ONTIC : Online Network TraffIc Characterization
Projet FP7-ICT
Début : 1/2/2014 – Fin : 31/1/2017

Philippe Owezarski – LAAS-CNRS

L'objectif du projet ONTIC était de proposer des mécanismes efficaces pour l'identification et la classification du trafic des réseaux de communication. Cette classification est essentielle pour de nombreuses tâches d'ingénierie et de gestion des réseaux. Au cours du projet, des mécanismes de fouille de données originaux et de caractérisation du trafic ont été proposés. Ils exploitent des approches d'analyse « big data » et des paradigmes de traitements distribués dans le « cloud ». Une partie de ce projet concernait la détection proactive en temps-réel d'anomalies et d'attaques. C'est cette partie qui sera présentée dans l'exposé proposé.

L'objectif est donc d'analyser le trafic collecté sur les réseaux de communication pour caractériser et classifier les anomalies qu'il contient. L'idée consiste à concevoir une méthodologie d'analyse du trafic fonctionnant de façon autonome et en temps réel, et ce sans aucune connaissance préalable du trafic. Ceci s'oppose aux techniques actuelles qui reposent sur les compétences d'experts et sont donc lentes et coûteuses et laissent les systèmes informatiques sans protection face à de nouvelles attaques pendant de longues périodes. Ces anomalies de trafic impliquent également la mobilisation parasite de ressources réseaux, d'éventuels saturations et ainsi une perte de qualité pour les flux légitimes mis en concurrence.

Les outils proposés pour ce faire reposent sur de nouveaux algorithmes d'apprentissage automatique non supervisé basés sur des techniques de partitionnement. L'exposé s'attachera à présenter :

- Les problématiques associées à l'utilisation d'algorithmes de clustering sur du trafic haut débit et extrêmement variables dans le temps ;
- Les solutions et notamment l'algorithme ORUNADA (Online Real-time Network Anomaly Detection Algorithm) ;
- Leur implémentation en utilisant les facilités offertes par les librairies Hadoop Spark et Storm, et leur exécution sur des gros cluster comme « Google Cloud » ;
- La construction de deux « ground truths » pour la validation d'outil de détection d'intrusion et d'anomalies. Ces deux « ground truths » reposent sur de grandes quantités de trafic collectées au cours du projet sur des réseaux commerciaux. Le premier consiste juste à identifier les anomalies contenues dans ce trafic, notamment en faisant appel à du « crowd sourcing ». Le second contient en plus des anomalies synthétiques.
- La validation d'ORUNADA notamment sur ces deux ground truths.