SCHOEN Adrien

Rapport de réunion n°6

Date de la réunion : 05/03/2021

Date de la prochaine réunion : 12/03/2021

Ordre du jour : Traitement des données

1. Ce que le stagiaire a dit

J'ai re-téléchargé les données manquantes du dataset

J'ai ensuite classé les différents pcap en plusieurs catégories en m'aidant de l'IP du terminal ayant servi à créer ce pcap et de la date. Je savais à quel jour était programmé quelle attaque contre quel terminal. J'ai donc classé le pcap de la machine correspondant à la victime en fonction de l'attaque de la journée.

Cette classification n'est pas suffisante, car au cours de la journée d'attaque, la victime échange avec le reste du réseau de manière totalement normale, donc le pcap contient aussi des flux parfaitement bénins.

J'ai donc divisé les pcap en fonction du couple IP source/ IP destination grâce à Splitcap, afin de pouvoir extraire le pcap qui ne contient uniquement que les flux entre l'attaquant et la victime.

Tous les paquets de ce pcap extrait seront considérés comme malveillant

J'ai fait ainsi pour toutes les attaques et j'ai ensuite divisé chaque pcap en plusieurs fichiers .bin qui contiennent chacun la suite hexadécimale correspondant à un paquet.

Mon programme charge directement ces fichiers binaires et les transforme en images de la forme souhaitée, si nécessaire en utilisant du bourrage ou une réduction.

Beaucoup d'attaques sont difficilement perceptibles au niveau paquet et nécessite d'avoir une vision plus haut niveau. Nous essayerons ainsi d'abord de générer du trafic classique.

2. Ce que les encadrants ont ajoutés

Il faut commencer à générer des paquets provenant des pcap bénins.

Ne pas hésiter à rentabiliser le temps d'attente pour faire le point sur la biblio, lire de nouvelles études, préparer le rapport etc.…

Pour l'évaluation des méthodes de génération : il faut que les données générées nous permettent d'entraîner des modèles assez simplement, mais il faut aussi qu'elles puissent être transmises sur un réseau.

Si j'ai besoin de ressource de calcul, je peux demander un accès à un cloud

3. Ce qu'il faut faire pour la prochaine séance

Commencer la génération, peut être avec des auto encoder.