Comparing Internet Traffic between Weekday and Weekend

Naomi Giertych

The purpose of this analysis is to gain a better understanding of how traffic may differ between the day of, shortly after, and a couple of weeks after Microsoft’s Security Patch on Tuesday, April 10, 2012. One might expect that traffic on the day of the security patch release to be very different from traffic on other days month. To answer this question, I examine the time dependence of different traffic types and level of activity based on traffic destination on Tuesday April 10, Saturday April 14, and Wednesday April 25, 2012 using the data from unassigned IP addresses.

Each communication over the internet has two identifying factors: the information for where it is going and where it is from (the IP/source/port) and the type of communication (UDP and TCP). UDP packets are used for more complex transactions such as videos where users can deal with being frustrated when all the information is not received. TCP packets are used for simpler, but more important transactions where information lost could result in catastrophe such as email. Ports send and receive the information contained in the communication packets.

To obtain an initial understanding of the time dependencies between the different days, I make a time plot the traffic, TCP, UDP, and number of unique sources over minutes in the day and examine their corresponding marginal densities. Below are the graphs for the overall traffic on each of the days.

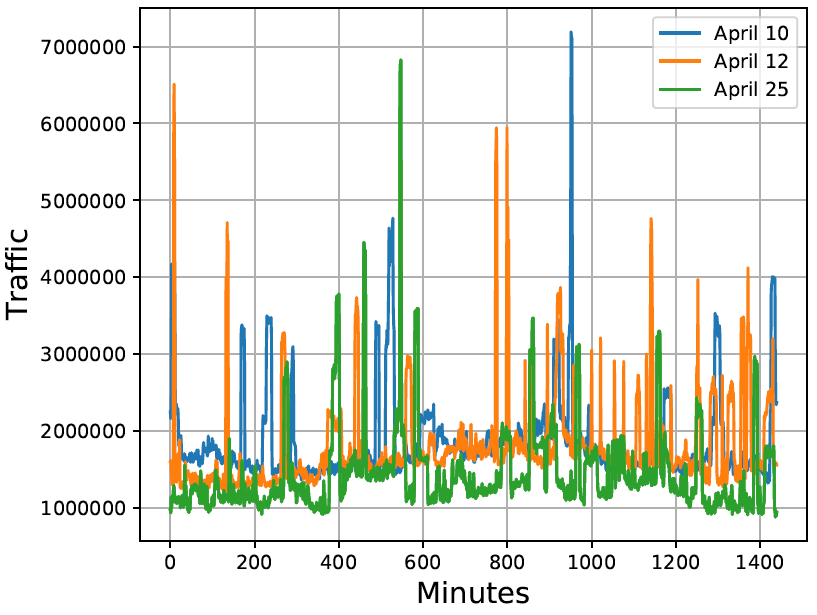


Figure 1: Overall Traffic

The amount of traffic on each of the days seems to exhibit differing amounts of time dependence. On April 10th and 25th, the trends appear to come in blocks of about 6 hours with a clear rise in traffic as the workday begins, a decline after lunch, and a spike in the evening. However, on April 14th, there is a longer period of increased traffic with more frequent large spikes in the evening than either of the weekdays. Below are the graphs for TCP and UDP.

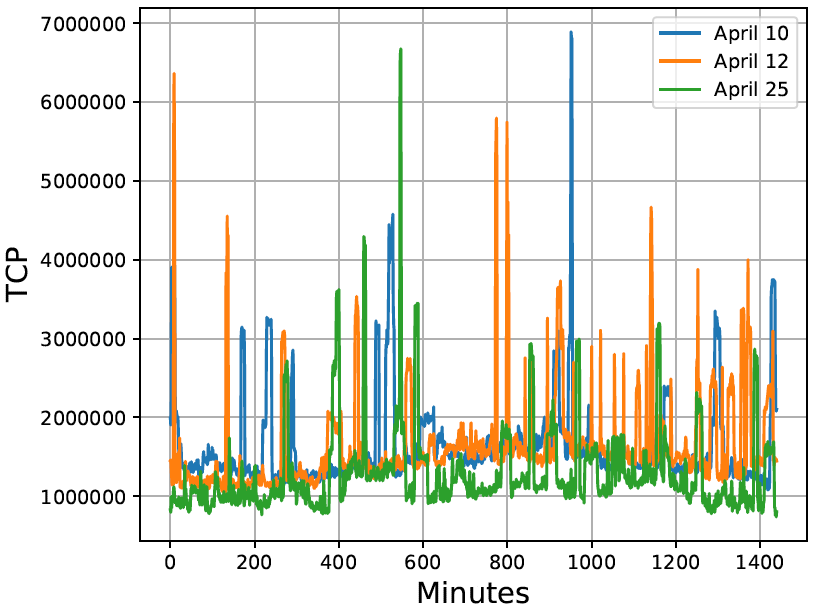


Figure 2: TCP Traffic

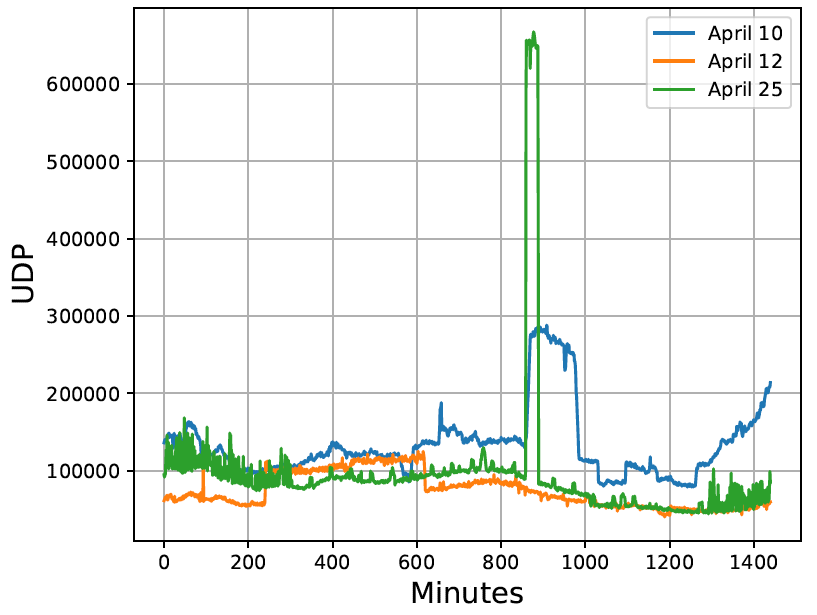


Figure 3: UDP Traffic

The TCP traffic is very similar to the overall traffic; this is unsurprising, however, since this makes up the vast majority of internet traffic. On the other hand, UDP is has a very different pattern. The peaks of activity and the trends in traffic are longer lasting. Interestingly, even though the amount of time people spent using UDP was longer on Saturday April 14, 2012, the total amount of traffic was lower than either of the weekdays.

To gain a better understanding of the time dependence, I calculate the tau autocorrelation for the overall traffic, TCP, UDP, and sources. Below are the corresponding plots for the tau auto correlation.

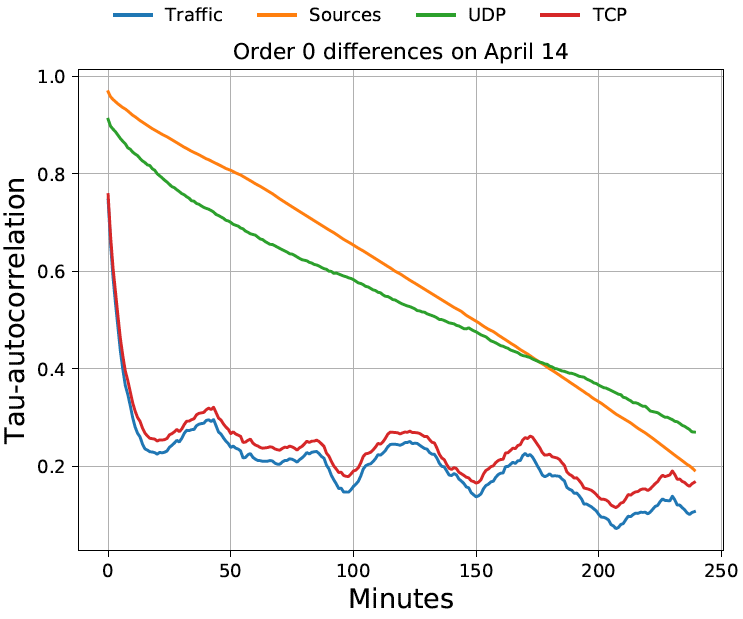
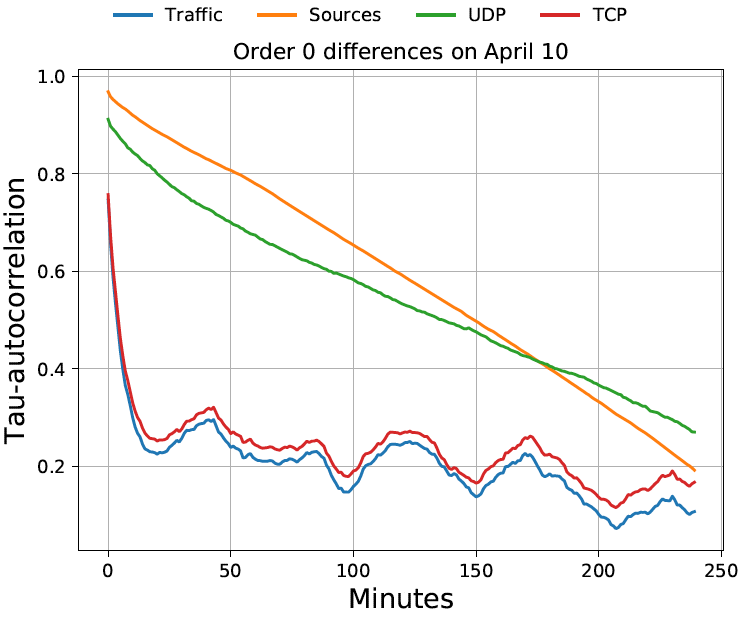


Figure 4: Tau Autocorrelation, No Differencing

Based on Figure 4, the overall traffic and sources have a long-range time dependence that decreases from highly autocorrelated to somewhat correlated for all of the day. Whereas, the UDP and the TCP traffic have a very high short-range time dependence but only a small amount of dependence after about an hour. This suggests that people may focus on something for about an hour, but change activities after.

Finally, I want to understand the differences in port usage between April 10th, 14th, and 25th because one might expect that malicious traffic would increase after the security patch release. Potentially malicious traffic would likely concentrate around a few ports that have been discovered to be vulnerable. Therefore, one would expect that the dispersion across ports, or the entropy, to be smaller during high periods of malicious activity. The normal range of activity has an entropy value roughly between 2 and 9. Below I graph the destination port entropy for the different days.

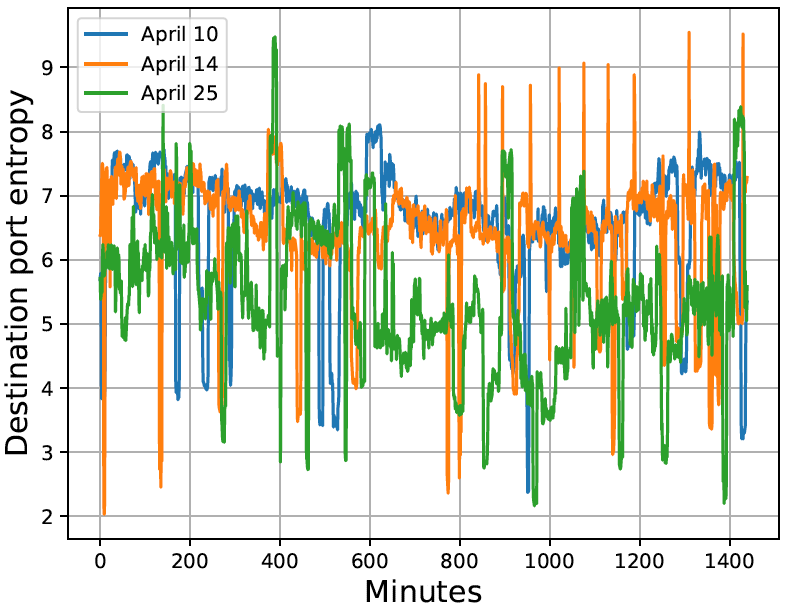


Figure 5: Destination Port Entropy

The entropy is similar between April 10th and April 14th, but, the entropy on April 25th is much more volatile. Additionally, the overall entropy on April 25th is around 6 which is lower than the overall entropy of around 7 on April 10th and April 14th. This suggests that there was slightly more concentration between a few ports on April 25th compared to April 10th and 14th, which is the opposite of what I expected. However, none of the graphs indicate a particularly low concentration of traffic, indicating potentially malicious activity.

Overall the days are fairly similar. None of the days seem to indicate large changes in potentially malicious activity. The amount and type of traffic was similar across the days. More indicatively, the estimated entropy was not strikingly different across the days.