الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٠80 ماي 5491٠ بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

MEMOIRE

Pour l'obtention du diplôme **d'ingénieur d'état**
Filière: **Informatique**
Spécialité:: **Système d'Information et Web (SIW)**

Thème

Anomaly Detection using AutoEncoders: The advanced Persistent Threats case

---

Présenté par:
BOUDOUARA Nadjet
LAIB Oumaima

Soutenu le 04 Juillet 2022 Devant le jury composé de:

Dr. KECHAR Mohamed	President
Dr. BENABDERRAHMANE Sid Ahmed	Encadreur
Pr. BENSLIMANE Sidi Mohamed	Co-Encadreur
Dr. KHALDI Miloud	Examinateur

Année Universitaire : 2021/2022

Abstract

Advanced persistent threats (APTs) are long-lasting cyber-attacks aiming at stealing vital data from target businesses. According to security experts, blocking all APTs is impossible, which emphasizes the necessity of early identification and damage limitation studies. Provenance trace mining and whole-system provenance tracking are deemed promising because they can aid in the discovery of causal linkages between activities and the detection of suspicious event sequences as they occur.

We provide here, a machine learning-based solution, that combines between the Auto-Encoders together with the Rule mining, for detecting genuine APT-like cyber attacks using provenance traces that use OS-independent indicators representing process activity.

The APTs and their anomaly scores in the models which were learnt from traces are used to rank anomalous processes. The results are subsequently provided as implications, which would help leveraging the causality for explaining the identified anomalies. Our proposed models were very competitive compared to existing approaches when tested on Transparent Computing program datasets (DARPA).

Keywords: Anomaly detection, Machine learning, Deep learning, web and cyber security, Advanced persistent threats , Big Data, web technologies,Autoencoder.

Résumé

Les menaces persistantes avancées (APT) sont des cyberattaques de longue durée visant à voler des données vitales aux entreprises cibles. Selon les experts en sécurité, le blocage de tous les APT est impossible, ce qui souligne la nécessité d'études précoces d'identification et de limitation des dommages. L'extraction de traces de provenance et le suivi de provenance de l'ensemble du système sont jugés prometteurs car ils peuvent aider à découvrir des liens de causalité entre les activités et à détecter des séquences d'événements suspects au fur et à mesure qu'ils se produisent.

Nous fournissons une technique non supervisée pour détecter les agressions authentiques de type APT à l'aide de traces de provenance qui utilisent des indicateurs indépendants du système d'exploitation représentant l'activité du processus.

Les vrais positifs et leurs scores dans tous les modèles appris à partir des traces sont utilisés pour classer les processus anormaux. Les résultats sont ensuite fournis sous forme d'implications, ce qui aide à tirer parti de la causalité pour expliquer les anomalies identifiées puisqu'elles sont interprétables. Notre stratégie a battu les approches concurrentes lorsqu'elle a été testée sur des ensembles de données du programme informatique transparent (DARPA).

Mot Clé : Détection d'anomalies, Machine learning, Deep learning, web et cybersécurité, Menaces persistantes avancées, Big Data, technologies web, Autoencodeur.