Proposition de stage recherche Master 2

Sujet : Approche hybride pour la détection de rançongiciels

Encadrants:
- Nga Nguyen, enseignant-chercheur, HDR, DVRC
- Christophe Rodrigues, enseignant-chercheur, DVRC
- En collaboration avec l'entreprise SitInCloud, spécialistes en Cybersécurité et IA

Possibilité de poursuivre en thèse doctorale avec l'entreprise SitInCloud, Devinci Research Center (DVRC) Paris La Défense et l'École Doctorale Sciences des Métiers de l'Ingénieur (SMI).

Veuillez envoyer votre candidature (un CV détaillé, une lettre de motivation et des bulletins de notes à nga.nguyen@devinci.fr et christophe.rodrigues@devinci.fr) .

Contexte :

L'utilisation de l'intelligence artificielle (IA) pour la cybersécurité ou plus spécifiquement pour la détection des logiciels malveillants est devenue un sujet de premier plan. Dans ce contexte, l'entreprise SitInCloud a développé Owlyshield, un des meilleurs modèles IA sur le marché pour la détection des rançongiciels, en disposant de plus de 115000 exécutables pour l'entraînement, un modèle avec plus de 3 millions(?)de paramètres et un taux de précision de 97%. Cependant, les logiciels malveillants ne cessent d'évoluer avec des techniques de plus en plus sophistiquées afin de déjouer les systèmes de détection.

L'objectif de ce stage est d'améliorer l'approche hybride qui combine l'analyse statique et l'analyse dynamique du code pour avoir des modèles plus fiables. L'analyse statique se base actuellement sur i) des caractéristiques extraites d'exécutables comme le nombre de sections, leurs tailles, les entropies, etc. et ii) des images qui représentent des exécutables. Des biais existent sur les images utilisées pour les réseaux neuronaux convolutifs (CNN) car les exécutables de maliciels sont généralement plus petits et compressés. Quant à l'analyse comportemental ou dynamique, nous souhaitons extraire plus d'informations des séries temporelles, ce qui permettra une meilleure analyse du temps d'exécution du code. Enfin, une pondération des décisions statiques et dynamiques devrait être faite en fonction de la quantité de données comportementales disponibles.

Mots-clés : Deep Learning, Cybersécurité, Python, Rust, Analyse de programme