FlowPic: Encrypted Internet Traffic Classification is as Easy as Image Recognition

Tal Shapira talshapira1@mail.tau.ac.il

Yuval Shavitt shavitt@eng.tau.ac.il

School of Electrical Engineering, Tel-Aviv University

AI Week
Nov. 17-21, 2019
Tel Aviv University

Introduction

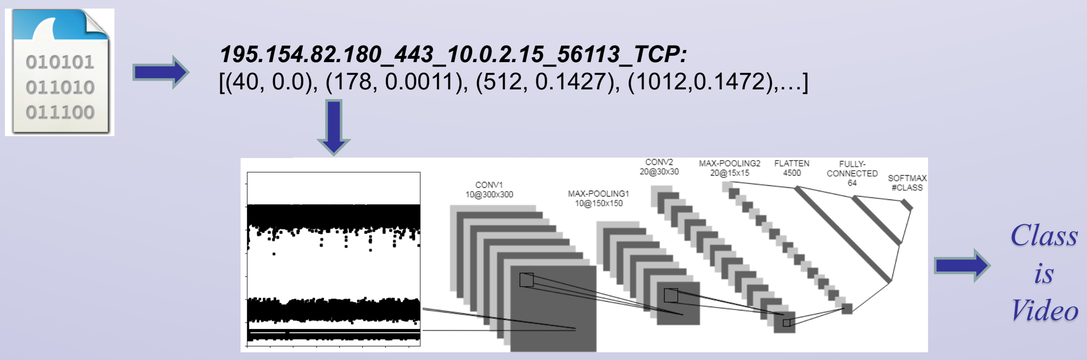
•Given a tap on a communication channel, our aim is to classify the flows even if they are passing through VPN or ToR.
•Our goal is to propose **a generic approach** for Internet traffic classification:
- We use the exact same architecture for all the experiments.
- Dealing with all kinds of Internet traffic classification problems.

•**Using all time and size related information available** in a network flow, instead of just using information from manually extracted features.
•Dealing with **only a time window of a unidirectional flow** instead of the entire bidirectional session.
•**Do not rely on the packet payload content:**
- Do not breach privacy.
- Minimalizing storage requirement.
- Classifying VPN and Tor traffic.

•Capturing an intrinsic characteristic of a category behavior, regardless of the encryption technique and a specific user.

Approach

1. Extract records from each flow, which comprised of a list of pairs, {IP packet size, time of arrival} for each packet in the flow.
2. Split each unidirectional flow to equal blocks (15/60 seconds).
3. Generate 2D-histogram. For simplicity, we set the 2D-histogram to be a square image.
4. Feed a LeNet-5 style Convolution Neural Network.



FlowPics Exploration

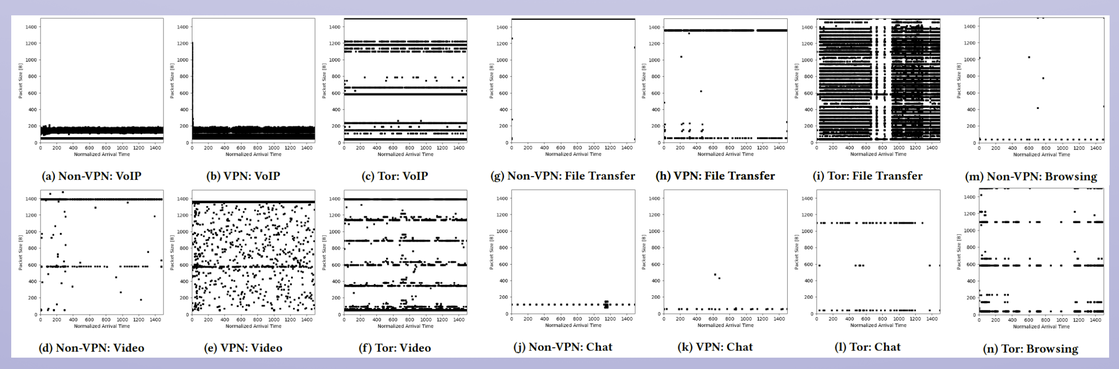


Figure 1: Examples of *FlowPics* for VoIP, video, file transfer, chat and browsing over non-VPN, VPN and Tor.

Figure 2: Examples of *FlowPics* for several video applications.

Experiments and Results

We use labeled datasets of packet capture (pcap) files from the Uni. of New Brunswick: "ISCX VPN-nonVPN traffic dataset" (ISCX-VPN)[15] and "ISCX Tor-nonTor dataset" (ISCX-Tor) [21], as well as our own small packet capture (TAU), and conduct different types of experiments; (1) multiclass classification experiments over non-VPN/VPN/Tor and merged dataset, (2) class vs. all classification experiments, (3) application identification, and (4) classification of an unknown application.

Problem	FlowPic Acc. (%)	Best Previous Result	Remark
Non-VPN Traffic Categorization	85.0	84.0 % Pr., Gil *et al.* [22]	Different categories. [22] used unbalanced dataset
VPN Traffic Categorization	98.4	98.6 % Acc., Wang *et al.* [7]	[7] Classify raw packets data. Not including browsing category
Tor Traffic Categorization	67.8	84.3 % Pr., Gil *et al.* [22]	Different categories. [22] used unbalanced dataset
Traffic Categorization over Merged Dataset	83.0	No previous results	
Non-VPN Class vs. All	97.0 (Average)		
VPN Class vs. All	99.7 (Average)		
Tor Class vs. All	85.7 (Average)		
Encryption Techniques	88.4	99. % Acc., Wang *et al.* [7]	[7] Classify raw packets data, not including Tor category
Applications Identification	99.7	93.9 % Acc., Yamansavascilar *et al.* [9]	Different classes

No previous results

Table 1: A summary table of our results for different traffic classification problems, with comparison to best known previous results over the ISCX dataset.

Class	Accuracy (%)			
VoIP	Training/Test	Non-VPN	VPN	Tor
	Non-VPN	**99.6**	99.4	48.2
	VPN	95.8	**99.9**	58.1
	Tor	52.1	35.8	**93.3**
Video	Training/Test	Non-VPN	VPN	Tor
	Non-VPN	**99.9**	98.8	83.8
	VPN	54.0	**99.9**	57.8
	Tor	55.3	86.1	**99.9**
File Transfer	Training/Test	Non-VPN	VPN	Tor
	Non-VPN	**98.8**	79.9	60.6
	VPN	65.1	**99.9**	54.5
	Tor	63.1	35.8	**55.8**
Chat	Training/Test	Non-VPN	VPN	Tor
	Non-VPN	**96.2**	78.9	70.3
	VPN	71.7	**99.2**	69.4
	Tor	85.8	93.1	**89.0**
Browsing	Training/Test	Non-VPN	VPN	Tor
	Non-VPN	**90.6**	-	57.2
	VPN	-	-	-
	Tor	76.1	-	**90.6**

Table 2: Class vs. all classification accuracy performances.

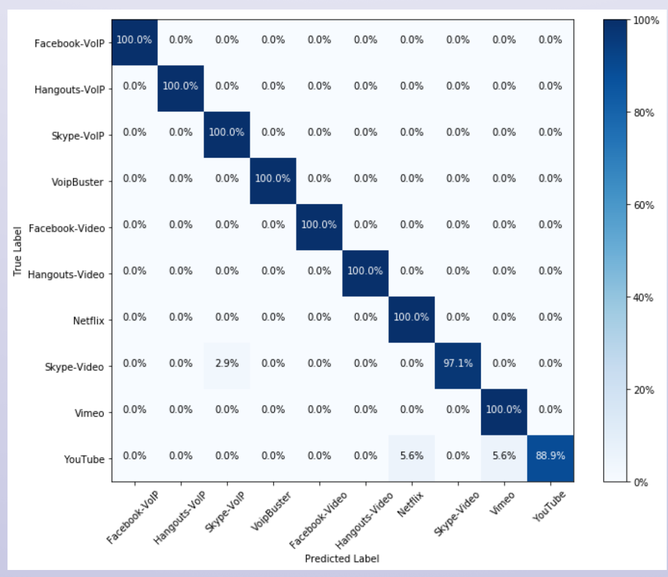


Figure 3: A confusion matrix of the VoIP and video applications identification problem.

Conclusions

- A **generic approach** for Internet traffic classification:
  - We use the exact same CNN architecture for all the experiments.
  - **Dealing with all kinds of Internet traffic classification problems**; traffic categorization, application identification, encryption techniques.
- We are the first to show:
  - **Identifying traffic category of an unfamiliar application**, by learning samples of other applications of the same traffic category.
  - Classifying different Internet traffic categories that pass through different encryption techniques by learning traffic that pass throw other encryption techniques.
- **Fast & simple**: Classification is done after the first **15 seconds of a unidirectional flow .**
- **No need for manual extraction of features.**
- Good results for traffic over VPN and Tor.

References

[1] P. Velan, M. Cermák, P. Celeda, and M. Drašar, **"A survey of methods for encrypted traffic classification and analysis,"** International Journal of Network Management, vol. 25, no. 5, pp. 355–374, 2015.

[7] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, **"End-to-end encrypted traffic classification with one-dimensional convolution neural networks,"** in 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), July 2017, pp. 43–48.

[10] B. Yamansavascilar, M. A. Guvensan, A. G. Yavuz, and M. E. Karsligil, **"Application identification via network traffic classification,"** in 2017 International Conference on Computing, Networking and Communications (ICNC), Jan 2017, pp. 843–848.

[15] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, **"Characterization of encrypted and vpn traffic using time-related features,"** in Proceedings of the 2nd International Conference on Information Systems Security and Privacy - Volume 1: ICISSP,, INSTICC. SciTePress, 2016, pp. 407–414.

[21] A. H. Lashkari, G. D. Gil, M. S. I. Mamun, and A. A. Ghorbani, **"Characterization of tor traffic using time based features,"** in Proceedings of the 3rd International Conference on Information Systems Security and Privacy - Volume 1: ICISSP,, INSTICC. SciTePress, 2017, pp. 253–262.

You can find more details in - https://ieeexplore.ieee.org/document/8845315