Realistic Website Fingerprinting Attacks

The purpose of our study is to determine if we can create a method that will be able to find when another website begins, hence the overlap (see Proposed Algorithms section for more information). The graphs on the right show how our algorithms performed on the data. Occurrence is a numeric value that is used to describe which packet sizes to dismiss from our first positive packet. We do not want packets that are very general to be considered for tests because common packet sizes will be general and not tell us about when a website is visited. We recorded the percent correct, which is measured by how often the first positive packet was found ±1 second from the desired time. This was to show graphically where there would be a good cutoff, meaning where we achieve the highest percent correct. This translates to finding a peak where the most common and uninformative packet sizes are and should be removed. As shown in figure 1, our percent correct was not very meaningful and increased as the more packet sizes were removed. We continued by calculating the average time, which was measured by the time from the start of the first unique positive packet was found. This result was recorded from every instance of the datasets so we could be as general as possible. In summary, the average time starts high and sharply declines before leveling off and passes through the time of the overlap. From the graph, there is no anomaly near the time we want and there is not much meaning to be gathered from it. Looking forward we would want to implement an approach like this using a machine learning model, and try to see if we can detect which website is overlapped.