ISCX
Information Security
Centre *of* Excellence

Characterization of Encrypted and VPN Traffic Using Time-Related Features

Gerard D. Gil, Arash H. Lashkari, MSI Mamun, Ali A. Ghorbani
Information Security Center of Excellence (ISCX), UNB

UNB

Abstract

Traffic characterization is one of the major challenges in today's security industry. The continuous evolution and generation of new applications and services, together with the expansion of encrypted communications makes it a difficult task. Virtual Private Networks (VPNs) are an example of encrypted communication service that is becoming popular, as method for bypassing censorship as well as accessing services that are geographically locked. In this paper, we study **the effectiveness of flow-based time-related features to detect VPN traffic and to characterize encrypted traffic into different categories**, according to the type of traffic e.g., browsing, streaming, etc. We use two well-known machine learning techniques (C4.5 and KNN) to test the accuracy of our features. Our results show high accuracy and performance, confirming that time-related features are good for encrypted traffic characterization.

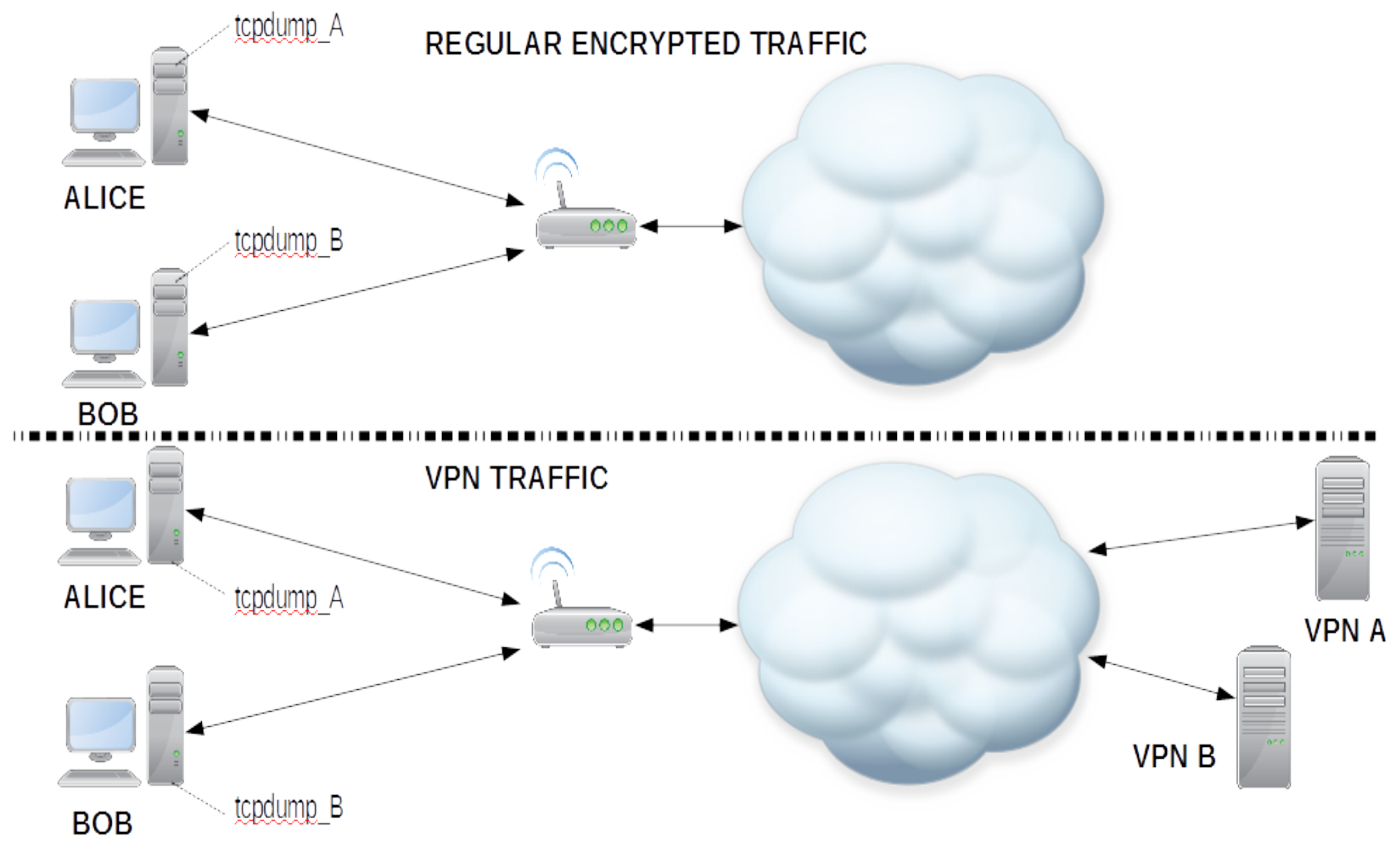
Time Based Features

FEATURE	DESCRIPTION
duration	Duration of the flow
fiat	Forward Inter Arrival Time (mean, std, max, min)
biat	Backward Inter Arrival Time (mean, std, max, min)
flowiat	Flow Inter Arrival Time (mean, std, max, min)
active	The amount of time a flow was active (mean, std, max, min).
idle	The amount of time a flow was idle (mean, std, max, min)
fb_psec	Flow Bytes per second
fp_psec	Flow Packets per second

- **Flow** set of packets sharing: {**source** IP, source port, destination IP, destination port , Protocol}

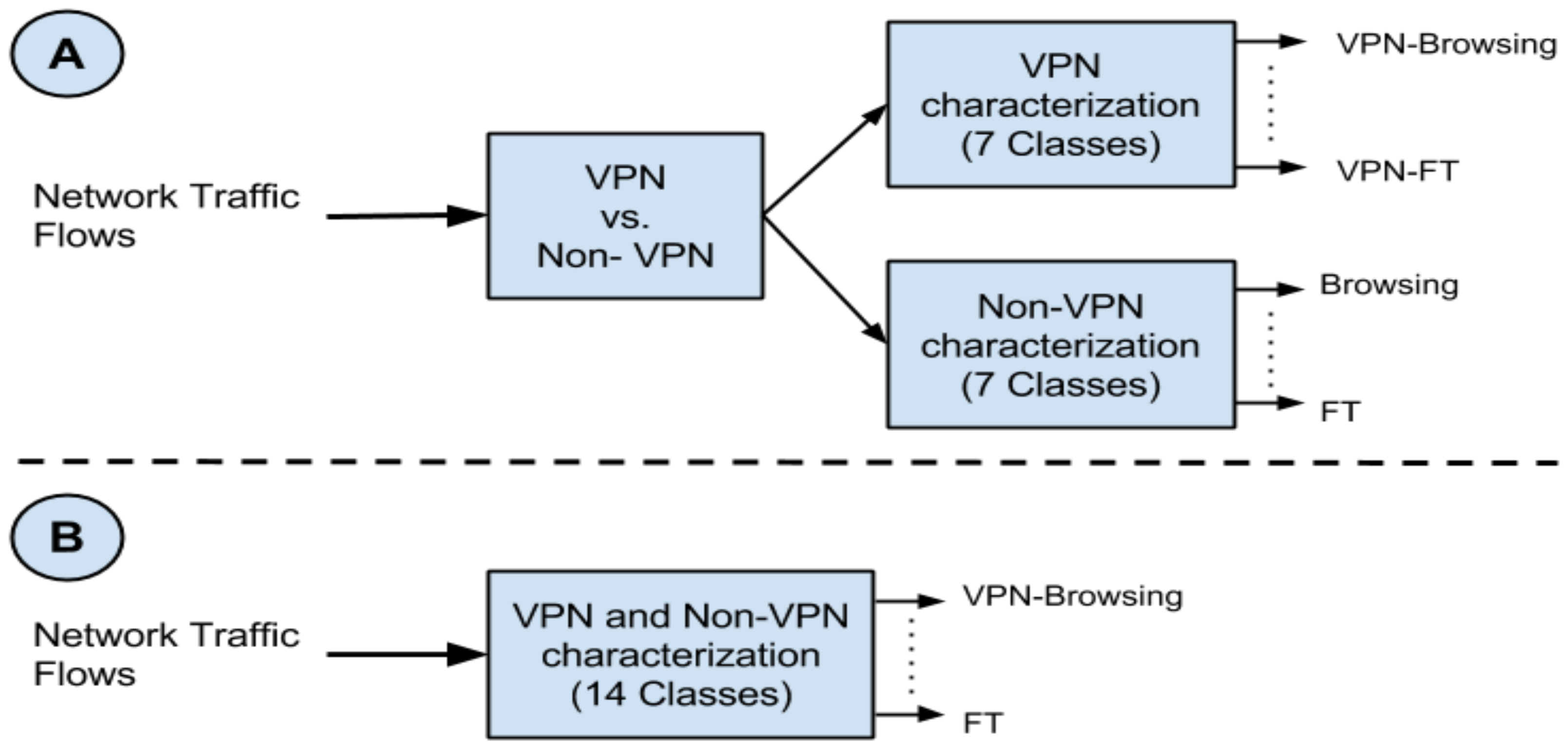
Dataset Creation:

TRAFFIC	APPLICATIONS
Web Browsing	Firefox and Chrome
Email	SMTPS, POP3 and IMAPS
Chat	ICQ, AIM, Skype, Facebook and Hangouts
Streaming	Vimeo and Youtube
File Transfer	Skype, FTPS and SFTP using Filezilla
VoIP	Facebook, Skype and Hangouts voice calls
P2P	uTorrent and Transmission (Bittorrent)

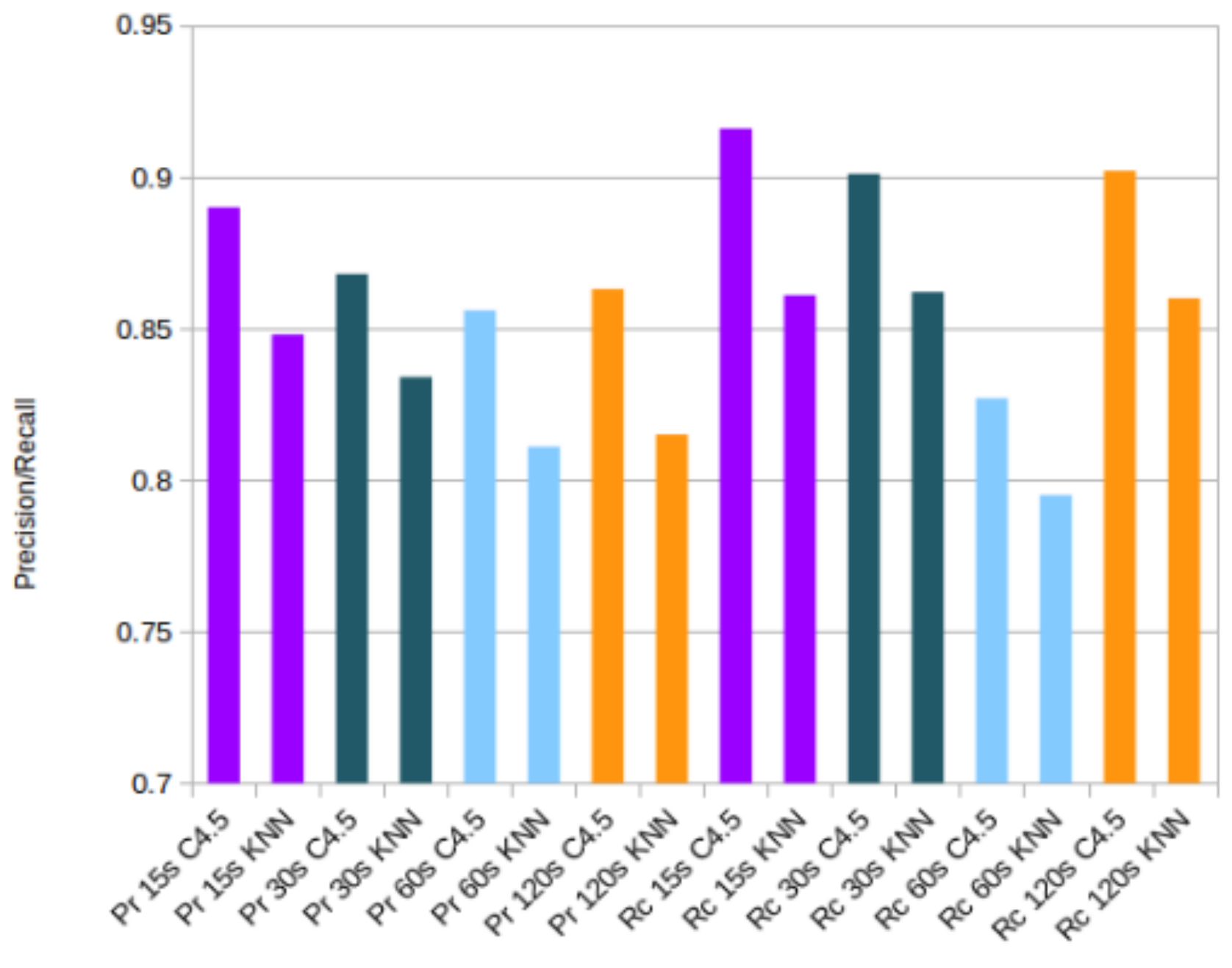


- We selected 4 different flow timeout values in this research: 15ms, 30ms, 60ms and 120ms

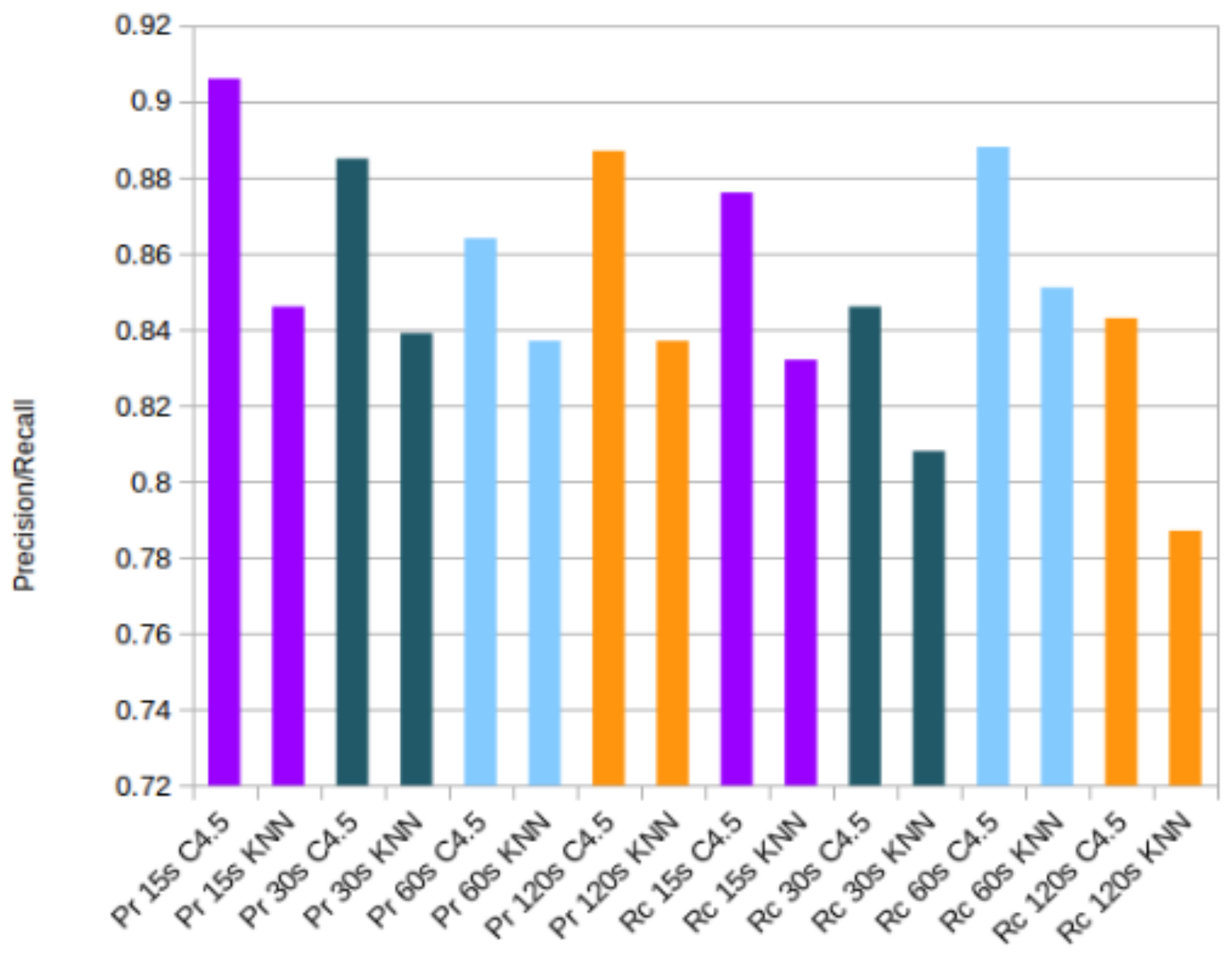
Experiment


A
Network Traffic Flows
VPN vs. Non- VPN
VPN characterization (7 Classes)
VPN-Browsing
VPN-FT
Non-VPN characterization (7 Classes)
Browsing
FT
B
Network Traffic Flows
VPN and Non-VPN characterization (14 Classes)
VPN-Browsing
FT

Scenario A VPN vs. NON-VPN

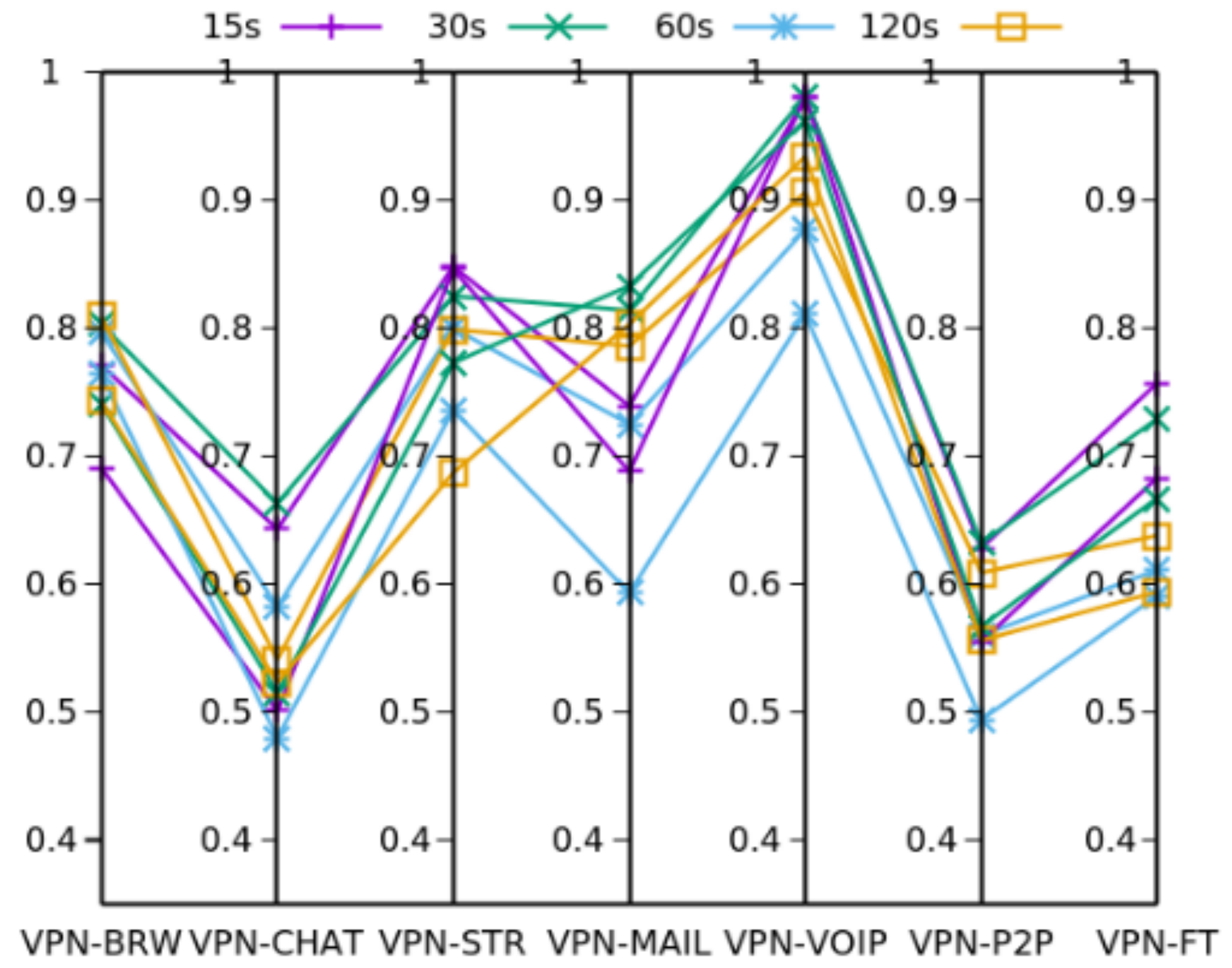


Precision and Recall VPN class

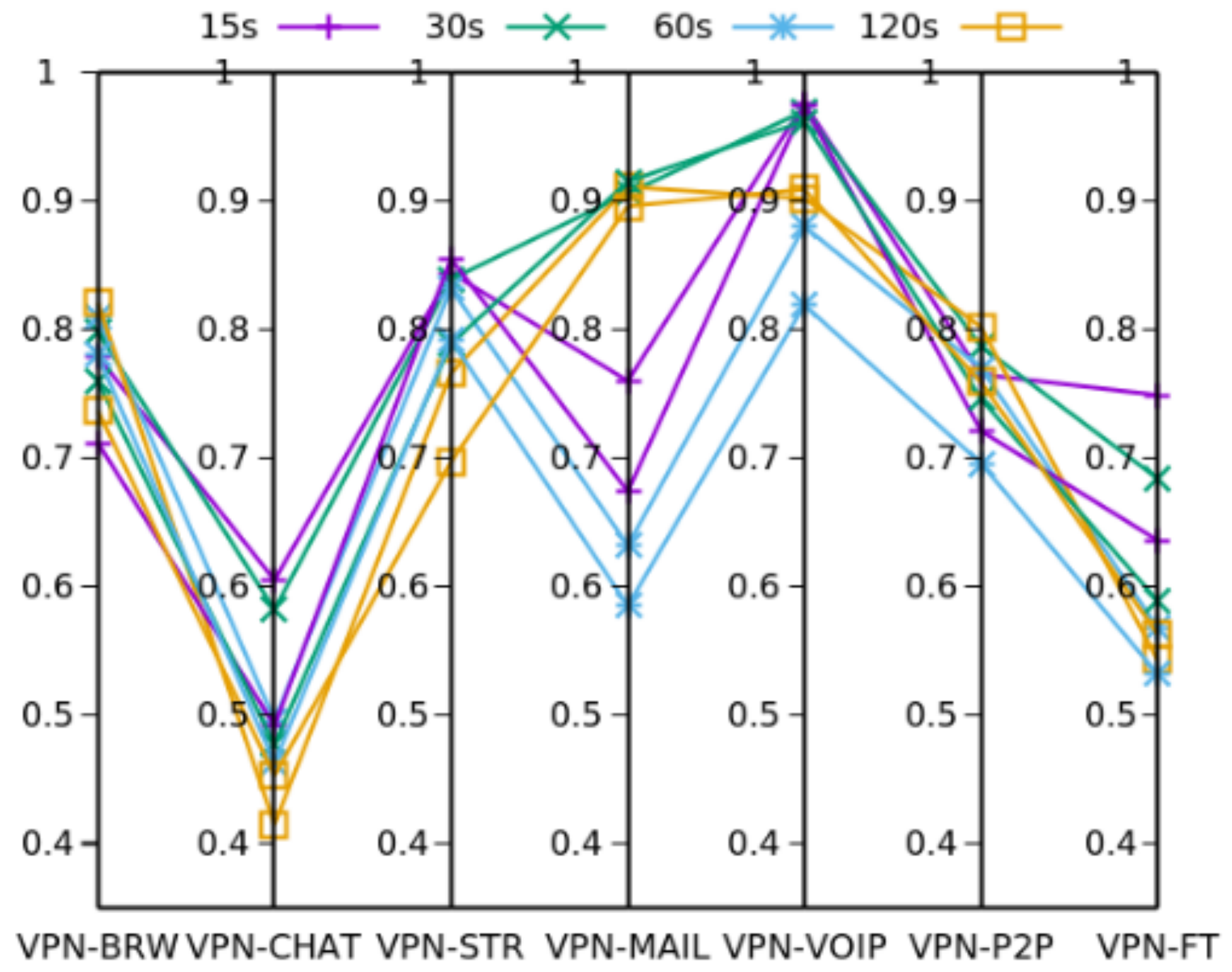


Precision and Recall NON-VPN class

Scenario B: VPN Characterization

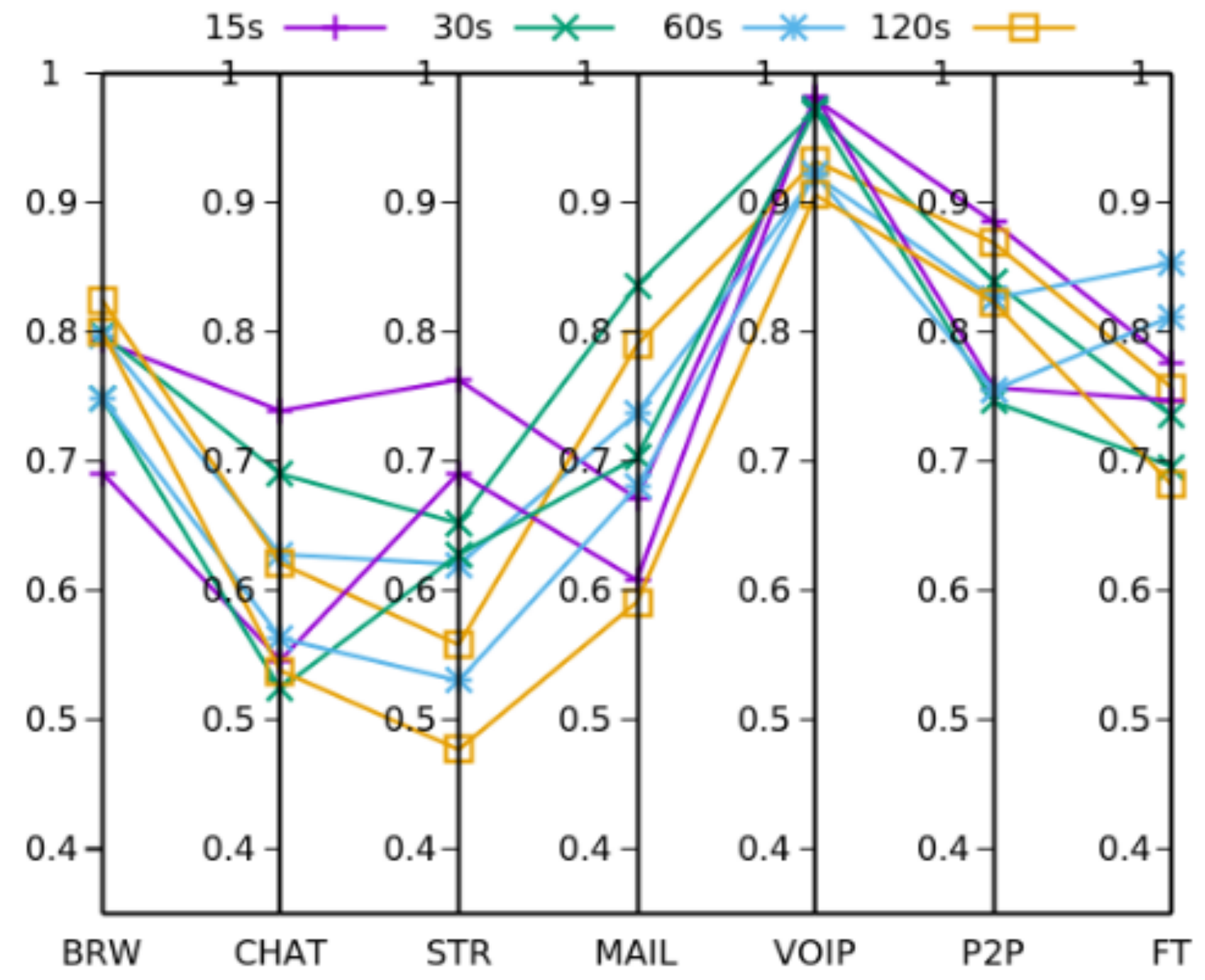


Precision (C4.5 and KNN)

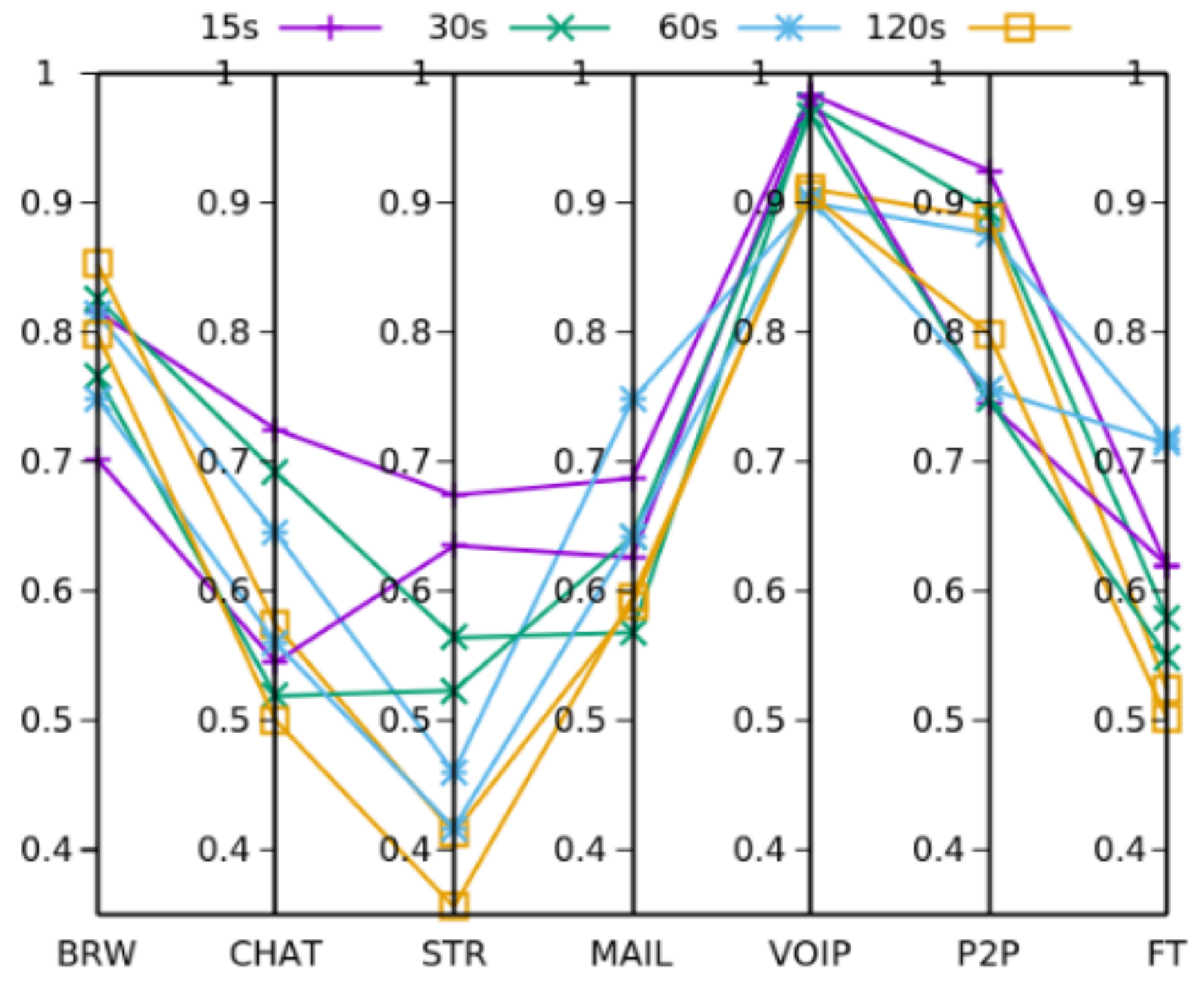


Recall (C4.5 and KNN)

Scenario B: NON-VPN Characterization



Precision (C4.5 and KNN)



Recall (C4.5 and KNN)

Conclusion & Future Direction

- Our classifiers perform better when the flows are generated using shorter timeout values, which contradicts the common assumption of using 600 *ms* as timeout duration.
- Future work: we plan to expand our work to other applications and types of encrypted traffic, and to further study the application of time-based features to characterize encrypted traffic.