Detection and Characterization of Tor Encrypted Traffic

Arash Habibi Lashkari, Gerard Drapper Gil, Mohammad Mamun, Ali A. Ghorbani

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

ABSTRACT

Traffic classification has been the topic of many research efforts, but the quick evolution of Internet services and the pervasive use of encryption makes it an open challenge. Encryption is essential in protecting the privacy of Internet users, a key technology used in the different privacy enhancing tools that have appeared in the recent years. Tor is one of the most popular of them, it decouples the sender from the receiver by encrypting the traffic between them, and routing it through a distributed network of servers. In this paper, we present a time analysis on Tor traffic flows, captured between the client and the entry node. We define two scenarios, one to detect Tor traffic flows and the other to detect the application type: Browsing, Chat, Streaming, Mail, Voip, P2P or File Transfer. In addition, we publish the Tor labelled dataset we generated and used to test our classifiers.

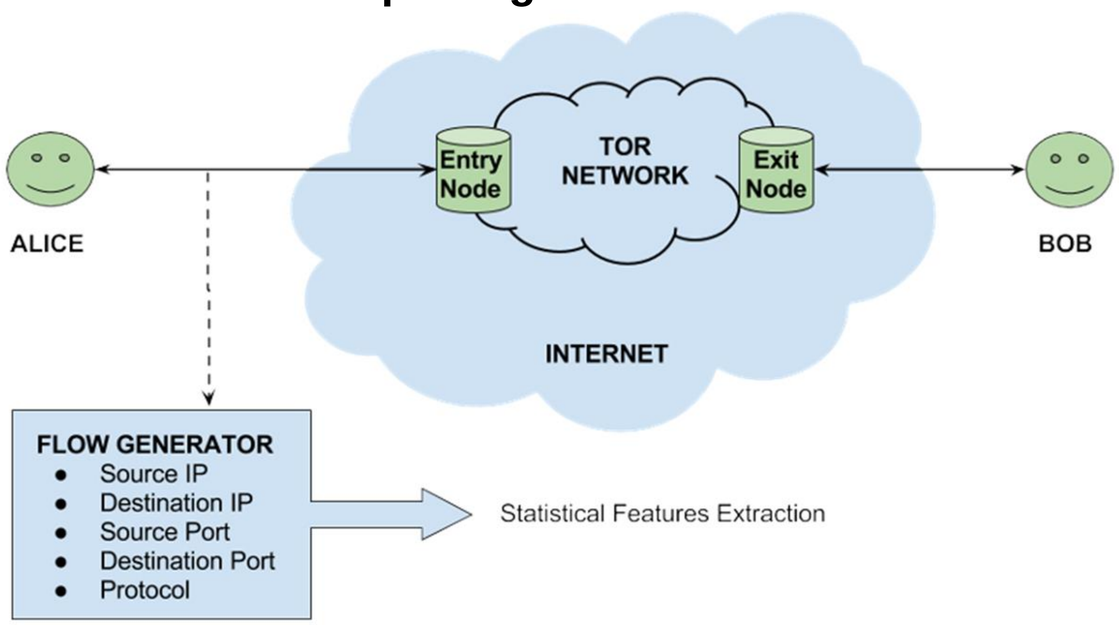
Previous Research

Research	Type	Features	Applications	Protocols
Alsabah et al.	Cell Information	Cell information such as circuit lifetime, cell inter-arrival time, number of cells sent recently from the network packets	-	-
Bai et al.	Network Traffic	Traffic features such as packet length and frequency of the packets' sending time	-	-
Luo et al.	Network Traffic	Burst volumes such as total size of all packets and directions	-	P2P, FTP, IM, Web
Our proposed method	Network Traffic	23 Time-related features	Browsing, Email, Chat, Audio, Video, File Transfer, P2P, VOIP	HTTP, HTTPS, Web, SMTP/S, PoP3/SSL, P2P, IMAP/SSL, SFTP, FTPS

Applications

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)

Capturing Scenario



Dataset Contents

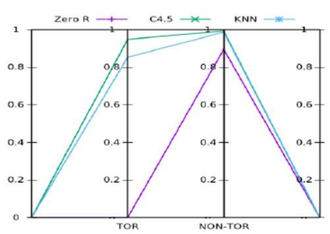
	Scenario A			Scenario B								
	TOR	NOTOR	Total	Bro	Ema	Chat	Aud	Vid	FT	VoIP	P2P	Total
10s.	8044	59790	67834	1604	282	323	721	874	864	2291	1085	8044
15s.	5631	48123	53754	1194	194	249	510	617	590	1544	733	5631
30s.	3130	43892	47022	694	111	153	332	364	311	790	375	3130
60s.	1723	41376	43099	411	60	90	190	196	165	413	198	1723
120s.	969	38285	39254	239	34	151	119	105	86	225	110	969

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 before going idle (mean, std, max, min)
idle	The amount of time a flow was idle before going active (mean, std, max, min)
fb_psec	Flow Bytes per second
fp_psec	Flow Packets per second

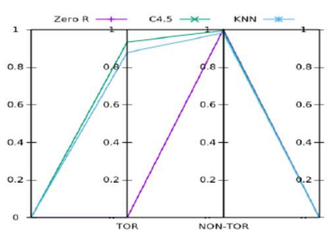
Results:

Scenario A

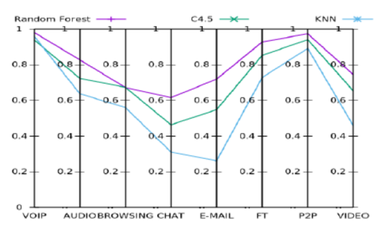
Scenario B



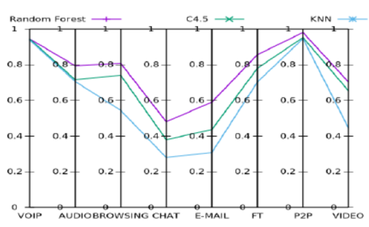
(a) Tor VS. Non-Tor Precision



(b) Tor VS. Non-Tor Recall



(c) Tor Characterization Precision



(d) Tor Characterization Recall

Conclusion and Future Works:

- TOR classifier obtains more than 90% accuracy by just four features Min_flowait, Std_biat, Mean_biat, Max_biat
- Further study on the other features such as flow-based and packet-based
- Extend the work to other types of encrypted traffic
- Extend the TOR dataset by adding more applications and other scenarios