. Both scenario configuration

Parameter	1st	2nd	3rd	4th
STA number	200	200	200	100
Packets/s	1000	1000	1000	2000
Traffic rate	8 Mbps	8 Mbps	8 Mbps	16 Mbps
Distance	200 m	250 m	150 m	200 m
Packets number	4000	4000	1000	1000

Table 5. Accuracy of different models

Different model	Each transmission	Each packet
KNN	36.5%	52.4560%
Random forest	58.2756%	81.5371%
Naive bayes	60.0064%	83.7732%
Ensemble	58.2155%	81.4883%
Discriminant analysis	46.4382%	51.6129%

Naïve Bayes is highest. It can reach 83.77%. It is not a bad accuracy. However, the accuracy is not particularly high. The reason may be that the test set is obtained by adjusting the simulation parameters, which is relatively new to the training set. And the default model parameters are used in our training process. Therefore, the classification accuracy is not that high.

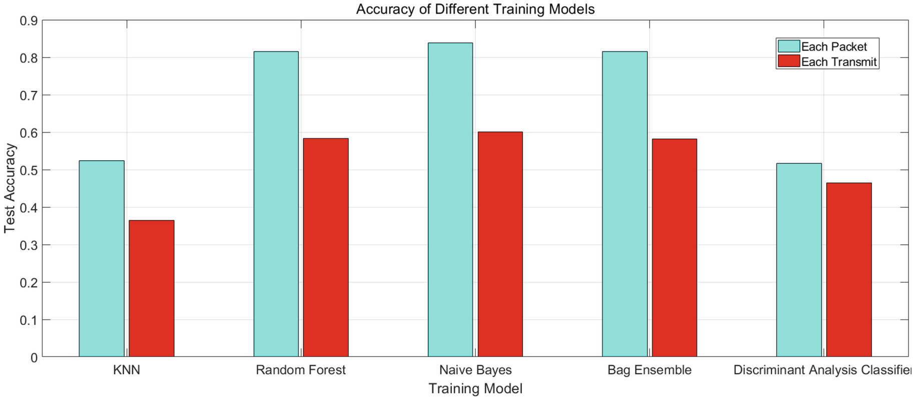

Fig. 3. Accuracy of different models

Figure 3 shows the accuracy comparison of the two training sets. It can be seen that the training set of each packet is generally better than that of each transmission. The reason might be that the total amount of information is the same in both cases. However, due to the small number of training set records in per packet case, the information content of each record is larger. Therefore, the classification accuracy of each packet case is higher (Table 6).

As can be seen from the above table, the accuracy of the model has been greatly improved after feature extraction. The accuracy of Random Forest model can reach 99%. This may because the key features of the data set are extracted. It greatly improves the classification ability of the model.

Table 6. Accuracy after features extraction

Different model	Each packet	Features extraction
KNN	52.4560%	79.5846%
Random forest	81.5371%	99.3385%
Naive bayes	83.7732%	—
Ensemble	81.4883%	95.1472%
Discriminant analysis	51.6129%	—

5 Conclusions and Future Work

In order to improve the comprehensive performance of WLAN, this paper proposes a new idea. We apply machine learning technology to WLAN. If the node can obtain current real-time access environment state, it can choose different optimization strategies accordingly. Thus, the overall performance of WLAN can be improved.

In this paper, access environment state is divided into four categories based on machine learning. They are perfect environment, serious collision, deep fading and both deep fading. Through our experiments, it is found that the accuracy of Naive Bayes model can reach 83%. And the accuracy of Random Forest model can reach 99% after feature extraction.

Due to the limited time and the author's level, there are still some parts to be improved in this paper. There is only ACK information used in the data set in this paper. It may lead to that the overall identification accuracy is not very reliable. In future studies, the author will use more information for training, such as packet time interval, CW size and other parameters. At the same time, select the appropriate model parameters. The accuracy and reliability should be further improved.

Acknowledgment. This work was supported in part by the National Natural Science Foundations of CHINA (Grant No. 61771390, No. 61871322, No. 61771392, No. 61271279, and No. 61501373), the National Science and Technology Major Project (Grant No. 2016ZX03001018-004), and Science and Technology on Avionics Integration Laboratory (20185553035).

References

1. Zheng, Y., Shi, T., Xu, X., Yuan, H., Yao, T.: Research on WLAN planning problem based on optimization models and multi-agent algorithm. In: 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), Ningbo, pp. 249–254 (2017)
2. Zhou, R., Li, B., Yang, M., Yan, Z., Zuo, X.: QoS-oriented OFDMA MAC protocol for the next generation WLAN. Xibei Gongye Daxue Xuebao/J. Northwest. Polytechnical Univ. **35**, 683–689 (2017)

3. Jiang, S.: State-of-the-art medium access control (MAC) protocols for underwater acoustic networks: a survey based on a MAC reference model. IEEE Commun. Surv. Tutor. **20**(1), 96–131 (2018)
4. Firdaus, F., Ahmad, N.A., Sahibuddin, S.: Effect of people around user to WLAN indoor positioning system accuracy. In: 2017 Palestinian International Conference on Information and Communication Technology (PICICT), Gaza City, pp. 17–21 (2017)
5. Chen, Y., Chen, I., Shih, K.: An In-band full duplex MAC protocol with interference free for next generation WLANs. In: 2018 International Conference on Electronics Technology (ICET), Chengdu, pp. 407–410 (2018)
6. Tsurumi, R., Morita, M., Obata, H., Takano, C., Ishida, K.: Throughput control method between different TCP variants based on SP-MAC Over WLAN. In: 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taichung, pp. 1–2 (2008)
7. Kim, J.D., Laurenson, D.I., Thompson, J.S.: Adaptive centralized random access for collision free wireless local area networks. IEEE Access **7**, 37381–37393 (2019)
8. Qu, Q., Li, B., Yang, M., et al.: An OFDMA based concurrent multiuser MAC for upcoming IEEE 802.11ax. In: Wireless Communications & Networking Conference Workshops. IEEE (2015)
9. Yang, M., Li, B., Feng, G., Yan Z.: V-CNN: When Convolutional Neural Network encounters Data Visualization (2018)
10. Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imag. **23**(4), 501–509 (2004)
11. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 832–844 (1998)
12. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1226–1238 (2005)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 770–778 (2016)
14. Tao, D., Li, X., Wu, X., Maybank, S.J.: General tensor discriminant analysis and gabor features for gait recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(10), 1700–1715 (2007)