Internship title: Graph modeling of log files for attack detections in IoT applications

Reference: CyberSecGraph (to be quoted in all correspondence)

Location: ERIC Laboratory, Porte des Alpes Campus, Bron. Tel: 04 78 77 31 54

Internship supervisors (Emails):

- Mohamed-Lamine Messai (mohamed-lamine.messai@univ-lyon2.fr)
- Hamida Seba (hamida.seba@univ-lyon1.fr)
- Nouria Harbi (nouria.harbi@univ-lyon2.fr)
- Fadila Bentayeb (fadila.bentayeb@univ-lyon2.fr)

Key words : Anomaly detection, security, IoT, graphs

Internship for a Master 2 student (or equivalent)

Duration: 5-6 months. The gratification will be around 550 €/month.

Desired period: from February 2022

Computer security is an important issue in a context where more and more devices (objects) are connected to the Internet. These connected objects form an Internet of Things (IoT) Network. In this context, there are several cybersecurity challenges and barriers to overcome: attack detection, big data analysis, extraction and learning from multiple and heterogeneous sources, ..., etc.

In general, the cybersecurity of computer networks is ensured by various tools such as intrusion prevention systems (IPS), intrusion detection systems (IDS), antivirus, anti-trojan as well as SIEM (Security Information and Event Management) systems used to collect, analyze, monitor data related to the security of systems. A weakness of these protection tools is that, although they detect a malicious action or an attack, they do not have the capacity to link it to another detected attack and therefore potentially miss a higher-level attack. In an IoT context, a cybercriminal who tries to compromise connected objects will interact with the target objects from the recognition phase until the exploit phase. These interactions are recorded in log files. Log files contain information about almost all events that occur in a system, depending on the level of

logging. For this, the deployed logging infrastructure collects, aggregates, and automatically stores the log files continuously produced by connected objects [1, 2]. The volume of trace files collected can quickly reach a large size (this depends on the collection periodicity and the number of connected objects of the IoT network). The storage and processing of this large amount of data is made possible with tools from Hadoop ecosystem (HDFS for distributed storage and solutions for distributed processing such as Spark, Pig, Giraph,, etc). The use of graph databases to represent and analyze the security of networks in general and IoT networks in particular has gained increasing attention in research in recent years [3, 4]. The use of graph-oriented databases is a promising solution that can improve the work of SOC (Security Operation Centers) analysts who wish to have tools that make the links between the different events in log files. Graph-oriented databases allow correlation between all malicious actions or attacks, using a graphic representation of the links between these malicious actions [5, 6].

The main objective of this internship is to model log files in a graph database in order to analyze this data to detect cyber-attacks. Achieving this goal requires accomplishing the following points :

- Modeling log files of a dataset using a graph-oriented database (under Neo4j for example).
- Identify and study the various cyber threats that can be detected by analyzing log files.
- Analysis of the database for identified attacks. Machine learning on a dataset for the early detection of a targeted attack (eg. DDoS).

To apply : The candidate must have advanced skills (M2 level) in computer science (data science, machine learning and notions of graph theory and computer security are highly desirable). Please send, before 31/01/2022, your application with a CV, a cover letter, as well as your grades for the current academic year and last year to mohamed-lamine.messai@univ-lyon2.fr

References

[1] Landauer, M., Skopik, F., Wurzenberger, M., & Rauber, A. (2020). System log clustering approaches for cyber security applications: A survey. *Computers & Security*, *92*, 101739.

[2] Wu, P., Lu, Z., Zhou, Q., Lei, Z., Li, X., Qiu, M., & Hung, P. C. (2019). Big data logs analysis based on seq2seq networks for cognitive Internet of Things. *Future Generation Computer Systems*, *90*, 477-488.

[3] Noura, H. N., Salman, O., Chehab, A., & Couturier, R. (2020). DistLog: A distributed logging scheme for IoT forensics. *Ad Hoc Networks*, *98*, 102061.

[4] Jia, Y., Xiao, Y., Yu, J., Cheng, X., Liang, Z., & Wan, Z. (2018, April). A novel graph-based mechanism for identifying traffic vulnerabilities in smart home IoT. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. pp. 1493-1501.

[5] Hofer, D., Jäger, M., Mohamed, A., & Küng, J. (2020, November). On Applying Graph Database Time Models for Security Log Analysis. In *International Conference on Future Data and Security Engineering* (pp. 87-107). Springer, Cham.

[6] Schindler, T. (2018). Anomaly detection in log data using graph databases and machine learning to defend advanced persistent threats. *arXiv preprint arXiv:1802.00259*.