# Findings:

* The use of information gain rate in this paper evaluates feature importance, enabling dual dimension reduction for enhanced classification efficiency and essential feature impact.
* Greater distance between points increases reverse influence, requiring consistent sample behavior for accurate classification; this method contrasts with traditional averages by emphasizing holistic decision-making.
* The GR-AD-KNN Algorithm has better detection performance than the traditional TAD-KNN algorithm based on the average Euclidean Distance.
* This paper used the offset increment average distance concept, improvements are made to the measured point, acknowledging varied effects of distant samples on decisions, enhancing algorithm stability, and alleviating "small group classification disadvantages."
* By emphasizing holistic decision-making and utilizing offset increment average distance, this approach improves classification accuracy, algorithm stability, and addresses challenges with distant sample effects and small group classification.

My Idea:

1. No information about detection accuracy was provided, and it was evaluated using an IPv4 KDD Cup 99 dataset even though IPv6 is totally different from IPv4.
2. DL based techniques can be used.
3. A single ML based technique has become insufficient.
4. Multiple Dataset could be used.

“Dual Dimensionality Reduction”

The information gain rate evaluates feature importance, enabling "double dimension reduction" for optimized Euclidean distance calculations and enhanced impact of essential features in classification; the dual dimensionality reduction method reduces dimensions and boosts efficiency in later classification algorithms.

In 2009. M. Yang and T. Li proposed a feature selection framework, using support vector machine (SVM) and particle swarm optimization (PSO), determined the best detection feature, and realized the effect of DoS attack detection on IPv6 networks.

Features are selected, and discrete types of sub features are filtered and aggregated. At the same time, the weight of different features is reflected through the information gain rate, which is used in DoS attack detection to improve the adaptability and detection efficiency of the detection system [20, 21].

The information gain rate is used as the evaluation index of feature importance.

Based on the information gain rate, the function of “double dimension reduction” is realized for the feature.

The information gain rate is used to optimize the Euclidean distance calculation of the algorithm ways to improve the influence of essential features and their sub features in classification decision-making.

Compared with the traditional ordinary average distance algorithm, the above decision method adds each type of sample point and set the idea of “holistic decision-making.”

The experiment is mainly divided into two parts. The first part is to realize the double dimensionality reduction of features and calculate the information gain rate of the first level and second-level features. When calculating the information gain rate of the first-level feature to achieve the dimensionality reduction function, this paper adopts the continuous type.

In the classification algorithm, in terms of experimental control settings, this article will only compare the GR-KNN algorithm with the weighted optimization of the Euclidean distance and the GR-AD-KNN algorithm.

Finally, to test the overall improvement effect of selecting sample points on the overall improvement of the experimental classification results by comparing the traditional TAD-KNN based on average distance decision making.

The classification capabilities of the Traditional Average Distance-KNN algorithm and the GR-AD-KNN algorithm are used to count the detection results of attack types with weak detection capabilities to evaluate the improvement and stability of the classification performance of the algorithm.

It is recognized that the effect of different long and short distance sample points on decision-making is different, and the algorithm’s stability is improved.

The problem of “small group classification disadvantages” is alleviated.