Research Internship Proposal

Title: Addressing the scarcity of labeled data for Network Intrusion Detection through Active and Few-shot Learning

Supervision: Kevin Jiokeng, LIX, École Polytechnique (kevin.jiokeng@polytechnique.edu)

In front of the proliferation of Network Intrusion attacks leveraging more and more sophisticated techniques, incorporation of Artificial Intelligence and Machine Learning is becoming the *de facto* standard to secure Network infrastructures. Research literature has indeed rapidly shifted from rule- and signature-based systems to AI/ML powered ones [1], trying to make the most from recent advances in this field.
Unfortunately, unlike in other fields like computer vision, full adoption of AI/ML for Network Intrusion Detection (NID) still faces the challenge of the absence of (good quality – and labeled) data, as literature still lacks large annotated datasets comprising rich enough diversity [2, 3]. Existing datasets are either too small or too specific to a given category of attacks or network setup. To face this challenge, recent research works have exploited either oversampling strategies as well as Active Learning and Few-shot Learning [4, 5].

While these approaches improve NIDS performance, we argue that careful combination of these learning schemes could become a real game changer in the field. To this end, the goal of this internship is to propose, develop and evaluate a hierarchical learning scheme well adapted to NIDS. Evaluation will be done with existing datasets as well as custom ones collected on real hardware available in our lab.

Expected candidate skills:

The most important skill for this internship is **to be eager to learn while trying new solutions.** On top of that, the following skills would be strongly appreciated.
- Hands-on experience and strong skills in Machine and Deep Learning. Knowledge of modern learning schemes such as Active and Few-shot Learning, Autoencoders and Transfer Learning would be appreciated.
- Strong programming skills in any common language such as C++, Python, Java, etc.
- Knowledge of network protocols functioning would be appreciated

This internship can lead to a PhD thesis. Funding already available.

References

[1] Dongqi Han et al. 2021. DeepAID: Interpreting and Improving Deep Learning-based Anomaly Detection in Security Applications. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21). ACM, New York, NY, USA, 3197–3217. https://doi.org/10.1145/3460120.3484589
[2] G. Apruzzese, P. Laskov and A. Tastemirova, "SoK: The Impact of Unlabelled Data in Cyberthreat Detection," 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P), Genoa, Italy, 2022, pp. 20-42, doi: 10.1109/EuroSP53844.2022.00010.
[3] Jordan Holland, Paul Schmitt, Nick Feamster, and Prateek Mittal. 2021. New Directions in Automated Traffic Analysis. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21). Association for Computing Machinery, New York, NY, USA, 3366–3383. https://doi.org/10.1145/3460120.3484758
[4] Liang, Junjie, et al. "FARE: enabling fine-grained attack categorization under low-quality labeled data." Proceedings of The Network and Distributed System Security Symposium (NDSS). 2021.
[5] Y. Zhang, J. Niu, G. He, L. Zhu and D. Guo, "Network Intrusion Detection Based on Active Semi-supervised Learning," 2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), Taipei, Taiwan, 2021, pp. 129-135, doi: 10.1109/DSN-W52860.2021.00031.