Detection and characterization of multi-layer encryption traffic (VPN through Tor) using AI

Kay Boldt, Arash Habibi Lashkari and Ali A. Ghorbani

Canadian Institute for Cybersecurity (CIC), University of New Brunswick (UNB)

Abstract

The usage of Tor or VPN to protect one's data or identity is common nowadays. But it can be detected by SIEMs, and there is also the possibility to recognize and classify different applications used through Tor or VPN. Therefore it is of interest if the usage of multiple layers of encryption still can be detected and further if it is still possible to classify the traffic. Therefore in this project first of all traffic of common applications without additional encryption will be created and captured. In the second step, the traffic of the applications will be sent through a VPN over Tor. From this created dataset features are going to be extracted and fed into machine learning. First, the detection of the usage of multi-layer encryption is of interest. Second, the classification of the traffic for the case where no additional encryption is used and for the case where VPN+Tor is used is compared.

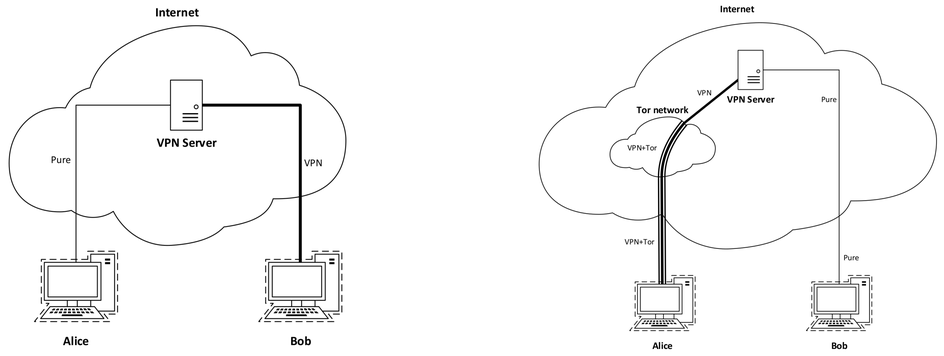
Previous Research

Authors	Paper	Year	Achievement
P. Mayank and A. K. Singh	Tor traffic identification	2017	Tor traffic detection
Zhihong Rao et al.	Tor anonymous traffic identification based on gravitational clustering	2018	Tor traffic detection without the need of training data
Gerard Draper-Gil et al.	Characterization of Encrypted and VPN Traffic using Time-related Features	2016	VPN traffic detection and classification using only time related features
Arash Habibi Lashkari et al.	Characterization of Tor Traffic using Time based Features	2017	Tor traffic detection and classification using only time related features
K. Shahbar and A. N. Zincir-Heywood	Benchmarking two techniques for Tor classification: Flow level and circuit level classification	2014	Comparison of flow- and cell-based Tor detection and classification

Traffic and Applications

TRAFFIC	APPLICATION
Browsing	Mozilla Firefox
Video chat	Hangout using Chrome
Video streaming	Youtube using Mozilla Firefox
Audio streaming	Spotify
SFTP file transfer	Filezilla
P2P file transfer	qBittorrent

Capturing Scenario



Pure VPN + Tor

Traffic captured on Alice's network interface

Flows in the dataset for different flow timeouts and traffic types

Traffic type	Number of flows per flow timeout 15s	45s	75s
Pure	208 905	140 551	106 519
VPN+Tor	301 486	150 626	100 620
Audio Streaming 30min Spotify - Pure	9 034	8 249	6 888
Audio Streaming 30min Spotify - VPN+Tor	16 607	10 770	8 490
Browsing 30min Firefox - Pure	36 887	33 092	30 568
Browsing 30min Firefox - VPN+Tor	29 606	18 738	13 640
P2P File Transfer qBittorrent - Pure	27 451	16 508	12 509
P2P File Transfer qBittorrent - VPN+Tor	60 500	23 615	14 724
SFTP File Transfer Filezilla - Pure	35 173	32 830	20 150
SFTP File Transfer Filezilla - VPN+Tor	85 666	40 815	26 529
Video Chat 30min Chrome Hangout - Pure	90 784	42 545	30 247
Video Chat 30min Chrome Hangout - VPN+Tor	58 566	31 456	20 332
Video Streaming 30min Firefox Youtube - Pure	9 576	7 327	6 157
Video Streaming 30min Firefox Youtube - VPN+Tor	50 541	25 232	16 905

Results:

Random Forest – Differentiation between Pure and VPN+Tor traffic

	Best 10 features Flow timeout 15sec	45sec	75sec	Best 20 features Flow timeout 15sec	45sec	75sec	Best 53 features Flow timeout 15sec	45sec	75sec
TP Rate - Pure	0.881	0.965	0.969	0.991	0.972	0.994	0.994	0.996	0.997
TP Rate - VPN+Tor	0.967	0.965	0.972	0.997	0.971	0.995	0.997	0.997	0.997
TP Rate - Weighted Avg.	**0.931**	**0.965**	**0.970**	**0.994**	**0.971**	**0.994**	**0.996**	**0.997**	**0.997**
FP Rate - Pure	0.033	0.035	0.028	0.003	0.029	0.005	0.003	0.003	0.003
FP Rate - VPN+Tor	0.119	0.035	0.031	0.009	0.028	0.006	0.006	0.004	0.003
FP Rate - Weighted Avg.	**0.084**	**0.035**	**0.029**	**0.007**	**0.029**	**0.006**	**0.004**	**0.003**	**0.003**

Random Forest – Classification of applications within Pure traffic

	Best 10 features Flow timeout 15sec	45sec	75sec	Best 20 features Flow timeout 15sec	45sec	75sec	Best 55 features Flow timeout 15sec	45sec	75sec
TP Rate - Audio Streaming 30min Spotify - Pure	0.174	0.229	0.295	0.419	0.419	0.495	0.593	0.616	0.598
TP Rate - Browsing 30min Firefox - Pure	0.753	0.711	0.773	0.840	0.851	0.879	0.883	0.902	0.916
TP Rate - P2P File Transfer qBittorrent - Pure	0.766	0.696	0.707	0.933	0.894	0.912	0.974	0.973	0.968
TP Rate - SFTP File Transfer Filezilla - Pure	0.794	0.764	0.798	0.961	0.936	0.963	0.986	0.992	0.990
TP Rate - Video Chat 30min Chrome Hangout - Pure	0.994	0.987	0.985	0.996	0.989	0.989	0.997	0.992	0.993
TP Rate - Video Streaming 30min Firefox Youtube - Pure	0.136	0.188	0.175	0.363	0.309	0.309	0.448	0.401	0.370
TP Rate - Weighted Avg.	**0.813**	**0.750**	**0.765**	**0.900**	**0.864**	**0.873**	**0.929**	**0.916**	**0.906**
FP Rate - Audio Streaming 30min Spotify - Pure	0.013	0.023	0.025	0.014	0.014	0.017	0.010	0.011	0.012
FP Rate - Browsing 30min Firefox - Pure	0.094	0.134	0.140	0.058	0.102	0.107	0.049	0.069	0.087
FP Rate - P2P File Transfer qBittorrent - Pure	0.043	0.039	0.036	0.013	0.011	0.008	0.004	0.003	0.002
FP Rate - SFTP File Transfer Filezilla - Pure	0.049	0.084	0.064	0.011	0.019	0.011	0.003	0.003	0.003
FP Rate - Video Chat 30min Chrome Hangout - Pure	0.011	0.007	0.007	0.007	0.005	0.005	0.005	0.003	0.002
FP Rate - Video Streaming 30min Firefox Youtube - Pure	0.014	0.024	0.024	0.015	0.018	0.017	0.013	0.015	0.015
FP Rate - Weighted Avg.	**0.037**	**0.061**	**0.061**	**0.018**	**0.033**	**0.037**	**0.013**	**0.020**	**0.028**

Random Forest – Classification of applications within VPN+Tor traffic

	Best 10 features Flow timeout 15sec	45sec	75sec	Best 20 features Flow timeout 15sec	45sec	75sec	Best 48 features Flow timeout 15sec	45sec	75sec
TP Rate - Audio Streaming 30min Spotify - VPN+Tor	0.280	0.410	0.527	0.441	0.513	0.597	0.464	0.479	0.523
TP Rate - Browsing 30min Firefox - VPN+Tor	0.492	0.557	0.560	0.458	0.540	0.567	0.467	0.574	0.620
TP Rate - P2P File Transfer qBittorrent - VPN+Tor	0.647	0.638	0.707	0.594	0.719	0.809	0.679	0.761	0.828
TP Rate - SFTP File Transfer Filezilla - VPN+Tor	0.589	0.649	0.687	0.621	0.685	0.742	0.672	0.772	0.809
TP Rate - Video Chat 30min Chrome Hangout - VPN+Tor	0.844	0.855	0.885	0.876	0.917	0.946	0.929	0.963	0.973
TP Rate - Video Streaming 30min Firefox Youtube - VPN+Tor	0.684	0.619	0.612	0.728	0.733	0.707	0.759	0.741	0.730
TP Rate - Weighted Avg.	**0.639**	**0.656**	**0.686**	**0.657**	**0.716**	**0.751**	**0.706**	**0.759**	**0.782**
FP Rate - Audio Streaming 30min Spotify - VPN+Tor	0.019	0.042	0.058	0.025	0.042	0.051	0.022	0.030	0.033
FP Rate - Browsing 30min Firefox - VPN+Tor	0.062	0.082	0.080	0.057	0.062	0.066	0.053	0.066	0.072
FP Rate - P2P File Transfer qBittorrent - VPN+Tor	0.122	0.074	0.060	0.109	0.068	0.054	0.094	0.058	0.047
FP Rate - SFTP File Transfer Filezilla - VPN+Tor	0.122	0.106	0.090	0.129	0.078	0.051	0.101	0.060	0.041
FP Rate - Video Chat 30min Chrome Hangout - VPN+Tor	0.048	0.046	0.035	0.044	0.035	0.027	0.031	0.018	0.014
FP Rate - Video Streaming 30min Firefox Youtube - VPN+Tor	0.077	0.068	0.055	0.065	0.058	0.048	0.064	0.059	0.054
FP Rate - Weighted Avg.	**0.088**	**0.075**	**0.065**	**0.085**	**0.060**	**0.048**	**0.071**	**0.049**	**0.042**

Conclusion:

- The usage of Tor can be correctly classified with an true positive rate of 99.7%
- Applications within Pure traffic can be classified with an true positive rate of 92.9%. Some applications like Audio or Video Streaming are significantly worse to detect.
- Applications within VPN+Tor traffic can be classified with an true positive rate of 78.2%. Again some applications like Audio Streaming and Browsing are hard to detect while Video Chart achieves an TP rate of 97.3%.

Future Work:

- Extend scenarios/dataset
- Test other machine learning algorithms/frameworks
- Explore new techniques to process the traffic for ML