Proposition de stage recherche en CyberSécurité

Construction d'un jeu de données évolutif, étiqueté et interrogeable de malware Android

Valérie Viet Triem Tong et Jean François Lalande
EPC INRIA CIDRE,
Campus de Rennes de CentraleSupelec

Depuis plusieurs années, des membres de l'équipe Inria Cidre s'intéressent à l'étude des codes malveillants sous Android dont une majeure partie est présentée sur le site du projet Kharon [**?**, **?**]. Dans ce cadre, nous faisons régulièrement des expérimentations qui visent à éprouver nos algorithmes (détection, caractérisation). Contrairement à la grande majorité des communautés scientifiques, il n'existe que très peu de *jeu de données* publiques de malware Android qui permettraient en particulier de comparer des résultats, mener rapidement des expérimentations pertinentes [**?**, **?**]. Ce manque de jeu de données s'explique par plusieurs facteurs :

- les malware Android, comme plus généralement les codes malveillants sont des données sensibles qui ne peuvent pas être stockées et distribuées facilement
- ces malware sont rapidement obsolètes car ils attaquent des vulnérabilités dans des applications ou des versions d'OS qui sont régulièrement mis à jour : les attaques perpétrées par ces malware deviennent rapidement non observables expérimentalement, si l'on ne s'assure pas d'une reproductibilité précise de l'environnement adéquat pour que la malveillance ait lieu.
- ces malware sont généralement mal ou peu étiquetés. On constate que les caractéristiques pertinentes pour l'étude d'un code malveillant (est il chiffré, est il obfusqué, est il un cheval de Troie, est il un dérivé de *cryptolocker*?) ne sont en général pas fournis. Il devient alors très difficile de qualifier la performance d'une expérience, de manière automatique, sans investiguer manuellement chaque échantillon.
- le temps nécessaire pour réaliser des expériences sur une large quantité de malware est prohibitif, lorsqu'on s'approche d'un environnement "réel", par exemple si l'on se passe d'un émulateur Android.

En conséquence, chaque chercheur de cette communauté passe régulièrement beaucoup de temps à construire un jeu de données permettant d'obtenir des résultats cohérents.

Pour répondre à cette problématique, nous pensons qu'il est possible de définir une méthodologie de construction de jeu de données à la demande.

Dans le stage, nous proposons tout d'abord d'explorer la nature des jeux de données existants, de répertorier les types d'expériences qui sont réalisés par les chercheurs en sécurité et de définir un modèle idéal répondant aux besoins identifiés. Dans un second temps, on s'intéressera à créer une méthodologie pour rendre la construction, l'utilisation et la valorisation d'un tel jeu de donnée. Il pourra s'agir de construire réellement le jeu de donnée pour une utilisation interne mais aussi de proposer des outils permettants à d'autres chercheurs la construction d'un jeu de donnée pour leur propre usage. Dans sa globalité, le

stage permettra à la fois d'explorer les enjeux scientifiques, le cadre légal et les difficultés techniques liées au domaine particulier de l'expérimentation en sécurité informatique.

Contacts

N'hésitez pas à nous contacter pour tout renseignement supplémentaire.
jean-francois.lalande@centralesupelec.fr et valerie.viettriemtong@centralesupelec.fr

Références

[1] Adrien Abraham, Radoniaina Andriatsimandefitra, Adrien Brunelat, Jean-François Lalande, and Valérie Viet Triem Tong. GroddDroid : a Gorilla for Triggering Malicious Behaviors. In *10th International Conference on Malicious and Unwanted Software*, pages 119–127, Fajardo, Puerto Rico, October 2015. IEEE Computer Society. Best Paper Award.

[2] Nicolas Kiss, Jean-François Lalande, Mourad Leslous, and Valérie Viet Triem Tong. Kharon dataset : Android malware under a microscope. In *The Learning from Authoritative Security Experiment Results (LASER) workshop*, pages 1–12, San Jose, United States, May 2016. USENIX Association.

[3] Jean-François Lalande, Valérie Viêt Triem Tong, Mourad Leslous, and Pierre Graux. Challenges for Reliable and Large Scale Evaluation of Android Malware Analysis. In *International Workshop on Security and High Performance Computing Systems*, Orléans, France, jul 2018. IEEE Computer Society.

[4] Mourad Leslous, Valérie Viet Triem Tong, Jean-François Lalande, and Thomas Genet. GPFinder : Tracking the Invisible in Android Malware. In *12th International Conference on Malicious and Unwanted Software*, Fajardo, Puerto Rico, October 2017. IEEE Conputer Society.