Wi-Fi Frame Classification and Feature Selection Analysis in Detecting Evil Twin Attack

Md. Asaduzzaman
Dept. of Computer Sc. & Engg.
Military Institute of Sc. and Tech.
Dhaka-1216, Bangladesh
asadbd45@gmail.com

Mohammad Shahjahan Majib
Dept. of Computer Sc. & Engg.
Military Institute of Sc. and Tech.
Dhaka-1216, Bangladesh
smajib@yahoo.com

Md. Mahbubur Rahman
Dept. of Computer Sc. & Engg.
Military Institute of Sc. and Tech.
Dhaka-1216, Bangladesh
mahbubkucse@gmail.com

***Abstract*—Wi-Fi are mostly used components for connecting to the internet today. Nowadays Wi-Fi can be found in work, home or even in bus and train. While using internet with these access points, the connection between user and server is barely secure. Attackers can harvest data, as well as modify or drop data by impersonating himself as a legitimate access point. Also attacker can acquire the credentials of a legitimate access point from the users by impersonating himself as the legitimate access point and forcing the legitimate access point to be stopped for time being. This attack is known as *Evil Twin* attack. In this paper the traffics of both legitimate access point and rogue access point are analyzed. The detection is concluded with 91.2367% accuracy. Wi-fi frames of both APs are captured and features are extracted. Best 10 features are selected to increase the accuracy using chi-square test, information gain, gain ratio and tree based random forest. Several algorithms are used to classify the frames; among those J48 decision tree algorithm gives the highest accuracy.**

***Index Terms*—Evil Twin Attack, Machine Learning, WiFi Frame Analysis, Beacon Frame, Rogue Access Point**

I. INTRODUCTION

Wi-fi is a mostly used technology to connect to the internet today. It has been analyzed that Wi-Fi traffic from both mobile devices and Wi-Fi-only devices together will account for more than half (51 percent) of total IP traffic by 2022, up from 43 percent in 2017 [1]. Nowadays the easy access to the internet through Wi-Fi has brought the world closer to us. In a study as of 2013, it is found that during a 12 hours travel of 82.76km in Paris, 21649 access points(APs) were identified, 55.4% of which were from wireless Internet service providers (WISPs) [2]. But there is a question about the data security those are passing through these APs. Many researchers have shown that IEEE 802.11i standard cannot prevent various Denial of Service (DoS) attacks including de-authentication, disassociation and memory/CPU DoS attacks [3] . Md Waliullah et al. conducted an experimental study analysis on IEEE 802.11 Wireless Local Area Network and analyzed 11 security issues [4]. The attack has a great impact on social life. Attackers can harvest data, including sensitive credentials and mislead traffic to another malicious site. Using ETA (Evil Twin Attack), attackers can take over victim's device permanently by malware infection. Which may be escalated to bigger cyber crimes.

978-1-7281-7366-5/20/$31.00 ©2020 IEEE

Evil twin is a wireless attack where an attacker creates a fraudulent access point to eavesdrop or conduct more complex attacks. The attacker can escalate privilege by connecting himself to the wi-fi by deauthentication attack and hence conduct MITM on the connected users 1(b). Attackers can also create a public free Wi-Fi in order to harvest the users' data which is given in figure 1(c).

There are three significant sorts of frames in 802.11: data frame, control frame and management frame. Data frames transfers data in the frame body and control frames helps the data frames to be delivered. Management frames are used by stations to connect and disconnect to a BSS (Basic Service Sets). Almost all APs broadcast their beacon frames which include frame information, radiotap header, 802.11 beacon frame, 802.11 radio information and wireless LAN for 802.11. In this research unnecessary features are filtered out from the data. Different algorithms are leveraged to select the most effective attributes by attribute evaluator [5]. Chi squared test is also carried out to evaluate the feature selection with the help of sci-kit learn [6]. Top ten most effective features are selected and different classification algorithms are applied to classify the frames into two classes i.e LAP and RAP. Popular algorithms are used directly from weka [5] and the results are analyzed. It is found that J48 decision tree gives the best result. Further the most effective features were analyzed those are responsible for the AP to be evil twin.

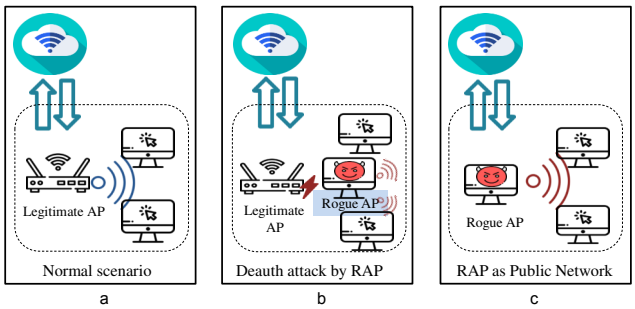

Fig. 1. Normal AP and evil twin attack scenario.

This paper is organized as follows: Sect. II reviews related work in brief. We describe our proposed scheme in Sect.

III. Section IV reports our experimental results and analysis. Conclusion and future work of this paper will be presented in Sect. V.

II. EXISTING SCHEMES FOR DETECTING EVIL TWIN ATTACK

In Evil twin attack or Wi-Fi intrusion, end devices are failed to ensure confidentiality, integrity and availability of the data of a user. Many researchers devised concepts and performed analysis regarding the detection methodology of evil twin APs. Daisuke Takahashi et al. focused on a survey on a user fingerprinting technique of IEEE 802.11 wireless LAN traffic and also summarized some of the researches on IEEE 802.11 network characteristic analysis to figure out rogue APs and MAC protocol misbehaviors [7]. Prof. Sandeep Vanjale et al. performed rogue AP detection using a database of known APs set by administrator [8]. It could detect APs of same SSID with different RSSI level but it was unable to detect free evil Wi-Fi. Mayank Agarwal proposed an IDS for detecting the evil twin attack and showed how it works in different cases regarding different number of legitimate and rogue APs [9]. But the algorithm fails for the case of one or multiple evil twin AP with no genuine AP. Alex Burns et al. analyzed the feature of evil twin attacks and proposed a bi-directional traceroute-based detection scheme [10]. Qian Lu et al. proposed a novel client-based solution to detect evil twin attack by monitoring WLAN frames and calculating correlation coefficients [11]. But the process checks for same SSID at first and if an rogue AP does not have the same SSID, it is classified as no evil twin. Kuo et al. proposed a user-side ETA detection method which uses time delay statistics of tcp connection termination [12]. The proposed scheme has some limitations that includes ET AP and victim AP must use same ISP and the detection process can be affected by many factors such as RSSI, internet speed etc. Harold Gonzales et al. proposed a method to detect evil twin attacks using contextual information [13]. S. et al. proposed and implemented an api with three modules that was able to detect ETA [14]. The detection process follows a mac database with respect to SSID name. These researches describe the detection process of ETA in different cases but none of these includes frame analysis. However, there are some papers those describe the detection process using machine learning and statistical techniques. Chao Yang et al. proposed to exploit fundamental communication structures and properties of evil twin attacks in wireless networks and to design new active, statistical and anomaly detection algorithms by differentiating the wireless hops (one or two hops) [15]. Thanthrige et al. used machine learning techniques to train and test the model with public Aegean Wi-Fi Intrusion Dataset (AWID) [16]. They also showed the attribute selection and feature reduction. The analysis was performed solely based on the AWID dataset. The model gave 94.97% accuracy using Random forest algorithm. Aminanto et al. focused on optimizing the impersonation attack detection which leverage Artificial Neural Network (ANN) for the feature selection and apply Stacked Auto Encoder (SAE) as a classifier for AWID Dataset [17].

In sum, there are a number of existing ETA detection methods but each of those are effective in some limited cases. The detection method and feature analysis is not integrated yet. Moreover no work was conducted in ETA detection by setting up evil twin AP, capturing the frames and analyzing the frames using machine learning techniques. Thus this research work will focus to carry out an ETA detection scheme by capturing the Wi-Fi frames, training a machine learning model using captured frames and finally detecting rogue AP by classifying the unknown frames. This research will also focus on classification and feature selection analysis without solely depending on the public dataset.

III. WORKING APPROACH

The working approach is carried out in several steps. The steps include capturing Wi-Fi frame, feature engineering, training machine learning model, finally testing the model by unknown data. The working approach is depicted in figure 2.

A. Capturing Wi-Fi Frame

A machine's interface can sniff wireless packets in two modes, i.e promiscuous mode and monitor mode. Promiscuous mode is normally used for packet sniffing that takes place on a router or on a computer. Whereas monitor mode allows packets to be captured without having to associate with an access point or ad hoc network first. Monitor mode can be used to create fake APs and also sniff the frames of the APs. There are a number of open source tools to create fake APs. Wireshark is a useful tool to capture frames. The captured frames can be exported as cap, pcap or json file. Contents of json file are easier to read and analyze. The json file conains a number of fields those are described in wireshark document [18].

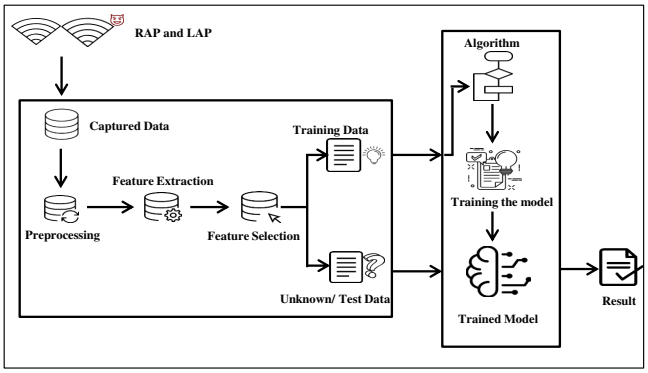

Fig. 2. Working approach.

B. Feature Engineering

The captured Wi-Fi frames contain malformed fields and unnecessary details. The data is to be preprocessed in order to make it ready for the operation. For this reason *preprocessing* step is carried out. In this step data is cleaned and labeled with respective classes (i.e RAP snd LAP) by filtering the frames using known SSIDs. *Attribute selection* is performed using some methods, i.e Chi squared test [19], Tree based

random forest, Information Gain and Gain ratio [20]. In case of information gain, The expected information needed to classify a tuple in D is given

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{1}$$

Here p_i is the nonzero probability that an arbitrary tuple in D belongs to class C_i and is estimated by $|C_i, D|/|D|$. After the partitioning, if we need $Info_A(D)$ information to arrive at an exact classification for attribute A ,

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_i|}{|D|} \times Info(D_j) \tag{2}$$

The term $|D_j|/|D|$ acts as the weight of the jth partition. $Info_A(D)$ is the expected information required to classify a tuple from D based on the partitioning by A. Finally we obtain the information gain for attribute A by calculating

$$Gain(A) = Info(D) - Info_A(D). \tag{3}$$

To calculate gain ratio for an attribute, we need to calculate split information for that attribute first. The value of the potential information generated by splitting the training dataset, D, into v partitions, corresponding to the v outcomes of a test on attribute A will be

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2(\frac{|D_j|}{|D|}) \tag{4}$$

The gain ratio is defined as

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)} \tag{5}$$

We can also select attributes using chi-square test using the formula of Chi Square test,

$$\chi^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k} \tag{6}$$

Here O_k denotes the observed value, and E_k denotes the expected value for feature X. Here, total number of instance is n.

C. Developing the Machine Learning Model

There are a number of algorithms in weka those can be used to find out the algorithm that gives the best output. The model can be evaluated using the evaluation measures defined as-

Accuracy	$\frac{TP+TN}{P+N}$
Error rate	$\frac{FP+FN}{P+N}$
Sensitivity	$\frac{TP}{P}$
Specificity	$\frac{TN}{N}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$

TP, TN, FP, P, N refer to the number of true positive, true negative, false positive, positive, and negative instances respectively. After having known the best algorithm, the proposed detection framework can be set up. Sci-kit learn [6] can be used to implement the detection scheme.

IV. EVALUATION AND ANALYSIS

An *experiment* is conducted to create fake AP and capture the Wi-Fi frames prior to the evaluation and analysis. In the next step, *frame analysis* is performed and unnecessary attributes are eliminated from the frames. *Feature selection* is carried out on the attributes. Analysis is conducted among various feature selection algorithms. Finally, *Classification and performance evaluation analysis* is carried out based on the selected features.

A. Experiment Setup

The experiment setup is carried out by configuring two machines for creating fake AP and another two machines for capturing Wi-Fi frames.There are also a number of LAPs around the configured machines. Fake APs are created by the attacker machines and other two machines are of victims'. Some open source tools are used for creating fake APs, e.g wifiphisher, 3vilTwinAttacker, Fake-AP and hostapd. In attacker machine, network interface is set to the monitor mode. In order to capture Wi-Fi frames, wireshark is used. Airmon-ng and airodump-ng are also used for this purpose. Also deauthentication is performed by the attacker's machines. In both cases, the frames are are captured. The experiment setup is shown in the figure 1 (b) and (c). The frames are saved as *JSON* file.

B. Preprocessing and Frame Analysis

As there are both legitimate and rogue APs in the experiment area, the exported data contains both of the frames. Some python scripts are written to automate the tasks of preprocessing and filtering the data. SSID of RAPs are known,so frames of the RAPs are separated from the frames of LAPs. The necessary fields are taken to an arff file with proper labeling (i.e. RAP and LAP). A total number of 38 attributes are chosen for attribute selection process.

C. Feature Selection and Analysis

In this stage, for chi square test and tree based random forest based feature selection, sklearn [6] is used. For information gain and gain ratio weka [5] is used. Top ten attributes are chosen to work for the classification process. In feature selection using chi-square test, different attributes along with their scores are shown in Table I. Using tree based random forest, the selected attributes are shown in figure 3. Using extra tree classifier, information gain and gain ratio for selecting the attributes give almost similar results. Features that have major effects in detecting evil twin attacks are radiotap_mactime, caplength, legth of the frame etc.

From the analysis, it is clear that almost all methods give similar results. It can be assumed that top ten features will be effective for the classification process.

D. Classification and Analysis

The data contains 5911 instances those are classified with the selected features. For the classification purpose naive bayes algorithm, decision tree algorithm (J48 and random forest)

TABLE I
ATTRIBUTES SCORE USING CHI-SQUARE FEATURE SELECTION.

Specs	Scores
radiotap.mactime	3.221466e+10
number	1.797014e+06
time_epoch	1.164053e+04
time_relative	1.775851e+03
radiotap.datarate	1.680321e+02
time_delta_displayed	1.359962e+02
len	2.078071e+00
cap_len	2.078071e+00
radiotap_length_length	0.000000e+00

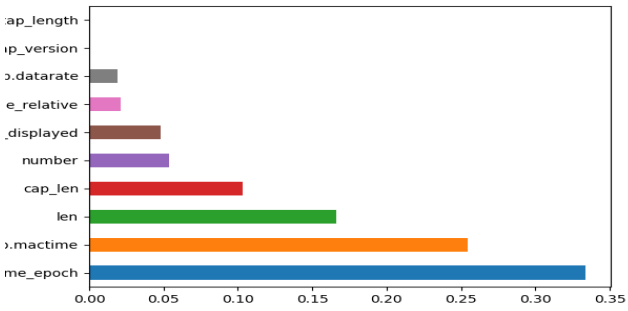

Fig. 3. Feature selection using Tree based random forest.

and rule based decision table algorithm are used. In all of the cases, 10 fold cross validation is used in order to increase the efficiency. The classification accuracy and error rate for different algorithms are shown in Table II along with weighted average of sensitivity and specificity.

TABLE II
CLASSIFICATION RESULTS.

Algorithm	Accuracy(%)	Error rate(%)	Sensitivity	Specificity	Precision	Recall
Naive Bayes	63.864	36.136	0.639	0.247	0.754	0.639
J48	91.2367	8.7633	0.912	0.143	0.917	0.912
Random Forest	83.7929	16.2071	0.838	0.190	0.838	0.838
Rule Based	91.2198	8.7802	0.912	0.154	0.922	0.912

From the obtained result it is found that decision tree algorithm works best for the given dataset. J48 algorithm gives the best result with the accuracy of 91.2367%.

V. DISCUSSION AND CONCLUSIONS

In this paper, a thorough analysis on the Wi-Fi frame is conducted to detect evil twin attack. Starting from the experimental set-up of fake AP, an overall scenario of an evil twin attack is described which also includes frame analysis, feature selection, frame classification in order to detect the fake APs. Decision tree algorithm gives the best result by detecting the attack with 91.2367% accuracy. The results of the framework will help the security researchers to work further in detecting evil twin attacks. Due to some malformed data obtained from the environment, the accuracy rate decreases. Also no analysis could be conducted using AWID dataset.

In future, the experiment will be conducted on bigger dataset. The fake AP data will be collected from an isolated environment. The results will be compared with the AWID dataset. Deep learning will be applied to increase the accuracy.

In this era of technological evolution, variety is being introduced in technology every day. Attacks are getting more sophisticated day by day. Attack on Wi-Fi routers is beyond our control. Analysis and works of this paper will help to deal with the attackers and make the task harder for them.

REFERENCES

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022 ," White Paper, Cisco, 2019.

[2] Y. G.-D. J. M. S. N. Vinícius F. S. Mota, Daniel F. Macedo, "On the feasibility of wifi offloading in urban areas: The paris case study." IEEE, 2013.

[3] L. Wang and B. Srinivasan, "Analysis and improvements over dos attacks against ieee 802.11 i standard," in *2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing*, vol. 2. IEEE, 2010, pp. 109–113.

[4] A. B. M. M. Md Waliullah and M. S. Rahman, "An experimental study analysis of security attacks at ieee 802.11 wireless local area network," *International Journal of Future Generation Communication and Networking*, vol. 8, no. 1, pp. 9–18, 2015.

[5] M. A. H. Eibe Frank and I. H. Witten, *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016*. MORGAN KAUFMANN, 2016.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] Y. Z. P. C. H.-H. C. Daisuke Takahashi, Yang Xiao, "Ieee 802.11 user fingerprinting and its applications for intrusion detection," *Computers & Mathematics with Applications*, vol. 60, pp. 307–318, 2010.

[8] D. P. Prof. Sandeep Vanjale, "A novel approach for elimination of rogue access point in wireless network." IEEE, 2014.

[9] S. B. Mayank Agarwal 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, pp. 130–145, 2018.

[10] X. D. L. Z. Alex Burns, Longfei Wu, "A novel traceroute-based detection scheme for wi-fi evil twin attacks." IEEE, 2017.

[11] Y. Z. X.-J. L. Y. Z. Y. L. Qian Lu, Haipeng Qu, "A passive client-based approach to detect evil twin attacks." IEEE, 2017.

[12] E.-C. Kuo, M.-S. Chang, and D.-Y. Kao, "User-side evil twin attack detection using time-delay statistics of tcp connection termination," 02 2018, pp. 211–216.

[13] J. L. D. M. D. S. Harold Gonzales, Kevin Bauer, "Practical defenses for evil twin attacks in 802.11." IEEE, 2010.

[14] H. S., K. Abdus Sattar, B. Sriramulu, and V. Rao, "Improving wi-fi security against evil twin attack using light weight machine learning application," *Compusoft*, vol. 8, 03 2019.

[15] Y. S. Chao Yang and G. Gu, "Active user-side evil twin access point detection using statistical techniques," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 1638 – 1651, 10 2012.

[16] U. Thanthrige, J. Samarabandu, and X. Wang, "Machine learning techniques for intrusion detection on public dataset," 05 2016, pp. 1–4.

[17] M. Aminanto and K. Kim, "Detecting impersonation attack in wifi networks using deep learning approach," 03 2017, pp. 136–147.

[18] "Wireshark; Display Filter Reference: Frame ," Available on: https://www.wireshark.org/docs/dfref/f/frame.html, [Online; Last accessed: 30 November 2019].

[19] K. yi and J. Beheshti, "A comparative study on feature selection of text categorization for hidden markov models," 06 2004.

[20] J. Han and M. Kamber, "Data mining: concepts and techniques morgan kaufmann," vol. 54, pp. 336–341, 01 2006.