INF-6422: Rapport laboratoire 1

Rendu le Mardi, 19 Janvier 2016

Thomas Luinaud, Paul Berthier

Table des matières

Mise en contexte

1.1

L'épidémiologie pourrait être définie comme l'étude des rapports existant entre les maladies ou tout autre phénomène biologique, et divers facteurs susceptibles d'exercer une influence sur leur fréquence, distribution et évolution. Entre d'autres mots, l'épidémiologie s'intéresse aux facteurs qui influencent la santé des populations. Plus particulièrement, l'épidémiologie s'intéresse, entre autre, à étudier la dynamique de propagation des maladies infectieuses afin d'établir des stratégies de prévention et d'intervention permettant de diminuer l'impact sur la santé publique. À cet effet, la modélisation mathématique s'est révélée particulièrement intéressante afin de simuler des scénarios épidémiologiques, d'évaluer les risques associés et de quantifier l'efficacité et l'impact de différentes méthodes d'intervention et de prévention. Plusieurs approches peuvent être retenues, telles que les simulations numériques, les modèles déterministes ou encore les modèles stochastiques. Chaque approche présente des avantages et des inconvénients. Il convient donc de choisir la méthode la plus appropriée en fonction des questions de recherche auxquelles vous souhaitez répondre.

1.2

Appliquée à la sécurité informatique, l'épidémiologie pourrait être vue comme l'étude des différents facteurs qui influencent la fréquence, la distribution et l'évolution des logiciels malveillants. Plus particulièrement, l'approche épidémiologique a inspiré de nombreux travaux de recherche portant sur l'étude de la propagation des logiciels malveillants. Le présent laboratoire vous permettra de vous familiariser avec certaines approches mathématiques fréquemment utilisées afin de modéliser la propagation de logiciels malveillants au sein d'un réseau.

Question 2

Une approche très répandue dans l'étude de la propagation des logiciels malveillants consiste à développer un modèle déterministe basé sur les concepts de compartiments et de règles [2]. Les compartiments servent à diviser la population étudiée en différentes classes et les règles à définir les conditions de transition entre chacune des classes.

1.1

En vous basant sur l'article « Optimising Networks Against Malware » [3], quel modèle comportemental (SI, SIS, SIR) s'appliquerait et pourquoi ? Justifiez votre réponse en expliquant quel modèle s'applique et pourquoi les autres modèles ne s'appliquent pas.

Dans un premier temps

Question 3

Lors de la question précédente, vous avez développé un modèle théorique basé sur un système d'équations différentielles. Heureusement, il existe une solution analytique à ce système afin de représenter le nombre de machines infectées en fonction du temps :

$$I(t) = \frac{I_0 N}{(N - I_0)e^{-\lambda t} + I_0} \tag{1}$$

Modèle stochastique

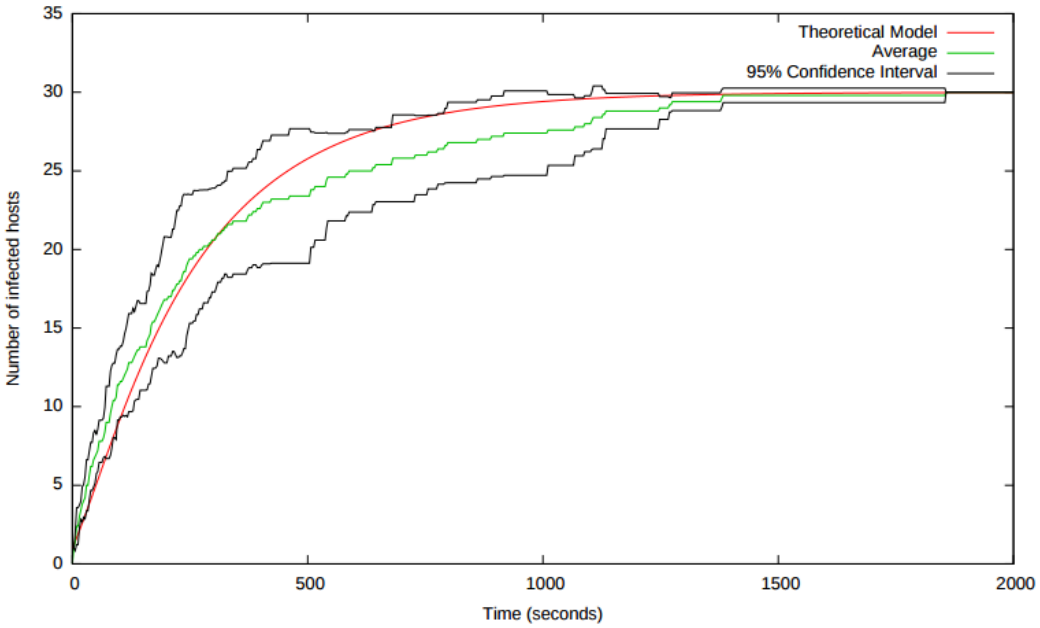
Dans l'article « Optimising networks Against Malware », l'auteur utilise un modèle stochastique basé sur les chaînes de Markov afin de modéliser la propagation d'un vers dans un réseau .

4.1

Expliquez les caractéristiques d'un modèle stochastique et pourquoi ce type de modèle s'applique dans le contexte de l'article. Est-ce qu'une approche déterministe aurait été préférable ?

Un modèle stochastique repose sur des variables aléatoires représentant l'évolution possible d'un système au cours du temps. Ce type de modèle s'applique très bien dans le contexte de l'article car on modélise l'évolution de l'infection d'un système de machines au cours du temps, chaque ensemble de machines infectées étant représenté par un état, avec une certaine probabilité p de passer à un autre état au temps $t+1$ (chaines de Markov).
Une approche déterministe n'aurait pas été réalisable, car les résultats de l'expérience ne sont pas fixes, et donc pas reproductibles. En effet, l'évolution de l'infection dépend de la rapidité à laquelle la première machine infectée réussit à atteindre le *gateway* afin d'atteindre l'autre sous-réseau. Cela se vérifie sur la 1, où l'on voit que l'intervalle de confiance de l'expérience est très large.

FIGURE 1 – Expérience avec G=4



Performance et optimisation

Toujours dans l'article « Optimising networks Against Malware », l'auteur étudie l'effet de la topologie du réseau sur la vitesse de propagation d'un vers informatique.

5.1

Appliquez le concept du double-tétraèdre étudié en classe à l'article. Situez votre double-tétraèdre dans un contexte d'optimisation, toujours en vous référant à l'article. Expliquez quels sont les aspects de votre double-tétraèdre qui sont fixes, et ceux qui sont modifiés.

5.2

Quels autres aspects/méthodes (autres que la topologie du réseau) pourraient être étudiés dans le cadre d'un problème d'optimisation où l'objectif est de limiter la vitesse de propagation d'un logiciel malveillant dans un réseau ? Donnez au minimum deux exemples.

5.3

Il a été démontré qu'une plus grande biodiversité au sein d'un écosystème permettait de ralentir la propagation des virus ou des bactéries dangereuses. Comment pourriez-vous appliquer le concept de diversité au sein d'un réseau informatique ? Quelle(s) forme(s) prendrait cette diversité ?