Week 8 writeup

Lecture 1

We started out with a brief history of some of the classics, starting with phishing. We were showed some examples and the humorous “Taliban trojan”. Botnets have also climbed in popularity but have recently been stiff-armed pretty hard by security experts.

Next were important tools for deriving value from messaging data: DIG, GREP/SED/AWK, and Regex are all classics that I’ve had personal experience with but he listed some other good ones too. The most important aspects of these tools is that they allow for parsing, grouping, aggregation, and outlier identification which are all key to figuring out what is spam and what is normal. Next we entered lab1 which delved into putting these ideas into practice.

Lecture 2

We went into the “data scientific method” which explains in steps:

1. Start with data
2. Develop intuitions about the data and the questions it can answer
3. Formulate your question
4. Leverage your current data to better understand if it is the right question to ask. If not, iterate until you have a testable hypothesis
5. Create a framework where you can run tests/experiments
6. Analyze the results to draw insights about the question.

Next was the classification lab which was pretty much the same as what we already did on Tuesday. Very fun assignment by the way!