2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering

Unauthorized Access Point Detection Using Machine Learning Algorithms for Information Protection

Doyeon Kim
Department of Computer Engineering
Sejong University
Seoul, Korea
rlaehdus2003@gce.sejong.ac.kr

Dongil Shin
Department of Computer Engineering
Sejong University
Seoul, Korea
dshin@sejong.ac.kr

Dongkyoo Shin*
Department of Computer Engineering
Sejong University
Seoul, Korea
shindk@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 (Access Point) 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. It is necessary to detect unauthorized Aps for protection of information. In this paper, we use RTT (Round Trip Time) value data set to detect authorized and unauthorized APs in wired / wireless integrated environment, analyze them using machine learning algorithms including SVM (Support Vector Machine), C4.5, KNN (K Nearest Neighbors), and MLP (Multilayer Perceptron). Overall, KNN shows the highest accuracy.**

Keywords—detection, unauthorized AP, machine learning, protection

I. 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, making it difficult to check every one. And even if you are tethering like a hotspot using authorized WiFi, identification is difficult unless you look directly at the AP (Access Point) list and look at the settings closely [1].

In a wireless local area network (WLAN), an access point is a station that transmits and receives data (sometimes referred to as a transceiver). An access point connects users to other users within the network and can also serve as the point of interconnection between the WLAN and a fixed wire network.

However, due to various smart devices, the existence of unauthorized AP has become unavoidable. Usage is also irrelevant, because there are no regulations or provisions relating to unauthorized APs, such as hotspots, as well as public places. This provides a very weak point to wireless networks. The network can be harmed by stealing or gleaming information of other users who have access to unauthorized APs, and because PCs can also be hacked [2]. Research to detect log AP and its risks have been actively studied until recently. Various methods of research are currently in progress, addressing various aspects of the issue[3]-[8]

In order to prevent such damage, it is necessary to ascertain which AP is an illegal AP. Experiments on various algorithms are needed to identify this with high accuracy [9]-[11].

*Corresponding Author: Dongkyoo Shin

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 results obtained are compared, to show which algorithm is more accurate.

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

II. RELATED STUDIES

Fig. 1. Authorized AP

Fig. 2. Unauthorized AP

The configuration of authorized and unauthorized AP is shown in Figures 1 and 2. The configuration of the authorized AP is to use the device by receiving the radio signal of the AP. On the other hand, 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 Figure 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]

Due to the relay AP structure as shown in Figure 2, the RTT difference with the authorized AP occurs. Among the papers that detect AP using difference of these RTT values, H. Han's method [2] is to apply the straightness to the straight line by

2324-9013/18/31.00 ©2018 IEEE
DOI 10.1109/TrustCom/BigDataSE.2018.00284

IEEE computer society

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.

Various algorithms are applied for classification of rogue APs even in unplanned situations.[9] [10]

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 [11].

Algorithm studies related to unauthorized AP detection have existed. Research aims to create a separate algorithm and use it to identify malicious wireless networks [12]. However, in this paper, the experiment was applied to the machine-learning algorithm, without making a separate detection algorithm. And by confirming the result, it is shown that the method for identification of unauthorized AP through the proved algorithm can be easily and variously applied.

The machine-learning algorithms and their features used in this paper for classification are as follows:

- SVM(support vector machine) : Based on a given set of data, we create a non-probabilistic binary linear classification model that determines which classifications of new data should be broken down and used to represent boundaries in the space in which data is mapped. The SVM algorithm is the algorithm that finds the boundary with the largest width[13].
- C4.5 : It is one of the algorithms for classifying and predicting data by making a decision tree. It is an algorithm that complements the limit of the existing ID3 (Iterative Dichotomizer 3) algorithm. The C4.5 algorithm uses the concept of information entropy to create a decision criterion and uses it to classify the sample set most effectively [14].
- KNN(k-nearest neighbors algorithm) : As a type of map learning, the input consists of the k closest training data in the feature space, and if used for classification purposes, the object is the object assigned to the most common item among the k nearest neighbors and classified by majority vote[15].
- MLP(multilayer perceptron) : The hidden layer is added between the input layer and the output layer, and supervisory learning is performed using the back propagation algorithm, so that data that cannot be linearly separated can be classified[16].

III. System Configuration and Dataset Extraction

For detection of unauthorized AP, we developed the "Intelligent Wireless AP Detection System" as shown in Figure 3. Data are collected from the authorized APs and unauthorized Aps and constructed as data sets. Data sets are analyzed by applying machine-learning algorithms including SVM (Support Vector Machine), C4.5, KNN (K Nearest Neighbors), and MLP (Multilayer Perceptron).

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

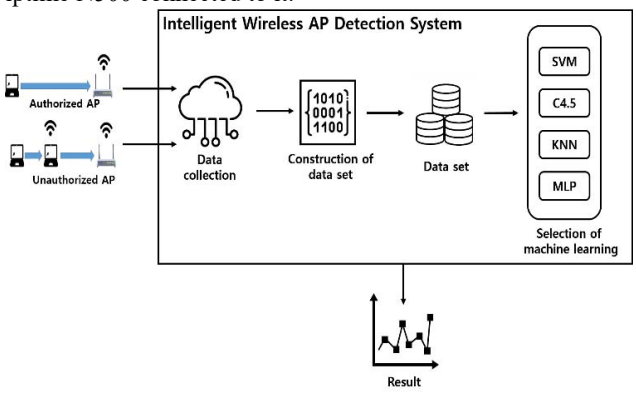

Fig. 3. Intelligent Wireless AP Detection System

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. 802.11n is used for wireless communication protocol that is the most widely used in real world.

Attributes of data set are shown in Figure 4. The measured value of RTT probe (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.

Mean probe	Mean DNS	Variance probe	Variance DNS	standard deviation probe	standard deviation DNS	Mean probe –Mean DNS
11	4	98	38	9.899465	6.164414	-7
11	4	0	96	0	9.797959	-7
				. . . .		
1	1	1	5	1	2.236068	0

Fig. 4. Attributes of data set

2300 items of data were measured in each of the authorized and unauthorized Aps; the remaining 20% (460) of the measured data were excluded and only the remaining 80% (1840) data were used.

IV. Experimental Results

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), C4.5, KNN (K nearest neighbors) and MLP (Multilayer Perceptron). The experimental results for each classification algorithm are shown in Table 1.

TABLE I. Experimental Result for Algorithms

Algorithms / Accuracy	SVM	C 4.5	KNN	MLP
True Positive	40	92.9	92.9	84.5
False Positive	0	9.1	8.5	8.4
Total Correctness	70	92.9	84.1	88

In the results of each algorithm, C4.5 and KNN algorithm are the most accurate in TP (True Positive) and SVM is the most accurate in FP (False Positive). Figure 7 compares the experimental results using ROC (Receiver Operating Characteristic) curve. As you can see in Figure 5, the KNN algorithm is the closest to this type, followed by the C4.5 in the ideal form. Overall, KNN showed the highest accuracy, in terms of overall accuracy. The reason why the KNN algorithm showed good performance is that the distribution of the data is more concentrated, because it is concentrated or distributed in a certain place like the classification of the KNN algorithm, rather than the decision tree or SVM. It seems to be a good reason to give a good result in this experiment, by classifying the distributed like KNN, rather than sorting by specific value or constant value like decision tree and SVM. In the case of MLP, we added a hidden layer and predicted a good result by back propagation. However, it seems that the number of hidden layers is small and the data value is insufficient for MLP experiment.

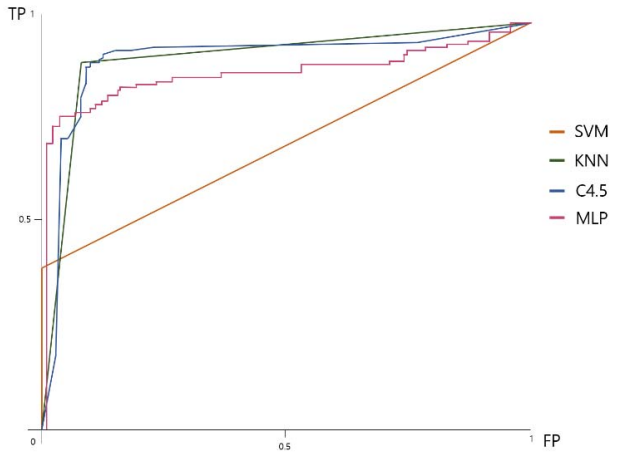

Fig. 5. ROC Curve for machine-learning algorithms

V. CONCLUSION

In this paper, we showed that the difference between authorized and unauthorized APs can be classified by machine-learning algorithms. If we detect the attacks from an unauthorized AP, we can disconnect it for protection of the system.

The methods in this paper will be applied to the protection of information,, including personal lifelog data. As a future research, we will design the protection scheme of personal lifelog data which are collected from smart devices analyzed using intelligent algorithms.

ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01059253).

REFERENCES

[1] S. Jana and S.K. Kasera. "On fast and accurate detection of unauthorized wireless access points using clock skews," *IEEE Transactions on Mobile Computing*, Vol. 9, No. 3, pp. 449-462, 2010.

[2] H.Han, et al. "A timing-based scheme for rogue AP detection," *IEEE Transactions on parallel and distributed Systems,* Vol. 22, No.11, pp. 1912-1925, 2011.

[3] F. Awad, M. Al-Refai, and A. Al-Qerem. "Rogue access point localization using particle swarm optimization," in *8th International Conference on Information and Communication Systems (ICICS),* Irbid, Jordan. May 2017. doi: 10.1109/IACS.2017.7921985

[4] S. Liu, Y. Liu, and Z. Jin. "Attack behavioural analysis and secure access for wireless Access Point (AP) in open system authentication," in *13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, Valencia, Spain, June 2017. doi: 10.1109/IWCMC.2017.7986377

[5] F. Awad, et al. "Access point localization using autonomous mobile robot," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Aqaba, Jordan, July 2017. doi: 10.1109/IWCMC.2017.7986377

[6] M. Agarwal, S. Biswas, and S. Nandi. "An Efficient Scheme to Detect Evil Twin Rogue Access Point Attack in 802.11 Wi-Fi Networks," *International Journal of Wireless Information Networks,* Vol 25. No. 3, pp 120-135, 2018.

[7] V. Modi, and C. Parekh. "Detection & Analysis of Evil Twin Attack in Wireless Network," *International Journal of Advanced Research in Computer Science* Vol. 8, No. 5, pp. 774-777, 2017.

[8] B. Pradeepkumar, et al. "Predicting external rogue access point in IEEE 802.11 b/g WLAN using RF signal strength," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI),* Udupi, India, Sept. 2017. Doi: 10.1109/ICACCI.2017.8126135

[9] A. Este, F. Gringoli, and L. Salgarelli. "On the stability of the information carried by traffic flow features at the packet level," *ACM SIGCOMM Computer Communication Review* Vol. 39. No. 3, pp. 13-18, 2009.

[10] L. Watkins, R. Beyah, and C. Corbett. "A passive approach to rogue access point detection," in *IEEE Global Telecommunications Conference 2007*, Washington, DC, USA, Nov. 2007, pp. 355 - 360.

[11] Peng, Lizhi, et al. "Early Stage Internet Traffic Identification Using Data Gravitation Based Classification," In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech*, Auckland, New Zealand, Aug. 2016. doi: https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.98

[12] C. Yang, Y. Song, and G. Gu. "Active user-side evil twin access point detection using statistical techniques," *IEEE Transactions on Information Forensics and Security,* Vol. 7, No.5, pp. 1638-1651, 2012.

[13] V. Vapnik, "Support Vector Machine," in *The nature of statistical learning theory*, Springer science & business media, 2013.

[14] J. R. Quinlan, *C4. 5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., 1993.

[15] N. S.Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, Vol. 46, No. 3, pp. 175-185, 1992.

[16] D.E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning internal representations by error propagation," in *Parallel distributed processing: explorations in the microstructure of cognition*, vol. 1, pp. 218-362, MIT Press, 1986.