ABSTRACT

In the development of the Internet network technology is now widely discusses phenomena atapun attack threat to a computer or server. Lots of various types of threats on a computer in an Internet network such as DoS (Denial of Service), DDoS (Distributed Denial of Service), a flash crowd, and so on. Therefore, to facilitate the retrieval of information in order to conform with the desire, the need for grouping in the traffic anomalies to identify the types of new attacks. In grouping the traffic anomalies, in this final project a clustering algorithm is an algorithm CURE who have achievements to handle large amounts of data, and also works by measuring the distance between a representative of the traffic with a list of points that have been previously selected cluster. However, the attention in the analysis process of grouping these anomalies is the problem of labeling and validation of each object on the results of the clustering process.

These problems needs a strategy in special labeling technique and a validation to analyze the results of the information obtained to fit the desires, needs to grouping in traffic anomalies. By validating the cluster we will get the optimal number of clusters in the analysis of traffic anomalies in this case is the method of clustering CURE (Clustering using Representatives). Results of the validation will explain how the quality of the cluster and each object using silhouette technique index. The main objective in the application of this validation is a modification of the algorithm CURE with the main focus of the issue of labeling each object in each cluster and also the validation of the results of CURE clustering algorithm.

Results from this research, the algorithm can detect anomalous traffic CURE well and get the best value validation using silhouette technique. From the analysis of the results of clustering algorithms CURE CURE algorithm validation values obtained using the technique Dataset KDDCUP'99 silhouette on the average values obtained the highest silhouette with accuracy 97.96%, and the average value of 0.7642748 silhouette cluster. At Darpa Dataset Week 5 Friday with a value of 98.56% accuracy, and the average value 0.763525532 silhouette cluster.

Keywords: traffic anomaly , clustering, cluster validation, CURE, Silhouette Coefficient