SCHOEN Adrien

Rapport de réunion n°15

Date de la réunion : 07/05/2021

Date de la prochaine réunion : 14/05/2021

Ordre du jour : Description Flux UDP et rapport bibliographique

1. Ce que le stagiaire a dit :

J'ai expliqué ce dont on avait besoin pour décrire un flux UDP. Le but est d'avoir un descripteur de flux pour pouvoir générer des flux et forger les headers des paquets de ces flux.

Dans un paquet UDP, les seuls champs dépendants de caractéristiques de flux se trouvent dans le header, donc la description d'un flux et la génération de header sont deux faces d'un même problème.

On se concentre sur les champs d'un header UDP qui ne peuvent être décrits autrement que via la connaissance du flux. Il s'agit de quatre champs : L'Identification number du paquet, le TTL du paquet, le port source et le port destination. Tous les autres champs peuvent soit être donnés par l'utilisateur, soit être déduit automatiquement.

Les quatre caractéristiques précédentes sont donc des caractéristiques du flux dans lequel est généré le header. A ces quatre caractéristiques s'ajoutent quatre autres : La durée du flux, le nombre d'octet et de paquet du flux ainsi que le Timestamp du premier paquet du flux.

A l'aide de ces 8 paramètres, nous pouvons décrire les flux et générer les header de tous les paquets. Je suis en train d'agréger les flux en fonction de ces 8 paramètres pour pouvoir générer des séquences de ces 8 paramètres.

J'ai ensuite résumé les recherches bibliographiques de cette semaine et de la semaine dernière. Pour rappel, on cherchait à répondre à deux questions :

1)Comment modéliser les différents niveaux d'information du trafic ?
2)Comment générer les paquets sans perte de l'information sur la structure ?

1) On va utiliser la méthode de pcapgan : On commence par générer diverses caractéristiques du flux, ce qu'ils appellent options. Et ensuite on génère les paquets en remplaçant les bytes correspondant aux caractéristiques générées à l'étape précédente par leur valeur.

2) On peut soit essayer de représenter la séquence dans une sorte d'embedding (Seq2Seq, Bert etc.) pour ensuite essayer de générer ces représentations, mais aussi essayer de directement générer des séquences (SeqGAN).

2. Ce que les encadrants ont ajouté :

Pour le header UDP la plupart de ces méthodes semblent exagérées. On peut essayer de générer les caractéristiques exposées par simple calcul à partir des paquets générés.

La question se pose pour la longueur, le nombre d'octet et la durée, ou on devra sans doute apprendre des distributions.

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

Agglomérer les flux en caractéristique pour le dataset du POC

Voir la génération de matrice de temps dans PCAPGan et essayer de s'en inspirer.