Proposition de sujet de thèse : Classification et détection des programmes malveillants (malwares) dans les applications mobiles.

Responsable : Akka Zemmari

lieu : LaBRI

téléphone : 05.40.00.60.93

e-mail : zemmari@labri.fr

équipe : Combinatoire et Algorithmes, thème : Algorithmes distribués.

Mots-clés : Machine learning, apprentissage statistique, processus stochastiques, reverse engineering.

Résumé

L'objectif de cette thèse est d'appliquer les techniques d'apprentissage machine (Machine Learning) à la classification des applications mobiles et à la détection de programmes malveillants (Malware).

Description du sujet

Ces dernières années, les applications mobiles ont connu un grand essor et plusieurs équipements utilisés dans la vie quotidienne en sont équipés. On parle désormais de l'internet des choses (Internet of Things ou IoT). Ces applications, par leur large utilisation, font l'objet de développements malveillants et de production d'applications dites malicieuses. Ces applications peuvent être téléchargées depuis des dépôts connus et contrôlés mais également depuis des dépôts plus ou moins ouverts.

Une des tâches fondamentales consiste à détecter les applications dites malicieuses et les différencier de celles bienveillantes (benign). Autrement dit, il s'agit de classifier ces applications.

Pour ce faire, une première tâche consiste à faire du reverse engineering afin d'analyser le code des applications considérées. Ensuite, à utiliser des techniques de classification issues du domaine de l'apprentissage machine (Machine Learnig).

Travail à réaliser

Le but de cette thèse est d'utiliser les techniques d'apprentissage machine afin de classifier les applications mobiles. Il s'agit dans un premier temps de faire du reverse engineering afin de pouvoir analyser le code de ces applications. Ensuite, nous étudierons les différentes techniques de machine learning telles que la classification naïve bayésienne, la machine à vecteur support (SVM), etc. afin d'en choisir les plus appropriés et les plus efficaces.

La thèse débutera donc par une étude de l'état de l'art. En effet, le domaine de la détection des applications malveillantes (malware) a été le sujet de plusieurs recherches ces dernières années. Ensuite, le doctorant se formera aux techniques de reverse engineering et aux techniques de classification par apprentissage machine.

Dans une deuxième phase, le doctorant devra proposer des solutions et les tester sur des jeux de donnés (datasets) connus et mis à disposition de la communauté des chercheurs.

Les travaux devront donner lieu au développement d'outils logiciels permettant de faire des analyses statiques et/ou dynamiques.

Références

- *DRACO : DRoid Analyst COmbo, An Android Malware Analysis Framework.* S. Bhandari, R. Gupta, V. Laxmi, M.S. Gaur, A. Zemmari. In 8th International Conference on Security of Information and Networks (SIN 2015).
- *CONFIDA : identifying covert feature misuse with CONtrolFlow Inter-component Dependence analysis in Android apps.* A. Bharmal, M. Conti, P. Faruki, M. S. Gaur, V. Laxmi, A. Zemmari. To appear.
- *Android Security : A Survey of Issues, Malware Penetration, and Defenses.* P. Faruki, A. Bharmal, V. Laxmi, V. Ganmoor, M. S. Gaur, M. Conti, M. Rajarajan. IEEE Communications Surveys and Tutorials 17(2) : 998-1022 (2015)
- Malware Genome Project. `http://www.malgenomeproject.org`.
- *Marvin : Efficient and Comprehensive Mobile App Classification Through Static and Dynamic Analysis.* M. Lindorfer, M. Neugschwandtner, and C. Platzer. In Proceedings of the 39th Annual International Computers, Software & Applications Conference (COMPSAC), 2015.