INF-6422: Rapport laboratoire 1

Rendu le Mardi, 19 Janvier 2016

à *Francois Labrèche*

Thomas Luinaud, Paul Berthier

Table des matières

1. 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.

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

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 :

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

4. Modèle stochastique

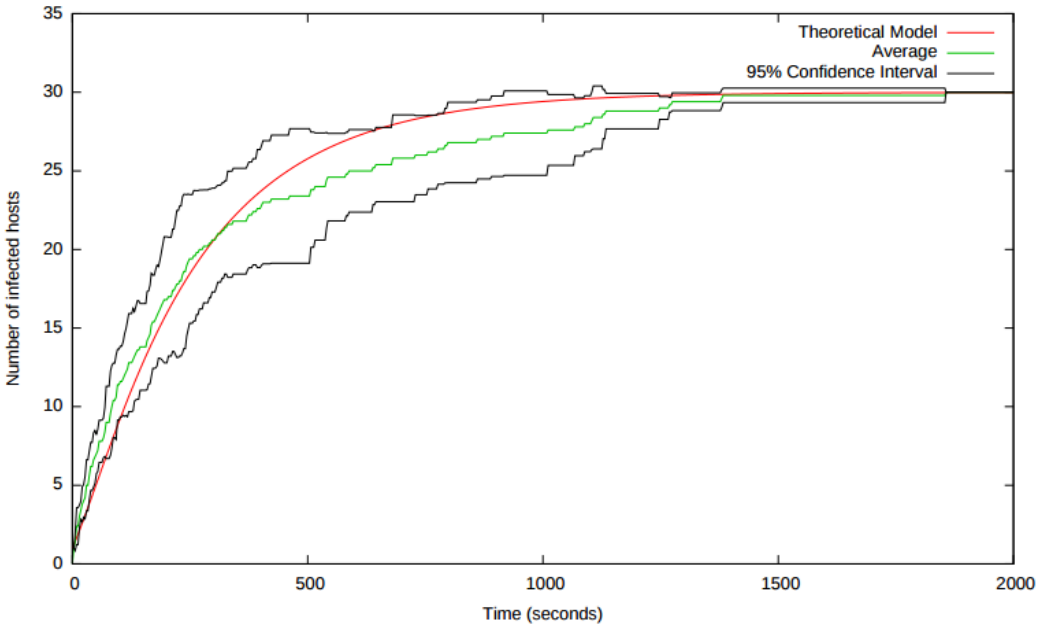
Dans l'article « Optimising networks Against Malware », l'auteur utilise un modèle stochastique basé sur les chaînes de Markov afin de modéliser la propagation d'un vers dans un réseau .

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



5. 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.

Le double-tétraèdre étudié en classe regroupe les propriétés suivantes :

- L'**environnement**, commun à l'attaquant et au défenseur
- Des **critères de performance**, propres à l'attaquant et au défenseur
- Les **caractéristiques** de l'attaque et de la défense

Les caractéristiques suivantes sont fixées :

- **Les caractéristiques de l'attaque** : On utilise le *Malware Emulation Framework*, qui produit un ver qui se reproduit en scannant les ip des machines voisines.
- **Les caractéristiques de la défense** : Il n'y en a pas. Dés que l'attaquant scanne l'ip d'une machine, celle-ci est infectée.
- **Le critère de performance de l'attaque** : La vitesse de propagation du virus sur le réseau.
- **Le critère de performance de la défense** : C'est également la vitesse de propagation du virus sur le réseau, mais dans le sens inverse (on souhaite que le virus se propage le plus lentement possible ...)

Le seul paramètre variable est l'environnement : on va faire varier la topologie du réseau en jouant sur le nombre de *gateway* interconnectant les différents sous-réseaux.

On est donc en présence d'un problème d'optimisation du critère de performance, c'est à dire la vitesse de propagation du ver sur le réseau (qui doit être la plus lente possible du point de vue du défenseur), la variable étant le nombre de *gateway*.

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.

Les autre méthodes suivantes pourraient être étudiées afin de limiter la propagation d'un logiciel malveillant dans le réseau :

- L'utilisation d'un *firewall*, filtrant l'utilisation de certains ports, traditionnellement utilisés pour transmettre des virus (telnet, ...)
- Le blocage de certains paquets ICMP (comme le *ECHO* utilisé par la commande *Ping*) afin d'empêcher l'attaquant de scanner les ports ouverts d'une machine.
- L'utilisation d'un antivirus ayant une base de donnée des signatures des virus connus, ou étant capable de reconnaître les comportements suspects des virus (ouverture de connexion sur des ports peu usités, ...)
- Demander à l'utilisateur une confirmation avant l'installation de tout programme en provenance du réseau.

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 ?

Le concept de diversité au sein d'un réseau informatique prend les formes suivantes :
- Différents *hardware* : Carte mère, processeur, carte graphique, carte réseau, ...
- Différents systèmes d'exploitation (Windows, Linux, Mac, ...)
- Différents logiciels ayant des fonctionnalités semblables
- Différentes versions des logiciels
- Différentes configurations des logiciels (ports utilisés, ...)

Une plus grande diversité permet de freiner la propagation d'un virus, car en général ceux-ci utilisent une faille exploitable uniquement sur une version précise d'un logiciel, et dans un environnement spécifique. Ainsi, une machine possédant une configuration n'étant pas vulnérable à la faille exploitée par le virus bloquera la propagation de celui-ci.

Cependant, une grande diversité des machines sur le réseau pose de grosses difficultés de gestion et de maintenance du parc informatique pour l'administrateur réseau. Cette technique de protection n'est donc pas vraiment exploitable à grande échelle ...