Journée des chercheurs CyberExcellence

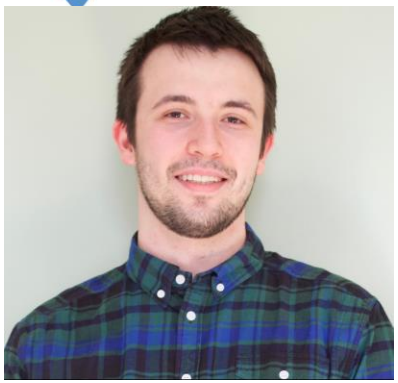
8 Novembre 2022

Charles-Henry Bertrand Van Ouytsel
Promoteur: Axel Legay

UCLouvain

icteam

Qui suis-je ?



Doctorant en 4ème année de thèse @UCLouvain
Twitter: @ChBertrandVo

Mon sujet : "Analyse et classification de malwares basé sur des méthodes de machine learning."

Centre d'intérêt : Analyse de malware, Machine Learning et exécution symbolique

UCLouvain

icteam

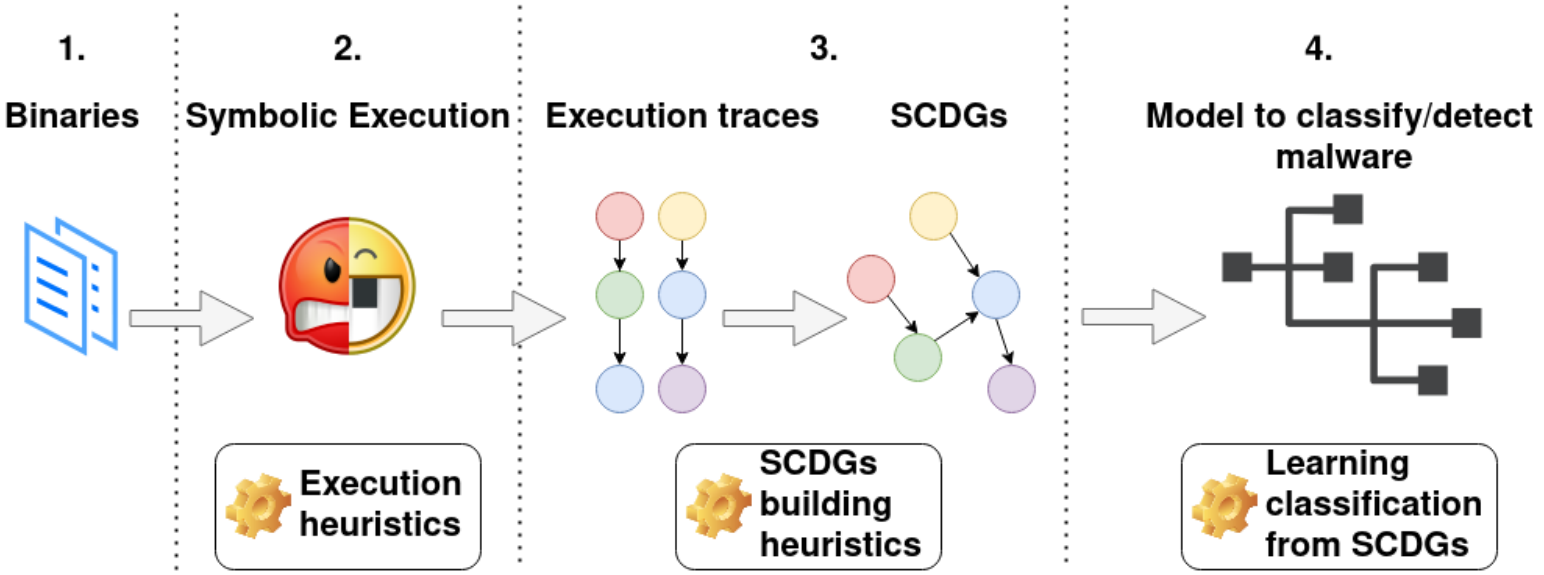
Différents projets

- SEMA : Symbolic Execution for Malware Analysis
- **Détection de packing par méthodes de machine learning**
- Federated Learning appliqué pour les classificateurs de malware

UCLouvain

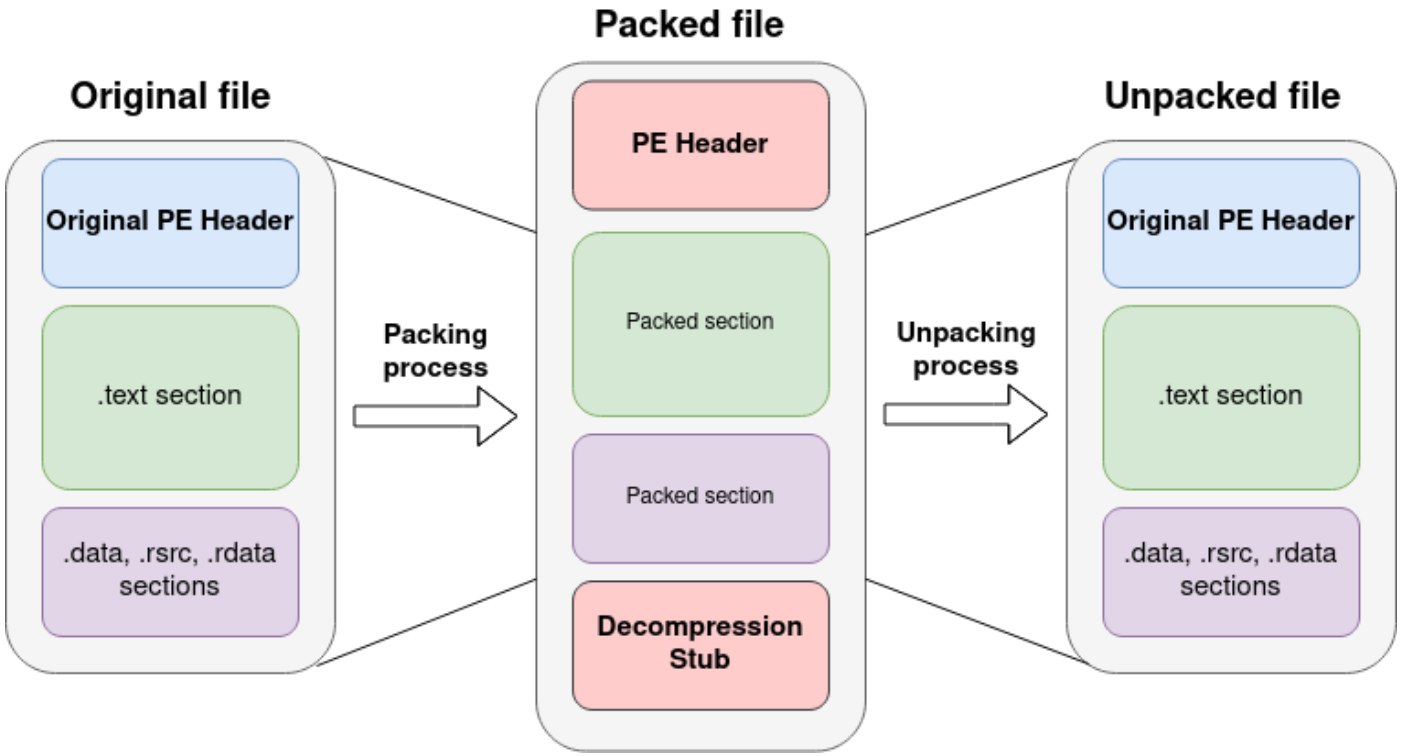
icteam

SEMA : Symbolic Execution for Malware Analysis



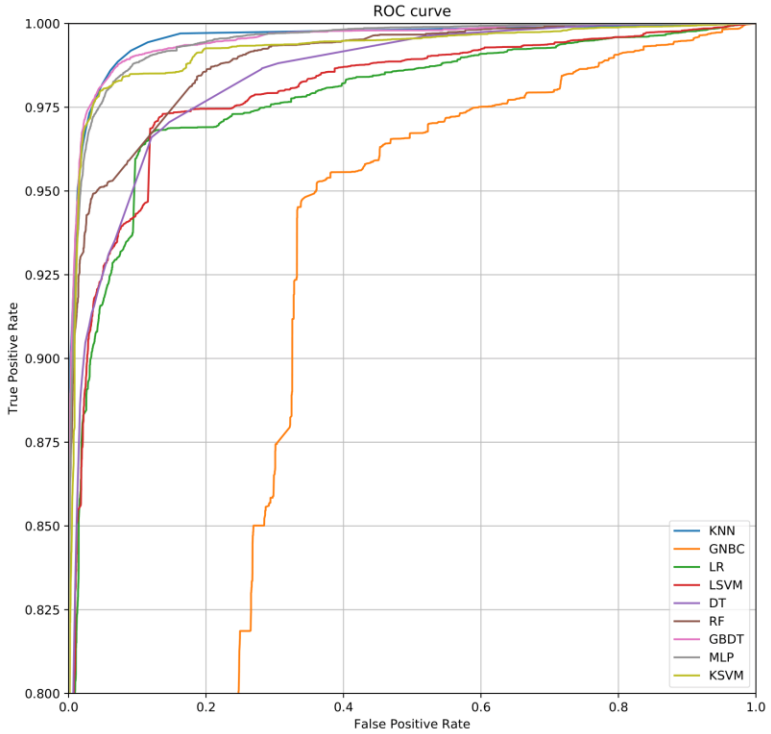
Packing : Qu'est-ce que c'est ?

Méthode d'**obfuscation** utilisées par les malware **mais aussi** par des logiciels légitimes pour protéger la propriété intellectuelle.



Détection grâce au Machine Learning

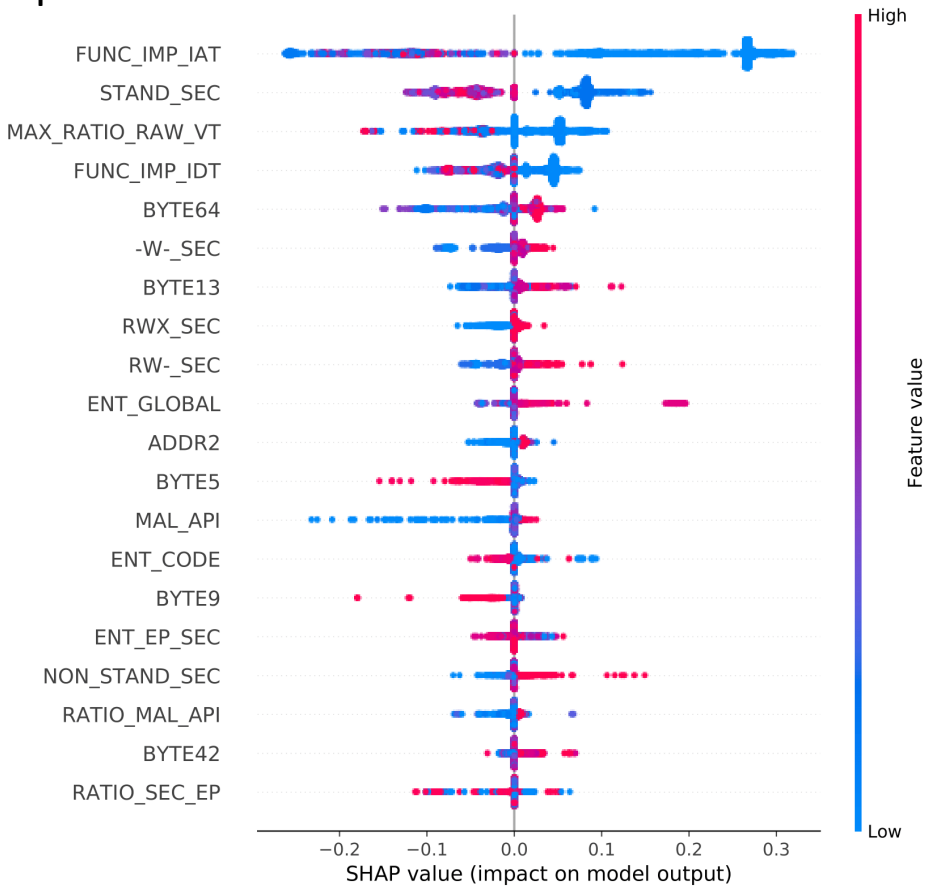
Utilisation de 119 features définies dans la littérature.
Elles sont relatives : aux sections, bytes, permissions, entropies, header,…
Différents modèles de machine learning investigués: KNN, arbres, SVM,...



Ref: BIONDI, Fabrizio, ENESCU, Michael A., GIVEN-WILSON, Thomas, *et al.* Effective, efficient, and robust packing detection and classification. *Computers & Security*, 2019, vol. 85, p. 436-451.

Interpréter les décisions des classificateurs

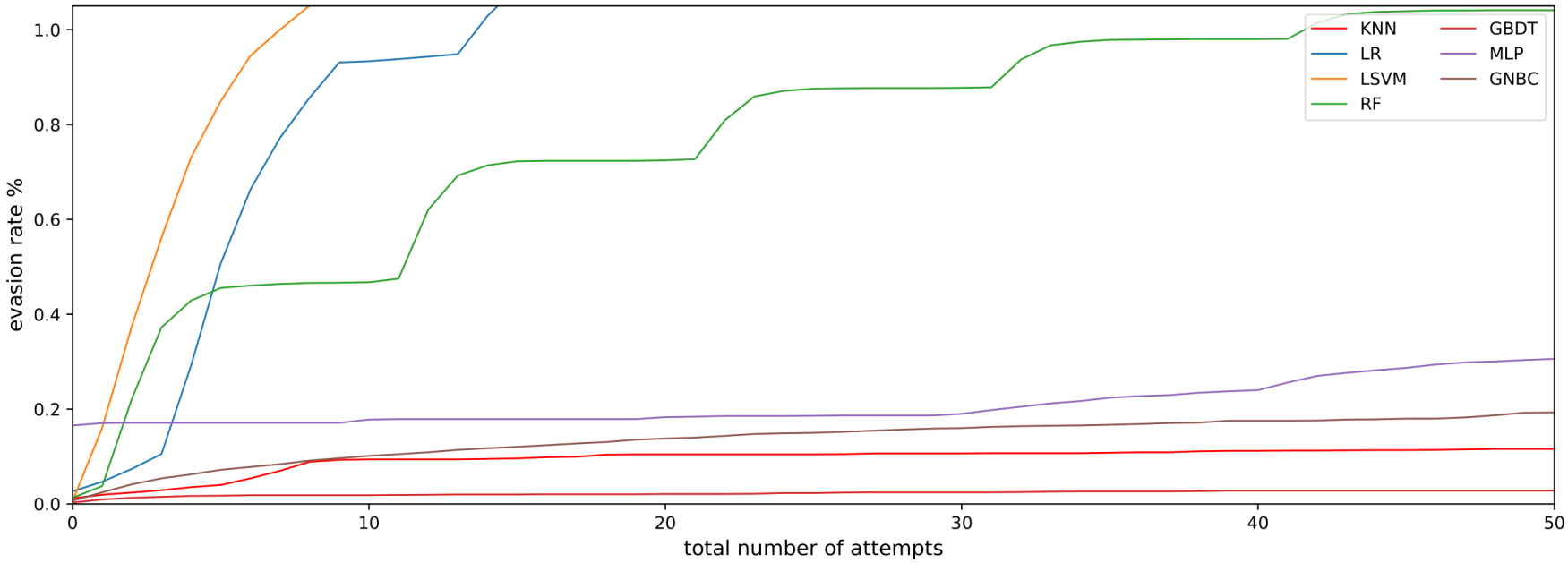
Utilisation de méthodes d'explicabilité (i.e: SHAP) pour comprendre les décisions et trouver d'éventuelles pistes à creuser.



Ref: SOTGIU, Angelo, PINTOR, Maura, et BIGGIO, Battista. Explainability-based Debugging of Machine Learning for Vulnerability Discovery. In : *Proceedings of the 17th International Conference on Availability, Reliability and Security*. 2022. p. 1-8.

Etudier les limites contre des adversaires

Utilisation d'un projet existant – MAB malware - pour muter des binaires et évader la détection des classificateurs.



Ref MAB-malware: SONG, Wei, LI, Xuezixiang, AFROZ, Sadia, *et al.* MAB-Malware: A Reinforcement Learning Framework for Blackbox Generation of Adversarial Malware. In : *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*. 2022. p. 990-1003.

Packing-Box

Problème récurrent dans le domaine : Répétabilité des expériences et Ground Truth

Notre proposition: un outil permettant de créer ces propres datasets et d'entrainer/tester directement des modèles

Alexandre D'Hondt, Charles-Henry Bertrand Van Ouytsel et Axel Legay

https://github.com/packing-box/docker-packing-box

Bientôt présentée à black hat

Et après ?

Différentes pistes pour le futur : Amélioration de SEMA, prédictions conformes, méthodes d'interprétabilités, nouveaux classificateurs, étendre le federated learning ...

Plusieurs TFEs en cours avec des étudiants: analyse de Macro Excel, extension de SEMA (visualisation, méthodes d'exploration,...), mutations de malware existants, extension de la packing-box...

Merci pour votre attention !

https://www.meetup.com/fr-FR/dcg3210/

SCAN ME