Testing DTF Method for Applicability in a Real Environment

E.C. Sasu*, O. Prostean**
* “Politehnica” University / PhD Student, Timisoara, Romania
** “Politehnica” University, Timisoara, Romania
cipisasu@ee-software.com, octavian.prostean@aut.upt.ro

Abstract—This paper demonstrates the effectiveness of DTF method, by examining real traffic, collected from a large number of computers. The Destination Traffic Fingerprint Method, described in detail by the authors in [1][2], focuses on the fact that most computers have a constant traffic to a specific set of IP destinations. This paper presents results of the DTF method applied to real traffic and proves that it is possible to extract a reference fingerprint based on the traffic to constant IP destinations.

I. INTRODUCTION

Even though the IT infrastructure has known a very important growth in the last years, the security issue still remains on top. Many researchers around the world are involved in different segments of network security and all of them are trying to find solutions to the vulnerabilities that are discovered.

From all the various types of network security issues, the authors of this paper are focused on a specific threat: penetration attack. This problem is very important because often happens that a company finds that an individual or a group of individuals has gain access to private data or services, within the company.

Some times, the problems are caused by people with no harm purpose, which simply try to benefit from different services offered by a company. But unfortunately, many times these attacks are caused by people with malicious intentions, which are trying to steal, alter and destroy data or services within the victim company.

Therefore, it is very important to detect all the intents of penetrating network security. A simple solution is to maintain a list of authorized MAC addresses and all the traffic that originates within the company, from an unauthorized MAC address is declared unauthorized. The main problem comes from the fact that the MAC addresses can be easily spoofed. In this case, a potential intruder can hide himself by spoofing a valid identity.

In the literature, there are a lot of solutions that provide the discovery of spoofed MAC addresses, like those in [4], [5], [6], [7], [8], [9]. The authors of this paper came with a new approach, regarding this subject. It is not our purpose in this work to describe the method in details, because we did this in [1] and [2]. That is why we will give only a brief description and after that we will focus on proving that this method really works.

Everything started with the observation that most of the computers are used by a single user, which usually has a few destinations to communicate in a constant manner. This destinations can be combined in a set of pairs P(IP, PP), where IP refers to the IP address of the destination and PP refers to the percentage of presence for the specified IP in the network traffic. The set of pairs can be used as a reference fingerprint for the computer and then, in real time, a system can record the actual fingerprint found in the network traffic and can compare it with the reference fingerprint, resulting a value named “Overall Degree of Recognition” which defines how much we can be sure that a network station is what it claims to be.

II. SCOPE OF THE DTF METHOD

Before we continue to describe the test results, it is very important to determine the scope of DTF method. Is it applicable on any computer?

The answer at the above question is “no, DTF method is not applicable on any computer”. In order to make it applicable, there are a few constraints:

− single user per network station
− similar daily activity

This two constraints determine the boundaries of the scope of DTF method. The first one, concerning a single user per network station can be extended to “multiple users per network station”, if all are using the same software applications, in similar ways.

Maybe we can think that the method has only a small scope because of the above constraints. But in fact, the scope is larger when we carefully analyze those who would benefit from the use of DTF. That is why we have to establish the potential beneficiaries.

To determine this, we keep in mind that we are focusing on the issue of MAC spoofing in network activity. In the most common cases, a regular person which has a personal computer, does not need special tools to provide the MAC spoofing signaling. Those that need this technique are the companies which are using infrastructure networks in order to provide the environment for their daily activities. They really need to know in every moment that only the authenticated users gain access in the network.

So, if we determined that DTF method is addressed mainly to the companies rather than individual persons, we have a real advantage from the fact that usually, the two constraints mentioned above are present in the most companies:

− usually each network station is managed by a single user
in the daily activity, the employees are working with the same software applications, installed for the purpose of the company.

Considering this, it is very possible to extract a set of pairs \( P(\text{IP}, \text{PP}) \) as a reference fingerprint, for each network station. This reference fingerprint can be used to determine if a MAC address which generates traffic is spoofed or not.

Even though we emphasized that the main utility of DTF method is for the companies, we still say that it may be applicable to many personal computers. This is because many persons use their computer in similar fashion every day. Therefore it is possible to extract the reference fingerprint from the daily traffic. Each user has an email address that is checked for new messages in a regular manner. An antivirus software can be installed and it will create a constant traffic with its update server. Also, other software applications might require Internet traffic.

As a conclusion, we can say that in most cases we can determine the reference fingerprint of a network station. If we think about a school laboratory or an Internet Cafe, probably DTF method will not work because the computers are used by many many users, in totally different ways. But this cases usually do not require MAC spoofing detection.

III. TESTING ENVIRONMENT

A. Describing the environment

The main objective of this paper is to demonstrate that DTF method can be applied. In other words, we want to demonstrate that computers do have constant traffic to some IP destinations and therefore, those destinations, together with their percentage of presence, can compose a reference fingerprint that differentiates the real network station from the spoofed ones.

As described in the above chapter, the main scope of DTF method relates to companies rather than personal computers because it is more likely to extract a reference destination fingerprint where users are determined to use frequently the same applications in their daily activity. But we also stipulated that it is possible to extract this fingerprint even from personal computers.

Therefore, our objective was to test the fingerprint generation process on personal computers. If DTF principles can be applied in this circumstances, then the results generated for companies have to be more reliable and they have to offer much better performances. If we can extract the reference destination fingerprint in a personal computer, where no rules are restricting the traffic, then it means that it has to be more certain that we can do the same with better results, in a company environment, where many regulations are limiting the usability of network stations.

So, for the validation test, a Packet Recorder Software Module was installed on about 100 different computers. This software module recorded the network traffic as it was seen at the network interface card level, using the driver described in [3]. Each recording has considered a different time interval, as described in Table 1.

After completion of the data collection, the purpose was to find the answer to the next question:

“Is it possible to identify IP destinations that are found constantly in the traffic?”

To answer this question, we generated the reference fingerprint at fixed consecutive time intervals, and our concern was to see if we can find IP destinations that are found in many or even in all the intervals.

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 50 hours</td>
<td>20,00%</td>
</tr>
<tr>
<td>50 – 100 hours</td>
<td>30,00%</td>
</tr>
<tr>
<td>100 – 150 hours</td>
<td>12,00%</td>
</tr>
<tr>
<td>150 – 200 hours</td>
<td>19,00%</td>
</tr>
<tr>
<td>200 – 250 hours</td>
<td>7,00%</td>
</tr>
<tr>
<td>250 – 300 hours</td>
<td>0,00%</td>
</tr>
<tr>
<td>Over 300 hours</td>
<td>11,00%</td>
</tr>
</tbody>
</table>

With these tests, we do not want to calculate the Overall Degree of Recognition. What we want is to find destination IPs that are present in many consecutive time units.

B. Notations and abbreviations

We note “TU” - a Time Unit interval and thus, the whole interval is separated in “N” consecutive TUs. For each TU, we extract the reference destination fingerprint.

After the generation of all “N” reference fingerprints, the result is given as a set of pairs defined as follows:

\[
M \{ P(TTUi, IPi), i=1, TOTALIP, TTUi>1 \}
\]

where:

- \( TOTALIp \) – total number of destination IPs, found in all intervals
- \( IPi \) – current IP address
- \( TTUi \) – total number of TUs where IP is found in the reference fingerprint

C. Test settings

The results shown in the next chapter are generated using the following test settings:

- One time unit has 8 hours (1 TU = 8h)
- The evaluated period consists in 24 consecutive time units
- all destination IPs that have less than 1% percentage of presence, are ignored and not included in the reference fingerprint
- if the time unit is divided in 4 equal subintervals, and the IP is not found in at least 2 subintervals, then that IP is ignored and not included in the reference fingerprint
IV. TEST RESULTS

A. General characterization of the collected data

Analyzing the collected data we found that 77% of the computers have relevant information, 19% of the computers do not record relevant information and 4% have no data.

By “relevant information” we understand MAC addresses with fingerprint generation succeeded. Therefore, all the “irrelevant information” refers to MAC addresses for which the fingerprint generation failed.

The high percentage for relevant data is very important in our study. This value communicates that it is very possible to extract the reference destination fingerprint even though the computers are not restricted by rules or regulations.

Before we go in details regarding this results, in the next section we will explain a few things about the representation of the results.

B. Understanding graphical representation

The results extracted for a 24 TU interval is shown in a diagram as in Fig. 2. The graphic shows a few IP addresses that are found in consecutive TU reference fingerprints. For each destination IP, we represent the percentage of the presence in each TU. In this way, a line in the diagram represents the variation of the percentage of presence from a time unit to another.

Each MAC address that was evaluated received a “local code” to facilitate the human recognition. This code can be any sequence of characters.

The Legend contains all destination IP addresses that are constantly found in the traffic. Each is represented with a different color and sorted in alphabetical order of the IP addresses.

One point in the diagram represents the percent of presence of a certain IP, as it is in the reference fingerprint for a Time Unit. So, each line in the diagram communicates that an IP address is found in consecutive Time Units and we understand from this, that the traffic to that IP can be declared as “constant”. Also, each line shows the variation in time of the percentage of presence. For best performances, this variations should not be very big, from a Time Unit to another.

The diagram shown in Fig. 2 can be interpreted as follows:

− during 18 TUs (144 hours), from all the destination IP addresses to which the MAC address is sending packets, 5 IP addresses are found in each Time Unit, meaning that they represent “constant traffic”
− IP 94.170.105.3 has a percentage of presence that varies between 95% - 100%. Also, IP 105.1.0.0 has a percentage of presence that varies between 90% - 95%.
Even though the other three IP addresses have smaller percentages of presence, they can be part of the reference fingerprint because it is not really important to have high rates of percentage of presence, but to have IP addresses that are constantly found in the network traffic.

C. Test results

When we look at the results extracted from the real traffic from all the evaluated computers, we are able to classify the destination IPs as follows:

- destination IPs present in traffic on many successive Time Units, with small variation of percentage of presence
- destination IPs present in traffic on many successive Time Units, with big variation of percentage of presence
- destination IPs present in traffic only occasionally or in a few consecutive Time Units

For all the diagrams, we keep in mind that the Time Unit number starts from 0, which means that if the record time ends at “K” Time Unit, we have a total of “K+1” Time Units for the evaluation.

The first class of traffic is presented in Fig. 3. There we can see a few examples of constant traffic, where the variations of the percentage of presence is small.

Fig. 3a shows that in all 21 Time Units we have 6 destinations present. The variations of the percentage of presence are about 10% for two of them and only about 5% for the other four. This means that if we establish a fingerprint using these IPs, it is very probably that the network station will be identified correctly with high rates.

In Fig. 3b we have 5 IP destinations with constant traffic, with bigger variations, but still they are limited around 15%.

The variations can be quite very small, as in Fig. 3c. There, excepting one IP which has very big variations, the other five destination vary in percentage of presence with only 2% - 3% for all 24 Time Units. The fingerprint will be very strong in this case.

Also, in another case shown in Fig. 3d, we see that a large number of destinations are constantly found during the 24 Time Unit evaluation time.

In Fig. 3e, the values for percentage of presence are small, but with small variations. This is still acceptable, because we are not seeking necessary to have high rates for percentage of presence. We are looking for constant traffic even if it has smaller values for the IP presence in traffic.

The second case, is the one where the computers do have traffic to some IPs in every Time Unit, but the percentages of presence have big variations from a Time Unit to another. A few examples can be found in Fig. 4. All the diagrams present very clear that we can identify destinations with traffic spread across a long time. The problem is that in those cases, the fingerprint generation module has to be fined tuned in order to achieve good performances in determining the Overall Degree of Recognition.

But, a careful look at the diagrams in Fig 4, reveals some destinations with small variations in percentage of presence. This means that the recognition can be improved if we emphasize this destinations over the others.
The last case mentioned at the beginning of this chapter, is the one that refers to the situation when we can not find destinations present across a long time. Instead, we see only local or temporary communications with different IPs. For those cases, the fingerprint generation process will fail and will not be possible to calculate the Overall Degree of Recognition.

V. CONCLUSIONS

The results of this test are very encouraging. This is because we proved that the computers really communicate in a constant manner with some destination IP addresses. If we focus on the proper generation of the reference fingerprints, we will obtain a good value for the Overall Degree of Recognition and in this way we can simply identify any attempt of MAC spoofing.

The test was done on personal computers, where no restrictions or regulations are found. This means that the involved cases represent a worse environment than the one found on companies. An individual person is supposed to use his computer in a totally unpredictable fashion, but still, the results show that the fingerprint generation process will succeed in many many cases.

Another aspect that has to be remembered is that a large number of the collected data contained traffic generated by torrent software. This applications are producing traffic to thousands, hundreds of thousands and even millions destinations. But, as we could see, DTF method could still be effective even in this environment, which is a very good remark about it.

DTF method allow easily the detection of inner or outer attacks. It has no importance if the network is wired or wireless, because the method can handle all the transmission mediums.

REFERENCES