Data Set Construction and Performance Comparison of Machine Learning Algorithm for Detection of Unauthorized AP

Doyeon Kim, Dongkyoo Shin(✉), and Dongil Shin

Department of Computer Engineering, Sejong University, 98 Gunja-Dong, Gwangjin-Gu, Seoul 143-747, South Korea
rlaehdus2003@gce.sejong.ac.kr,
{shindk, dshin}@sejong.ac.kr

Abstract. With the frequent use of Wi-Fi and hotspots that provide a wireless Internet environment, awareness and threats to wireless AP security are steadily increasing. Especially when using unauthorized APs in company, government and military facilities, there is a high possibility of being subjected to various viruses and hacking attacks. Therefore, it is necessary to detect and detect authorized APs and unauthorized APs. In this paper, to detect authorized APs and unauthorized APs, the characteristics of RTT (Round Trip Time) values are set as dataset and each machine learning algorithm SVM (Support Vector Machine), J48 (C4.5), KNN (K nearest neighbors), and MLP (Multilayer Perceptron).

Keywords: Machine learning · Wi-Fi · Algorithm · SVM · KNN · J48 MLP

1 Introduction

Due to the rapid development of devices using wireless networks, it is hard to find places without WiFi in our lives. WiFi is readily available in companies, cafes, military facilities, schools and public institutions. WiFi is used by many unspecified users, so it is difficult to check every one. And even if you are tethering like a hotspot using authorized WiFi, it is not certain to identify it unless you look directly at the AP list and look at the settings closely.

However, due to various smart devices, the existence of unauthorized AP has become indispensable. Usage is also irrelevant because there are no regulations or provisions on unauthorized APs, such as hotspots, as well as public places. This provides a very weak point to wireless networks. It can be harmed by stealing or gleaming information of other users who have access to unauthorized APs, and because PCs can also be hacked.

In order to prevent such damage, it is necessary to determine the more accurate illegal AP. Experiments on various algorithms are needed to identify high accuracy. In this paper, a dataset was created using RTT (Round Trip Time) values. The data set thus constructed is applied to the machine learning algorithm to obtain the result, and then the obtained results are compared to show which algorithm is more accurate.

J. J. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 474,
https://doi.org/10.1007/978-981-10-7605-3_144

In Sect. 2, we discuss the related research and the existing methods for unauthorized AP classification. In Sect. 3, we introduce the relationship between the experimental configuration introduction and the attribute values used in the data set. Section 4 analyzes the results of the experiment and Sect. 5 summarizes conclusions and future directions.

2 Related Research

The configuration of the authorized AP and the unauthorized AP is shown in Figs. 1 and 2. The configuration of the authorized AP is to use the device by catching the radio signal of the AP. On the other hand, the configuration of the unauthorized AP is configured such that the other AP receives the signal of the existing AP and receives the signal to construct a new AP so that it can be used by other users. As shown in Fig. 2, the new AP must have two wireless LAN cards. One LAN card receives a normal AP signal and the other generates a new AP based on the received signal [1].

Fig. 1. Authorized AP

Due to the relay AP structure in Fig. 2, the RTT difference with the authorized AP occurs. Among the papers that detect AP using difference of these RTT values, Han and Shen's method in [2] is to apply the straightness to the straight line by using difference of RTT value and standard deviation value. These seals are applied to the data distribution and classified. However, in this experiment, the α and β values of the linear equations for classification are not shown to be flexible by using fixed constants. In this paper, we try to classify by applying various algorithms in order to find algorithms that can be detected even in unplanned situations.

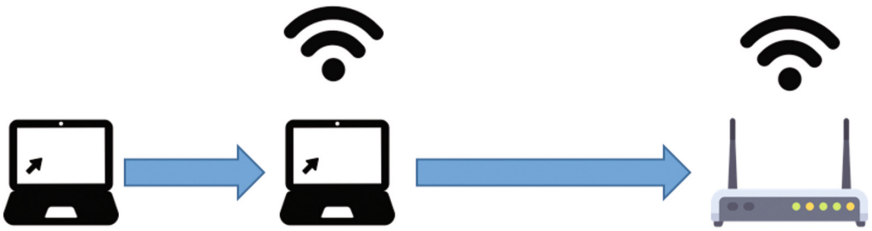

Fig. 2. Unauthorized AP

As a method of selecting feature points for RTT values, the difference, mean, variance, and standard deviation of delay times of each authorized and unauthorized AP are used [3, 4].

3 Experiment Configuration and Dataset Extraction

In this experiment, the Lenovo ideaPad Z400 touch device was used as a terminal PC for unauthorized PC use, and the router used for the authorized AP is Netgear GS608. The unauthorized AP has built a new AP using the LG XNOTE P210-GE30P and the iptime N500 connected to it.

Each data was measured and accumulated RTT value from the campus DNS server and AP using tracert.exe in the window10 environment, and only the hop counts which were measured except the hop that can not be measured were collected.

Also, the protocol used in network experiments can affect the results depending on which one is used. There may be a big difference in the communication protocol for each protocol, and the bandwidth and channel can also cause errors in the experiment. In this paper, we have experimented with 802.11n which is the most widely used.

The amount of data was 2300 data measured in each of the authorized and unauthorized APs, and the remaining 20% (460) of the measured data were excluded and only the remaining 80% (1840) data were used. Then, the measured value of RTTprobe (RTT to authorized AP) value, RTTDNS (RTT to DNS) value, RTTDNS - RTTprobe value, variance value and standard deviation were used as attribute values to classify data.

4 Experiment Result

In the experiments, the algorithms to be compared were selected from the classification - related algorithms among the machine learning algorithms. The algorithms are SVM (Support Vector Machine), J48 (C4.5), KNN (K nearest neighbors) and MLP (Multilayer Perceptron). The experimental results for each algorithm are shown in Table 1.

Table 1. Result algorithms

Accuracy	Algorithms			
	SVM	J48	KNN	MLP
[a]TP	40	92.9	92.9	84.5
[b]FP	0	9.1	8.5	8.4
[c]TC	70	92.9[2]	84.1	88

[a]TP: True Positive (%), [b]FP: False Positive (%), [c]TC: Total Correctness (%).

In the results of each algorithm, J48 (C4.5) and KNN algorithm are the most accurate in TP (True Positive) and SVM is the most accurate in FP (False Positive).

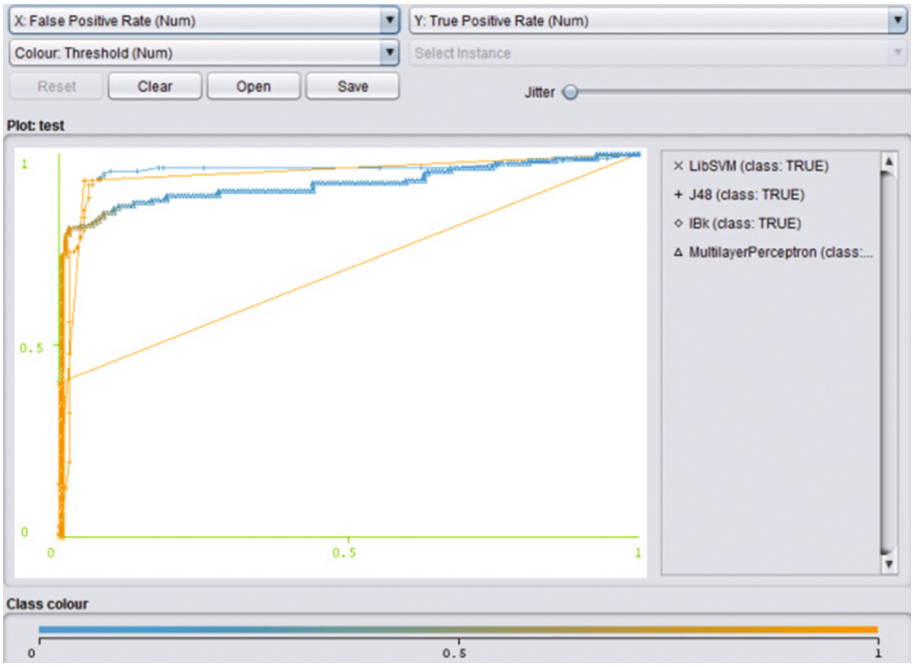

Fig. 3. Visualization of experiment results

Figure 3 compares the experimental results with graphs using the weka tool visualization function. The closer to the shape of ┌, the more ideal the result. As you can see in Fig. 3, the KNN algorithm is the closest to this type, followed by the J48 in the ideal form. Overall, KNN showed the highest accuracy when looking at overall accuracy.

5 Conclusion

In this paper, we can confirm that the difference between authorized and unauthorized APs can be classified by machine learning, and that there are some unstable parts in terms of accuracy. However, since the amount of data set and the attribute value used in the experiment are also related to this problem, it is necessary to accumulate a large amount of data sets in the future and to perform rigorous experiments by supplementing the property values. And I think that it is necessary to develop a program to discriminate the AP by using it.

Acknowledgments. This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract (UD160066BD).

References

1. Jang, R.-H., et al.: Analysis of time-based unauthorized AP detection methods according to hardware performance of unauthorized AP. J. Korean Inst. Commun. Inf. Sci. **40**(3), 551–558 (2015)
2. Han, H., et al.: A timing-based scheme for rogue AP detection. IEEE Trans. Parallel Distrib. Syst. **22**(11), 1912–1925 (2011)
3. Lee, J., Lee, S., Moon, J.: Detecting rogue AP using k-SVM method. J. Korea Inst. Inf. Secur. Cryptol. **24**(1), 87–95 (2014)
4. Kang, S., Nyang, D., Choi, J., Lee, S.: Relaying rogue AP detection scheme using SVM. J. Korea Inst. Inf. Secur. Cryptol. (JKIISC) **23**(3), 431–444 (2013)