QUIC Upper-Layer Application Classification

Abstract

Traffic classification, the process of categorizing network traffic into various classes, is essential for various purposes such as quality of service control, pricing, resource planning, malware detection, intrusion detection and more. In recent years, Deep Learning (DL) has emerged as a state-of-the-art technique for traffic classification due to its ability to automatically select features through training and its high capacity for learning complex patterns. DL can learn the nonlinear relationship between raw input and corresponding output without the need for feature selection and classification, but it requires a large amount of labeled data and computational power.

QUIC is a transport protocol designed to improve the performance of encrypted HTTP traffic. It is a secure, highly encrypted, multiplexed protocol that encrypts both the payload and most of the header. Identifying the upper layer application that used the QUIC transport layer is useful for load balancing, Quality of Experience (QOE), Cyber attack mitigation etc.

The goal of this project is to train a model to classify QUIC connection to several classes of upper-layer applications.

Project overview

In this project you will:

a. Develop a dataset of tagged pcap frames:
    a. Establish a QUIC connection to several known servers using known applications (like YouTube, Facebook, Netflix, Whatsapp/Whatsapp WEB/…) using headless chrome
    b. Establish visibility into the encrypted packets using wireshark.
    c. Sync. The data from the headless chrome log file and Wiresharc and extract the following features: (Connection ID, Time of arrival, Size, Direction: up/down stream, upper-layer application, Number of streams in the packet, number of objects in the packet, Estimated RTT)

b. Develop and train a model to classify a QUIC stream connection to one of the upper-layer application using the first 4 features)

Prerequisites

1. Introduction to computer networks (236334).
2. Machine Learning / Deep learning background. (contact Barak Gahtan for approval).

Notes

- The above project's aim is to produce in the long run an academic paper. The project can also be a good way to start a research proposal for an MSc.

Instructors

Barak Gahtan (barakgahtan@cs.technion.ac.il)

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

References:

[1] 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

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

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