ABSTRACT

In fact, large-sized data is inaccurate, incomplete and inconsistent. A bad quality data is not going to produce the results of a process that is not qualified. Why data is inaccurate, incomplete and inconsistent they are from human error and computer error during data entry. Besides that cause data to complete them are not consistent in naming rules and inconsistent in a format for charging. With the inconsistent data will make the data relevant to the possibility of unrecorded and difficult to understand. If the recorded data is inconsistent then the data will be automatically deleted.

In this final project, do the processing of raw data to the preprocessing stage using data reduction techniques that sampling and dimensionality reduction. The goal is to reduce the complexity of the data examined and the results of preprocessing the data obtained can be classified based on the algorithm needs vitality.Pada sampling process large amounts of data to be processed into new data randomly from the existing sample data. While in the process further dimensionality reduction, data that have a high dimensionality will be reduced to lower dimensionality that will get the output in the form of new features.

Data that will be examined in the form of raw data stream results conducted by NS-3. Data streaming is done by the NS-3 consists of normal attacks and anomalies.Data will be processed to the preprocessing stage, so it will gain new relevance traffic features. The results of this research obtained the complexity of each algorithm. With the results of such complexity, the complexity of a better scenario 1 with scenario 2. With the merger between Sampling + PCA, the obtained value of big-O notation O (n, p). N is the number of sampling and analysis of data p as the number of columns from PCA analysis.

Keywords: DDoS,Sampling,Dimensionality Reduction,Time Complexity,NS-3