Internship / Ph.D. Thesis Proposal 2024-2025

Title: Dynamic model serving for network traffic analysis

Host laboratory: LIP, ENS de Lyon, 46 allée d'Italie, Lyon, France

Advisors:

Francesco Bronzino MCF HDR ENS Lyon `francesco.bronzino@ens-lyon.fr`

External Collaborators: The thesis will be carried on within a collaborative project with the University of Chicago and ICSI at UC Berkeley.

Starting Date: The thesis will start in October 2025 and will be preceded by a Master internship.

Keywords: Traffic Analysis, Machine Learning, Network Systems,QoS inference

Description.

Network management increasingly relies on machine learning models to make predictions for network management tasks, including traffic classification, intrusion detection, and application quality of experience (QoE) inference. These tasks typically require real-time inference and decision-making. Unfortunately, existing models typically prioritize prediction accuracy without considering deployment constraints such as prediction latency and throughput. As a result, while machine learning has been shown promise for improving various aspects of network operations, deploying these models in real-world networks with tight performance constraints remains a challenge, especially when delays in decision-making can lead to substantial packet loss or degraded user experiences.

To balance trade-offs between model performance and system constraints, a range of models is available for each task. For instance, CATO [2] identifies models that are Pareto optimal in terms of both accuracy and system costs. However, deploying these models in real-world environments remains challenging. The users of these models have to decide a priori which model might best fit their deployment, configuring their measurement system accordingly [2, 1]. Yet, due to the dynamic nature of network traffic, which fluctuates continuously throughout the day, identifying a configuration that remains optimal under all conditions is challenging, if not impossible. To address this, we propose to develop system-driven solutions that dynamically select and serve traffic inference models from a pool of models with varying performance and accuracy characteristics. This project will explore several approaches, from enabling measurement infrastructures to prioritize processing the most informative flows—via admission control techniques tailored to network analytics—to designing a multi-stage, constraint-aware model selection mechanism that can adapt in real-time.

Candidate Requirements.

- The candidate should have completed a qualifying program by the starting date of the thesis.
- Comfortable speaking English or French (French is not required).
- Good understanding of computer networks protocols and systems (preferably both). Understanding of machine learning is a plus.
- Good proficiency with at least one programming language for systems, preferably Golang or Rust.

What to submit. An up to date CV, university transcripts, and a letter of motivation clearly stating what the motivations to work on the described subject. One or (preferably) two recommendation letters are also welcome and strongly encouraged.

References

[1] F. Bronzino, P. Schmitt, S. Ayoubi, H. Kim, R. Teixeira, and N. Feamster. Traffic refinery: Cost-aware data representation for machine learning on network traffic. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2021.

[2] G. Wan, S. Liu, F. Bronzino, N. Feamster, and Z. Durumeric. Cato: End-to-end optimization of ml traffic analysis pipelines. *arXiv preprint arXiv:2402.06099*, 2024.