Stage de Master 2 (5-6 mois)

Référence : CyberSecGraph (à rappeler dans toute correspondance)

Lieu : Laboratoire ERIC, Campus Porte des Alpes, Bron. **Tél :** 04 78 77 31 54

Responsables du stage (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)

Thématiques : Détection d'anomalies, sécurité, IoT, graphes

Type de stage : Fin d'études bac +5, Master 2

Durée : 5-6 mois

Période souhaitée : à partir de février 2022

Rémunération : Indemnités de stage légales (environ 550 € par mois)

Intitulé : Modélisation de fichiers de traces en graphes de données pour la détection d'attaques dans les applications IoT

Sujet : La sécurité informatique est aujourd'hui un sujet que nous ne pouvons laisser de côté. Les menaces se multiplient, les attaques évoluent notamment avec le déploiement massif d'objets connectés à Internet ce qui élargit la surface de vulnérabilité. De plus, d'autres préoccupations liées au piratage des objets connectés et aux craintes relatives à la vie privée ont attiré l'attention des chercheurs. Afin que les réseaux IoT arrivent à leur maturité, il existe plusieurs défis et barrières liés à la cybersécurité à lever : la détection d'attaques, l'analyse de données volumineuses, extraction et apprentissage à partir de sources multiples et hétérogènes, ..., etc.

D'une manière générale, la cybersécurité des réseaux informatique est assurée par différent outils tels que des systèmes de prévention d'intrusion (IPS), des systèmes de détection d'intrusion (IDS), des antivirus, des anti-trojan ainsi que des systèmes SIEM (Security Information and Event Management) utilisé pour collecter, analyser, surveiller des données concernant la sécurité des systèmes d'information. Une faiblesse de ces

outils de protection réside dans le fait que bien qu'ils détectent une action malicieuse ou une attaque, ils n'ont pas la capacité à la relier à une autre attaque détectée et donc ratent potentiellement une attaque de plus haut niveau.

Dans un contexte IoT, un cybrcriminel qui essaye de compromettre des objets connectés va interagir avec les objets cibles de la phase de reconnaissance jusqu'à la phase de l'exploit. Ces interactions sont enregistrées dans des fichiers de traces. Les fichiers de traces (journaux ou logs) contiennent des informations sur presque tous les événements qui se produisent dans un système, en fonction du niveau de journalisation. Pour cela, l'infrastructure de journalisation déployée collecte, agrège et stocke automatiquement les fichiers de traces produits en continu par les objets connectés [1, 2]. Le volume de fichiers traces collectés peut rapidement atteindre une grande taille (cela dépend de la périodicité de collecte et du nombre d'objets connectés dans le réseau IoT). Le stockage et le traitement de cette grande quantité de données est rendu possible avec des composants de Hadoop (HDFS pour le stockage distribué et des solutions pour le traitement distribué telles que Spark, Pig, Giraph, …, etc).

L'utilisation de base de données graphes pour représenter et analyser la sécurité des réseaux en générale et des réseaux IoT en particulier a gagné une attention croissante dans la recherche au cours des dernières années [3, 4]. L'utilisation de bases de données orientées graphes est une solution prometteuse qui peut améliorer le travail d'analystes SOC (*Security Operation Centers*) qui souhaitent avoir des outils qui font les liens entre les différents évènements sur des fichiers de traces. Les bases de données orientées graphes permettent de faire une corrélation entre toutes les actions malicieuses ou attaques, en utilisant une représentation graphique des liens entre ces actions malicieuses. Récemment, les systèmes de détection de botnet (réseaux de machines zombies où des ordinateurs contrôlés à l'insu de leurs utilisateurs par un cybercriminel) qui exploitent l'analyse des graphes de communication à l'aide de l'apprentissage automatique ont attiré l'attention pour faire face à la propagation de botnets. Des approches de modélisation et d'exploration de données basées sur des graphes ont été proposées et fournissent des résultats intéressants [5, 6].

L'objectif principal de ce stage est de modéliser les fichiers de traces sous forme d'une base de données graphes pour procéder à l'analyse de ces données afin de détecter des attaques cyber.

Atteindre cet objectif passe par l'accomplissement des points suivants :

- Modélisation de fichiers de traces d'un jeu de données par une base de données orientée graphes (sous Neo4j par exemple).
- Identifier et étudier les différentes cybermenaces qui peuvent être détectées par l'analyse de fichiers de traces.
- Analyse de la base de données pour les attaques identifiées.
- Faire de l'apprentissage automatique sur un jeu de données pour la détection en amont d'une attaque ciblée (ex. DDoS).

Profil du/de la stagiaire : Compétences avancées (niveau M2) en informatique (science de données, machine learning et notions de théorie des graphes et de sécurité informatique fortement souhaitées).

- **Merci d'adresser, avant le 31 Janvier 2022, votre candidature avec un CV, une lettre de motivation, ainsi que vos notes de l'année universitaire en cours et de l'année dernière à** mohamed-lamine.messai@univ-lyon2.fr

Références

[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*.