Analysing variation of 802.11 frame distribution and calculating an average person's wake up time on a weekday vs weekend

Sourav Mohapatra (5204887, Github: souravmohapatr) Tolga Parlan (5392578, Github: tolgaparlan)

March 1, 2021

https://github.com/souravmohapatra/wlan-packet-analysis

Abstract: *Observing the nature of utilization of the wireless transmission medium gives us information on where and how improvements to protocols and channels can be done. In this report we present our experimentation and analysis on the distribution of various packet types present in 802.11 WLAN spectrum.*

1 Introduction

The WiFi IEEE Standard specifies a lot of different types of packets, each with a different functionality aimed to address a particular aspect of the wireless communication. In a broad classification, the packets can be divided into three primary types - Management frames, Control frames and Data frames. Management frames perform supervisory functions; they are used to join and leave wireless networks and move associations from access point to access point. Control frames are used in conjunction with data frames to perform area clearing operations, channel acquisition and carrier-sensing maintenance functions, and positive acknowledgment of received data. Control and data frames work in conjunction to deliver data reliably from station to station.

Since these frames are so closely related to the protocol and the nature of WLAN traffic, analyzing their distribution with respect to parameters such as the time of the day and week and environment can give numerous interesting correlations. As a part of this report, we aim to find out a few of such relations and provide appropriate justifications.

2 Method

As a part of this experiment we aim to measure the following:

- Distribution of the 802.11 frames classified by their type and subtype.
- Correlating human behavior using properties and frequency of various frames.

2.1 Hypothesis

Due to the large variety in the types of frames, it is obvious that some of the frames will be more frequent than the others. The frequency of such frames can be hypothesized by looking at their functionality. We can reason for the following observation regarding the frame distributions.

- Control and Management frames will form the bulk of the frames distribution. Data frames, although are the frames actually carrying the data, should be exceeded by the number of frames involved in managing the channel access. In fact, this was the motivation behind the development of 802.11ax standard [1].
- The average number of packets over the air depends on the time of the day and the surrounding human environment. On a very broad term, the number of packets should be more during the day than during the night. And there would be variation within this observation too. Some of the frames would be more or less affected by the time of the day.

Apart from these primary objectives, we also plan to show any other interesting observation that we come up with while performing the analysis. For example, we aim to present any unique packets, their functionality and their correlation with the physical environment. With the above objectives in mind, we can present our test setup.

3 Test setup

We used two independent setups to collect the data. One of the setup was at a student housing building in Delft, Netherlands while the other setup was in a suburban area in Famagusta, Northern Cyprus. Both of the setups are laptops with their wireless cards acting as the medium of sniffing. The configuration of the wireless chip-sets are given as below.

- **Setup at Delft:**
  - Intel Wi-Fi 6 AX200
  - 802.11ax Dual Band 160Mhz
  - Ubuntu 20.04, Kernel version 5.8.0
- **Setup at Cyprus**
  - Intel Wireless-N 7260
  - 802.11bgn 2.4 GHz
  - Ubuntu 20.10, Kernel version 5.8.0

Along with the configurations we would like to note a few points about our setup, which might have affected our observations.

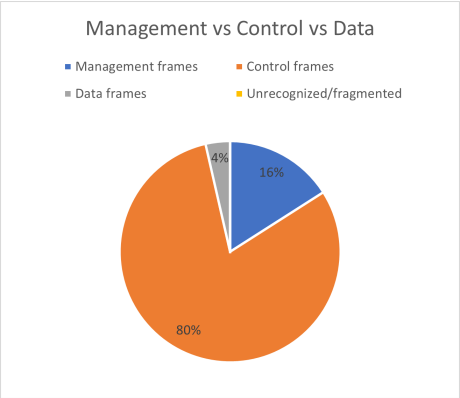

Figure 1: Distribution of 802.11 packets categorized according to their type (wlan.fc.type == 1)

- Since our scenario does not demand any peer device to remain in connected state, all of our packet captures are performed using *monitor mode*. The dwell time for each channel has been set at a random value between 1 and 10 seconds. There are various interesting channel hopping methodologies. In the method proposed by [2] the time spent in one channel is calculated according to the number of packets being received in that channel. Although it is a great way to ensure our capture time is not wasted on empty channels, it did not seem a good idea for our implementation. Our aim was to have an impartial view of the types of packets over the air, and if we change our dwell time based on channel congestion, we could miss out on packets that are sent more in non-congested environment (such as some of the action frames).

- Due to various logistical considerations, we could not bring the element of mobility into our experimentation. We did not have access to any embedded wireless devices and nor could take out our laptop for extended time. What we did try to exploit is the fact that our two setups are in different countries and different environments.

- All of our automated work is done using python wrappers over native Linux commands. The code is present in the Github repository mentioned above. None of the captured packets are made public in lieu of privacy consideration.

We took our measurements intermittently over a week. For the analysis of the packet distribution, we took a total sample of around 13.5 million packets. To understand the variation of the packets over time, we took two samples over two full days, one on a weekday and one on a weekend.

Using this test setup, we carry forward our experimentation and present the results and their interpretation in the next section.

4 Results and Evaluation

The first objective is to analyze the distribution of the frames categorized by their types.

4.1 Overall distribution of the frames

From a sample size of 13.5m packets, we can show the distribution of the various packets. In the figure 1 we can clearly observe that majority (around 80%) of the frame types are control frames.

What are control frames?: 802.11 Control Frames assist with the delivery of Data & Management frames. Unlike management & data frames, Control frames does not have a frame body. This makes it small in size and that can make their high number of occurrence feasible.

Why are they so frequent?: As defined above, control frames are used to control the transmission of the other two types of frames. This in itself gives a trivial explanation to the high percentage. Before a station can send/receive any actual data, it needs to understand if the transmission is possible. And it is the functionality of the control frames to ensure that the transmission goes smoothly. Acknowledgement frames also come under control frames. So for every data transmission, there is an acoompanying acknowledgement frame. In fact, the use of acknowledgement frames became so overwhelming that in 802.11n (HT) the concept of Block Acknowledgement came into the picture. Instead of sending an ACK frame for every data frame, the device can buffer ACKs and send a block of ACKs at one go.

What are management frames?: Management frames, as the name suggests, manage the network. They perform supervisory functions, for an example they are used to join and leave wireless networks and move associations from access point to access point (called roaming[1], introduced in 802.11k, 802.11r, 802.11v). It is natural that management frames take the second place as any device first must be introduced into the network before it can start the transmission.

Why are there so few data frames?: We see that the number of data frames are the lowest. This might be counter intuitive at first but we have to understand the fact that this distribution doesn't take into account the size and the airtime of each particular packet. In general, data frames are much larger than control or management frames. The amount of actual data carried over by a data frame is more.

4.2 RTS and CTS take up the majority of the traffic in a densely occupied environment

With the overall distribution discussed, we can look at the distribution of each individual packet types. The distribution can also be visualized though histograms given in the figures 2 3. We can see that RTS and CTS control frames take up almost 50% of the packet distribution in figure 2, which represent a student dorm with busy wifi traffic. In the meantime, figure 3 shows that in a suburban area where most residents are not home

[1]Interesting fact - Roaming is the method which enables seamless connection to Eduroam throughout the various scattered APs broadcasting the same SSID. The device recognizes if there is a decrease in the signal strength (among other parameters) and performs a switch to a better network, if available. And this is done oblivious to the user, which makes it seamless.

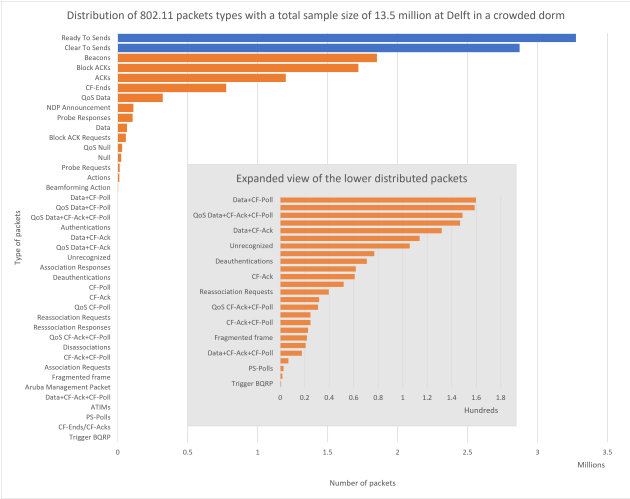

Figure 2: Exact distribution of each type of 802.11 packets in Delft

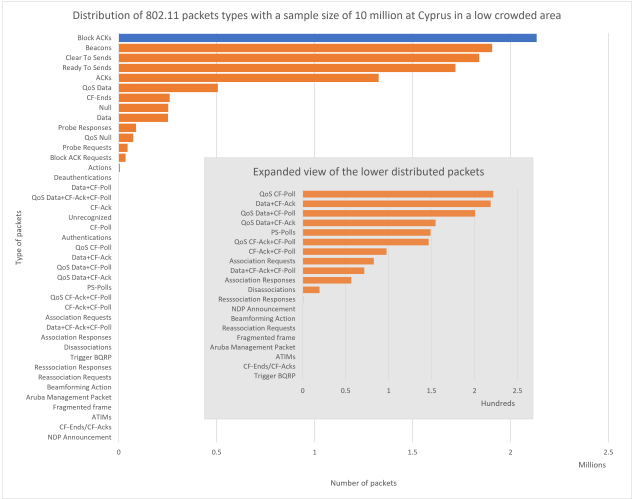

Figure 3: Exact distribution of each type of 802.11 packets in Cyprus

during the day, Beacons and ACKs form the majority. To understand this pattern, we need to look at how a WiFi medium is accessed.

How is wireless medium accessed?: When a node wants to transmit data to another node, it sends out a RTS 'Request to Send' packet. The receiver node replies with a packet called CTS 'Cleared to Send' packet. After the transmitter node receives the CTS packet, it transmits the data packets. Encoded within the RTS/CTS packets is a duration field. The duration field is set such that the data transmission can be completed within the designated time period. If a transmitter node does not receive a CTS packet it understands that the medium is busy and thus enters into an exponential back off mode.

Is this observation justified?: Since RTS and CTS are so closely coupled with the ability to use the transmission medium, every node sends out RTS and CTS before every transmission. This justifies their large percentage when a lot of transmission occurs. On the other hand, if the wireless medium is relatively contention free, RTS and CTS numbers will be lower.

What are the next most frequent packets?: From the histogram for dense environment, we can see that Beacons follow RTS and CTS. Then it is followed by Acknowledgement frames. The high number of beacons are justified as they are broadcast, independent and periodic transmission done by (almost) every AP. They are the identifying tokens that client nodes need to initiate connection to an AP. If there are plenty of Access Points but relatively few active Mobile Stations, Beacons can easily become the most common packet, as observable in our suburban environments measurements from Cyprus where most residents are away during the day, presumably working. Acknowledgement frames too justify their existence by the fact that they are sent for every data transmission. Acks are followed by CF-Ends (Control Frame end) and QoS Data (actual data packets). As we go further down the frequency distribution, we see that the number of packets for other types fall off drastically. We can safely say that more than 90% of the transmission in the air consists of just 5-6 types of packets.

Why are the bottom 5 packets so rare?: The 802.11 set of protocol is vast and has a lot of different functionalities. Since our WiFi adaptor in Cyprus was relatively old and did not support the newer 802.11 protocols, the bottom 5 packet types were newer frames of this type. The below are the frames in decreasing order of rareness which we actually observed.

- *Trigger BQRP*: This is an 802.11ax trigger frame. Trigger frames have been introduced in 802.11ax standard to help in allocation of resource units for OFDMA transmission. BQRP (Bandwidth Query Report Poll) is sent by the AP to the clients to send a report on the occupancy of a certain channel bandwidth. Since this is a Wi-Fi 6 frame, which is not yet widely deployed, it is very rare. In fact, we could only get *one* such frame in the sample size of 13.5 million.

- *PS-Polls*: These are power management frames. They are used by devices to fetch data from the AP in case of going into a sleep state.

- *Data/CF-Ack/CF-Poll/Combination of these*: Most of these frames have been replaced by their QoS equivalent. Although these are valid frames, they can be mostly found in legacy devices only. Thus the low numbers.

- *Reassociation Requests*: Mobile stations moving between basic service areas within the same extended service area need to re-associate with the network before using the distribution system again. Since our setup is stationary, and the scenario is rare, the frame is also rare.

- *Deauthentications*: This basically shows the number of times we turn our wifi off or manually disconnect from the network. Most of the times, our client node goes out of the range and then the AP performs an internal disconnection without

the explicit deauthentication frame, justifying it's low number. Relatively higher (but still very low) number of deauthentication frames were observed in Cyprus, which can be explained by the experimenter's house having two separate but overlapping extended service areas which needs manual disconnection and connection to the router which gives better service in certain rooms.

4.3 When do people wake up?

The great thing about analyzing wireless packet captures is that they correspond to actual human behavior. In this part of the experiment, we try to find a correlation between the types of 802.11 frames over the air with the sleep pattern of people.

What is the experiment setup?: This experiment includes packet captures over two full days. The first day is a weekday, Thursday, 24^{th} of Feb and the second day is a weekend, Sunday 28^{th} of Feb. The sniffing setup was kept in the middle of the common kitchen area of a dorm. Apart from the day of collection, every other parameter were kept the same. We present the results in the figure 4.

Analyzing the results: In figure 4 we can see the variation of the three types of frames - Management (in red), Control (in green) and Data (in black) against the time of the day over two days (weekday and weekend). The lines are 10 sample moving average of the number of packets per 10 minutes plotted on a log scale. The first thing to note is that the number of management frames stay constant thought the period. Management frames are largely made up of beacons among other frames. We saw from the distribution in figure 2 that the beacons are in the top 3 frames. And the number of beacons remain constant if the number of beacon-ing APs are constant, which is the case in any residential area. The variation in Data frames are caused by nodes sending/receiving data. It can increase in two cases - number of such nodes are high or a single node is transmitting a lot of data. Thus we cannot, to a degree of certainty, correlate data frames to the number of active nodes as even a single user with high data usage can skew this heavily. The most important trend is that of Control frames. An increase in control frames, by definition, directly indicates an increase in network activity. Thus we can use this variation to correlate to a physical aspect of the people using the network such as their sleep pattern. In the figure, we can see that at some point in the morning, the control frames start increasing and cross the number of management frames. We can take this threshold of crossing the management frames to be the point when a average person wakes up.

Weekday vs Weekend: To make the observation and correlation more concrete and interesting, we performed the analysis for data on a weekday and on a weekend. As we can see from the figure, the *average wake-up time on a weekday is around 8:00 AM* while that in *a weekend is 9:00 AM* - a increase of 1HR on average, which is pretty accurate when compared to a real life scenario.

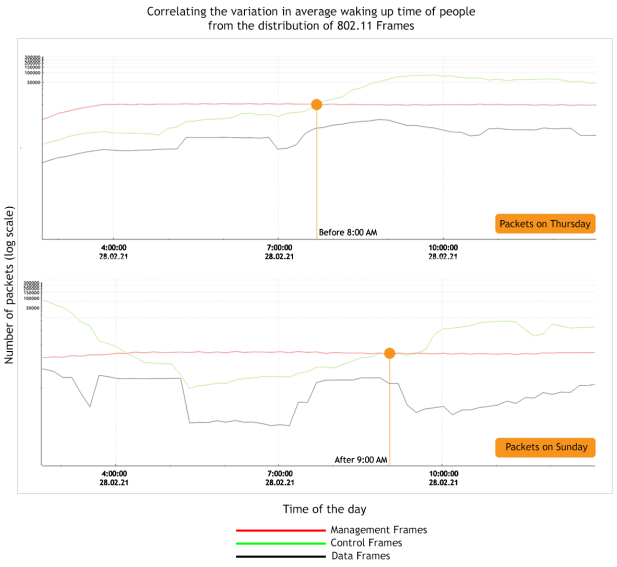

Figure 4: Comparing the sleep pattern of people in a student housing using the nature of packet distribution. The spike in control frames indicate increased activity, which can be correlated with the average number of people using their devices.

4.4 WiFi as a behavior predictor

The above simple experiment showed how the wireless medium can be analyzed to understand a simple human behavior. There can be a wide range of correlation that can be done. With a bigger sample set and intelligent data science methods, we can correlate even the minute human behavior patterns such as average times spent on attending meetings, preferred time for streaming media, average dinner time etc.

5 Conclusion and Reflection

In this report we presented our findings and analyzed the results of how different frames in 802.11 occupy the wireless medium. We correlated an average person's sleep cycle with that of the frame distribution which showed an example of how we can predict/learn about crowd behaviour using WiFi. In the process of doing so, we learnt about the functioning of WLAN, the nature of wireless medium and how 802.11 protocol works.

References

[1] B. Bellalta. "IEEE 802.11ax: High-efficiency WLANS". In: *IEEE Wireless Communications* 23.1 (2016), pp. 38–46. DOI: 10.1109/MWC.2016.7422404.

[2] U. Deshpande, T. Henderson, and D. Kotz. "Channel Sampling Strategies for Monitoring Wireless Networks". In: *2006 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*. 2006, pp. 1–7. DOI: 10.1109/WIOPT.2006.1666486.