Towards Designing Robust and Efficient Classifiers for Encrypted Traffic in the Modern Internet

Xi Jiang[1], Shinan Liu[1], Saloua Naama[2], Francesco Bronzino[3], Paul Schmitt[4], and Nick Feamster[1]

[1]University of Chicago

[2]Université Savoie Mont Blanc

[3]École Normale Supérieure de Lyon

[4]University of Hawaii, Manoa

1 Abstract

Over the past several decades, the Internet infrastructure has evolved in many ways and one notable trend is encrypted transport which renders conventional traffic classification methods increasingly less effective. In this position paper, we argue that existing classifiers for encrypted network traffic are suffering from crucial problems associated with low robustness against model drifts and inadequate efficiency for real-life deployment. We propose potential solutions to these challenges by reducing the feature space required for such classifiers and exploiting robust network-level features across multiple datasets across time and space.

2 Introduction

Network traffic classification is a common network management task that involves inferring Internet services and applications. Efficiently and accurately classifying network traffic allows network operators to perform a wide range of essential network operations, including capacity and resource planning, quality of service (QoS) monitoring, traffic prioritization, malicious traffic detection, etc [3, 17, 30, 32, 35–38]. Conventional approaches to traffic classification often rely on network features handcrafted from expert knowledge [27,33,40]. More recent efforts have applied machine learning (ML) to perform classification, using both classical-learning-based [6, 10, 12, 18, 20, 21, 29] and deep-learning-based methods [2, 8, 11, 22, 23, 31, 34, 39, 42, 44, 46]. These methods have generally performed well when applied to curated datasets and evaluated in specific contexts—moreover, they have frequently depended on domain-specific features, including IP addresses and information that is available in unencrypted packet payloads.

However, the rise of encrypted network traffic [4, 9, 13, 15, 19, 25, 26, 28, 43] now threaten the effectiveness of long-established network traffic classification methods. In this position paper, we examine the challenges associated with designing traffic classifiers that are robust and efficient against pervasive encryption of the application and transport layers. Based on these observations, we present an opportunity for the network research community to re-examine this critically important space, to develop new methods for traffic classification that are robust in the face of encryption, and more accurate and efficient on modern network traffic. We also suggest several possible solutions to these challenges.

3 Why are current encrypted traffic classifiers not enough?

Existing classifiers focus on accuracy but not efficiency. Increasing utilization of different network traffic encryption schemes alter the feature space of ML-based traffic classifiers by (1) *reducing the usefulness of affected features* or (2) *shifting the feature importance distribution*, and the majority of the existing classifiers attempt to address these issues by relying on complex deep-learning based models to avoid manually articulating informative features [8, 11, 18, 22, 34, 39, 46]. Unlike traditional methods that are heuristics-based [1, 27, 33, 40, 45] or classical machine-learning based [5, 6, 6, 18, 20, 21] which usually depend on a few pre-selected components of the traffic flows, the complex nature of these deep-learning models also means that they typically require lengthy network traffic inputs, such as the entirety of the packet headers, to make traffic classification decisions accurately. Unfortunately, in a real-world deployment setting such as an Internet Service Provider (ISP), capturing and storing large portions of the traffic flows on a large scale can introduce high overheads in terms of system costs, such as memory requirements, and as well as unnecessary delays to network traffic. Moreover, it is crucial for network administrators to make classification decisions quickly so that appropriate follow-up actions can be taken and considering a broad set of network traffic features can slowdown the inference speed of such classifiers which further reduces their efficiency.

Classifiers evaluated using closed-world datasets are not robust against model drift. While most existing classifiers designed for encrypted network traffic show promising results when evaluated with closed-world datasets, such classifiers often fail to remain robust when given newer network traffic received at different times or locations. To illustrate this issue, we conducted a sample study to collect TLS encrypted traffic across a wide range of applications at two different locations and times (two years apart), and split the collected traffic into two different datasets (*old* and *new*) accordingly. Our study shows that while we can train ML-based traffic classifiers to perform well on the *old* dataset, the performance of such classifiers degrades severely when applied directly to the *new* dataset, even though both datasets contain traffic from the same set of applications. More generally speaking, while many existing encrypted traffic classifiers are evaluated using well-known datasets such as ISCX VPN-NonVPN [14] and UNIBS-2009 [16], these classifiers are not robust against the above-mentioned model drift as such closed-world datasets are not necessarily sufficient to describe what the most up-to-date Internet traffic actually looks like.

4 What are some plausible solutions?

Utilize classical machine-learning methods to reduce feature space to improve efficiency. While deep learn-based approaches seem to be the mainstream approach for designing classifiers for encrypted network traffic, we found that we can utilize classical machine-learning methods to reduce the number of features to consider while obtaining reasonably good classification results. Reducing the feature space while maintaining the classification accuracy can effectively lower the relevant system cost for classifier implementers, because they need to preserve less traffic information. A plausible way to reduce the feature space is to rank network-level features according to the feature importance as interpreted by the models and neglect features that are less informative (or have negative impacts on classifier performance). Evaluated using prominent datasets, including the QUIC dataset [41], the ISCX VPN-NonVPN traffic dataset [14], and our collected TLS encrypted traffic flows (which include *video streaming* [7], video *conferencing* [24], and *social media* applications), our results show that we can arrive at relatively similar performance when providing the models with just the top few features (packet header fields) compared to all features. At the same time, we observe a reduction in inference time needed to arrive at classification decisions as fewer features (i.e. fewer matrix multiplications) are being considered.

Perform statistical analysis on multiple datasets to locate features robust against model drift. While training and evaluating models based on a single closed-world dataset can lead to classifiers that are not robust to potential model drift, we can try to identify features that remain consistently robust across datasets and exploit these features when designing classifiers. Here we define a set of features to be *robust* when models trained and validated using this set of features can achieve similar performance when tested on a new dataset that it has never seen before. One reasonable way to obtain this set of features is through statistical analysis/comparison across datasets and finding network-level features with relatively consistent values and distributions (for each predicting application/service) across the datasets. Providing the models with this set of robust features allows us to avoid environment-specific features that are over-fitted to a particular dataset which can be easily rendered ineffective by model drift.

5 References

[1] G. Aceto, A. Dainotti, W. De Donato, and A. Pescapé. Portload: taking the best of two worlds in traffic classification. In *2010 INFOCOM IEEE Conference on Computer Communications Workshops*, pages 1–5. IEEE, 2010.

[2] I. Akbari, M. A. Salahuddin, L. Ven, N. Limam, R. Boutaba, B. Mathieu, S. Moteau, and S. Tuffin. A look behind the curtain: traffic classification in an increasingly encrypted web. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(1):1–26, 2021.

[3] F. Baker, B. Foster, and C. Sharp. Cisco architecture for lawful intercept in ip networks. *Internet Engineering Task Force, RFC*, 3924, 2004.

[4] D. Belson. Akamai state of the internet report, q4 2009. *ACM SIGOPS Operating Systems Review*, 44(3):27–37, 2010.

[5] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamatian. Traffic classification on the fly. *ACM SIGCOMM Computer Communication Review*, 36(2):23–26, 2006.

[6] L. Bernaille, R. Teixeira, and K. Salamatian. Early application identification. In *Proceedings of the 2006 ACM CoNEXT conference*, pages 1–12, 2006.

[7] F. Bronzino, P. Schmitt, S. Ayoubi, G. Martins, R. Teixeira, and N. Feamster. Inferring streaming video quality from encrypted traffic: Practical models and deployment experience. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(3):1–25, 2019.

[8] Z. Bu, B. Zhou, P. Cheng, K. Zhang, and Z.-H. Ling. Encrypted network traffic classification using deep and parallel network-in-network models. *IEEE Access*, 8:132950–132959, 2020.

[9] K. Cho. Trends in japanese residential traffic. *ISOC Panel on Internet Bandwidth: Dealing with Reality*, 2009.

[10] K. C. Claffy. Internet traffic characterization. 1995.

[11] S. Cui, B. Jiang, Z. Cai, Z. Lu, S. Liu, and J. Liu. A session-packets-based encrypted traffic classification using capsule neural networks. In *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 429–436. IEEE, 2019.

[12] C. Dewes, A. Wichmann, and A. Feldmann. An analysis of internet chat systems. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, pages 51–64, 2003.

[13] A. Dhamdhere and C. Dovrolis. Ten years in the evolution of the internet ecosystem. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 183–196, 2008.

[14] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani. Characterization of encrypted and vpn traffic using time-related. In *Proceedings of the 2nd international conference on information systems security and privacy (ICISSP)*, pages 407–414. sn, 2016.

[15] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse? In *International Conference on Passive and Active Network Measurement*, pages 1–10. Springer, 2008.

[16] F. Gringoli, L. Salgarelli, M. Dusi, N. Cascarano, F. Risso, and K. Claffy. Gt: picking up the truth from the ground for internet traffic. *ACM SIGCOMM Computer Communication Review*, 39(5):12–18, 2009.

[17] H. Jiang, A. W. Moore, Z. Ge, S. Jin, and J. Wang. Lightweight application classification for network management. In *Proceedings of the 2007 SIGCOMM workshop on Internet network management*, pages 299–304, 2007.

[18] T. Karagiannis, K. Papagiannaki, and M. Faloutsos. Blinc: multilevel traffic classification in the dark. In *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 229–240, 2005.

[19] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. *ACM SIGCOMM Computer Communication Review*, 40(4):75–86, 2010.

[20] T. Lang, G. Armitage, P. Branch, and H.-Y. Choo. A synthetic traffic model for half-life. In *Australian Telecommunications Networks & Applications Conference*, volume 2003, 2003.

[21] T. Lang, P. Branch, and G. Armitage. A synthetic traffic model for quake3. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 233–238, 2004.

[22] M. Lotfollahi, M. Jafari Siavoshani, R. Shirali Hossein Zade, and M. Saberian. Deep packet: A novel approach for encrypted traffic classification using deep learning. *Soft Computing*, 24(3):1999–2012, 2020.

[23] Q. Ma, W. Huang, Y. Jin, and J. Mao. Encrypted traffic classification based on traffic reconstruction. In *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 572–576. IEEE, 2021.

[24] K. MacMillan, T. Mangla, J. Saxon, and N. Feamster. Measuring the performance and network utilization of popular video conferencing applications. In *Proceedings of the 21st ACM Internet Measurement Conference*, pages 229–244, 2021.

[25] G. Maier, A. Feldmann, V. Paxson, and M. Allman. On dominant characteristics of residential broadband internet traffic. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, pages 90–102, 2009.

[26] O. Malik. Wholesale internet bandwidth prices keep falling. *GigOM Blog, http://gigaom. com*, 2008.

[27] A. W. Moore and K. Papagiannaki. Toward the accurate identification of network applications. In *International workshop on passive and active network measurement*, pages 41–54. Springer, 2005.

[28] W. B. Norton. Video internet: The next wave of massive disruption to the us peering ecosystem (v1. 3), 2008.

[29] V. Paxson. Empirically derived analytic models of wide-area tcp connections. *IEEE/ACM transactions on Networking*, 2(4):316–336, 1994.

[30] V. Paxson. Bro: a system for detecting network intruders in real-time. *Computer networks*, 31(23-24):2435–2463, 1999.

[31] V. Rimmer, D. Preuveneers, M. Juarez, T. Van Goethem, and W. Joosen. Automated website fingerprinting through deep learning. *arXiv preprint arXiv:1708.06376*, 2017.

[32] M. Roesch. Snort-the de facto standard for intrusion detection/prevention, 2005.

[33] S. Sen, O. Spatscheck, and D. Wang. Accurate, scalable in-network identification of p2p traffic using application signatures. In *Proceedings of the 13th international conference on World Wide Web*, pages 512–521, 2004.

[34] T. Shapira and Y. Shavitt. Flowpic: Encrypted internet traffic classification is as easy as image recognition. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 680–687. IEEE, 2019.

[35] P. Singhal, R. Mathur, and H. Vyas. State of the art review of network traffic classification based on machine learning approach. *International Journal of Computer Applications*, 975:8887, 2013.

[36] R. Sommer. Bro: An open source network intrusion detection system. *Security, E-learning, E-Services, 17. DFN-Arbeitstagung über Kommunikationsnetze*, 2003.

[37] R. Sommer and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy*, pages 305–316, 2010.

[38] L. Stewart, G. Armitage, P. Branch, and S. Zander. An architecture for automated network control of qos over consumer broadband links. In *TENCON 2005-2005 IEEE Region 10 Conference*, pages 1–6. IEEE, 2005.

[39] B. Sun, W. Yang, M. Yan, D. Wu, Y. Zhu, and Z. Bai. An encrypted traffic classification method combining graph convolutional network and autoencoder. In *2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC)*, pages 1–8. IEEE, 2020.

[40] G. Szabo, I. Szabo, and D. Orincsay. Accurate traffic classification. In *2007 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pages 1–8, 2007.

[41] V. Tong, H. A. Tran, S. Souihi, and A. Mellouk. A novel quic traffic classifier based on convolutional neural networks. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2018.

[42] M. Wang, K. Zheng, D. Luo, Y. Yang, and X. Wang. An encrypted traffic classification framework based on convolutional neural networks and stacked autoencoders. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 634–641. IEEE, 2020.

[43] J. Wu, Z. M. Mao, J. Rexford, and J. Wang. Finding a needle in a haystack: Pinpointing significant bgp routing changes in an ip network. In *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2*, pages 1–14, 2005.

[44] H. Yao, C. Liu, P. Zhang, S. Wu, C. Jiang, and S. Yu. Identification of encrypted traffic through attention mechanism based long short term memory. *IEEE Transactions on Big Data*, 2019.

[45] S.-H. Yoon, J.-W. Park, J.-S. Park, Y.-S. Oh, and M.-S. Kim. Internet application traffic classification using fixed ip-port. In *Asia-Pacific Network Operations and Management Symposium*, pages 21–30. Springer, 2009.

[46] W. Zheng, J. Zhong, Q. Zhang, and G. Zhao. Mtt: an efficient model for encrypted network traffic classification using multi-task transformer. *Applied Intelligence*, pages 1–16, 2022.