Paper Summary
Class-of-Service Mapping for QoS

Sarthak
2020CS10379

May 2024

Problem Statement

The paper addresses the challenge of mapping traffic from various applications to different Quality of Service (QoS) classes using k-Nearest Neighbors(k-NN) and Linear Discriminant Analysis(LDA) algorithms.

Motivation

- Different QoS requirements & relative importance of applications, e.g. Transactional applications having higher importance
- Traditional port-based application classification being unreliable due to applications using dynamic or non-standard ports and the use of port 80 for non-web applications
- Payload inspection for classification being computationally expensive and impractical for high-speed links, especially with encrypted traffic

Key Ideas

The paper proposes a statistical signature-based framework for traffic classification. Instead of relying on port numbers, the method uses characteristics of traffic flows to form signatures that identify how applications use the network. These signatures are insensitive to application-layer protocols, allowing classification based on usage patterns like interactive use or bulk data transfer.

Framework

Statistics Collection

This involves placing monitors in the network and collecting appropriate statistics of the traffic from certain aggregates. Then, a vector of statistics ($\boldsymbol{S}^C(i)$) for each connection is formed and used to update the statistics of each aggregate that connection is involved in. Ideally, statistics should be chosen that can be updated on-line in a streaming fashion. Implying no need to store data per packet, but rather per connection.

Classification

Two common ML Algorithms were used for classification :

- **k-NN** : Take $\hat{G}(X)$ to be the class G_i of the data point X_i which minimizes the distance $||X_i - X||$. The k nearest neighbors essentially 'vote' on the class of the observation), to enhance its robustness. Good on low dimensional data.
- **LDA** : Choose the class with maximal conditional probability, given the feature vector X. For any data point X we then choose the class with the largest $\delta_g(X) = Pr(g|x = X)$. Decision boundaries formed are linear. Can be generalized to QDA.

Rule Creation

Post classification, a set of rules was created for each mapping $A_i \longrightarrow C_j \longrightarrow Q_k$. Note that the error rate of the classification algorithm is in forming these rules, not in classification of packets in the actual QoS implementation.

Contributions

The results show the method's feasibility and potential for real-world application, including its ability to classify even fine-grained traffic classes indicating effectiveness of Machine Learning methods to solve learning problems in Networks.

Strengths

- Collected statistics could be used for other planning problems
- Provide substantial cost savings for network operators
- Better than unreliable traditional port-based classifiers, computationally expensive payload inspectors

Weaknesses

- Care must be taken in choice of algorithm used for classification, e.g. LDA being error prone (57%) and 3,5-NN having least error (14%)
- Requires manually choosing the aggregates and statistical features