INF6422: TP1

pour le Mardi 2 Février 2016

Philippe Troclet (1815208) et Alexandre Mao (1813566)

Table des matières

1 Modèle déterministe

1 . 1 Choix d'un modèle comportemental

En regardant tout d'abord les 3 modèles, nous pouvons constater les caractéristiques suivantes pour chaque modèle :

- Modèle SI : Le modèle SI (avec S représentant le nombre de machines saines et I le nombre de machines infectées) s'intéresse à l'évolution d'une épidémie ,dans une population à une ou plusieurs machines infectées, et à sa propagation auprès du reste de la population en ne prenant en compte que le fait que les machines infectées vont infecter les machines saines.
- Modèle SIS : Le modèle SIS (avec S représentant le nombre de machines saines et I le nombre de machines infectées) s'intéresse à l'évolution d'une épidémie dans une population à un ou plusieurs facteurs pathogènes. On considère dans ce cas que le facteur pathogène n'est que temporaire et se guérit au bout d'un laps de temps donné. Une machine infectée aura le pouvoir de "se guérir" toute seule pour repasser dans l'état sain.
- Modèle SIR : Le modèles SIR (avec S représentant le nombre de machines saines, I le nombre de machines infectées et R le nombre de machines qui ont été guéries et immunisées) représente l'évolution d'une épidémie dans une population où on va introduire auprès de la population un vaccin contre le ou les facteur(s) pathogène(s). Une machine infectée pourra alors recevoir le vaccin ou l'antidote et se retrouver immunisée contre l'agent pathogène.

En nous basant sur l'article "Optimising Networks Against Malware", le modèle comportemental qui s'appliquerait serait le modèle SI. En effet, dans l'article, les auteurs s'intéressent à l'évolution d'un ver(agent pathogène) dans un réseau avec un nombre défini de machines (population étudiée). Les machines ne possédant pas de système immunitaire qui pourrait être représenté par un anti-virus, ou un logiciel interne qui scannerait le système pour rechercher et éliminer des élément suspects ou des facteurs externes. Les machines ne peuvent se débarrasser du ver de façon autonome, nous ne pouvons donc pas être dans le cas SIS. Et il n'y a pas de référence dans l'article à des mises à jours du système ou d'un anti-virus éventuel, ou éventuel à la mise en place de patch pour combler une quelconque faille, nous pouvons ainsi conclure qu'il n'y a pas d'injection de remèdes contre les vers et que les machines n'ont nullement été guéries et immunisées contre cet agent pathogène. Nous ne pouvons donc pas nous trouver dans le cas du modèle SIR. Cet article étudie simplement l'évolution de l'infection d'un ensemble de machines dans différentes configurations données, nous pouvons en déduire que c'est bien le modèle SI qui s'appliquerait. Notons que cette conclusion est légitime de part l'absence de guérison (pas d'anti-virus) ainsi que l'impossibilité pour une machine de se déconnecter, ou de quitter le réseau. Pour connaître de façon plus précise les hypothèses utilisées dans le cadre de l'article, on pourra se reporter au premier paragraphe de la partie 2.3 du dit article, intitulé : *"Markov process Model".*

1 . 2 Identification des équations différentielles

Dans la question précédente, nous avions déterminé qu'un modèle comportemental de type SI s'appliquait à l'étude de la propagation des logiciels malveillants. Sachant qu'un ordinateur peut être sain ou infecté, et uniquement l'un de ces deux états, on a la première relation :

$$S + I = N$$

Où N est la taille de la population et S la taille de la population saine. Tandis que I est la taille population infectée. Si de plus, on note λ le nombre de contacts par machine par unité de temps, On a alors que $\lambda \cdot I$ représente le nombre de machines qui ont été atteintes par une machine infectée entre deux unités de temps. Sachant que $\frac{S}{N}$ est la proportion de machines saines à l'instant courant, on peut approximer le nombre de machines nouvellement infectées entre deux pas de temps par $\lambda \cdot I \cdot \frac{S}{N}$. Ce qui nous donne l'équation différentielle suivante :

$$\frac{dI(t)}{dt} = \lambda \cdot I(t) \cdot \frac{S(t)}{N}$$

On peut alors remplacer $S(t)$ par $N - I(t)$, et on obtient :

$$\frac{dI(t)}{dt} = \lambda \cdot I(t) \cdot \frac{N - I(t)}{N}$$

De part la relation entre S et I, on a :

$$\frac{dS(t)}{dt} + \frac{dI(t)}{dt} = 0$$

De ce fait, $\frac{dS(t)}{dt} = -\lambda \cdot (N - S(t)) \cdot \frac{S(t)}{N}$.

2 Simulation numériques

2 . 1 Etude de I et S en fonction du temps

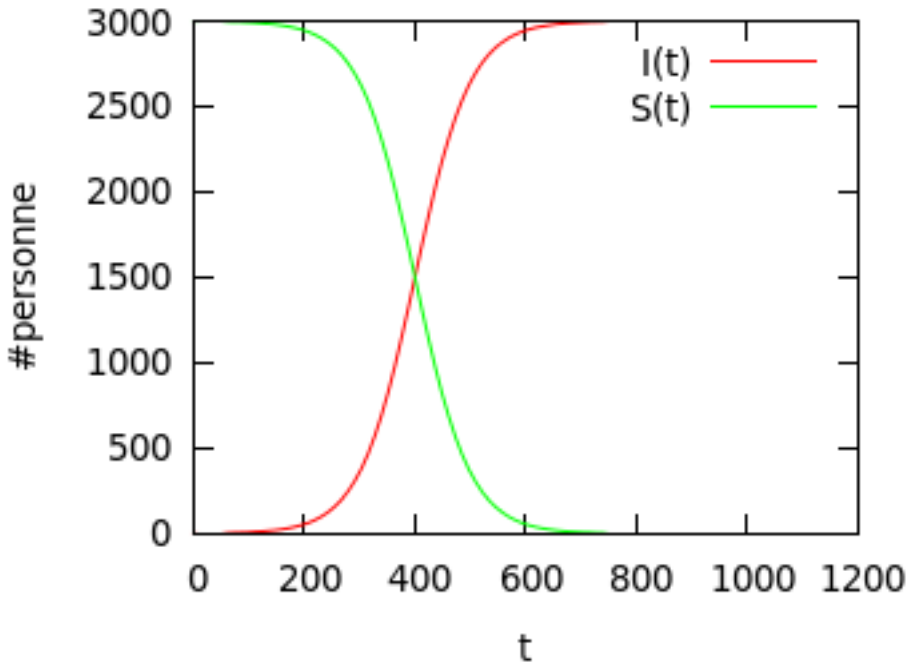

FIGURE 1 – évolution de la population en fonction du temps

On peut remarquer sur la figure 1 que la fonction I converge rapidement vers son maximum. Cela via de la convergence rapide de l'exponentielle décroissante vers sa limite. Ainsi, en environ 700 secondes, l'intégralité des machines vulnérables est atteint. Il aura donc fallu moins de 12 minutes au ver pour contaminer l'ensemble du réseau pourtant composé de 3000 machines.

2 . 2 Etude de l'influence du paramètre λ

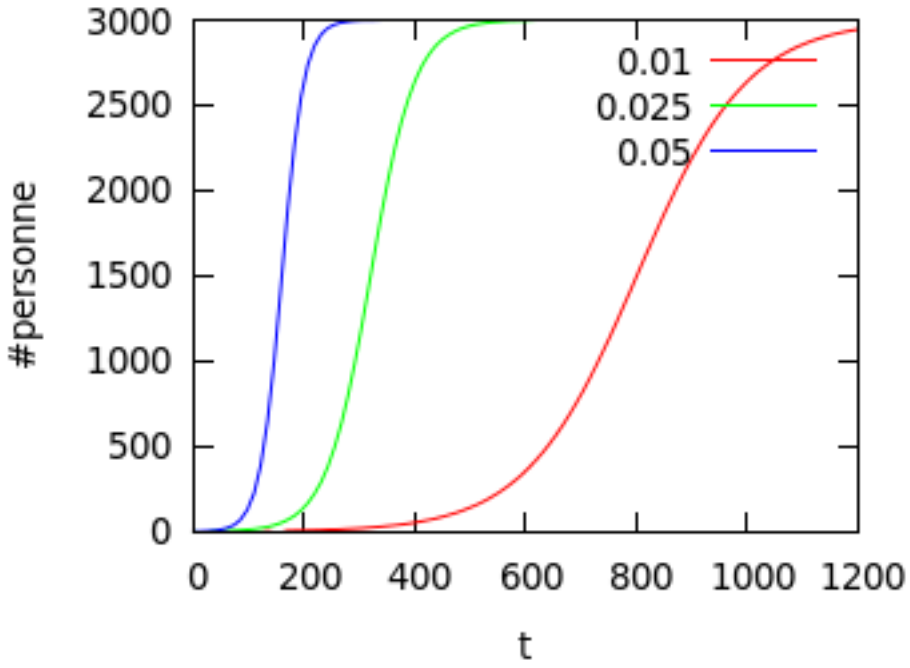

FIGURE 2 – influence du paramètre λ

Dans l'équation différentielle présentée précédemment, le paramètre λ correspond au nombre de contacts par machines par unité de temps. De ce fait, plus ce paramètre est élevé, plus une machine peut atteindre d'autres machines entre deux pas de temps. Ainsi, si λ, augmente le nombre de machines qu'une machine infectée peut contaminer augmente. Une illustration graphique de ce phénomène est visible sur la figure 2. On y voit en effet que la croissance de la courbe correspondant à $\lambda = 0.05$ a une plus forte croissance que les autres courbes (correspondant à $\lambda = 0.025$ et $\lambda = 0.01$). On peut penser aussi que le paramètre λ traduit la probabilité d'infection au cours du temps, et en ayant un nombre de contact plus important entre les machines, cette probabilité augmente.

Avant de clore cette partie sur le modèle déterministe, il est important de noter que les équations établies dans la première partie ne traduisent pas la possibilité que deux machines infectées tente de contaminer la même machine. En effet, l'équation :

$$\frac{dI(t)}{dt} = \lambda \cdot I(t) \cdot \frac{S(t)}{N}$$

Implique que parmi toutes les machines contactées, la proportion de machines vulnérables ($\frac{S(t)}{N}$) sera infectée. Or il est possible qu'une même machine soit atteinte par deux infectées distinctes. La solution de l'équation est donc un majorant de I. En revanche, cette approximation est acceptable pour les grands réseaux (pendant le début de l'infection) où il est peu probable que deux infectées initient une communication avec une même machine saine de part l'effet de nombre. Toutefois, sur de petits réseaux, ces équations de font pas nécessairement sens.

3 Modèle stochastique

3 . 1 Caractéristiques d'un modèle stochastique

Modèle stochastique : Un processus stochastique est la représentation de l'évolution discrète ou en temps continu d'une variable aléatoire. Dans un modèle stochastique, on s'intéresse à l'évolution de manière probabiliste du modèle. Dans notre cas, l'espace est dénombrable, on a donc un processus discret. On étudie ici l'évolution de la variable nombre de machines infectées en fonction du temps avec un pas de temps de 1 seconde ainsi que la probabilité à un instant donné, de pouvoir infecter une machine saine. Une approche déterministe n'aurait pas été préférable, car l'évolution de l'infection dépend de la probabilité à contaminer une nouvelle machine au cours du temps. Dans l'article, l'évolution de l'infection est étudiée selon 3 scénarios, dont deux possèdent un sous-réseau intermédiaire(la partie J du réseau). Or un modèle déterministe ne pourrait prendre en compte correctement la probabilité de l'infection de la passerelle (gateway) entre les deux sous-réseaux(I et K) ni la manière aléatoire de choisir les adresses à tester pour la propagation de l'infection.
De ce fait, une approche déterministe ne pourrait capturer les caractéristiques du phénomène étudié dans cet article. Une telle approche, donnerait en effet les mêmes résultats pour les trois scénarios, (Car étant par essence déterministe, elle ne peut saisir le côté aléatoire du phénomène). Ainsi, l'auteur ne pourrait conclure sur l'impact de la topologie d'un réseau sur la propagation du vers.

4 Performance et optimisation

4 . 1 Application du concept du double-tétraèdre

En nous situant dans un contexte d'optimisation, toujours en référant à l'article, nous pouvons situer notre double-tétraèdre dans un contexte d'optimisation. Il nous faut prendre en compte les différentes stratégies possibles pour éviter l'infection et la propagation des vers au niveau de l'attaqué. Alors que L'attaquant va chercher lui à infecter le plus grand nombre de machines, et le plus vite possible.

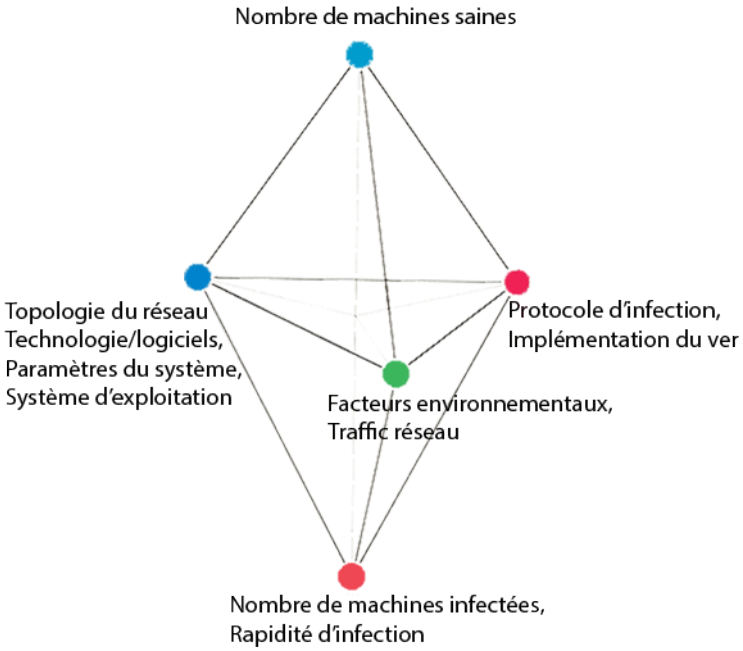

FIGURE 3 – concept du double tétraèdre appliqué à l'article

Les aspects du tétraèdre qui sont fixes sont : L'environnement externe sur lequel nous ne pouvons pas agir. C'est-à-dire les facteurs extérieurs dont qui ne sont pas modifiables ni par l'attaqué ni par l'attaquant. Dans le cas présent, un premier élément serait le réseau externe : aucun des deux partis n'a de contrôle sur l'acheminement des paquets. On peut également ajouter certains éléments de l'environnement interne : les ordinateurs du sous-réseaux. En effet, l'article ne s'intéresse qu'à l'effet de la topologie du réseau sur la propagation du ver. Ainsi, l'idée de modifier les ordinateurs afin de les protéger est hors champ. Par exemple, l'idée de fermer les ports non-indispensables est pertinente, mais ne rentre pas dans le cadre de cet article. De plus, il semble naturel, dans le cadre de cet article, de supposer que l'attaquant ne peut altérer les ordinateurs afin de les rendre vulnérables à l'infection. (Même si, dans le cadre général, quand on fait une analyse de risque, il faut considérer le cas d'une attaque physique). On peut donc ajouter ces derniers à l'environnement.

Les aspects du tétraèdres qui sont modifiés sont :

— Au niveau de l'attaqué, on va modifier les aspects de design et d'opération du réseau : La topologie du réseau peut être modifiée de telle sorte à ralentir la propagation d'une éventuelle infection.
— Au niveau de l'attaquant, on va modifier les aspects de design et d'opération du ver : Les caractéristiques du ver de telle sorte à ce qu'il se propage plus vite, qu'il soit moins rapidement détectable et tout cela pour chercher l'augmentation du nombre de machines infectées. (Comme exemple de modification de design, on peut considérer l'ajout d'une hitlist ou encore une meilleure répartition des adresses à attaquer entre les vers)(Ces concepts sont introduits dans l'article *how to own the internet on your spare time*, l'une des références de *Optimizing network against malware*).

4 . 2 Compléments d'études

Les autres aspects/méthodes qui 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 sont :

— Mise en place d'un IDS pour détecter les trafics étranges sur le réseau
— Mise en place d'anti-virus sur les machines
— Fermer les ports inutiles sur les machines, ne garder que les ports nécessaires actifs
— Essayer de détecter le ver à l'infection, par exemple si l'infection se fait par mail, le vers peut être détecter par un scan du mail
— Utilisation de différents systèmes d'exploitation et de logiciels (afin que les machines ne soient pas toutes vulnérables aux mêmes attaques)
— Éducation des utilisateurs sur les comportements à risque.
— Mise à jour régulière des logiciels pour combler des failles potentielles

4 . 3 La diversité dans un réseau informatique

On peut appliquer le concept de diversité au sein d'un réseau informatique en utilisant des systèmes d'exploitation différents(Windows, Linus, Mac OS) selon le rôle de la machine dans le réseau. Cette diversité prendrait la forme d'une diversité au niveau logiciel(client de mail différent, système d'exploitation, ...), les failles de sécurité ne seront plus les mêmes. Il pourrait aussi y avoir une diversité au niveau du matériel utilisé pour avoir des drivers différents, et éventuellement limiter le nombre de machines pouvant être infectées par une faille d'un driver. L'idée serait donc d'avoir des machines les plus différentes possibles. En effet, les systèmes d'exploitations différents ne fonctionnent pas de la même façon, les logiciels et les programmes ne se lancent pas de la même manière. Par conséquent les failles ne sont pas les mêmes, ce qui obligerait l'attaquant à créer un ver beaucoup plus complexe qui pourrait être capable d'infecter des systèmes très différents. On peut également envisager une diversité des utilisateurs, au sens où il y aurait des utilisateurs à faibles privilèges et des administrateurs. Les utilisateurs à faibles privilèges ne pourraient lancer que certains programmes, mais ne pourraient pas lancer d'installation sans intervention de l'administrateur. Ainsi, il faudrait que le ver infecte un administrateur afin de pouvoir se propager efficacement. (On supposerait que les administrateurs sont moins nombreux et mieux formés que les autres). On pourrait penser même à une hiérarchisation des droits d'accès des employés selon leur poste et leur besoin dans le réseau. L'idée étant d'attribuer aux employés juste les droits sur les fichiers/systèmes dont ils ont besoin pour leur travail et rien de plus. Par exemple, en plus de la séparation du réseau selon le secteur dans l'entreprise, les employés n'auront pas accès, ni ne pourront se connecter aux machines qui ne sont pas dans leur secteur de travail.