Rapport d’activité journée 2 Hackathon

Ce rapport tente de reprendre les travaux et décisions du groupe aujourd’hui.

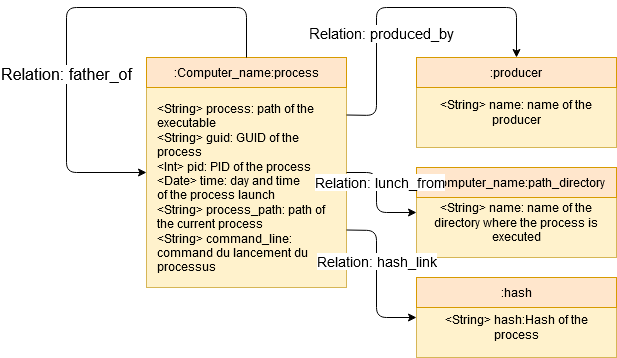
**Architecture du graphe**

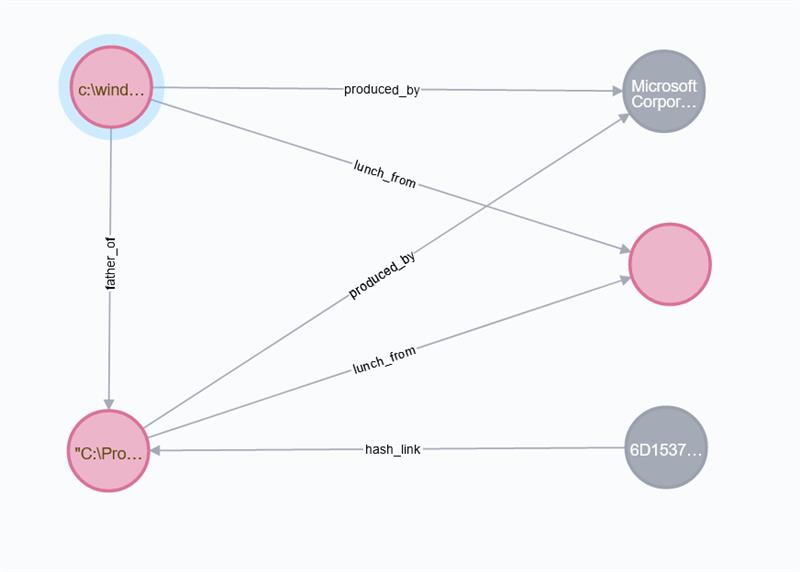
Nous avons abandonné les nœuds de type ‘Menace’ car ils ne représentaient rien... ils n’étaient pas utiles.

Nous avons décider de créer, en plus des nœuds processus, des nœuds pour :

1. Les répertoires depuis lesquels des processus s’exécutent (en tout 47 emplacements)
2. Les Hash des processus, afin de voir directement quels sont les processus que l’on peut déjà labelliser comme ‘anormal’ (242 au total, pas tous anormaux).
3. Les Editeurs, afin d’en extraire plus facilement des statistiques

On peut donc schématiser notre architecture (pour l’instant !!) comme suit :



Et si l’on s’amuse à isoler quelques processus on doit avoir une représentation qui ressemble à ça : 

Avec :

* En rouge les nœuds de Processus
* En gris (en haut à droite) le nœud de l’éditeur
* En gris (en bas à droite) le Hash.

Les hash sont maintenant annotés dans la base grâce à VirusTotal.

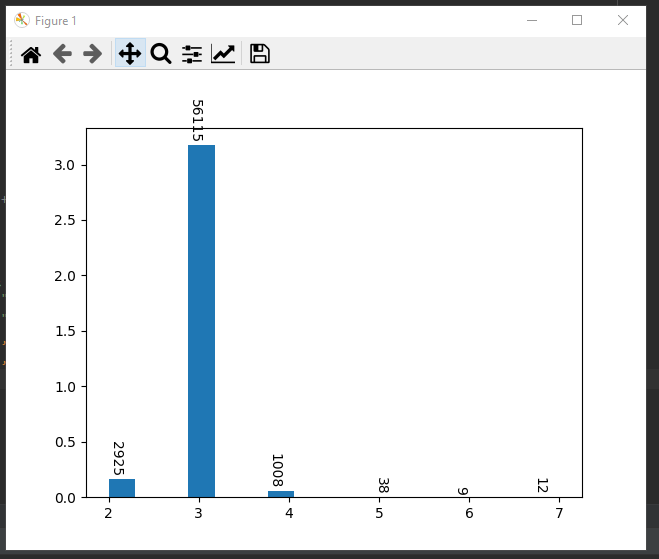
**Indicateurs et métriques**

Pour l’instant cette architecture doit nous permettre faire plusieurs calculs ou recherches sur les données. Notre idée est de construire plusieurs métriques telles que :

1. Pour un processus, la profondeur de son dossier d’exécution dans l’arborescence des fichiers
2. Pour un processus, le nombre d’autres processus qui s’exécutent aussi dans le même dossier
3. Pour un processus, son nombre moyen (ou autre agrégation) d’enfants à chacune de ses exécutions
4. Pour un processus, le nombre de dossiers différents dans lesquels il peut être amené à s’exécuter
5. Pour un processus, sa fréquence d’exécution
6. Pour un Hash donné, le nombre de processus qui le référencent

Nous ne savons pas exactement si ces métriques seront pertinentes, mais nous voulons tout de même les essayer. Ces indicateurs risquent de donner beaucoup de faux positifs ?

Actuellement, on peut déjà dresser les distributions des deux premières métriques :



Sur cet histogramme nous avons la distribution de la profondeur des dossiers d’exécution des processus. L’écrasante majorité des 60107 processus s’exécutent à une profondeur de 3.

Pour l’indicateur des dossiers les plus communs, on trouve que 53221 processus s’exécutent dans le dossier “C:\Windows\System32\”.

Nous continuons à écrire les fonctions et requêtes pour les indicateurs manquants. Par la suite, l’idée est d’utiliser ces indicateurs dans des fonctions d’anomalies que nous définirions nous-même ou bien d’utiliser des algorithmes de détection d’Outliers, mono ou multivariés.

**Représentation graphique**

Nous sommes en train de développer une interface graphique afin de visualiser le plus efficacement possible les données que nous manipulons.

**Quelque(s) interrogation(s) :**

Y a -t- il une différence entre process\_path et process ?

Nous avons essayer de mettre des métriques qui prennent en compte vos remarques et conseils du premier rapport, avez-vous des commentaires sur ceux ? En voyez-vous d’autres ?