Explaining the Role of t-SNE and Brute-Force Optimization in Attack Classification

1 Introduction

The CIC IoT 2023 dataset presents a multi-class classification challenge with significant overlap among attack types. Initial explorations using t-SNE for dimensionality reduction and visualization indicated clusters of separable classes. However, model performance on subsets derived from t-SNE clusters was suboptimal. Consequently, a brute-force approach was employed to evaluate all possible attack type combinations, identifying subsets that maximized classification performance metrics. This paper aims to mathematically explain this scenario.

2 t-SNE as a Visualization Tool

2.1 Mathematical Basis of t-SNE

t-SNE is a nonlinear dimensionality reduction technique that minimizes the Kullback-Leibler (KL) divergence between pairwise similarities in high-dimensional and low-dimensional spaces:

$$C = \sum_{i \neq j} P_{ij} \log \frac{P_{ij}}{Q_{ij}}, \tag{1}$$

where P_{ij} and Q_{ij} represent the probabilities of similarity in the high-dimensional and low-dimensional spaces, respectively. P_{ij} is computed using a Gaussian kernel in high-dimensional space, while Q_{ij} uses a Student t-distribution in 2D space.

2.2 t-SNE Limitations

While t-SNE is effective in revealing local structure and potential clusters, it distorts global relationships and class densities. This can lead to misleading interpretations of separability, particularly for overlapping or poorly defined classes. For example, clusters appearing distinct in 2D may not exhibit the same separability in high-dimensional space.

2.3 Relevance to Attack Classification

t-SNE provided preliminary insights into potential separable groups among attack types. These visualizations informed initial feature subset selection, narrowing the search space for combinatorial exploration.

3 Brute-Force Subset Optimization

3.1 Problem Definition

Let $A = \{A_1, A_2, \ldots, A_n\}$ represent the set of attack types. The goal is to identify the subset $S^* \subseteq A$ that maximizes a performance metric $f(S)$ (e.g., accuracy, F1 score):

$$S^* = \arg\max_{S \subseteq A} f(S). \tag{2}$$

Given n attack types, the number of possible subsets is $2^n - 1$, excluding the empty set.

3.2 Algorithmic Approach

The brute-force method evaluates all subsets S for model performance, identifying the subset S^* with the highest metric. This exhaustive exploration compensates for t-SNE's limitations, ensuring that all potential combinations are considered.

3.3 Mathematical Insights

- **High-Dimensional Separability:** Subsets with better performance exhibit lower pairwise overlap in feature space, measurable by metrics like silhouette score or Davies-Bouldin index.
- **Information Gain:** The incremental information contributed by subset S can be quantified as:
$$IG(S) = H(C) - H(C|S), \tag{3}$$
where $H(C)$ is the entropy of class labels, and $H(C|S)$ is the entropy given subset S.

4 Experimental Results

4.1 t-SNE Visualization

4.2 Performance Comparison

Table 1 summarizes the results of the brute-force optimization for different subsets of attack types. These results demonstrate the superiority of exhaustive exploration over t-SNE-informed subsets.

Subset	Accuracy	Recall	F1 Score	CPU Time (s)
DDoS+Recon	99.3	99.21	99.3	6.08
DDoS+Web+Mirai	99.41	99.41	99.41	9.56
DoS+Recon	99.62	99.83	99.62	6.81

Table 1: Performance metrics for selected subsets of attack types.

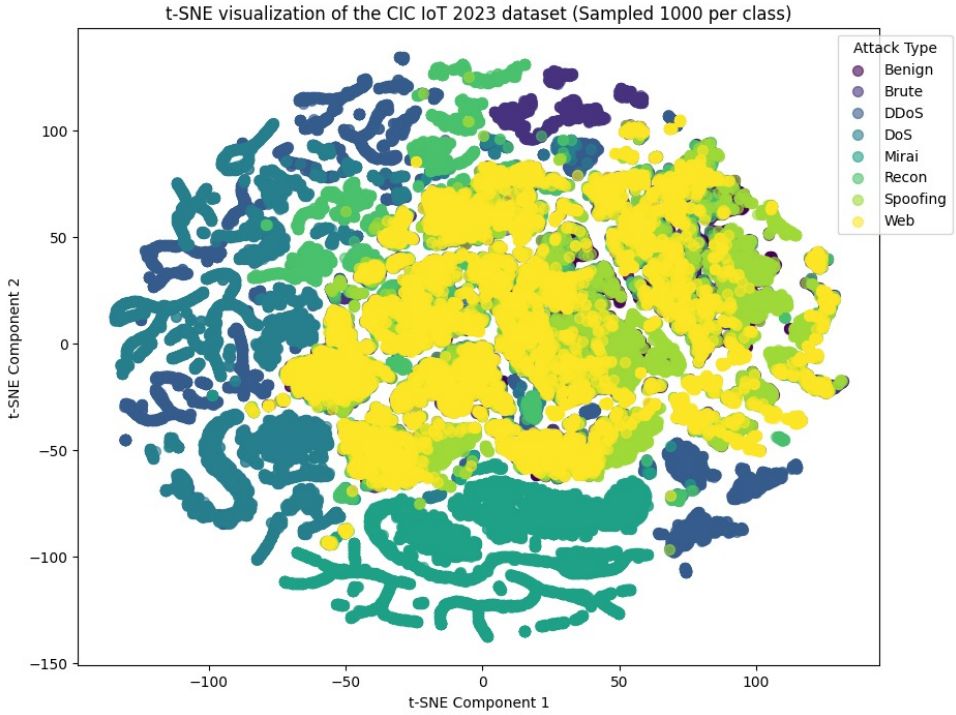

Figure 1: t-SNE visualization of the CIC IoT 2023 dataset. Each point represents a sample, color-coded by attack type.

5 Conclusion

This study demonstrates the complementary roles of t-SNE and brute-force optimization in attack classification. While t-SNE aids in visualizing potential class separability, brute-force ensures rigorous exploration of feature subsets, leading to optimal performance. Future work could explore hybrid approaches combining visualization and algorithmic optimization for enhanced efficiency.