

Rapport de réunion n°9

Date de la réunion : 26/03/2021

Date de la prochaine réunion : 02/04/2021

Ordre du jour : Prise en main Igrida

1. Ce que le stagiaire a dit

J'ai réussi à faire tourner mon code sur Igrida. Il a fallu pour cela que j'apprenne à utiliser les différents outils (pour charger des librairies, voir l'utilisation des ressources, etc.). J'ai aussi mis en place le multiprocessing

L'entraînement sur Igrida est toutefois encore trop long. Nous disposons pour répondre à ce problème, d'énormément de RAM, ce qui nous permettrait de charger le dataset directement dans la mémoire vive et ainsi de réduire le temps de chargement des données (l'étape la plus longue de l'entraînement).

Le temps de chargement des données est estimé à deux heures et j'espère que ce temps sacrifié accélérera l'ensemble du processus d'entraînement.

J'ai commencé à regarder des articles sur la génération sous respect d'une grammaire par autoencoder, mais il a fallu que je comprenne déjà bien ce qu'était une grammaire

2. Ce que les encadrants ont ajoutés

En informatique, une grammaire est un ensemble d'éléments permettant de décrire un langage, où un ensemble de valeurs doivent respecter une syntaxe correcte. Même si ça semble un peu poussé pour ce que l'on souhaite faire ; générer des paquets en connaissance de la grammaire nous dispenserait d'apprendre certaines règles et ainsi de nous concentrer sur l'apprentissage de la distribution des valeurs.

On peut aussi regarder du côté des DSL (domain specific language). Ce sont des ensembles de règles grâce auxquelles on pourra essayer de générer nos paquets à l'aide de fonctions spécifiques à notre problème (la génération de trafic réseau). Il y a quelques posters qui ont été distribués au cours de la séance sur ce sujet-là

On pourra plus tard penser à fusionner les méthodes de génération haut niveaux avec celle de bas niveaux que je suis en train de mettre en œuvre.

Il faudra essayer de voir si on ne peut pas décomposer la génération de paquet en deux parties : une génération des entêtes et une génération du payload en concordance avec les entêtes générés.

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

Continuer l'entraînement sur Igrida

Suivre la formation sur Igrida du 2 avril

Revoir la métrique d'évaluation des modèles afin d'y inclure une composante jugeant de la diversité des paquets générées.