LAB OF COMPUTER COMMUNICATION & NETWORKING
Networking Education, Research and Innovation
Computer Science Department, Technion - Israel Institute of Technology

Deep Learning based QUIC Traffic Classification

Abstract:

Traffic classification, the categorization of network traffic into appropriate classes, is important to many applications, such as quality of service (QoS) control, pricing, resource usage planning, malware detection, and intrusion detection. Because of its importance, many different approaches have been developed over years to accommodate the diverse and changing needs of different application scenarios. In particular, the growing trends of Internet traffic encryption and an increase Virtual Private Networks (VPNs) and The Onion Router (ToR) usage, raise additional challenges to network traffic classification.

Traffic classification techniques have evolved significantly over time. The first and easiest approach is to use port numbers. However, its accuracy has been decreasing because newer applications either use well-known port numbers to disguise their traffic or do not use standard registered port numbers. Despite its inaccuracy, the port number is still widely used either alone or in tandem with other features in practice. The next generation of traffic classifiers, relying on payload or Deep Packet Inspection (DPI), focuses on finding patterns or keywords in data packets. These methods are only applicable to unencrypted traffic and has high computational overhead.

As a result, a new generation of methods, based on flow-statistics, emerged. These methods rely on statistical or time series features, which enable them to handle both encrypted and unencrypted traffic. These methods usually employ classical Machine Learning (ML) algorithms, such as random forest (RF) and k-nearest neighbor (KNN). However, their performance heavily depends on the **human** engineered features, which limit their generalizability.

Deep Learning (DL) obviates the need to select features by a domain expert because it automatically selects features through training. This characteristic makes DL a highly desirable approach for traffic classification, especially when new classes constantly emerge and patterns of old classes evolve. Another important characteristic of DL is that it has a considerably higher capacity of learning in comparison to traditional ML methods, and thus can learn highly complicated patterns. Combining these two characteristics, as an end-to-end approach, DL is capable of learning the non-linear relationship between raw input and corresponding output without the need to break the problem into the small sub-problems of feature selection and classification. To achieve this goal, DL requires sufficient labeled data and adequate computation power.

In this proposed project, we will deploy a framework for traffic classification task, including data collection and cleaning, feature selection, and model selection. We will use an innovative approach: The Internet traffic flows will be transformed into FlowPic images and from this point, we will take advantage of current advances in the field of image recognition using DL methods, and design a Convolutional Neural Network (CNN) architecture to classify the traffic.

Instractor:

Eran Tavor (tavran@cs.technion.ac.Il)

References:

• The Applications of Deep Learning on Traffic Identification
https://www.blackhat.com/docs/us-15/materials/us-15-Wang-The-Applications-Of-Deep-Learning-On-Traffic-Identification-wp.pdf

• FlowPic: Encrypted Internet Traffic Classification is as Easy as Image Recognition –
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8845315&tag=1

• Seq2Img: A sequence-to-image based approach towards IP traffic classification using convolutional neural networks -
https://ieeexplore.ieee.org/document/8258054