Rapport sur le stage de de mobilité entrante effectué du 25 septembre 2019 au 23 Décembre 2019

Dans le laboratoire :

Lab-STICC UMR CNRS 6285, Université de Bretagne Occidentale, 29200 Brest

Stage de Doctorat

Kherbache Meriem
Doctorante en troisième année
Spécialité Informatique
Dirigée par : Dr AMROUN Kamal
Codirigé par : Dr ESPES David

Financé par l'université
Abderrahmane Mira, Bejaia.

جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa

Remerciements

Je tiens à commencer ce rapport de stage par des remerciements, à l'université de Bretagne Occidentale de m'avoir permis d'effectuer ce stage, à ceux qui m'ont beaucoup appris au cours de ce stage, et même à ceux qui ont eu la gentillesse de faire de ce stage un moment très profitable.
Aussi, je remercie Monsieur David Espes, mon directeur de stage qui m'a formé et accompagné tout au long de ce stage avec beaucoup de patience et de pédagogie. Enfin, je remercie l'école doctorale de m'avoir permis d'effectuer un stage au sein de leur établissement au cours de ces deux mois.

Introduction

Du 25 septembre 2019 au 23 décembre 2019, j'ai effectué un stage au sein de Lab-STICC UMR CNRS 6285, Université de Bretagne Occidentale.

Mes travaux de thèse concernent l'hybridation des méthodes de fouilles de données et d'apprentissage automatique pour la détection d'intrusions dans un réseau. Elle est dirigée par Dr. Kamal AMROUN – Maitre de Conférences (HDR) à l'université de Abderrahmane Mira.

Les travaux effectués et les apports du stage

Cette thèse vise à contribuer à l'amélioration des méthodes d'évaluation des systèmes de détection d'intrusion. Cette thématique se compose de trois problématiques :

- Réduction du nombre de caractéristiques d'un jeu de données : l'algorithme de Machine Learning utilise un ensemble de caractéristiques pour identifier si un trafic réseau est bénin ou malicieux. Les caractéristiques utilisées vont donc avoir un impact direct sur l'efficacité de l'algorithme à classifier correctement le trafic mais également sur la rapidité de classification de ce dernier. En effet, il faut pouvoir trouver un sous-ensemble optimal de caractéristiques qui permettra de trouver le meilleur compromis entre efficacité et rapidité.
- Identification d'attaques complexes à signaux faibles : avec l'amélioration des mécanismes de sécurité actuellement déployés, les attaques actuelles essaient d'être le plus discrètes possibles. Pour cela, ces attaques vont être réalisées sur une longue période temporelle et ressembler le plus possible à un trafic bénin. Actuellement les algorithmes de Machine Learning détectent particulièrement bien les attaques à forts signaux (tels que les attaques par Dénis de Service ou Dénis de Service Distribués). Il est donc nécessaire de faire évoluer les algorithmes de classification afin qu'ils puissent détecter de manière efficace les attaques complexes.
- Utilisation d'algorithmes d'apprentissage non-supervisés : il existe deux classes d'algorithmes d'apprentissage : les algorithmes supervisés et les algorithmes non-supervisés. Durant la phase d'entrainement, les algorithmes supervisés reposent sur un étiquetage précis du trafic (bénin, malicieux) pour construire leur modèle de détection. Avoir accès à un trafic étiqueté est souvent une tâche fastidieuse et onéreuse car elle repose sur la connaissance d'experts en sécurité. De même, la diversité des environnements de production rend peu générique de telles approches. A contrario, les approches non-supervisées reposent sur la similarité du trafic pour les catégoriser sans avoir recours à des informations complémentaires lors de l'entrainement du modèle. Une telle souplesse permet de ne pas avoir recours à des experts onéreux et de pouvoir s'adapter facilement à des environnements de production très différents. Cependant, l'efficacité limitée de ces algorithmes font qu'ils sont utilisés uniquement dans des environnements très spécifiques. Afin de disséminer au mieux la détection d'intrusion par analyse comportementale et d'en maîtriser les coûts d'exploitation, il est essentiel de pouvoir reposer sur des algorithmes non-supervisés.

➢ **Objectifs du stage :**

Pour résoudre ces problématiques, nous avons déjà résolu la première et la deuxième problématique en proposant une démarche rigoureuse couvrant l'ensemble des étapes de l'évaluation d'un IDS dans le stage de l'année passée au niveau de Laboratoire Lab-STICC à l'Université de Bretagne Occidentale (UBO). Ce stage est une continuité de ces travaux, Nous avons traité la troisième problématique qui est de proposée une méthode non-supervisé visant à améliorer les IDSs.

Les objectifs du stage sont les suivants :

1. Rédaction et soumission d'un article à une conférence internationale **DAT2020,** basée sur une amélioration d'une méthode non supervisé (K-means Clustering) dans les IDSs.
2. Finalisation et soumission de l'article sur la sélection des caractéristiques basée IDS dans une revue internationale nommée, **Journal of Information Security and Applications (JISA).**
3. Proposition d'une nouvelle méthode de sélection de caractéristiques basée sur la corrélation entre les caractéristiques, Tester la méthode proposée sur différent jeu de données (NSL-KDD, CICIDIS2017), pour montrer la faisabilité de notre approche.

➢ **Les travaux effectués :**

Mes travaux de recherche sont actuellement dans une phase de finalisation. En effet, ce stage m'a permis traiter les différents jeux de données vu la disponibilité des machines puissante et de finaliser mes travaux, également de proposer une nouvelle méthode de sélection de caractéristiques basée sur la corrélation entre les caractéristiques à publier prochainement.

➢ **Les apports du stage :**

Ce stage m'a permets d'avancer sur deux plans, sur l'état d'avancement de ma thèse, pour cela j'ai pu rejoindre le laboratoire LabSTICC afin de travailler avec mon co-directeur de thèse Monsieur David ESPES, le laboratoire m'a donné accès à une machine de forte capacité mémoire afin de traiter mes différents jeu de données. Sur le plan personnel, en effet j'ai pu acquérir une certaine autonomie, et un goût prononcé pour le travail. Cela m'a aussi permis de développer une certaine aisance quant à mener à bien un projet, me voir confier des responsabilités, ainsi que d'analyser un problème donné et chercher des solutions adéquates.

Université de Bretagne Occidentale
Lab-STICC UMR CNRS 6285
6 Av. le Gorgeu - CS 93837
29238 BREST CEDEX 3 - France
Tél.:33(0)2 98 01 61 26

Visa du laboratoire d'accueil

Emmanuel RADOI
Directeur du Lab-STICC/UBO

Visa du directeur d'accueil

David ESPES

Signature

Kherbache Meriem