# ET4394 Wireless Networking - Wireshark Assignment

Sjors Nijhuis 4157478
Kevin van der Mark 4154509

**This report describes the Wireshark project required for the course Wireless Networking, ET 4394, at the Delft University of Technology.**

## Introduction

The aim of the project is to extract information on chip set and vendor type in WiFi systems (W13). The assignment is chosen based on the slides "Slides_Wireless_Networking_P_Pawelczak_Project_Division_2018" found on the Brightspace page of the course.

## Goal

The goal of the project is to find out where most Apple iPhone users and where most Android users get on and off the train on the route between Delft and Amsterdam. This is done by searching for patterns in the types of mobile phones used based on obtained MAC addresses and their respective usage based on the amount of packets send/received in different locations. This should lead to a demographic overview where locations can be pin pointed where users of the different brands get on and off the train. By measuring where the device send their first and last packet, the routes users take can be determined. Combining this with the amount of packets send per user, an average network usage can be calculated per vendor type.

## Software

The software used for "sniffing" is TShark combined with a Python script. The data is afterwards analyzed and processed using Matlab. All files used can be found in the Github repository of our group: `https://github.com/1424Bravo/Wireless.git`.

### Capturing

When capturing for extensive time a capture file can get quite large. Since only basic packet data is needed for this application, only MAC addresses and relative packet time are saved to a JSON file. This is done using the capture filters command: ***-e eth.addr -e frame.time_relative***. This reduces the size of one received packet to approximately 250 bytes (in JSON format) instead of a whole packet, which can be up to 1.5 megabyte.

### Processing

Processing is done in **Python**. A script has been made to parse the **JSON** file to a number of dictionaries, which can be easily accessed and searched. From here, several statistics can be extracted, such as the number of addresses from a certain vendor, the number of active clients on the network, and the first and last time an address was seen on the network.

These results are plotted by a MatLab script in order to generate graphical representations of the resulting data.

## Measurements

To measure differences or patterns in network use of various vendors, the active time on the network and the number of transmitted packets is measured. Therefore, for each packet on the network, the MAC addresses and (relative) time are stored. The active time and number of packets can be easily calculated based on these measurements. For a train network, clients are expected to enter and exit the network only when the train is at a station. It is expected that many clients enter and exit the network simultaneously, but that the total number of active clients only slightly changes, since many people step in and out of the train at roughly the same moment.

## Results

At the moment of writing this report, no live measurements have been obtained yet. This is due to the fact that no suitable WiFi dongle has been available so far. The deadline for this project has been overlooked and replaced with a deadline actually stated for the SDR project, March 15th. This report is to show the project has never the less been taken seriously and real measurements and conclusions will be obtained.

The analysis of the data as explained in previous sections has up to this point in time been done on test data obtained via the website `https://wiki.wireshark.org/SampleCaptures`.

Figure 1 shows the number of MAC addresses on the network per vendor. The most active vendor on the network for this measurement is by far Hewlett Packard. This is due to the fact that this is a sample pcap file obtained on a server network.

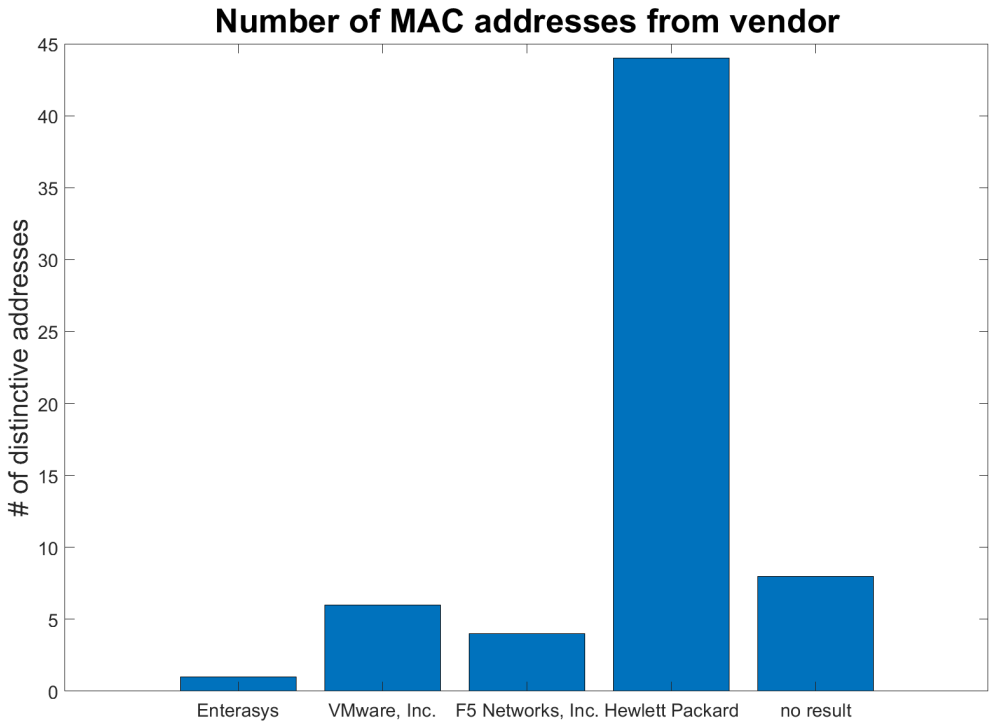


Figure 1: The number of active MAC addresses on the network per vendor.

Figure 2 shows the total number of active clients on the network. The first 50 seconds of the graph are not very relevant. Devices already connected to the network started sending their first packets in this period which looks like a huge amount of new connections but actually is a measurement error due to the start up phase. The same yields for the last 50 seconds; quite a lot of clients did not send data for a while, and are therefore marked as inactive.

Figure 3 shows the average time that hosts of a certain vendor types spend on the network. In this sample data most hosts are active for a long period, since a lot of network equipment is not often connected or disconnected.

In figure 4 the average number of packets from hosts of a certain vendor type is shown. A lot of 'no result' hosts are sending lots of traffic. This traffic is probably generated, and thus no vendor exists from this data. Expected is that some vendors have a tendency to transmit more data - for instance for finding locations or getting the local weather. However, the difference in transmitted packets are now far larger than would be expected in a train measurement, since the experiment is conducted for a longer period.

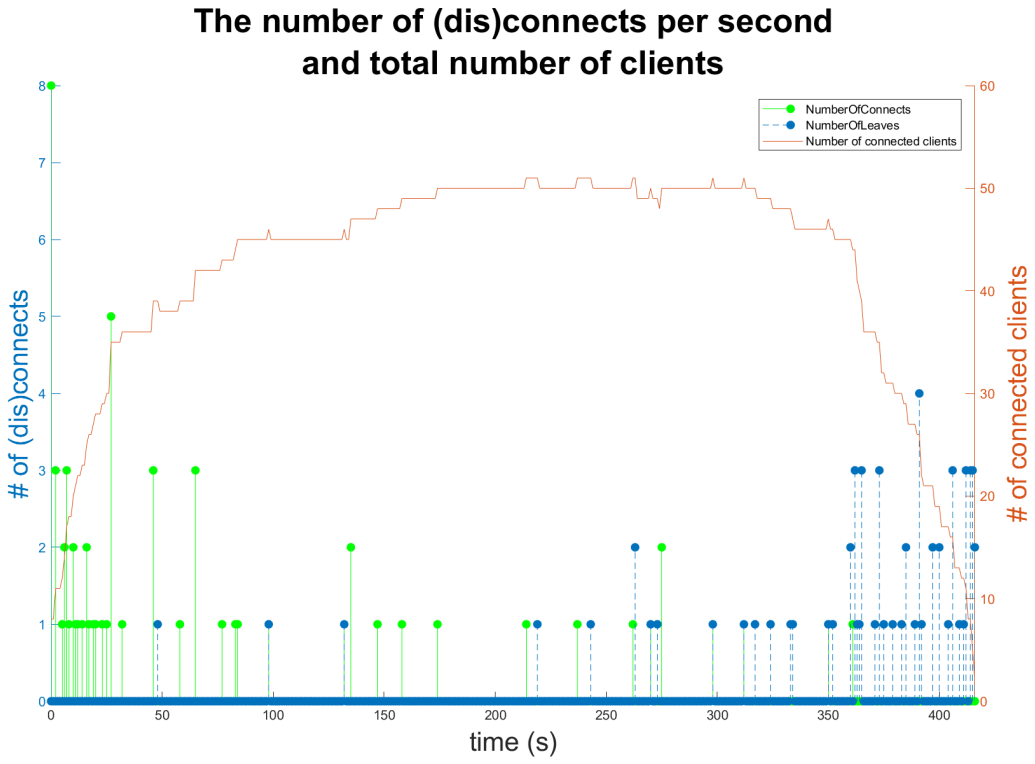


Figure 2: The number of connects and disconnects of clients per second and the total number of clients that are connected to the network.

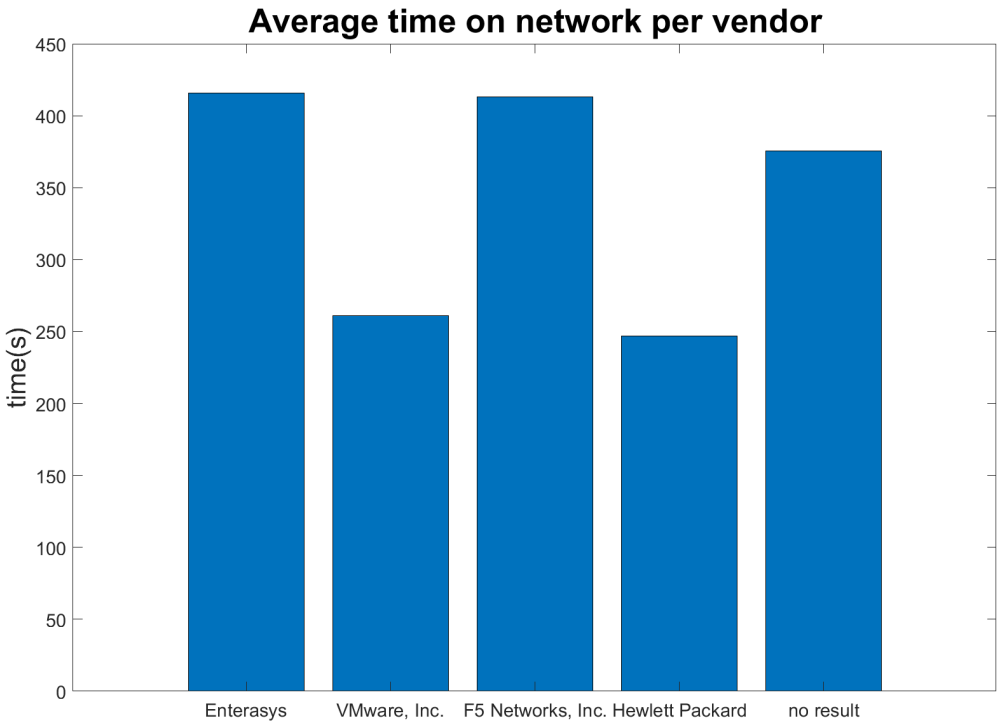


Figure 3: The average time a host spends on the network per vendor.

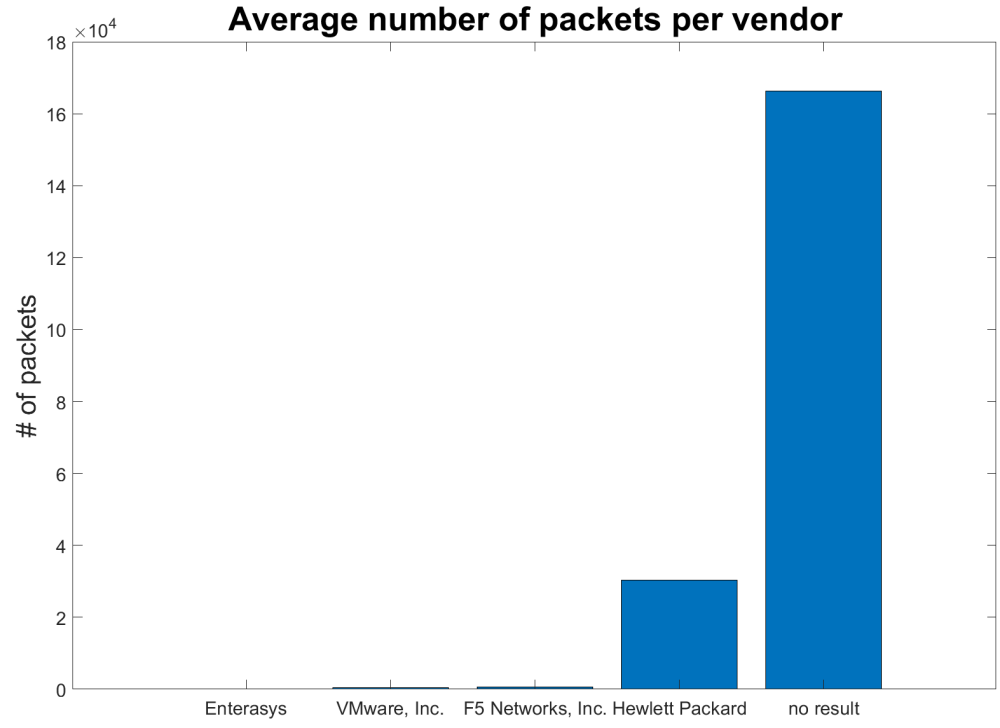


Figure 4: The average number of packets a single address transmits per vendor.