Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Laurent Vanbever
Networked Systems Group

Anonymized traffic trace collection in the data plane

Semester/Master thesis proposal

Most network researchers need traffic traces from real networks in order to evaluate and verify their ideas. For example, the network community has recently developed many artificial intelligence-based techniques that require accurate and real data to work well [7]. Unfortunately, getting traces from real networks is rather difficult. Operators are often unwilling to share traffic traces from their network for obvious privacy reasons. Only a small number of organizations such as [2, 5] periodically publish a limited number of anonymized traffic traces.

To convince more network operators to make traffic traces publicly available, we are currently designing a traffic collection framework that is easy to deploy, flexible at runtime and ensures privacy. The key idea of our collection framework is to use the capabilities of newest programmable switches. Last semester, a student started to develop the framework in P4 [6] during his semester thesis [4]. Based on some operator input, the framework generates optimized P4 code which automatically collects, anonymizes and forwards traffic in order to generate traces. The system even partially runs on a Barefoot Tofino [1] switch.

In this thesis we would like to extend the existing framework in multiple ways. First, as the available memory and processing capabilities of data-plane devices are limited, the current framework produces highly optimized but static P4 code. However, the observed network traffic is constantly changing whereas the P4 code used in a Tofino switch cannot be changed at runtime. The student is therefore expected to improve the framework so that it can adapt its behavior according to the traffic, and this at runtime, without having to update the data-plane implementation, recompile the P4 code or restarting the device. Second, to compensate for the limited processing capabilities of data-plane devices (e.g., no cryptographic hash functions) the current framework uses a simplified algorithm to anonymize the traces compared to state-of-the-art approaches (e.g., [3]). The student is expected to theoretically explore what the tradeoffs are between the different implementations. Is our current anonymization easier to break/reverse? Third, the current framework can only anonymize IPv4 addresses. Therefore, the framework must be extended to support IPv6 addresses as well. Finally, the entire framework should be adapted to the Tofino-specific P4 specifications to show that it is practicable and that the framework can generate anonymized traces for Tbps of traffic.

Milestones

- Get familiar with the existing framework/code basis;
- Extend the existing framework such that operators can add/remove new prefixes as soon as the observed traffic mix changes;
- Theoretically analyze the achieved anonymization guarantees. How difficult would it be to break the traffic anonymization? Is our approach comparable to existing solutions?
- Extend the framework such that it works for IPv6;
- Test the full framework on a Tofino switch in our lab.

Requirements

- Knowledge in P4 programming;
- Communication Networks (227-0120-00L), or equivalents.

Contact

- Thomas Holterbach, thomahol@ethz.ch
- Tobias Bühler, buehlert@ethz.ch
- Prof. Dr. Laurent Vanbever, lvanbever@ethz.ch

References

[1] Barefoot Tofino P4 switch. `https://www.barefootnetworks.com/products/brief-tofino/`.

[2] The caida anonymized internet traces dataset. `http://www.caida.org/data/passive/passive_dataset.xml`.

[3] Cryptography-based prefix-preserving anonymization. `https://ant.isi.edu/software/cryptopANT/index.html`.

[4] A framework for collecting data traffic from real networks. `https://nsg.ee.ethz.ch/fileadmin/user_upload/theses/traces_collector_thesis_proposal.pdf`.

[5] Mawi working group traffic archive. `http://mawi.wide.ad.jp/mawi/`.

[6] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, and D. Walker. P4: Programming protocol-independent packet processors. *SIGCOMM Comput. Commun. Rev.*, 2014.

[7] A. Gupta, C. Mac-Stoker, and W. Willinger. An effort to democratize networking research in the era of ai/ml. In *Proceedings of the 18th ACM Workshop on Hot Topics in Networks*, 2019.