Détection d'anomalies de sécurité par les graphes en environnement d'Industrie 4.0

Màj : 15/1/2019
Etablissement de rattachement : Université de Strasbourg
Candidat :
Directeur de thèse : Pierre Parrend (Enseignant-Chercheur, ECAM Strasbourg-Europe)
Ecole doctorale : Mathématiques, Sciences de l'Information et de l'Ingénieur - responsable Pr. T. Noel
Unité d'accueil : Laboratoire ICube (Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie ; UMR7357)

Contexte

La transformation numérique, loin d'être un phénomène passager, constitue une révolution technologique très forte, en particulier pour les entreprises industrielles. L'objectif de cette thèse est de créer un logiciel pérenne de détection d'anomalies de sécurité pour l'Industrie 4.0, et d'enrichir le modèle d'écosystème immunitaire artificiel développé au sein du laboratoire ICube en raffinant les travaux de détection de graphes d'anomalies pour leur donner une expressivité suffisante pour traiter les enjeux du projet. L'approche mise en œuvre intègrera les modèles de graphes de scénarios d'attaque abstraits créés par le laboratoire ICube. Elle aura pour objectif de formaliser les graphes d'attaques en intégrant le modèle des Stream Flows et de proposer des algorithmes de caractérisation et de détection d'anomalies exploitant ces graphes.

Objectifs

L'objectif du projet est de renforcer la continuité confidentialité/intégrité/disponibilité en identifiant des graphes d'anomalies au sein de logs systèmes. Ces anomalies peuvent être des anomalies de sécurité ou des défaillances systèmes.

Ce projet s'appuiera sur les Outils SimSC et Morwilog développés au sein de l'équipe CSTB du laboratoire ICube. Il s'agit d'extraire des traces informatiques (log) pour l'identification de scénarios d'usage (utilisateur; machine) à des fins d'analyse d'anomalies.

L'approche consiste à mettre en œuvre l'analyse de graphes complexes, par des algorithmes de recherche d'anomalies indépendants du cas d'application. Ces algorithmes seront des modèles 'white box' à forte sémantique, par opposition aux réseaux de neurones, qui fonctionnent en mode 'black box'. Selon l'avancement du projet, des collaborations internes au laboratoire peuvent être envisagés avec l'approche IA/EA (Intelligence artificielle/Evolution Artificielle) qui couple l'extraction de motifs par réseaux de neurones et la génération de solutions par évolution artificielle.

Le Livrable de la thèse comportera un ensemble d'algorithmes de détection d'anomalies et des bibliothèques logicielles de détection d'anomalies pour l'Usine du Futur.

Les cas d'applications sont:

- La cybersécurité pour l'usine connectée et les infrastructures critiques
- La détection d'anomalies de production
- La supervision de Cloud

Organisation des travaux

Les phases du projet seront : 1) identification de scénarios de défaillances ; 2) création d'un modèle de détection d'anomalies comportementales par approche stochastique ; 3) optimisation et l'évaluation de la contribution proposée.

Le projet de thèse débutera par l'identification de scénarios de défaillances au travers de logs systèmes, dans un environnement d'industrie 4.0. Il s'agit d'adapter le modèle SimSC/Morwilog pour l'extraction, c'est à dire l'identification et apprentissage, de ces scénarios de défaillances. Cette phase pourra durer un semestre.

La deuxième phase de la thèse consistera en la création d'un modèle de détection d'anomalies comportementales par approche stochastique. Il s'agit ici de formaliser les Graphes de Scénarios d'Attaques Abstraits (AASG), par exemple sur la base de la formalisation des Stream Flows, et de définir différentes métriques pertinentes (critères) de distance afin d'identifier les scénarios d'usage anormaux, représentés sous forme de graphes de logs, par anomalie globale puis anomalie locale. Des algorithmes génétiques pour la détection d'anomalies par minimisation multicritère des fonctions de distances définies précédemment seront proposés en privilégiant les algorithmes parallèlisables. L'exécution des algorithmes de détection en environnement distribué sur GPGPU sera envisagée si cette approche est pertinente dans le cadre de la contribution. Cette phase pourra durer 18 mois.

La troisième phase de la thèse consistera en l'optimisation et l'évaluation de la contribution proposée, ainsi qu'en le raffinement des modèles : nouveaux modèles comportementaux ; amélioration des stratégies de détection d'anomalies par algorithmes génétiques. L'évaluation inclue la finalisation de l'architecture à plugin ; un audit de code ; un audit de sécurité ; le déploiement dans un deuxième environnement de test (partenaire industriel, partenaire académique). Cette troisième phase est planifiée sur la troisième année de thèse.

Un semestre est planifié pour la finalisation et la rédaction de la thèse de doctorat.

Du temps sera conservé pour la participation à des projets de R&D partenariaux dans le cadre du déploiement de la plate-forme 'Usine du Futur' de l'ECAM Strasbourg-Europe. Une phase sera dédié au déploiement de l'environnement de test 'IT 4.0' : Infrastructure IT et sa connexion au système de production du plateau technique 'usine 4.0' ; Exécution de cas d'anomalies représentatifs

Collaborations

Des collaborations pourront être envisagées selon l'avancement du projet avec l'Unitwin UNESCO « Campus Numérique des Systèmes Complexes » et les partenaires de la communauté Artificial Life and Robotics (Universités de Hiroshima, Tokyo, Beppu, Japon) pour l'étude de l'émergence, le Pôle sécurité Défense et la Chaire de Sécurité des Systèmes d'Information du Conservatoire National des Arts et Métiers pour l'étude des enjeux de cybersécurité, l'équipe Complex Networks du LIP6 pour la modélisation des Stream graphs.

Références

Latapy, M., Viard, T., & Magnien, C. (2018). Stream graphs and link streams for the modeling of interactions over time. Social Network Analysis and Mining, 8(1), 61.
J. Navarro, A. Deruyver, P. Parrend, A Systematic Survey on Multi-step Attack Detection, Computers and Security, Elsevier (IF : 2.849, SNIP : 2.217, SJR : 0.866), page 102, 2018
P. Parrend, F. Guigou, J. Navarro, A. Deruyver, P. Collet, Artificial Immune Ecosystems: the role of expert-based learning in artificial cognition, Journal of Robotics, Networking and Artificial Life, Atlantis Press, page 5, mars 2018

P. Parrend, P. David, F. Guigou, C. Pupka, P. Collet, The AWA Artificial emergent aWareness Architecture model for Artificial Immune Ecosystems, IEEE Congress on Evolutionary Computation 2017, Special Session on Artificial Immune Systems: Algorithms, Simulation, Modelling & Theory, San Sebastian, Spain, juin 2017

F. Guigou, P. Collet, P. Parrend, The Artificial Immune Ecosystem: a bio-inspired meta-algorithm for boosting time series anomaly detection with expert input, EvoApplications, 20th European Conference on the Applications of Evolutionary Computation, Amsterdam, Netherlands, avril 2017

F. Guigou, P. Parrend, P. Collet, An artificial immune ecosystem model for hybrid cloud supervision, Complex System Digital Campus'15, Tempe, AZ, United States, septembre 2015

J. Navarro, A. Deruyver, P. Parrend, Morwilog: an ACO-based System for Outlining Multi-Step Attacks, IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2016), Athènes, Greece, décembre 2016