Extraction automatique des TTP d'un code binaire d'un programme

Offre de thèse

Jean-Yves Marion
Jean-Yves.Marion@loria.fr
LORIA, Université de Lorraine

Dylan Marinho
Dylan.Marinho@lip6.fr
LIP6, Sorbonne Université

1. Contexte

Aujourd'hui, la détection des programmes malveillants est réalisée par des modèles neuronaux, suppléant les règles syntaxiques de type Yara. Si cette approche fonctionne correctement sur un périmètre assez large de menaces, la pression et les capacités offensives sont telles qu'il faut faire des avancées scientifiques pour dépasser le plafond de verre des défenses actuelles contre les malwares.

2. Sujet

2.1 Problématique

Ce sujet de thèse s'inscrit dans le thème général de la lutte contre les malwares, et plus particulièrement dans le domaine de l'analyse de codes binaires de programmes obfusqués avec une application directe, mais potentielle, à la détection de comportement suspect.

La plupart du temps, seul le code binaire des programmes malveillants, comme les rançongiciels, sous Windows/Linux/MacOS est accessible. Ce dernier doit être analysé pour comprendre les intentions de la charge finale de l'attaque. Cette tâche, fastidieuse et chronophage, est réalisée par des experts en rétro-ingénierie. Un triage est fait auparavant en essayant d'apparenter un programme malveillant à une famille connue de telle sorte à diminuer le nombre d'analyses. Le résultat est une liste de tactiques, de techniques, et de procédures (TTP) qui sont implantées dans le malware, ce qui va alimenter la *Cyber Threat Intelligence* (CTI) par la suite.

2.2 Objectif

L'objectif de cette thèse est d'extraire les tactiques, techniques et procédures (TTP) du code binaire d'un programme malveillant.

Dans le contexte du projet DefMal, cette thèse contribue à la fois aux travaux sur la rétro-ingénierie et à ceux sur la détection. L'approche envisagée est de commencer par une analyse dynamique pour extraire le graphe de flot de contrôle et le graphe d'appel des fonctions, ainsi que différentes informations, en particulier les modifications de registres du système, les créations de threads, les informations sur les communications avec le *Command & Control C2*.

La partie Loria (équipe Carbone) du projet DefMal a mis à disposition un service d'analyse dynamique qui fournira toutes ces informations. Ensuite, nous utiliserons des heuristiques d'identification de fonctionnalités. Ce travail nécessitera d'accroître les approches par différents biais, notamment par analyse symbolique dynamique ou par IA générative.

2.3 Questions de recherche.

1. Étant donné un code binaire d'un programme, comment identifier un motif qui correspond à une procédure ?
2. Comment d'un graphe de procédures définir une technique, puis une tactique ?
3. Comment à partir de l'extraction de TTP conclure que le comportement d'un programme est potentiellement malveillant ?

2.4 Résultats attendus

Les résultats seront publiés dans les meilleures conférences possibles. Certaines parties des travaux devraient être applicables assez vite, et des prototypes seront développés et validés de manière incrémentale en suivant les avancées scientifiques. L'outil pourra aussi être présenté dans les conférences plus techniques comme SSTIC, BotConf ou BlackHat. Le prototype final a pour vocation à être une brique dans la plateforme d'analyse de DefMal qui enrichira les TTP associés aux malwares et la CTI qui en résulte.

3. Organisation et support

Une réunion hebdomadaire est organisée en visio avec le doctorant et des réunions avec l'ensemble de l'équipe sont faites toutes les deux semaines.

Ce poste sera affecté dans une zone à régime restrictif (ZRR) au Loria. Le doctorant bénéficiera du savoir-faire de l'équipe et pourra échanger et travailler en collaboration avec les ingénieurs et les post-docs de l'équipe. Le doctorant aura accès au Laboratoire de Haute Sécurité (LHS). Enfin, le laboratoire offre un environnement scientifique épanouissant avec de nombreux séminaires et une association des doctorants.

4. Candidature

- La thèse peut débuter à tout moment à partir de maintenant, et au plus tard en juin 2026 - la date limite de candidature est fixée à mars 2026.
- Pour postuler, merci de contacter Jean-Yves Marion (Jean-Yves.Marion@loria.fr) et Dylan Marinho (Dylan.Marinho@lip6.fr).