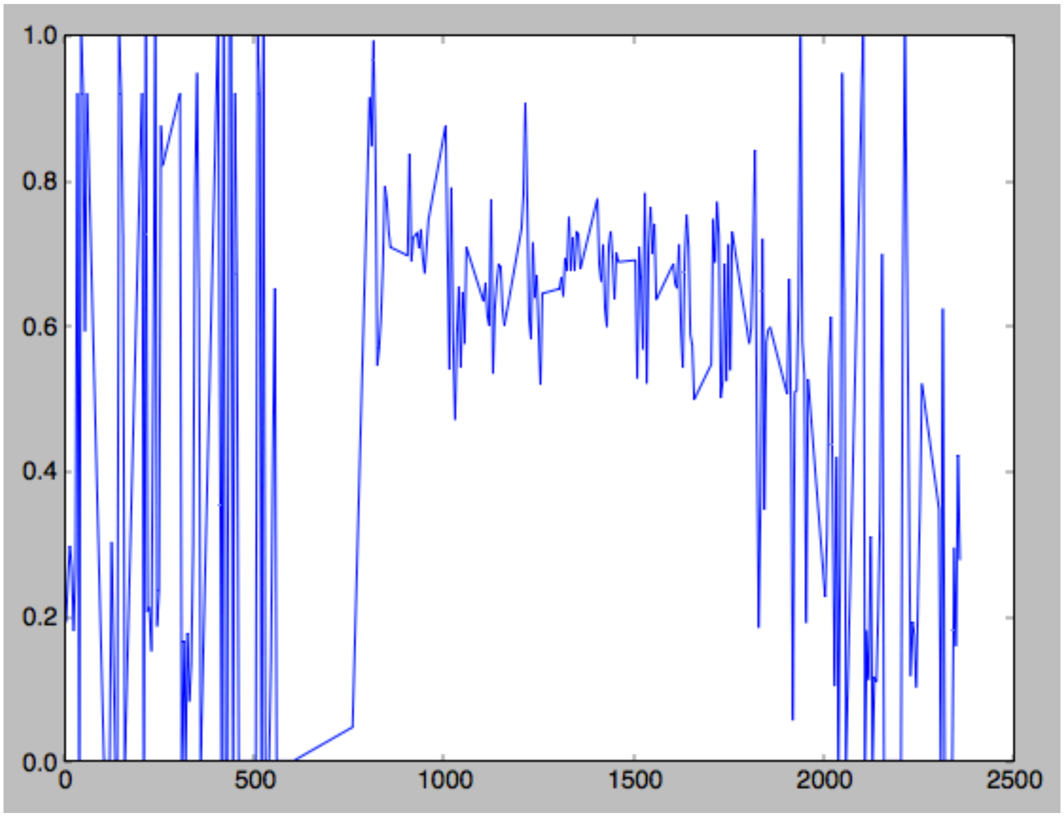
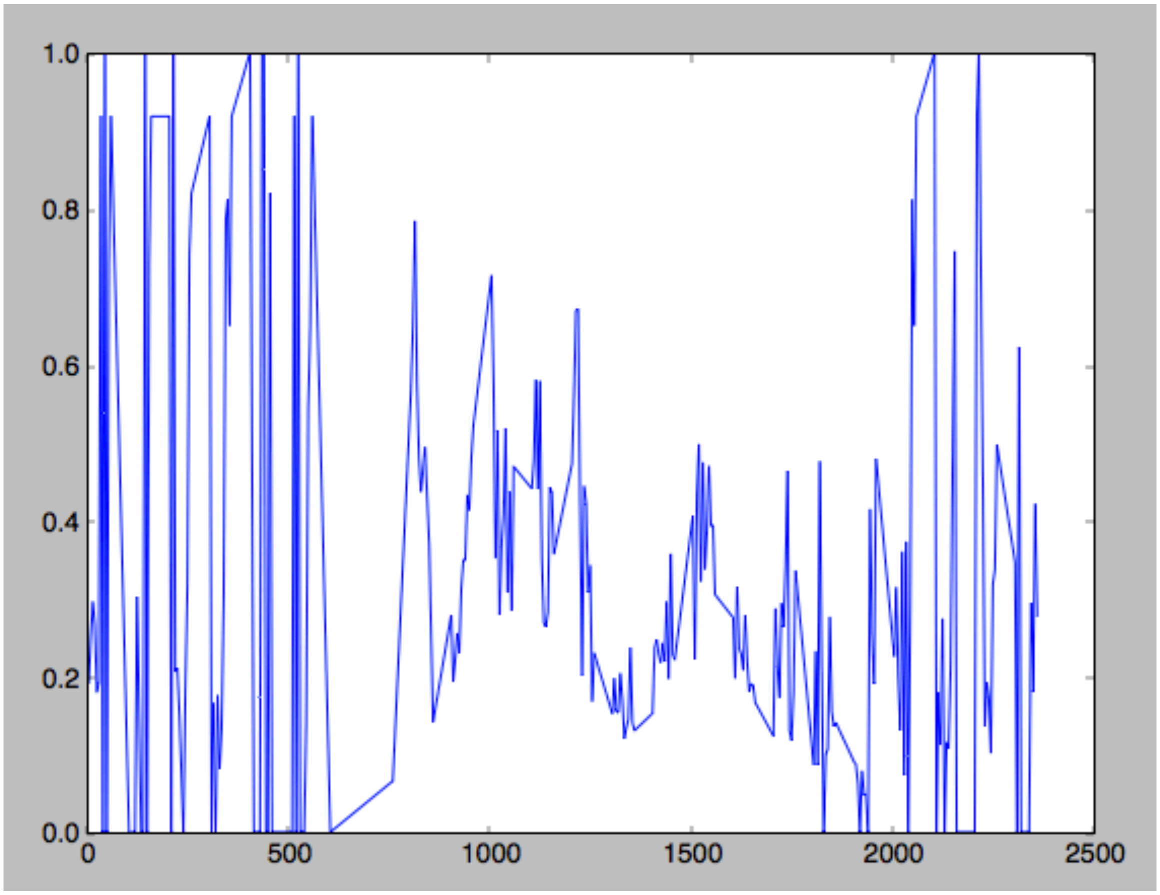
Attacks:
1) 8:12:16 (rows 2100 to 2623 in the log)
This was clearly some sort of DDoS attack. Over 523 requests were sent within 2s from a variety of source IP addresses that all used the same suspicious ecr/i:rxx service.

2) 11:55:15 (rows 14143 to 14157 in the log)
This was clearly some sort of malicious probing attack. The accesses were all in quick succession (within 1s) of each other. In addition, the source IP address 001.002.003.004 was clearly fabricated.

Analysis of Merits of Entropy:
In general, entropy was not a good indicator for detecting these attacks. Firstly, entropy (roughly) gives a measure for how many distinct values there are for a random variable in a given time interval, which is great for detecting DoS attacks (entropy of source IPs would decrease as many would be from the same source). However, it is not useful for detecting anomalous accesses. For example, for our two attacks, it would be more useful to detect that ecr/i:rxx is unusual for our web server and that 001.002.003.004 is unlikely to be a real IP address. In the first case, entropy of the source IP addresses did not help (see first figure) because the attack was distributed across multiple machines. In the second case, the attack was only over a short interval so it was not very impactful on the overall entropy measurement. The following figures display the entropy over time (5 minute intervals) for source IP address and serv., respectively. Note that time 1000 corresponds to 10:00 and so forth. Also, entropy was normalized so that it always was between 0 (all the same) and 1 (all distinct).

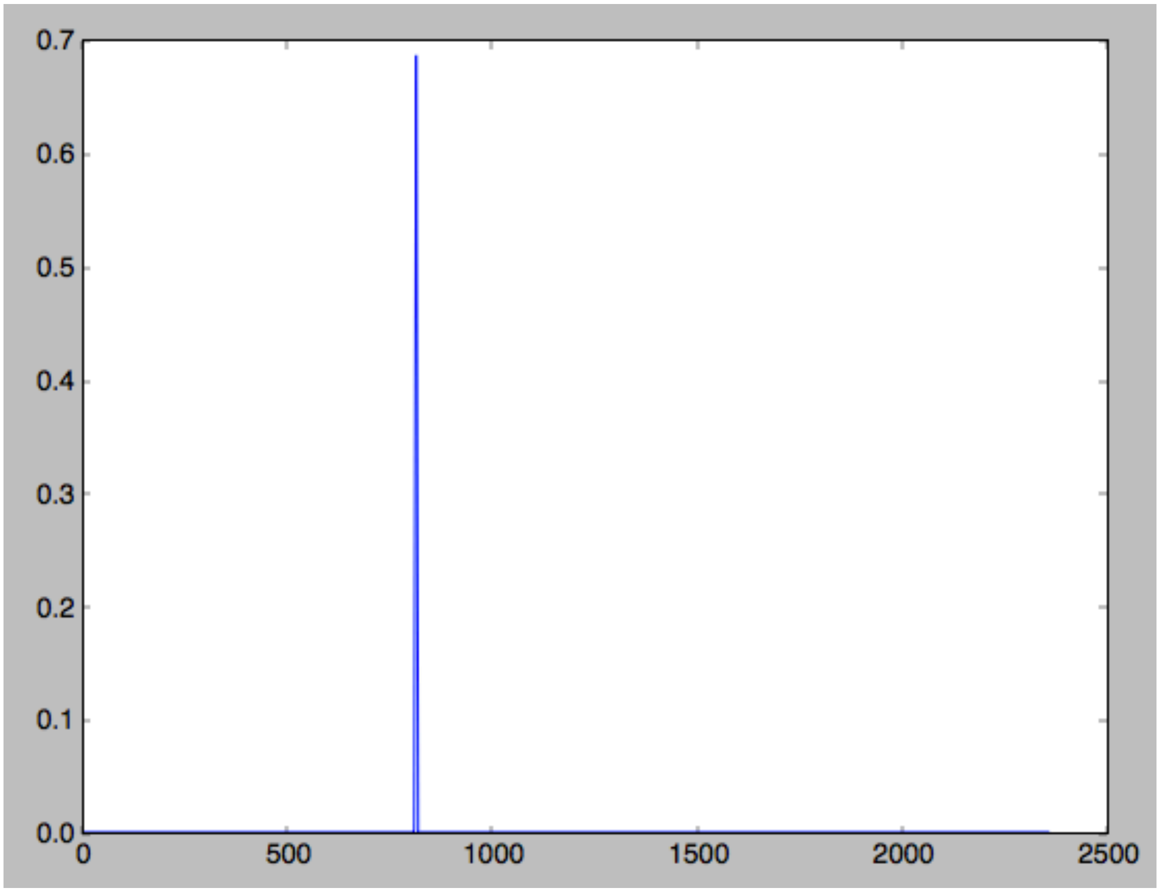




As you can see, there is too much noise in the plots to discern anomalies. In particular, the data is especially noisy before 8pm and after 6pm because server accesses are sparser in those regions.

However, it should be noted that there was one entropy measurement that proved to be extremely successful. By analyzing the entropy of the access durations, we were able to pinpoint the exact moment that the first attack happened. This is because the requests in this attack lasted over 30 seconds whereas everything else only lasted 1. The following plot shows this.

In conclusion, we found that entropy can only be used for anomaly detection for certain types of attacks, and in general, it is better to not have too much sparsity in the data, as this generates a lot of noise.


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
500
1000
1500
2000
2500