**ISBN** 978-981-14-1684-2
**Proceedings of 2019 the 9th International Workshop on Computer Science and Engineering**
(WCSE 2019 SUMMER)
*Hong Kong, 15-17 June, 2019, pp. 760-766*
doi: 10.18178/wcse.2019.06.113

# A Method of Fingerprint Legitimacy Discrimination Based on Fuzzy Matching Algorithms for Terminal

Ziang Lu[1+], Lu Chen[1], Mu Chen[1], Yong Li[1]
[1] Institute of Information and Communication, Global Energy Interconnection Research Institute,
State Grid Key Laboratory of Information & Network Security,
Nanjing Jiangsu, China
E-mail: luziang@geiri.sgcc.com.cn

**Abstract.** Nowadays, increasing number of network security events occur according to some illegal terminal counterfeiting intruding into the enterprise's internal network. The traditional terminal access system only uses some easy feature information (such as MAC address, IP address, host name, etc.) of the terminal equipment as the admission certificate to judge the legitimacy of terminals. In this paper, a method based on fuzzy approximate matching algorithm is proposed to identify the legitimacy of fingerprint features of terminal equipment. This method combines the basic characteristics of equipment and the behavior characteristics of equipment to generate device fingerprints, and uses the fuzzy approximate matching algorithm to compare the similarity between device fingerprints and fingerprints recorded in fingerprint database, so as to determine whether the device to be tested is a legitimate device. Finally, experiments show that the accuracy of the algorithm is 94.5%.

**Keywords:** Terminal; behavior characteristics of equipment; device fingerprint; fuzzy matching

## 1. Introduction

At present, with the rapid development and wide application of computer and network technology, information office almost covers the production and management of enterprises. However, the popularization of network terminals not only brings great accuracy and convenience in work, but also brings serious security risks to the company's information security. There are many examples of hackers attacking enterprise network by terminal. The most typical way is to implant Trojan and virus into company computer terminal through USB or web embedded Trojan so as to obtain company computer operation privilege and remote control to steal enterprise confidential data information. If the network terminal administrator often scans and patches the vulnerabilities of the devices in the network, kills Trojans and viruses and monitors the network implementation with the firewalls, intrusion detection system (IDS) and intrusion prevention system (IPS) deployed in the network[1], to a certain extent, it can greatly reduce the possibility that the terminal devices in the company network are subjected to external intrusion.

However, compared with hackers attacking enterprise networks by common means, terminal counterfeiting is a more direct and lethal way. There are some authentication mechanisms for the computer terminals in the network. For example, when the switch opens the port security mode, only the computer terminals with the designated MAC address can exchange data through the switch port. In other case, when the computer enters the network, it must pass password authentication or certificate authentication to prove its legitimacy to the network. But these typical network equipment access mechanisms are too simple in the face of the high level network intrusion technology and terminal counterfeiting technology. Terminal MAC address[2] and IP address can be easily modified and forged. Password or other kinds of certificate information also can be stolen and used by illegal persons. Not only about that, the rampancy of computer

+ Corresponding author. Tel.: 15850505659
*E-mail address*: liyong@geiri.sgcc.com.cn

terminal counterfeiting into the network also benefits from the exposure of the access interface. Through investigation, we know that in many enterprises public business hall, there might be some internal network entrance open to the external people without any protection, such as the illegal use of small wireless hub, cable network interface exposing in public areas, etc. These kinds of security vulnerabilities open another door to illegal hackers to intrude into enterprise information network.

In order to prevent the aforementioned counterfeit terminal devices from entering the network, a terminal access system based on device fingerprint is needed to identify the identity legitimacy of the terminal. Wang Yanhao et al. [3] identify Android intelligent terminal on the premise of zero privilege. Because the application scenario studied is based on Android applications, it is necessary to obtain the implicit identifier inside the Android smart terminal. Based on the physical layer, application layer and user layer, the appropriate hidden identifier is selected to form device fingerprint, and the corresponding fingerprint identification algorithm is proposed to complete terminal identification. Yin Xinming et al.[4] proposed a scheme based on device fingerprint decision tree classification algorithm to solve the problem of video private network equipment access detection. The scheme designs device fingerprint based on operating system fingerprint according to the characteristics of video equipment. Meanwhile, the method of collecting and storing device fingerprint of video private network is proposed. Through the decision tree classification algorithm of device fingerprint, the untrusted devices in the video private network can be detected and alerted in time. It plays an important role in preventing illegal intrusion and data leakage of video private network.

This paper will propose a method to judge the legitimacy of LAN terminal equipment based on device feature fingerprint. In this method, the basic characteristics of mobile terminal devices are acquired and dynamic behavior characteristics of devices are deeply digged. The similarity between the fingerprints of the terminal devices to be detected and legitimate terminal devices are matched in order to determine whether the terminal connected to the network is legitimate.

# 2. Fingerprint Characteristics of Equipment

## 2.1. Basic characteristics of equipment

The basic characteristics of the equipment are some inherent or relatively fixed configurations and information of the terminal itself. The basic characteristics of the equipment are supposed to be unique and unchangeable. The most typical basic features of devices include MAC address, IP address (static IP or the fixed IP in DHCP lease time[4]), operating system, open port, host name, browser information, etc. Because the basic features of the above devices are relatively easy to obtain, it’s simple to generate device fingerprints by this way. However, the basic features of equipment are easily imitated and forged, and the fingerprints generated by the basic features of equipment are very unreliable.

## 2.2. Equipment behavior characteristics

Equipment behavior characteristics refer to the characteristics generated by the terminal in the process of network operation. Typical behavior characteristics of networking devices include data flow, access behavior, access time, online and offline time, service status, web page title, etc. Monitoring the device's behavior for a certain time can depict the reasonable interval of the device's behavior. For example, assume that the function and environment of the device have not changed significantly, if the flow of a device at a certain time in a certain day has an obvious abnormal deviation from that of the past at the same time, then it can be judged that the device's behavior is abnormal and may be maliciously attacked or taken place by impersonation. Device fingerprints generated by device behavior characteristics can effectively capture hidden device behavior anomalies, and it is difficult to imitate the behavior characteristics. At the same time, there is a certain probability that the legitimate equipment with occasional abnormal behavior anomalies is defined as illegal equipment.

## 2.3. Acquisition of fingerprint characteristics of equipment

In this paper, a method combining the basic characteristics of equipment with the behavior characteristics of equipment is proposed to generate the fingerprint characteristics of equipment. The

acquisition methods of fingerprint features of terminal devices connected to LAN are mainly divided into active acquisition and passive acquisition. Active acquisition mainly use active inquiry to scan the network, automatically discover the connection terminals, construct identifiable inquiry frames to interact with the connection terminals to extract the terminal device features and protocol features. Passive acquisition is the way that features such as protocol, port, address and other information are acquired by analyzing the image data stream and parsing the packets. In this paper, throughout combining active and passive acquisition, some typical basic features and equipment behavior characteristics mentioned above equipment can be collected, as shown in Table 1.

Table 1: List of fingerprint characteristics

| Characteristics | Method | Label | Access Path | Type |
|---|---|---|---|---|
| Basic Characteristics of Equipment | Active | Open Ports | Invoke nmap script and use TCP's SYN scanning mode[6] to detect the open service port of the target machine and write to port_list file | File |
| | A/P | IP Address | Discovery of new IP records by remote scan method or get flow | Char |
| | Passive | MAC Address | Passive listening via trunk sense method | Char |
| | A/P | Host Name | Active scanning by NetBIOS or analysis of host field in HTTP message | Char |
| | A/P | OS and Version | Active scanning by NetBIOS method or analysis of the UA-OS field in HTTP message | Char |
| | Passive | Browser | Obtain by parsing the User-Agent field in HTTP message | Char |
| Behavior Characteristics of Equipment | Passive | Size of Data Flow | Create a new Bridge_Model object, mount it in KERNEL, then use IP as hash index to calculate flow size | Float |
| | Passive | List of Access Behaviors | Obtain the frequency of IP address access from data flow and calculate the statistical characteristics of IP address, such as time series, mean, variance, etc | Char List |
| | Passive | List of Access Time | Construct a new AccessTime object and then invoke the sniffer() method continuously in unit time. By comparing the time settings, the clustering characteristics of access time are obtained | Char List |
| | Active | List of Online/Offline Time | Execute Ping script to scan the remote detection continuously. If there is no ICMP REPLY[7], the offLine() method will be invoked to record the offline time; if there is a response, onLine() method will be used to record the online time | Int List |
| | Passive | Service State | Get the well-known port number in the IP response and request package. Then parse the data flow to parse the protocol message of the port access flow. Record the port number to the list if the parsing result is consistent with the corresponding protocol of the well-known port | Int List |
| | Passive | Title of Web Page | Obtain it by resolving Title field in HTTP message by parsing data flow. | Char |

There are two stages to verify the legitimacy of the identity of the terminals entering the network: the stage of precise matching of basic characteristics and the stage of fuzzy matching of behavioral characteristics.

- Firstly, according to the basic characteristics of the device, the MAC address of the network-entry terminal is matched accurately, and the corresponding terminal information entries are queried and obtained in the access database. Then, the IP address, port information, host name, operating system and version, browser information of the network-entry terminal are acquired and compared with the input terminal information entries to ensure the network-entry terminal equipment. The basic configuration and features are exactly the same as those of database entry.
- Because illegal intruders can forge illegal terminals into legal terminals by configuring the same basic characteristics of the equipment, the identity of the terminal can not be determined as a legal terminal after the accurate matching of the basic characteristics of the equipment and authorizing the temporary access permission. It is necessary to have further monitor to the behavior characteristics of the terminal. Through the combination of active and passive methods, the data flow behavior, access behavior, online and offline time, service status, web pages title are monitored, and the results of

various indicators are analyzed and discriminated by fuzzy method. The incoming terminal that found abnormal behaviors will be regarded as illegal terminal and disconnected its network link.

# 3. Fingerprint Legitimacy Discrimination Based on Fuzzy Matching

## 3.1. Selection of terminal behavior characteristics

There isn't any agent in terminal, the categories of basic features that detected by bypass access equipment are few, so the basic features of equipment can be regarded as dictionary acquisition form (< key, value > key value pair). The terminal would be judged as a legal device only if the dictionary is pretty similar with the dictionary belonging to one of devices recorded in access database. Wang Yanhao et al.[3] proposed an association matching algorithm considering possible changes in device features. This method draws lessons from Eckersley's research work on desktop browser fingerprints, allowing only one feature to change in a certain extent in all device features[8]. This method can be applied to the case where there are many features of equipment with only one of them change.

The precise matching algorithm and association matching algorithm mentioned above are not applicable to the situation where multiple behavior characteristics of equipment change at the same time. In this paper, a fuzzy approximate matching algorithm is proposed, which compares the dynamic and numerical drift behavior features with those of the fingerprint database. Fuzzy approximate matching algorithm divides each dynamic behavior characteristic that may appear numerical drift into two indexes to measure. For terminals accessing LAN, the dynamic behavior characteristics can select the total flow characteristics of the terminal, the mathematical and statistical characteristics of accessing an IP address, the time characteristics of the terminal accessing an IP address, the online and offline time habits of the terminal, the flow characteristics of a well-known port and so on, such as Table 2 shows.

Table 2: Selection of terminal behavior characteristics

| Behavior Characteristics of Equipment | Parameter A | Parameter B |
|---|---|---|
| Uplink flow size in limited time | Mean | Variance |
| Downlink flow size in limited time | Mean | Variance |
| Access Behavior Characteristics of IP in limited time | Accessing frequency | Total flow size |
| Online and offline time list for every day | Online time | Offline time |
| Flow size of known ports in limited time | Mean value of flow size of the port | Mean value of flow size of the port |

Because the quantitative measurement of different indicators is not uniform. In order to compare the comprehensive similarity of the devices entering the network, it's necessary to normalize the offset of all indicators. The sum of the normalized results of each behavioral characteristic offset is the difference between the terminal behavioral characteristic fingerprint and the device fingerprint database. Because even if it is the same legitimate device, the data detected by its behavior characteristic index will be slightly deviated in a reasonable range after each online launch, when the similarity between the device characteristic fingerprint of the device to be tested and the device characteristic fingerprint of a device in the device fingerprint database is higher than a certain threshold, it is determined that the device is the same. The similarity formula of feature fingerprints is expressed as follows:

$$S_{\text{Fingerprint}} = 1 - \sum_{i=1}^{k} \sqrt{[1-(\frac{a_i}{a_{fpi}})]^2 + [1-(\frac{b_i}{b_{fpi}})]^2} \tag{1}$$

In algorithm 1, the principle of judging equipment matching with equipment fingerprint database equipment is that if and only if the similarity of each behavior index measurement value of the equipment to be tested and the corresponding behavior index value of the equipment to be matched in the equipment fingerprint database is greater than or equal to 80%, and the overall similarity is greater than or equal to 85%, the equipment matching is successful. It should be pointed out that the selection of thresholds 0.8 and 0.85 in the algorithm is obtained by collecting fingerprint data of test users over a period of time and analyzing the changes of list type data in the two fingerprints. At the same time, the overall similarity values are consistent

with those in references [3] and [8]. Every time the device accesses to the network, the behavior pattern generated is unique. After matching the average of the reference value of the new input device fingerprint database and the measured value of the device's behavior characteristics reduces the occurrence of false alarm in the process of fingerprint matching.

Algorithm 1: Fuzzy matching algorithms of terminal accessing

```
Check the number of previous successful access to the network for n-1;
if（n = 1）{
//Judged as new equipment
Obtaining the initial behavior characteristic index of item i of the equipment;
for（i = 1, i < I, i++）{
//To determine the legal domain Ai of Item i;
generate point pi（ai， bi）；
Ai is the area where the distance dpi from (0,0) to pi is shorter than 0.2;
}
}
if（n > 1）{
//The device has been successfully connected to the network N times
for（i = 0， i < I, i++）{
The weight value Wi of Item i of the device is obtained;
The latest measurement value Pi of index I is obtained from fingerprint database;
Ai is the area where the distance dpi from any point to pi is shorter than 0.2;
Obtain the measured value pnew of ith behavior characteristic index of the equipment;
if（pnew⊂Ai）{
Get the average value pi-new of pnew and pi;
}
else{
Fingerprint database matching failed, device is illegal;
return；
}
Calculate the distance di from pnew to pi;
//Computing the similarity Si between the device's behavior characteristics and the device's in fingerprint database;
Si = Si + Wi*di；
}
if（Si >0.15）{
Fingerprint database matching failed, device is illegal;
}
else{
for（i = 0， i < I, i++）{
pi = pi-new；
}
}
```

# 4. Experimental Verification

## 4.1. Experimental flow and data acquisition

This paper designs the experimental environment to verify the effectiveness of the above algorithm. Three local hosts respectively called Host C(192.168.0.1), Host D(192.168.0.2) and Host E(192.168.0.3) are the target hosts that the incoming devices may need to access. The device fingerprint matching module makes bypass connection on the switch. The PC A represents the legitimate terminals to be admitted, while the PC B indicates an attack terminal to imitate PC A.

In the first stage of the experiment, data of the device behavior characteristics of the legitimate terminal PC A should be collected so as to initialize the fingerprint description of PC A in the device fingerprint database. By learning the behavior pattern generated by the using PC A in one hour, the statistical characteristics of the upstream and downstream flow of the PC A, the access behavior and flow of the PC A to 192.168.0.1, 192.168.0.2 and 192.168.0.3, the login time of the terminal device, the statistical characteristics of flow of ICMP and UDP ports are obtained respectively, and the fingerprint of the device is generated and input into the fingerprint database of the device. In the second stage of the experiment, the abnormal behavior operations are defined and written in script. When the device runs the script, the abnormal traffic and abnormal access to different IP addresses can be generated randomly. In the third stage of the experiment, the PC B with the same type as PC A was camouflaged with identical basic features including port, IP address, MAC address, host name, version of OS and so on. Then abnormal behavior

operation scripts were executed on PC B. In the fourth stage, 200 groups of experiments (mixed 180 groups of PC A and 20 groups of PC B were conducted).

## 4.2. Analysis of experimental results

In order to visually reflect whether the fingerprint validity discrimination algorithm of the fuzzy matching can be effectively implemented, this experiment result applies $R_{Accuracy}$, $R_{FPR}$ and $R_{FNR}$ as the evaluation indexes to verify the effectiveness. The calculation formulas are as follows:

$$R_{Accuray} = \frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \quad R_{FPR} = \frac{N_{FP}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \quad R_{FNR} = \frac{N_{FN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \tag{2}$$

In the above formulas, $N_{TP}$ denotes the time of successful matching that PC A is allowed to access to the network, $N_{TN}$ denotes the time of failed matching that PC A is forbidden to access to the network. $N_{FN}$ denotes the time of failed matching that PC B is forbidden to access to the network, while $N_{FP}$ denotes the time of successful matching that PC B is allowed to access to the network. In this experiment, the number division of $N_{TP}$, $N_{TN}$, $N_{FP}$ and $N_{FN}$ is as shown in the table .

Table 3: Distribution of experiment result

| Result | $N_{TP}$ | $N_{TN}$ | $N_{FP}$ | $N_{FN}$ |
|---|---|---|---|---|
| Times | 169 | 20 | 11 | 0 |
| Average Similarity | 93.1% | 71.4% | 82.2% | 0 |

It can be seen from the Table 3 that the $R_{Accuracy}$, $R_{FPR}$ and $R_{FNR}$ of the fingerprint validity discrimination algorithm of the fuzzy matching terminal is respectively shown as 94.5%, 5.5% and 0%. Among them, the missing rate is perfect because the script execution on PC B produces more obvious anomalies, the average similarity is only 71.4%, which is significantly lower than 85% of the matching success criteria. Compared with the false alarm rate, the false alarm rate is more prominent. The main reason is the dynamic behavior characteristics of legitimate equipment will change greatly user low probability. Choosing more high-quality dynamic behavior characteristics as matching factors will remove more occurrence of false alarm.

# 5. Conclusion

In this paper, a method of fingerprint legitimacy discrimination based on fuzzy matching algorithms for terminal equipment is proposed. This method lists some useful device characteristics, which help LAN to estimate whether a terminal is legal or not to access. Finally, a experiment is designed to verify that the fuzzy matching algorithm does work, and the accuracy is up to 94.5%.

# 6. Acknowledgement

Supported by the science and technology project of State Grid Corporation of China: "Research on Key Technologies of Marketing Site Terminal Security Access" (Grand No. SGGR0000XTJS1800079)

# 7. References

[1] Wu Z, Xiao D, Xu H, et al. Virtual Inline: A Technique of Combining IDS and IPS Together in Response Intrusion[C]// International Workshop on Education Technology & Computer Science. 2009.

[2] Sinha S. MAC Address[M]// Beginning Ethical Hacking with Python. 2017.

[3] Wang Y H, Ma Y Y, Yang M, et al. Zero permission Android device identification based on implicit identifiers[J]. journal of southeast university, 2015.

[4] YIN Xinming,HU Zhengliang,CHEN Guoliang, et al.Research on IP Video Network Access Detection Based on Decision Tree Classification of Device Fingerprint[J]. Netinfo Security, 2016 (12) : 68-73.

[5] Khadilkar M, Feamster N, Sanders M, et al. Usage-based dhcp lease time optimization[C]// Acm Sigcomm Conference on Internet Measurement. 2007.

[6] Balram S, Wiscy M. Detection of TCP SYN Scanning Using Packet Counts and Neural Network[C]// IEEE

International Conference on Signal Image Technology & Internet Based Systems. 2008.

[7] Ravaioli R , Urvoy-Keller G , Barakat C . Characterizing ICMP Rate Limitation on Routers[C]// ICC 2015 - 2015 IEEE International Conference on Communications. IEEE, 2015.

[8] Eckersley P. How unique is your web browser [C] // Proceedings of the 10th International Conference on Privacy Enhancing Technologies. Berlin,Germany,2010: 1-18.