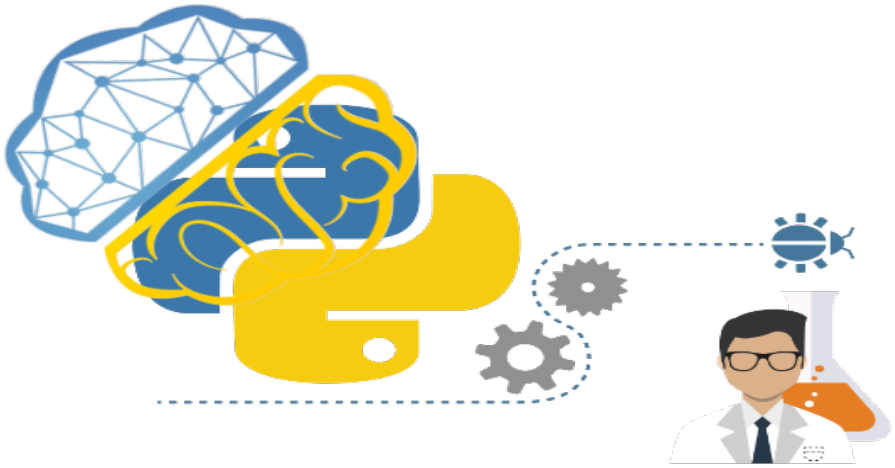
Université de Valenciennes et du Hainaut-Cambrésis

M2C - Malware Clustering et Classification



Vincent Romé - Axel Foulon - Julien Dupagny

— 9 Juin 2018 —

HISTORIQUE DES VERSIONS			
DATE	VERSION	ÉVOLUTION DU DOCUMENT	RÉDACTEUR
9/06/2017	0.1	Version préliminaire	Équipe complète

Table des matières

Introduction

Jusqu'à présent les systèmes de détection d'intrusion reposaient traditionnellement sur des signatures générées manuellement par des experts en sécurité, puis nous avons vu apparaître des systèmes permettant de détecter des patterns entre des jeux de données ceci à permis de générer automatiquement ces règles. Aujourd'hui avec l'apparition du "Big Data" des solutions comme le Deep Learning ou Machine learning sont souvent présentés comme les technologies pouvant révolutionner les systèmes de détections et les performances. Ces systèmes permettent de générer automatique un modèle de détection à partir de données et leurs capacité à généraliser les événements malveillants permettait en effet de détecter des éléments encore inconnus. L'objectif est d'ici comprendre le fonctionnement de ces algorithmes et de les appliquer au milieu de la sécurité informatique. D'exposer les résultats de notre recherche sur la détection de fichier PE (exécutable windows malveillants). Tout en gardant un regard critique sur les résultats et en essayant d'apporter des solution sur l'utilisation de tel algorithme en production.

Un modèle de détection supervisé est construit à partir de données labellisées fournies par l'expert : des événements bénins, mais aussi des événements malveillants pour guider le modèle de détection. L'algorithme d'apprentissage va automatiquement chercher les points permettant de caractériser chacune des classes ou de les discriminer pour construire le modèle de détection

Abstract

Contexte du projet