Titre du sujet : Un système d'aide à la décision pour l'optimisation des algorithmes d'IA dans l'analyse de cybersécurité à partir des Logs

Nom du directeur de thèse : NADIRA LAMMARI

Email du directeur de thèse : lammari@cnam.fr

Nom du co-encadrant : VERONIQUE LE GRAND

Email du co-encadrant : veronique.legrand@cnam.fr

Description du sujet de thèse :

Un système d'aide à la décision pour le choix et le paramétrage d'algorithmes de l'intelligence artificielle à des fins d'analyses complexes de la cybersécurité à partir de masses de logs

Contexte :

Dans le domaine de l'investigation numérique, se pose, à l'heure actuelle, la nécessité de réduire le nombre de faux-positifs et de faux-négatifs qui se produisent lors des analyses de sécurité complexes utilisant les journaux d'activité des systèmes d'information. Les investigateurs ne peuvent expliquer avec précision et pertinence les causes d'un incident de sécurité. L'explicabilité d'un incident de sécurité se trouve néanmoins à la croisée de plusieurs phénomènes : (a) La génération massive d'observations en situation, (b) la variabilité des comportements malveillants et (c) le développement important des outils d'analyse de sécurité à partir du machine learning.

Plusieurs difficultés techniques constituent un frein à l'explicabilité d'un incident de sécurité. Tout d'abord, l'analyse se fait au travers d'observations qui, à l'heure actuelle, sont de nature big-data, et, rendent particulièrement confuse l'analyse de l'expert de sécurité. Ensuite, la variabilité des attaquants entraine un besoin d'adapter les analyses des experts de sécurité aux situations, ce qui rend inopérant les modèles à base de statistiques.

L'utilisation de l'intelligence artificielle et en particulier du machine learning permet d'accompagner l'humain dans le traitement des big-data. On assiste actuellement à un déploiement très important de tels algorithmes, où chacun est adapté à des situations d'analyse spécifiques. Cependant, la performance de ces algorithmes repose sur leur paramétrage, ce qui requiert de fortes connaissances du domaine du machine learning pour les analystes de sécurité. Les méthodes permettant d'analyser ces différents algorithmes ainsi que les critères pour les adapter et les appliquer sont, de nos jours, peu explorés.

Objectif de la thèse :

L'objectif de cette thèse sera de mettre en place un état de l'art sur les rôles, usages et applicabilités de ces différents modèles d'algorithmes. Il s'agira notamment de dégager un système d'aide à la décision pour le choix et le paramétrage d'algorithmes de l'intelligence artificielle à des fins d'analyses complexes de la cybersécurité à partir de masses de logs. L'enjeu sera d'adapter les algorithmes de l'IA et d'en comprendre la finesse des paramétrages afin d'optimiser pleinement leur utilisation pour les analyses de la cybersécurité.

Ce système d'aide à la décision devra être intégré dans la plateforme Vulneris mise en œuvre dans le cadre du projet Fui HuMa.

Références bibliographiques

[1] Nadira Lammari, Véronique Legrand, Gloria Elena Jaramillo Rojas, Olfa Atig : An Ontology for cyber incident root cause analysis from event logs. Papier accepté à IBIMA 2018

[2] J. Navarro, V. Legrand, S. Lagraa, J. François, A. Lahmadi, G. De Santis, O. Festor, N. Lammari, F. Hamdi, A. Deruyver, Q. Goux, M. Allard, P. Parrend. "HuMa: A Multi-layer Framework for Threat Analysis in a Heterogeneous Log Environment", International Symposium on Foundations and Practice of Security, October 2017, Vol. LNCS, volume 10723, pp.144-159, Nancy, France, (DOI: https://doi.org/10.1007/978-3-319-75650-9_10)

[3] A. Chellam, L. Ramanathan, Ramani. S. Intrusion detection in computer networks using lazy learning algorithm. Procedia Computer Science, Volume (132), 2018. https://doi.org/10.1016/j.procs.2018.05.108

[4] A. Verma and V. Ranga. Statistical analysis of cidds-001 dataset for network intrusion detection systems using distance-based machine learning. Procedia Computer Science, Volume (125), 2018. https://doi.org/10.1016/j.procs.2017.12.09

[5] S. Soheily-Khah, P. Marteau, N. Béchet. Intrusion detection in network systems through hybrid supervised and unsupervised machine learning process: A case study on the iscx dataset. ICDIS 2018. https://doi.org/10.1109/ICDIS.2018.00043

[6] M. Aamir and S. M. A. Zaidi. Clustering based semi-supervised machine learning for DDoS attack classification. Journal of King Saud University -Computer and Information Sciences, 2019. https://doi.org/10.1016/j.jksuci.2019.02.003

[7] Y. Gu, Y. Wang, Z. Yang, F. Xiong, Y. Gao. Multiple-features-based semi supervised clustering DDoS detection method. Mathematical Problems in Engineering Journal, Volume 2017, 2017. https://doi.org/10.1155/2017/5202836

[8] M. Idhammad, K. Afdel, and M. Belouch. Semi-supervised machine learning approach for DDoS detection. Applied Intelligence, Volume 48(10), 2018. https://doi.org/10.1007/s10489-018-1141-2

[9] S. Su. Topical behavior prediction from massive logs. Big Data 2017, 2017. https://doi.org/10.1109/BigData.2017.8258363

Titre :"Expérimentation, programmation, étude et formalisation de machines d'exécution pour des programmes SugarCubesJS "