SUPERVIZ – Postdoctoral position on Test Data Generation using Traffic Morphing

Gregory Blanc
SAMOVAR, Télécom SudParis
Institut Polytechnique de Paris

start date: Autumn 2023

Context

Evaluation is a pillar for the certification of cybersecurity products. In particular, security monitoring, especially intrusion detection, has been struggling with a low evaluation reproducibility in terms of methodologies or platforms [1]. To guarantee a certain level of confidence, intrusion detectors should be tested with a sufficiently large and diverse amount of samples, representative of realistic behaviours, including both malicious ones (*sensitivity*) and benign ones (*specificity*). Other important aspects to assess is their scalability and their robustness (e.g., using *adversarial examples* [2]). When data is lacking or the quality is low, data augmentation may provide sufficient data for testing the ability of intrusion detectors to withstand the load and recognize new behaviours. Good quality data usually display properties of realism and diversity [6]. While the latter can be reasonably obtained using perturbation methods [7], it counteracts with the former as diverse samples may be unpractical [4]. This is due to an actual dichotomy between the feature space (on which pertubations generally apply) and the problem space where traffic really flows [5]. Additionally, it may be desirable to test data-driven intrusion detectors with real traffic, beyond the feature space. To that end, there exist perturbations that can be applied to the problem space [3], such traffic morphing [8].

Objectives.

This postdoctoral project proposes to generate new data samples to challenge network intrusion detection systems (NIDS). Perturbation methods may be used: they require existing traffic from which we derive new samples. We aim at triggering misclassification for both benign samples (*false positives*) and malicious samples (*false negatives*) in order to assess the robustness of the NIDS. Another important objective is the generation of problem-space samples, either synthetically or by perturbing existing traffic. Different approaches from both feature space and problem space will be assessed and compared in various settings (open model, closed model, model generation, model adaptation).

Activities. The project contributes to a better assessment of NIDS by:

- surveying the state of the art of NIDS evaluation methods
- surveying the state of the art of data perturbation at feature- and proble-space levels
- implementing a selection of promising approaches and comparing their performance in terms of evasiveness and realism, as well as scalability and cost-effectiveness

- integrating the test set generation method into the SUPERVIZ evaluation method

Research environment

The work is performed in the context of the SUPERVIZ project funded by France 2030, which gathers a community around the topics of supervision, including detection, response and validation. In particular, the offered position is supported by Work Package 5 which deals with the assessment and validation of NIDS.

Supervising team. SAMOVAR has a long expertise in intrusion detection in various environments as well as in security assessment. SCN team has collaborated on such topics in European projects. Additionally, in GRIFIN, the postdoctoral researcher will have opportunities to collaborate with esteemed academics in the UK (UCL, MDX) as well as integrate the results with dynamic SMEs (Montimage).

Application

Applications are to be sent before October 2023 to the position supervisor, or through the Recruitee portal. The application package comprises:

- an up-to-date resume
- a copy of the Ph.D diploma
- recommendation letters or a list of references

Required qualifications. Candidates fit the following profiles:

- Ph.D in computer science
- Experience with ML/DL libraries
- Knowledge of adversarial machine learning
- Proficiency in both written and spoken English

References

[1] S. Ayoubi, G. Blanc, H. Jmila, T. Silverston, and S. Tixeuil. Data-driven evaluation of intrusion detectors: A methodological framework. In G.-V. Jourdan, L. Mounier, C. Adams, F. Sèdes, and J. Garcia-Alfaro, editors, *Foundations and Practice of Security*, pages 142–157, Cham, 2023. Springer Nature Switzerland.

[2] B. Biggio and F. Roli. Wild patterns: Ten years after the rise of adversarial machine learning. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 2154–2156, 2018.

[3] D. Han, Z. Wang, Y. Zhong, W. Chen, J. Yang, S. Lu, X. Shi, and X. Yin. Evaluating and improving adversarial robustness of machine learning-based network intrusion detectors. *IEEE Journal on Selected Areas in Communications*, 39(8):2632–2647, 2021.

[4] M. A. Merzouk, F. Cuppens, N. Boulahia-Cuppens, and R. Yaich. Investigating the practicality of adversarial evasion attacks on network intrusion detection. *Annals of Telecommunications*, pages 1–13, 2022.

[5] F. Pierazzi, F. Pendlebury, J. Cortellazzi, and L. Cavallaro. Intriguing properties of adversarial ml attacks in the problem space. In *2020 IEEE symposium on security and privacy (SP)*, pages 1332–1349. IEEE, 2020.

[6] A. Schoen, G. Blanc, P.-F. Gimenez, Y. Han, F. Majorczyk, and L. Mé. Towards generic quality assessment of synthetic traffic for evaluating intrusion detection systems. In *RESSI 2022-Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information*, 2022.

[7] J. Vitorino, N. Oliveira, and I. Praça. Adaptative perturbation patterns: realistic adversarial learning for robust intrusion detection. *Future Internet*, 14(4):108, 2022.

[8] C. V. Wright, S. E. Coull, and F. Monrose. Traffic morphing: An efficient defense against statistical traffic analysis. In *NDSS*, volume 9, 2009.