

Project Title

Recognizing Applications in Traffic Flows Using Machine Learning

Project Participants:

First Name	Last Name	CNet ID	Project Role
Sam	Xie	muyanxiesam	Co-lead
Jillian	Chen	jianghui	Co-lead

Project Description

As network traffic becomes increasingly encrypted, traditional payload-based methods for application identification are losing effectiveness. This poses a key challenge for network management, security monitoring, and quality of service enforcement. Our project addresses this problem by applying machine learning techniques to classify applications using flow-level characteristics such as packet size, timing, and duration. Positioned at the intersection of networking and ML, this work builds on prior studies in encrypted traffic classification. Through this project, we aim to gain insights into the effectiveness of different machine learning approaches for application identification in encrypted traffic. We will explore supervised learning algorithms such as Random Forests, Support Vector Machines, and Neural Networks, and evaluate their performance using metrics including accuracy, precision, recall, and F1-score. Cross-validation will be used to ensure robustness, and the models will be tested on unseen data to assess generalization. Ultimately, our goal is to understand which features most strongly contribute to accurate classification and the challenges encryption poses for traffic analysis.

Data

The project will utilize a dataset of network traffic captures in PCAPNG format, specifically focusing on encrypted traffic. The dataset will be processed to extract relevant features for machine learning, including flow statistics and packet metadata. A link to the dataset is here: [Dataset Link](#).

Deliverables

We will submit a clean, fully executable Jupyter notebook with all analysis, model training, and evaluation code, verified using “restart kernel and run all.” All necessary data files will be included or clearly referenced. A concise Sphinx-formatted report will accompany the notebook, summarizing our methods, results, and key insights, with inline code for clarity and professional presentation.