6397 Homework 2

Host Intrusion Detection

This homework is based on the paper “A sense of self for Unix processes” and related work on n-grams. The purpose is to make sure that you read and understood the material. Keep in mind that the paper was written in 1996 and there have been a lot of advances in the area.

One issue that we discussed in class is the “precision” of the definition of self in the example (tables on page 122 of the paper or p. 40 of the slides). Please justify this concern. Find out all the extra 4-grams that was not part of the original training data. Give a quantitative measure on how serious this problem is.

You can also provide additional criticism or comments on the paper or n-gram approach. Your report should be written in word format if possible. There is no programming required but you can certainly include some analysis using available tools.