A Method for Detecting 802.11 Wireless Malicious Phishing Stations through Fingerprinting

Na Lv [1, 2, a)], Jianguo Jiang [1, b)], Haitao Zhu [1, c)]

[1] *Institute of Information Engineering, Chinese Academy of Science, Beijing, China*
[2.] *School of Cyber Security, University of Chinese Academy Science, Beijing, China*

[a)] lvna0206@sina.com
[b)] jiangjianguo@iie.ac.cn
[c)] zhuhaitao@iie.ac.cn

Abstract. Malicious phishing stations, which disguise as legitimate devices through MAC address forgery, constitute a lot of Wireless Local Area Network (WLAN) security threats, such as secret information theft, implantation of Trojans and backdoors, etc. In this paper, a passive method based on wireless fingerprinting for detecting malicious phishing stations is proposed. We design 11 dimensions of features of station's fingerprinting, which can be extracted from frames and packets on MAC layer and application layer of open system interconnection (OSI) protocol stack. We have monitored wireless traffic above 60 hours and collected more than 10GB data in a real scenario to fingerprint all stations for recognizing phishing stations. We also evaluate the performance of proposed method by considering precision, recall, false positives and false negatives. The results show that our method has good performance that can detect phishing stations effectively and our method is also scalable.

Key words: WLAN; MAC; OSI; legitimate devices; fingerprinting.

INTRODUCTION

In recent years, WLAN has been widely used in various types of scenes because of its flexibility, mobility, scalability and usability and become an essential part in modern life. Meanwhile, there are also amount of security attacks like eavesdropping, privacy disclosure, unauthorized accessing, masquerading and so on. Among these attacks, masquerading identity like node forgery and address spoofing is a major threat of WLAN security, which can be easily using for stealing and misusing confidential files, implanting Trojans, attacking internal network and starting more sophisticated attacks. So once the malicious phishing stations have accessed to the network, we must detect these counterfeit devices promptly to protect the security and serviceability of networks. Fingerprinting devices makes it possible to identify these illegal invaders.

Device fingerprinting is a process depicting each device by its characteristics, which is observable and accessible. Generally, most 802.11 networks operate in infrastructure mode as depicted in Fig1, using an AP, access point to manage wireless communications. STAs (stations) are wireless devices which contain IEEE 802.11- conformant PHY and MAC layer interfaces, such as PDAs, laptops or desktop PCs. In this paper, we proposed a completely passive method based on wireless fingerprinting to detect malicious phishing wireless stations accessing to the network by masquerading as legal stations through MAC address forgery. According IEEE 802.11 standards [1], a variety of features can be extracted and utilized. To make device fingerprinting effectively, the extracted features should be difficult to forge and stable when environment changes. Considering these two properties, we design and extracted multi-layers features to increase forgery difficulty and most of features are stable.

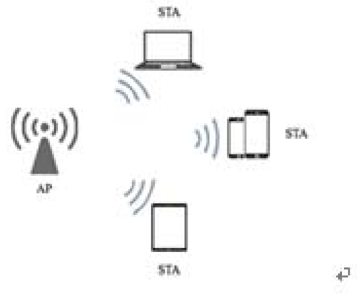

FIGURE 1. A simple architecture of infrastructure wireless network.

RELATED WORK

In this section, we have discussed some existing wireless device fingerprinting methods having been proposed as promising solutions to reducing the vulnerability of wireless network to node forgery or insider attacks [2-7]. [2] proposed a method to distinguish AP and clients based on clock skew. The method can fingerprint an AP easily but not applicable for fingerprint a wireless station because they used timestamp signatures extracted from beacon frames to fingerprint AP, but stations never send out beacon frames in infrastructure mode. A. Selcuk Uluagac et.al proposed a wireless device fingerprinting technique using wired-side observations in [3]. They captured time-variant behavior using statistical techniques to create unique device type signatures. [4] demonstrated an effectiveness technique using duplicate SIM and packet injection which is an active fingerprinting method. [5] employed a distribution-based measurement to obtain time signature of each wireless device and develop a decision-tree-based multi-level classifier for device fingerprinting. [6] introduced unsupervised learning method, presented a wireless device fingerprinting technique based on artificial neural networks. The parameters they used are transmission time and frame inter-arrival time. In [7], Jason Franklin et.al proposed a fingerprinting technique concerned with active scan function in wireless stations where stations send out probe request frames when actively scanning. They utilized difference among techniques that different drives perform probing and characterize the time deltas between probe requests to classify wireless devices drivers using Bayesian approach.

The methods above all mostly used single time-related feature to fingerprint wireless devices. In this paper, we aimed at identify phishing wireless stations and our approach employed mutli-dimensions features also including time features but also combining other information such as vendor specific identification to make fingerprinting more accurate.

PROPOSED APPROACH

Our malicious phishing stations detection system consists of four modules, monitor module, fingerprints extraction module, identification module and protection module and an extra fingerprint database as shown in Fig 2. Briefly, the whole technique includes two processes, 1) monitoring and extracting legitimate fingerprints and store them in fingerprint database, 2) monitoring and identifying in real time. We elaborate more on each module as follows.

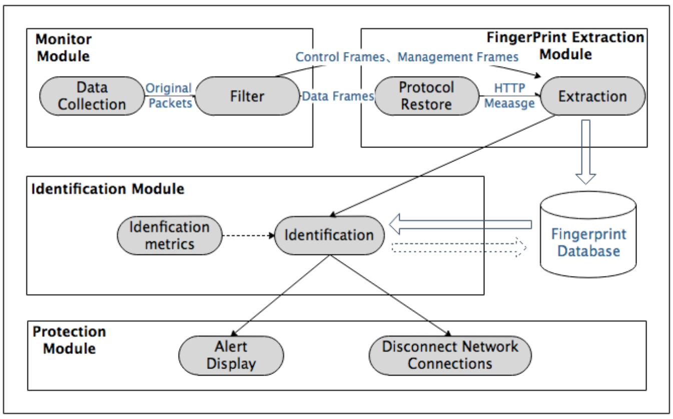

FIGURE 2. The framework of our malicious phishing stations detection system.

Monitor Module

We deploy sniffers in real test environment to capture wireless frames from the 802.11 spectrum, then filter useless frames like beacon frames which are not sent from stations and then forward the desired frames to fingerprint extraction module. In our deployment, we use laptops with wireless radios loaded with Linux operate system as sniffers and monitoring the testbed in our department building covering more than 50 students and teachers and use Libpcap [8] to collect frames from more than 300 devices.

Fingerprint Extraction Module

There are two parts in this module which are protocol restore sub-module and extraction sub-module. According 802.11 standards, each MAC frame has a mac header, a variable length of frame body, and a FCS containing an IEEE 32-bit CRC. There are three MAC frame types, management frames, data frames and control frames, which can be identified by type and subtype fields of frame header. Management frames are used to negotiation and control between STA and AP. Data frames are used to transmit data during communication. We extract some fields of these frames as features of stations. And furthermore, we record the earliest time and the latest time when each station starts and stop sending frames. These two times can be roughly considered as a station's working hours.

All features we extracted are shown in Table 1.

TABLE 1. Features for Fingerprinting Stations

Features	Location	Layer
RSSI	Radio Header of Packets	MAC Layer
Data Rate	Radio Header of Packets	MAC Layer
Length	Probe Request Frame Body	MAC Layer
Supported Rates	Probe Request, Association Request, and Reassociation Request frame body	MAC Layer
Extended Supported Rates	Probe Request, Association Request, and Reassociation Request Frame Body	MAC Layer
SSID	Probe Request Frame Body	MAC Layer
OUI	Probe Request, Association Request, and Reassociation Request Frame Body	MAC Layer
Browser	User-agent Field of HTTP messages (Data frame restore)	Application Layer
OS	User-agent Field of HTTP messages (Data frame restore)	Application Layer
Device	User-agent Field of HTTP messages (Data frame restore)	Application Layer
Work Time	Packet Header	Users' Behavior

Identification Module and Fingerprint Database

In this module, we calculate the similarity of real-time fingerprints and legal fingerprints in database. We

Weight assignment. To construct fingerprints, we adopt entropy-IDF method to assign weights to features. Entropy tells how much information there is in a feature, and IDF, which is inverse document frequency, is often used as a weighting factor in searches of user modeling. We also introduce a harmonic factor, which represents the difficulty of feature masquerading. The weight of feature f_i is calculate as:

$$\omega_i = \delta \cdot \sum -p_k lb(p_k) \times \log \frac{|N|}{|n_i|} \tag{1}$$

where δ is the harmonic factor of f_i We propose three kinds ways of feature counterfeiting which are hardware counterfeiting, software counterfeiting and behavioral counterfeiting and assign harmonic factor to each counterfeiting type according to the difficulty of counterfeiting.

And $\sum -p_k lb\left(p_k\right)$ calculates the entropy of f_i, p_k indicates probability of occurrence of each value of features, N is the total number of packets we have extracted features from, n_i is the number of packets containing feature f_i. Assuming M is the total dimensions of features, normalization of weights as

$$\omega_i^{'} = \frac{\omega_i}{\sum_{j \in M} \omega_j} \tag{2}$$

So the fingerprint can be presented as a set of features and their weights as $\left\{\left(f_1, \omega_1^{'}\right), \left(f_2, \omega_2^{'}\right), \ldots, \left(f_n, \omega_n^{'}\right)\right\}$.

Similarity Measurement. We employ three similarity calculation methods of Jaccard correlation coefficient, Euclidean distance and string pattern matching to calculate the similarity of different types of features.

Firstly, we use Jaccard correlation coefficient to measure the similarity of category features like supported rates, extended supported rates, oui, etc. as

$$sim\left(f_p, f_l\right) = \frac{\left|\left\{f_p^{\ 1}, f_p^{\ 2}, \ldots, f_p^{\ n}\right\} \cap \left\{f_l^1, f_l^2, \ldots, f_l^n\right\}\right|}{\left|\left\{f_p^{\ 1}, f_p^{\ 2}, \ldots, f_p^{\ n}\right\} \cup \left\{f_l^1, f_l^2, \ldots, f_l^n\right\}\right|} \cdot \omega_f^{'} \tag{3}$$

Let $f_l\left(f_l^1, f_l^2, \ldots, f_l^n\right)$ be the values of feature f in fingerprinting database and $f_p\left(f_p^{\ 1}, f_p^{\ 2}, \ldots, f_p^{\ n}\right)$ be the values of the feature f extracted from test station.

Similarly, we calculate the similarity of numeric features like work time and RSSI as

$$sim\left(f_p, f_l\right) = \sqrt{\sum_{i \in N}\left(f_l^i - f_p^{\ i}\right)^2} \cdot \omega_f^{'} \tag{4}$$

And the similarity of features of string form such as user-agent is

$$sim(f_p, f_l) = \begin{cases} 0, & if \ \ f_p = f_l \\ \omega_f^{'}, & otherwise \end{cases} \tag{5}$$

And at last, the overall similarity of two fingerprints, one of them is a legal fingerprint in fingerprinting database and the other is the test fingerprint constructed for test station in testbed is

$$sim = \sum sim\left(f_p, f_l\right) \tag{6}$$

Protection Module

When the calculation result of similarity is lower than the threshold we set, the test fingerprint is considered belonging to malicious phishing station, and the protection module presents alerts to the system administrator right now and will take autonomous action to disconnect the network of malicious stations.

EVALUATION

We evaluate the performance of our fingerprinting technique using four metrics, which are precision, recall, false positive rate and false negative rate defined as follows, $\Pr ecison = \frac{TP}{TP+FP}$, $\text{Re}\,call = \frac{TP}{TP+FN}$, $FPRate = \frac{FP}{FP+TN}$ and $FNRate = \frac{FN}{TP+FN}$, where TP, TN, FP, FN refer to true positives, true negatives, false positives and false negatives.

Evaluation Results

In the test, we constructed 1815 positive samples which perform malicious phishing stations and 165 negative samples which perform legitimate stations. Figure 3 shows the four-evaluation metrics where we test three thresholds, 0.7, 0.8 and 0.9 to observe the difference of evaluation results.

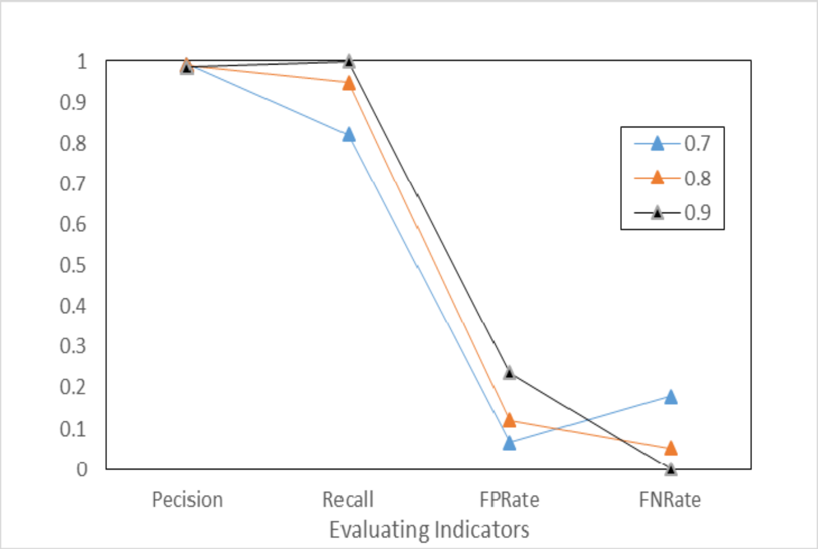

FIGURE 3. The evaluation results of our technique while we set the threshold to 0.7, 0.8 and 0.9.

When we set the threshold as 0.7, which means if the similarity of test fingerprint and legal fingerprint is lower than 70%, we think the station which the test fingerprint belonging to is a malicious station, as well as when the threshold is 0.8 and 0.9. For all three thresholds, we have got good precisions that is above 99%, which means that our approach can identify more than 99% of malicious stations. And as the threshold becomes larger, the recall rate increases and at the same time the false negative rate is lower, but the false positive rate also becomes higher. When the threshold is set to 0.9, the precision rate and the recall rate are both close to 100%, the false negative rate is close to 0, but the false positive rate reaches 23.6%, which means we recognize some legitimate stations as illegal ones. We think when the threshold is set to 0.8, we can achieve good evaluation results of all four metrics.

SUMMARY AND FUTURE WORK

In this paper, we proposed a multi-dimensions and multi-layer wireless station fingerprinting technique. We design and extract 11-dimension features of 802.11 frames to fingerprint wireless stations. We design a weight assignation measurement combining multiple factors and we employ multiple similarity approaches to measure closeness of these features. The evaluation results show our method have good performance. Also, our method is scalable that we can deploy more monitors to detect malicious stations in larger environment. In future work, we should consider optimization of algorithm parameters and introduce more features to fingerprint stations.

REFERENCES

1. IEEE-SA Standards Board, IEEE Std IEEE 802.11-2007 Information Technology-Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications, IEEE Computer Society, 2007.
2. S. Jana and S. K. Kasera, On Fast and Accurate Detection of Unauthorized Wireless Access Points Using Clock Skews, IEEE Transactions on Mobile Computing, 2009, pp. 449–462.
3. A. S. Uluagac et.al, Passive Technique for Fingerprinting Wireless Devices with Wired-side Observations, Communications and Network Security, 2013.
4. M. Kotha and M. Singirikonda, Unique Wireless Device Fingerprinting Technique for Secured Data Communication in Wireless Network, International Journal of Computer Applications, 2012, pp. 14-19.
5. Chao Shen et.al, Passive Fingerprinting for Wireless Devices: A Multi-Level Decision Approach, IEEE International Conference on Identity, 2017.
6. Kaushal Kumar et.al, An ANN Based Approach for Wireless Device Fingerprinting, IEEE International Conference on Recent Trends in Electronics, 2017.
7. Jason Franklin et.al, Passive Data Link Layer 802.11 Wireless Device Driver Fingerprinting, Conference on Usenix Security Symposium, 2006.
8. “TCP Dump/libpcap”, Available at http://www/tcpdump.org/