Aix*Marseille université
Collège doctoral

DE LA RECHERCHE À L'INDUSTRIE
cea
CADARACHE

Titre de la thèse
Vers une approche Big Data orientée processus pour la détection, la prévention et la gestion des cyberattaques

Mots-clés
Cybersécurité, Big Data, Process Mining, Apprentissage, SIEM/ICS

Equipe d'accueil AMU	***Equipe d'accueil CEA Cadarache***
Directeur de thèse : Omar BOUCELMA	Co-directeur de thèse : Cédric COCQUEBERT
ED de rattachement : 184, Mathématiques et Informatique	ED de rattachement :
Laboratoire : LSIS, UMR CNRS 7296	Dpt/Service/Laboratoire : DTSG/STIC/GI

Descriptif du sujet

L'objectif de la thèse est de proposer de nouveaux algorithmes et systèmes pour détecter, identifier et contrôler les cyberattaques. Ces algorithmes seront intégrés aux SIEM existants (Security Information and Event Management) ou alors dans des nouveaux systèmes en support aux activités des centres opérationnels de sécurité (SOC).

Le sujet s'attaque aux chantiers suivants : la collecte, la préparation des données qualifiées avec notamment la traçabilité, la modélisation de haut niveau des attaques, et le traitement des données avec des approches Big Data : stockage des données pertinentes, traitement de données en flux pour la détection en temps réel, apprentissage sur les scénarii et les données d'attaques, etc.

Concernant les attaques, au-delà de la prise en compte d'une typologie classique, nous proposons une approche où les attaques seront représentées comme des processus métiers, avec un modèle et un langage standard de haut niveau. Ainsi représentées, les attaques pourront donc être stockées dans une base de données, enrichir la connaissance métier des acteurs concernés, notamment par des méthodes de fouille (process mining) , et des simulations exécutables générées automatiquement sur des plateformes de test (hors SIEM en production). D'un point de vue stratégique, alors que certains travaux [1] et initiatives (www.enisa.europa.eu) plaident pour un échange d'informations dans le domaine, cette approche offre la possibilité de ne divulguer que ce qui est nécessaire à des partenaires (processus public/privé).

Le traitement des données soulève plusieurs questions : (1) l'intégration de plusieurs sources de données hétérogènes [2], (2) l'assurance de l'intégrité des données, et (3) la fouille de données massives. Concernant le point (1) nous comptons nous appuyer sur des méthodes et techniques développées dans le domaine de la gestion de données. Le point (2) peut amener à des solutions innovantes via la gestion de la Provenance (traçabilité) avec des modèles de contrôles d'accès [3]. Avec la multiplication des systèmes distribués, la provenance connaît un regain d'intérêt : par exemple, pour évaluer les risques de sécurité inhérents à ces systèmes [4], ou pour assurer l'intégrité des données [5], avec en ligne de mire, une évolution pour développer des systèmes « secure by design ». Enfin, le développement de méthodes de fouille de données mais aussi de processus [6] et de techniques d'apprentissage adaptées, permettra de détecter les cyberattaques mais pourra aussi aider à la mise en place des solutions d'aide à la décision pour les SOC.

Références

[1] Oscar Serrano, Luc Dandurand, Sarah Brown : On the Design of a Cyber Security Data Sharing System, WISCS 2014:

[2] C. Perez, P. Baldit, C. Cocquebert, J-P. Guichard : Countering the cyber threats: from behavioral analysis to dedicated Security Operations Center for industrial systems. AIEA Conf 2016

[3] Julien Lacroix, Omar Boucelma: Design and Implementation of a Trust Service for the Cloud. OTM Conferences 2015: 620-638.

[4] Mohsen Rezvani, Aleksandar Ignjatovic, Elisa Bertino, Sanjay Jha:
Provenance-aware security risk analysis for hosts and network flows. NOMS 2014: 1-8

[5] Wai-Kit Sze, R. Sekar: Provenance-based Integrity Protection for Windows. ACSAC 2015: 211-220

[6] Rafael Accorsi, Thomas Stocker : On the exploitation of process mining for security audits: the conformance checking case. ACM SAC 2012: 1709-1716.