Network Simulation for MAC Spoofing Detection, using DTF Method

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 simulating a network and testing how it responds to different situations. All the packets injected into the simulator are collected from real computers. The simulator allow to create a virtual network, with as many network stations as desired. The simulation proves that DTF method can quickly determine if a MAC address is or it is not what it claims to be.

I. INTRODUCTION

The Destination Traffic Fingerprint method was described in detail by the authors in [1] and [2]. The base idea is that a computer connected to a network, will generate traffic to a set of destinations. It is very possible that from all the IP destinations, to extract a few of them, characterized by the fact that they have a constant traffic in time.

The set of IP destinations that have a constant traffic in time can be collected into a reference fingerprint. In this fingerprint we record not only the IP, but also the percentage of presence in traffic of that IP. The reference fingerprint is compared with the actual fingerprint and based on the difference between them, the method establishes the Overall Degree of Recognition. This value represents how certain is that a MAC address is real or spoofed.

In this paper we focus on the demonstration that this method really works and it can be used in real networks. For this, a simulator was build in order to provide the test environment. We used real packets, collected from a certain number of network stations. This packets can be injected into the simulator and thus we can see how the MAC addresses are checked for consistency in real time.

Chapter II will present a few functions that allow us to see different aspects of the collected data. Chapter III will show the simulator and the way it works.

II. VIEWERS IN THE SIMULATOR

A. Viewing IP distribution over time

The DTF method tries to find a fingerprint based on the destinations that have a constant traffic. To demonstrate that a destination has a constant traffic, it is important to visualize the traffic. That is why we created a diagram that allows to view it like in Fig. 1.

We evaluate the traffic at a minute level. This means that our concern is to search if we have traffic to the evaluated destination in every minute. We are not interested to count the packets, but only to know if we had at least one packet sent to the destination in every minute. On OX axis we pointed the minutes, starting with 0. On OY axis, for each minute we have only two possible values: “0” if we have no traffic to the destination, or “1” if we have. When the source has stopped working for a time greater than a certain value, we consider it “stopped” and we mark the time on the diagram with a blue rectangle.

Using this tool it is very simple to study different situations found in the network traffic. A few of them are
discussed below.

The first type of traffic seen with the help of the visualizer, is shown in Fig. 2. The IP destinations that fit in this category have a continuous traffic directed to them from the MAC source. Excluding the time when the network station was in stop mode, almost all the time we have traffic. By “all the time” we understand “each minute”. Having this kind of destination in a reference fingerprint, makes it very powerful.

The second case refers to the destinations to which the traffic is not “continuous”, but instead is “constant” with high rates for the percent of presence, as in Fig. 3. This IPs are also very useful in a reference fingerprint precisely because of high rate of the percent of presence.

The third case is defined by destinations that have traffic at low rates for the percentage of presence, but distributed at constant intervals. Fig 4 shows a few examples of them. Even if we prefer high percentages of presence, we do not ignore this cases because of their nature of repeating. That is why they can be included in the reference fingerprints.

All three cases discussed above have a relevant importance in DTF method, and that is why the fingerprint generator module has to be able to detect them and isolate them among the others.

As the authors have proved by different tests, many computers do have traffic to destinations like these. So, the reference fingerprints can be very strong and any attempt to spoof the MAC addresses will be recognized and signaled by DTF method.

The power of DTF method comes from the fact that it is likely impossible for an attacker to generate on his computer the same reference fingerprint. In this way, the calculated Overall Degree of Recognition will have very poor values and the system will declare the MAC addresses as spoofed.
A problem that has to be solved is represented in Fig. 5 and refers to some “tricky cases”. We call them “tricky” because the IPs do not have a constant traffic even though the percentages of presence do have significant rates.

The first case is in Fig. 5 and has 100% presence in the first half of the interval and 0% presence in the second half. This will produce a 50% value for the percentage of presence which in fact is a very good value, but the traffic is not constant. The second case also shows a traffic only in a period of time between two stops of the MAC source.

This cases have to be identified very precisely and excluded from any reference fingerprint because otherwise will determine the system to raise many many false alarms.

The last case that we discuss here is concerning the destinations that do not have at all a constant traffic. They are represented by isolated local traffic. We can not find any rule of repetition because they rather have a sporadic nature than a repeating one. Fig. 6 displays a few examples.

**B. Viewing ODR distribution over time**

Another important utility help us to visualize the variation of the Overall Degree of Recognition of a certain MAC address, from the moment when it was found in the network until the end of the evaluated period.

In Fig. 7 we can see an example for the variation of the Overall Degree of Recognition. The MAC address that was observed had three stops in the evaluated period of time. Based on the reference fingerprint that it has, the system recognizes the network station as in the diagram. For a fast recognition we need a very strong fingerprint.

By a “strong fingerprint” we understand a fingerprint where the IPs that make it, have a strong rate of repetition. It is not necessary to have very high rates for the percent of presence. Instead, it is preferable to have constant rates for the repetition in traffic.

Depending on each situation, we need the system to raise a good conclusion in a shorter or longer time. This time will directly influence the type of IPs that can form a reference fingerprint.

The fingerprint generator module can be configured for best performances, in order to provide the expected speed of the MAC spoofing signalization.

![Fig. 5 Possible “triky cases”, with significant rates for the percentage of presence, but found only temporary](image1)

![Fig. 6 Isolated traffic](image2)

![Fig. 7 Variation of the Overall Degree of Recognition](image3)
C. Viewing fingerprint variation over time

This is another helpful utility in the process of studying DTF method. We need to be sure if we can establish a strong fingerprint in order to reduce false alarms and obtain a reliable detection system for the spoofed MAC addresses.

Therefore, we divide the hole evaluated time interval in smaller intervals, called “Time Units” or simply “TU”. In each Time Unit, the utility will generate the fingerprint using the corresponding traffic. At the end of the generation for all TUs, the utility will compare all of them and will generate a graphical representation in order to see the IPs that appear in consecutive Time Units.

The results are given in a specific format, as shown in Fig. 8. Each line in the diagram represents the variation of percentage of presence for an IP that is found in consecutive Time Units. In this way we can visualize the IPs that will produce the best performances if included in the reference fingerprints.

Each Time Unit is numbered with “0” at the beginning. We have valuable data starting with TU=0. Analyzing the results we can identify a few categories, mentioned in a brief description below.

The first one is the case where the destination IPs have small variations in time for the percentage of presence. This are the best cases and desirable to fill the reference fingerprints.

The second case show variations not so small, but still smaller enough to reach strong fingerprints. If bigger variations are present, the Overall Degree of Recognition can bring the system to raise false alarms.

We keep in mind that DTF method does not really need high rates for the percentage of presence. Speaking in terms of the variation of this percentages over time, it is rather important to have smaller variations, even though the percentages have small values.

III. NETWORK SIMULATION

A. Settings for the fingerprinting generation

The first thing to do before the simulation starts is to establish the settings for the fingerprint generator module. This settings affect the behavior of this module.

It is very important to determine a fine tuning, because the fingerprint generator module will automatically generate the fingerprints for the MAC addresses that are detected.

B. Defining the virtual network

The second thing necessary before the simulation starts is to define the virtual network. This task is done in the simulator by selecting the sources from where it will take the data.

The data can be collected from different computers, and added to a global database that keeps all of them together and does a few statistical queries for each individual database.

C. Loading data and simulating the network

After the above steps were done, the simulation may begin. The window that appears on the screen contains a few sections:

- Simulation Settings
- Network Zone
- MAC Table
- Actual Fingerprint IP List
The Simulation Settings zone is in the left-bottom part, and refers to the following aspects. In the first place, for each database that was chosen to be a source for the simulation process, we can set up the starting point. By default, the starting point is set to be the first minute in the database, but we can change that and force the simulating process to start using a different offset for each database.

In second place, the simulation can be set up to start and end at the right moment. Also, we can choose to reset or to keep the actual signatures when a network station stops. Each real minute will be simulated using a certain number of seconds, established by the user and the determination of a “stopped” network station is done by a parameter called “Maximum pending time”.

With all of these parameters set up, the simulation can start. First of all we press the button “Load data” in order to load the information in a cash memory for better performances and then START button effectively starts the simulation.

The Network Zone in the simulator is in the left-upper part and draws a graphical representation of the network, using relevant icons for each situation:

- unidentified source
- identified source my MAC address with three different subcategories:
  - real MAC
  - possible real MAC
  - spoofed MAC

The MAC Table keeps record of all MAC addresses which were found in the network and for each tells a few important information:

- MAC address
- Status: “active”, “pending” or “stopped”
- Total number of minutes from the first appearance
- Overall Degree of Recognition
- Total number of IPs in the fingerprint

Actual Fingerprint IP List shows all the IPs that form the fingerprint for the selected MAC address in the MAC Table. For each IP, the simulator tries to geographically locate the IP and tells the actual percentage of presence.

The simulation can be paused in any moment and then restarted from the point where it stopped. When the simulation is paused, we have the following functions available:

- generate the variation of the Overall Degree of Recognition for the selected MAC address in the MAC Table
- observe the presence distribution for the selected IP in the actual fingerprint
IV. CONCLUSIONS

In conclusion, the simulator is a very powerful tool that provides a very good testing environment for DTF method.

The data that is injected into the simulator is real data, collected from real network stations. From all the collected data, the simulator is able to use only those that are selected by the user.

All the viewer utilities are a good help in validating different decisions regarding DTF method.

Using this tool, it is possible to test the future work in this area, in order to prove the power of DTF method.

REFERENCES

