SUPERVIZ – Postdoctoral position on Explaining IDS Decisions through Visualisations

Gregory Blanc
SAMOVAR, Télécom SudParis
Institut Polytechnique de Paris

start date: Autumn 2023

Context

Intrusion Detection Systems (IDS) contribute to the protection of information systems and organisations. When security operations, including detection, are precisely adapted to the needs of the target entity, it helps prevent severe security incidents [1]. In a security operations centre (SOC), security analysts perform incident analysis and response with the support of appropriate visualisations [9]. Yet, IDS are dependent on event collection, especially when they are behaviour-based. As a matter of fact, research on data-driven intrusion detections traces its roots back to Lee and Stolfo's data mining approach [7]. With the recent progress in artificial intelligence research, machine learning (ML) and deep learning (DL) methods have been massively exploited to perform anomaly-based intrusion detection in complex and dynamic networks [3]. Although anomaly-based techniques theretically enable the detection of previously unseen attacks, they are particularly sensitive to (small) changes in the underlying data distribution [4]. Additionally, deep-learning-based anomaly detectors are often seen as blackboxes for the lack of interpretability of their decisions [10], further eroding trust in data-driven IDS.

Recently, explainability has been proposed as a means to increase the interpretability of deep learning methods [6] and often offer visual interpretations to human operators. But existing post-hoc methods, i.e. that explain the decisions of blackbox models, have been developed for image or text data, and is often inadapted for cybersecurity data [5].

Objectives.

This postdoctoral project proposes to generate new visualisations based of data-driven IDS alerts and explanations to support the operation of SOC, and improve the efficiency of monitoring sophisticated malicious campaigns. To that end, several technologies could be exploited or improved including causal relationships [11], knowledge graphs [8]. To further support attack investigation, provenance graphs may also be leveraged [2].

Activities. The project contributes to better supporting monitoring and attack investigation:

- surveying the state of the art of alert correlation and causality detection
- surveying visualisation, graphical and explanation methods to investigate not only decisions but also input data and model hyperparameters for deep-learning-based IDS

- proposing explanation-enhanced security visualisations and graphs to support human operators
- integrating visualisations and explanations into an IDS evaluation framework

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 4 which deals with the robustness and efficiency of security monitoring.

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 explainable artificial intelligence
- Proficiency in both written and spoken English

References

[1] ANSSI. Security incident detection service providers – Requirements reference document. `https://www.ssi.gouv.fr/uploads/2014/12/pdis_referentiel_v2.0_en.pdf`, Dec. 2017. Version 2.0.

[2] G. Brogi and V. V. T. Tong. Terminaptor: Highlighting advanced persistent threats through information flow tracking. In *2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5. IEEE, 2016.

[3] A. L. Buczak and E. Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*, 18(2):1153–1176, 2015.

[4] R. Chalapathy, N. L. D. Khoa, and S. Chawla. Robust deep learning methods for anomaly detection. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 3507–3508, New York, NY, USA, 2020. Association for Computing Machinery.

[5] F. Charmet, H. C. Tanuwidjaja, S. Ayoubi, P.-F. Gimenez, Y. Han, H. Jmila, G. Blanc, T. Takahashi, and Z. Zhang. Explainable artificial intelligence for cybersecurity: a literature survey. *Annals of Telecommunications*, pages 1–24, 2022.

[6] D. Gunning and D. Aha. Darpa's explainable artificial intelligence (xai) program. *AI magazine*, 40(2):44–58, 2019.

[7] W. Lee and S. J. Stolfo. Data mining approaches for intrusion detection. In A. D. Rubin, editor, *Proceedings of the 7th USENIX Security Symposium, San Antonio, TX, USA, January 26-29, 1998*. USENIX Association, 1998.

[8] L. Leichtnam, E. Totel, N. Prigent, and L. Mé. Sec2graph: Network attack detection based on novelty detection on graph structured data. In *Detection of Intrusions and Malware, and Vulnerability Assessment: 17th International Conference, DIMVA 2020, Lisbon, Portugal, June 24–26, 2020, Proceedings 17*, pages 238–258. Springer, 2020.

[9] P. Limousin, R. Azzabi, L.-P. Bergé, H. Dubois, S. Truptil, and L. Le Gall. How to build dashboards for collecting and sharing relevant informations to the strategic level of crisis management: an industrial use case. In *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pages 1–8. IEEE, 2019.

[10] S. Seng, J. Garcia-Alfaro, and Y. Laarouchi. Why anomaly-based intrusion detection systems have not yet conquered the industrial market? In *International Symposium on Foundations and Practice of Security*, pages 341–354. Springer, 2021.

[11] C. Xosanavongsa, E. Totel, and O. Bettan. Discovering correlations: A formal definition of causal dependency among heterogeneous events. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 340–355. IEEE, 2019.