SCIENCES SORBONNE UNIVERSITÉ

POLYTECH INGENIEURS PARIS-UPMC

Rapport d'ARE DYNAMIC

Le Virus Informatique

BAILLY Antoine
MATHEU Edouard
GUILY Florian
OUERTATANI Achref

A travers ce projet nous avons essayé de modéliser la propagation d'un virus informatique au sein d'un ensemble d'ordinateurs et de serveurs reliés en réseau interne et public. Pour cela, après quelques recherches parfois non abouties sur le phénomène de propagation d'un virus informatique, nous avons mis en place un modèle pour la propagation d'un virus informatique. Un réseau d'utilisateurs et de serveurs généré aléatoirement a été créé. L'équipe de développement s'est répartie les tâches en 2 groupes : l'un s'occupant de la création du réseau et de la propagation du virus, l'autre s'occupant de l'affichage. Ainsi, nous avons pu obtenir un réseau dans lequel nous pouvons observer la propagation d'un virus informatique au cours du temps.

Introduction:

Ce projet a pour objectif de modéliser la propagation d'un virus informatique au sein d'un ensemble d'ordinateurs et de serveurs reliés en réseau interne et public. Nous allons donc à travers ce projet essayer de répondre à la problématique suivante : Comment la propagation d'un virus informatique évolue-t-elle en fonction de divers facteurs ? Pour cela il faut savoir qu'il existe 3 types de virus: le virus classique qui se réplique dans un environnement informatique, le ver informatique qui se répand dans un réseau de machines et le cheval de Troie qui est plutôt une plateforme qui délivre un virus ou un ver. Ainsi, l'équipe de développement composée de MATHEU Edouard, GUILY Florian, OUERTATANI Achref et BAILLY Antoine, après avoir eu l'idée de plusieurs modèles de propagation (propagation entre les serveurs, propagation entre les utilisateurs , propagation en peer-to-peer), s'est mise d'accord sur le modèle à adopter.
Nous essaierons dans notre développement de répondre à ces questions:

- Comment modéliser le comportement d'un virus dans un environnement informatique ?
- Quels sont les facteurs de propagation d'un virus informatique ? Comment influencent-ils sa propagation ?
- Comment se protéger des virus informatiques ?

Ainsi nous allons voir dans un premier temps ce qu'est un virus informatique , puis dans un second temps nous verrons les étapes de notre modélisation, enfin nous verrons les limites de notre modèle avant de conclure.

auteur : Antoine Bailly
relecteurs: Edouard Matheu , Florian Guily, Achref Ouertatani

I) Le virus informatique

1) Qu'est ce qu'un virus informatique

Comment le virus peut-il atteindre un ordinateur? Le virus informatique affecte une machine de la même façon qu'un virus biologique affecte un corps. En effet, il existe 3 types de virus:

- Le virus classique qui se réplique dans un environnement informatique vise à endommager le système en se comportant comme des bouts de code inutiles qui viennent s'implanter dans le programme pour l'endommager.
- Le ver informatique est un programme autonome qui s'installe sur le disque dur d'un ordinateur (disque dur externe, disque amovible). Le ver informatique a besoin d'un hôte pour se reproduire, pour cela il se dissimule et se cache dans des fichiers et des codes exécutables contenus dans l'ordinateur cible. Les vers écrits sous forme de script peuvent être intégrés dans un courriel ou sur une page HTML. Le ver a aussi, mise à part sa reproduction, certains objectifs malveillants :
    - Espionner l'ordinateur qu'il a pour hôte
    - Détruire des données sur l'ordinateur où il se trouve ou y faire d'autres dégâts
    - Installer d'autres programmes nocifs : spywares, ou cheval de Troie, dans le but de capturer des mots de passe, des numéros de cartes bancaires…
    - Prendre le contrôle de l'ordinateur à distance afin d'en faire une passerelle, un relais pour envoyer de multiples requêtes vers un serveur internet dans le but de le saturer.
    - Diffère d'un virus : Un ver informatique a besoin d'un hôte pour se reproduire.
- Le cheval de Troie qui est plutôt une plateforme qui délivre un virus ou un ver; on donne l'exemple des jeux téléchargeables en ligne, ces jeux peuvent contenir le virus et à chaque fois qu'on joue, le virus se propage encore plus. Le virus informatique peut détruire des fichiers, ralentir les serveurs et les réseaux, donner un accès facile à l'espion grâce au backdoor.

auteur : Antoine Bailly
relecteurs: Edouard Matheu, Florian Guily, Achref Ouertatani

2) Notions fondamentales pour la mise en place de notre modèle

Nous avons d'abord fait des recherches sur l'organisation d'un réseau informatique ou plus largement "internet". Le virus se déplaçant dans cet environnement, celui-ci doit être conforme à la réalité.
Nous nous sommes donc mis d'accord sur le modèle suivant:
L'environnement sera composé de trois agents:

- les serveurs globaux (serveurs fournissants des fichiers ou pages web tel que google etc…), connectés à un grand nombre d'individus et à quelques des serveurs locaux.
- les individus ou usagers, connectés à un serveur local ou global (les deux sont impossible).
- les serveurs locaux, faisant la liaisons entre plusieurs individus et un serveur global.

Les serveurs globaux seront les centres névralgiques du réseau. Un grand nombre d'utilisateur y seront connectés ainsi que certains réseaux locaux. Ces réseaux locaux

sont utilisés pour modéliser une famille ayant plusieurs appareil inter-connectés ou une entreprise. Ces groupes d'utilisateurs restreint utilise la même connection à internet, représentée par nos réseaux locaux. Ces derniers servent donc de "passerelle" entre plusieurs utilisateur et les réseaux globaux. Quand aux utilisateurs, ils sont soit connectés à un serveur global, soit à un serveur local. Un individu ne peut être connecté à plusieurs serveurs globaux ou locaux.

Par la suite, nous avons cherché à savoir comment un virus se propage sur internet.
Nous avons donc trouvé trois principaux moyens de propagation :

- soit il affecte les agents qui sont connectés à lui
- soit il affecte quelqu'un qui n'est pas directement connecté à lui : peer-to-peer via mail ou échange de support de stockage externe, etc …
- si jamais il affecte un réseau local, toutes les personnes connectées à ce réseau local sont affectées.

Nous avons donc pris en compte ces trois moyens de propagation pour notre modélisation.
Nous avons aussi regroupé différentes variables liées à la protection en une seule : possession ou non d'un pare-feu, la vigilance de l'utilisateur vis-à-vis des virus (comportement sur internet par exemple) , etc..

auteur : Florian Guily
relecteurs: Antoine Bailly, Edouard Matheu, Achref Ouertatani

II) Etapes de la modélisation

Répartition des tâches :

Deux équipes ont été formées afin d'avancer plus rapidement. Edouard MATHEU et Achref OUERTATANI s'occupait de la création du réseau et de la propagation du virus informatique d'un côté et de l'autre Florian Guily et Antoine BAILLY s'occupait de l'affichage du réseau et de la visualisation du virus.

auteur : Antoine Bailly
relecteurs: Edouard Matheu , Florian Guily, Achref Ouertatani

1) Création du réseau et mise en place de la propagation du virus dans ce réseau

Pour la première équipe, la première étape consistait en la création du réseau informatique de manière procédurale.
Notre réseau devait être constitué par différents agents connectés entre eux.
Nous avons donc tout d'abord créé la population d'individus, représentant des ordinateurs personnels.
Cette population a été modélisé sous la forme d'un dictionnaire avec l'adresse d'un individu (un entier) comme clé et une liste comprenant différentes informations propres à l'individu:

- Infection : un booléen indiquant si l'individu est infecté ou non (True pour infecté)
- Antivirus : un booléen indiquant si l'individu possède un antivirus (True si l'individu en possède un)

- Protection : un flottant représentant la protection et la vigilance d'un individu de manière générale (protection par un pare-feu, comportements à risques sur internet, etc.). Il va de 0.0 à 1.0, 1.0 étant la protection maximale.
- Temps pour la désinfection : float directement lié à protection, indiquant le temps que l'individu prendra pour désinfecter son ordinateur manuellement (réinitialisation de l'ordinateur, envoi en réparation, etc.). Plus la protection de l'individu est élevé plus le temps sera court.
- Un ensemble des individus connectés à l'individu (ici vide, afin de respecter l'homogénéité des différents dictionnaires des agents dans le réseau, ce qui est nécessaire pour certaines fonctions)
- Une chaîne de caractères (adresse du serveur local connecté à l'individu s'il existe, vide sinon).
- Un ensemble (serveurs globaux connectés à l'individu).

Le nombre d'individus dans le dictionnaire est lié à une variable globale et les valeurs de la liste sont créées de manière procédurale :

- Tous les individus sont sains dans un premier temps (le « patient zéro » est sélectionné plus tard)
- Ils sont tous protégés par un antivirus (les individus non protégés sont sélectionnés plus tard)
- La protection est calculée aléatoirement
- Le temps de désinfection est inversement proportionnel à la protection
- Les individus ne sont pas connectés à quoi que ce soit pour l'instant (puisque les serveurs n'ont pas encore été créés.)

Nous avons ensuite créé deux autres dictionnaires, en se basant sur les mêmes principes (en changeant juste quelques conditions pour le calcul des valeurs, notamment en connectant les agents entre eux), pour créer les serveurs locaux et globaux. Pour enfin fusionner ces dictionnaires dans un dictionnaire général représentant le réseau.
Puis nous avons modélisé la propagation du virus en commençant par une fonction simulant un échange de données.
La structure principale de cette fonction était la suivante : On prend deux individus (en paramètre), si l'individu 1 est infecté et que l'infectiosité du virus est supérieure à la protection du deuxième individu, alors il devient infecté.

Nous avons utilisé cette fonction pour créer une fonction simulant une journée dans le réseau. Cette fonction effectue un certain nombre d'échanges de données, soit entre deux agents directement connectés dans le réseau, soit entre deux individus sélectionnés aléatoirement (en peer-to-peer, par mail, échange de support de stockage externe, etc.).

Nous avons aussi créé un agent antivirus, sous la forme d'une liste, composée d'un booléen indiquant si le virus est enregistré dans la base de données de l'antivirus, d'un ensemble composé des agents possédant un antivirus et d'un facteur efficacité qui indique le nombres personne au bout desquelles l'antivirus détecte le virus).
Ils nous a alors fallu intégrer l'antivirus dans différentes fonctions, tout d'abord dans la fonction « échange de données », où si le premier individu est infecté et le deuxième possède un antivirus, l'antivirus a une certaine probabilité de détecter le virus et l'enregistrer dans sa base de données.

Nous avons aussi sélectionné 24% des individus et des serveurs locaux pour indiquer qu'ils ne possédaient pas d'antivirus car d'après les derniers rapports des taux de protections en 2012 les statistiques montrent que 24% des PC dans le monde ne sont pas protégés contre les virus.
Enfin nous avons modifié la fonction « journée », pour qu'à la fin d'une journée, si l'antivirus a le virus enregistré dans sa base de données, il désinfecte tous les possesseurs d'antivirus, et les protège contre celui-ci en augmentant leur protection à 1.0.

Pour finir nous avons modélisé la désinfection manuelle des agents affectés, que nous avons intégré dans la fonction « journée » : Quand un individu est infecté, on lui attribue un décompte de jours au bout duquel il sera désinfecté. Á la fin d'une journée, ce décompte est décrémenté.

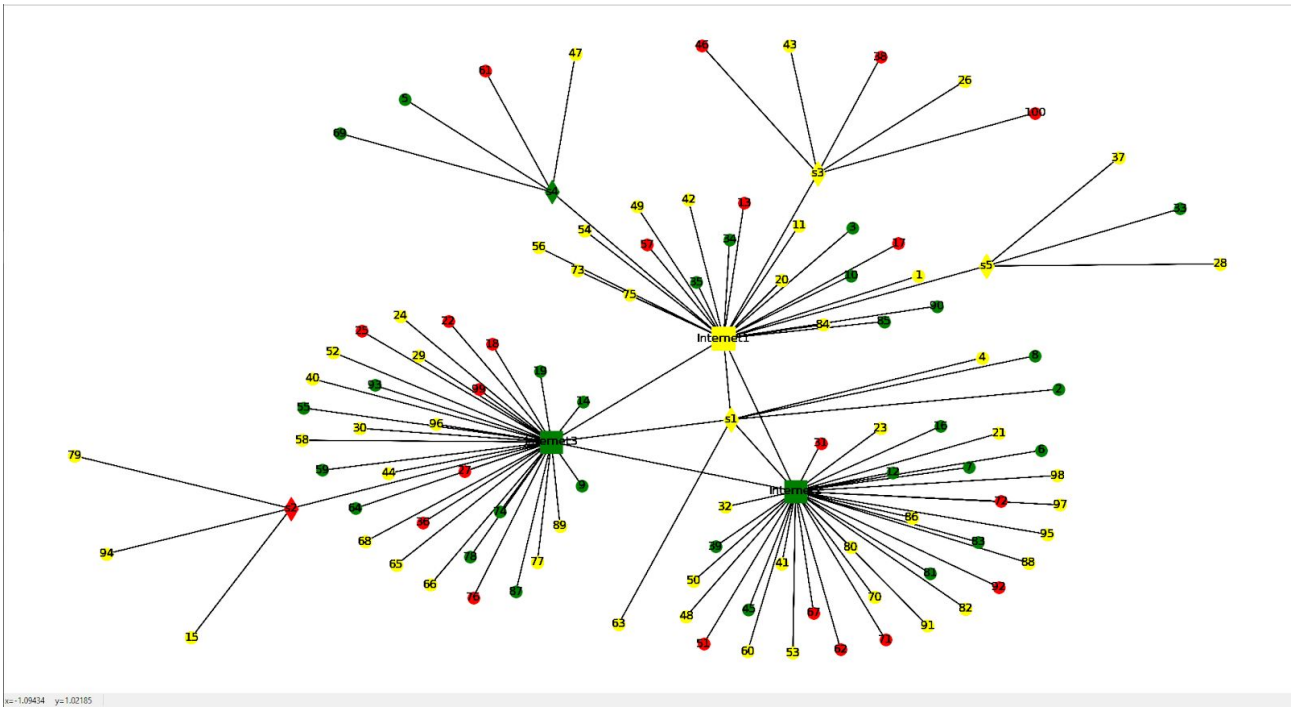
auteur : Edouard Matheu
relecteurs: Antoine Bailly , Florian Guily, Achref Ouertatani

2) Affichage et visualisation de la propagation du virus(sortie)

Pour l'équipe s'occupant de l'affichage durant le projet, il y a eu 2 semaines de recherche et de manipulation (prise en main) d'une bibliothèque permettant de visualiser un réseau. Celle qui a été retenue est la bibliothèque NetworkX. Cette bibliothèque permet la création de noeuds et de liaisons entre les noeuds ce qui est parfait pour modéliser notre réseau. Ainsi nous avons créer pour chaque utilisateur un noeud possédant les mêmes caractéristiques que les utilisateurs créer par l'autre groupe . Nous avons de même créé des noeuds pour les serveurs locaux et globaux (de formes différentes pour la visualisation).

Une fois les noeuds conçus nous avons reliés noeuds et serveurs par des liaisons en fonction de la base de donnée des connexions établies par l'autre groupe en leur donnant une forme/structure afin que cela soit plus visuel et plus simple à comprendre.

Par la suite afin de visualiser le virus, nous avons mis en place un « code couleur » pour que le virus laisse une trace : vert l'individu n'est pas infecté, rouge l'individu est infecté, jaune l'antivirus a "éradiqué" le virus. Ainsi l'autre groupe pouvait visualiser son travail, ce qui pouvait l'aider afin de se corriger ou avancer dans leur travail.
auteur: Florian GUILY ; relecteurs: Antoine BAILLY, Achref OUERTATANI, Edouard MATHEU



III)Resultat,analyse et remarque

1) Analyse des Expériences

Ainsi le modèle établit rendait des résultats différents et variés suite au rôle de l'aléatoire dans le modèle mais surtout le rôle des facteurs. Ensuite on a essayé de répondre à notre problématique et aux questions que l'on se posait initialement.

a) Influence du réseau sur la propagation du virus

Le premier réflexe que l'on a eu dans l'analyse de donnée était de voir l'influence de notre réseaux sur la propagation. On a donc fait varier les deux paramètres du modèle, le nombre des serveurs locaux et le nombres d'utilisateurs en %. (voir annexe 1 et 2)

Commentaire: On remarque que l'effet du nombre d'utilisateur ou d'agent ne peut pas apporter un gros changement au pourcentage d'infectés et du coup ce pourcentage reste constant. Pourtant les graphiques n'affichent pas une droite, les courbures sont dû à l'effet de l'aléatoire

Conclusion: Le nombre d'individus n'influence pas la propagation du virus et donc notre modèle pourrait éventuellement échapper à la limite d'effectifs.

Une animation de comparaison a été aussi effectuée pour montrer la différence entre un modèle de propagation par réseau uniquement et un modèle de propagation avec le peer to peer.

Conclusion: le peer to peer a permis de ne pas se bloquer et de se détacher du caractère géographique que le réseau lui impose.

b) Les caractéristiques du virus qui influent sa propagation

Le premier facteur qui saute au yeux et qui est évident et essentiel à étudier est l'infectiosité du virus. Ainsi, le premier graphique a servi à exploiter le nombre d'infectés en fonction de l'infectiosité du virus. (voir annexe 3)

Commentaire: Notre courbe a la forme d'une exponentielle croissante et cela décrit le comportement du virus.

Conclusion: Donc plus l'infectiosité du virus est élevée, plus les individus auront de chances d'être infectés.

On a ensuite vu qu'il était intéressant d'étudier si les différents types de virus pouvaient se propager différemment. On a donc choisi deux types de logiciels malveillants que l'on a rencontré pendant nos recherches : les virus classiques et les vers. (voir annexe 4)

Commentaire: La courbe de viralité du vers dépasse celle du virus classique mais d'une valeur très petite et dans certains cas les courbes sont confondues.

Conclusion: Au contraire des attentes les vers et les virus affectent le même taux de utilisateurs même s'ils suivent des chemin de propagation différents car les vers ciblent plus et se propagent plus rapidement sur les serveurs locaux. Donc finalement le type du virus n'affecte pas la propagation.

Finalement, toujours dans les variables du virus, on a fait varier le patient 0 entre un serveur global , un serveur local et un individu quelconque. Pour cela on a créé un graph avec 3 courbes qui indiquent le changement du patient 0 et décrivent la viralité du virus. (voir annexe 5)

Commentaire: Les deux courbes de serveurs locaux et d'utilisateurs sont confondues mais celle du serveur globale est supérieure au deux autres.

Conclusion: Le patient 0 peut être un facteur intéressant dans le cas où c'est un serveur global ce qui est conforme au modèle car d'un côté les serveurs globaux sont mieux protégés et d'un autre côté ils ont plus de connections. Ainsi la propagation du virus varie si le patient 0 et serveur globale ou non.

c) L'importance de l'antivirus

À un moment dans la modélisation nous avons vu qu'il fallait un moyen pour éradiquer le virus car même les virus les plus célèbres ont été éradiqués à un moment donné. On a donc fait varier l'efficacité d'un antivirus en fonction du nombre d'infectés. (voir annexe 6)

Commentaire: Notre courbe est une courbe croissante qui tend vers un maximum de nombre d'individus infectés. La courbe 2 le montre et la courbe 1 (zoom de courbe 2) montre plus le comportement de l'antivirus.

Conclusion: L'efficacité de l'antivirus traduit son taux de détectabilité afin d'ajouter la signature du virus dans sa base viral. Une fois que le virus est détecté le nombre des individus infectés diminue considérablement. Mais la remarque la plus importante ici c'est que si on dépasse un certain taux d'efficacité il est inutile de l'augmenter encore plus. Dans notre cas à partir d'une détection, au bout de 500 rencontres avec le virus, elle est considéré comme très efficace dans le cas où 76% de la population est couverte (voir source).

On a aussi fait varier la couverture d'un virus (les individus et serveurs locaux qui possèdent un antivirus) et on voit sur le dernier graphique l'évolution du nombre d'infectés (voir annexes 7)

Commentaire: Notre courbe est décroissantes mais on peut distinguer 3 parties.
première et troisième parties: décroissance faible; et deuxième partie: décroissance brutale.

Conclusion: Dans ce graphique on a considéré l'efficacité du virus égale à 1000 rencontres. La décroissance brutale décrit bien l'efficacité d'un antivirus et on peut même en déduire une valeur qui permettra de juger si un antivirus est efficace sur une population quelconque ou pas.

auteur : Achref Ouertatani
relecteurs: Edouard Matheu , Florian Guiiy, Antoine Baillyi

2) Limites du modèle

Notre modélisation de la propagation du virus au sein du réseau informatique ne peut pas être prise comme modèle de référence car il manque de nombreux éléments pour que ce modèle soit considéré comme complet notamment à cause du peu de données dont nous disposions sur le nombre d'infectés par des virus informatiques dans le monde notamment.

En effet, pour rendre la modélisation de ce virus possible dans le court laps de temps que nous avions et dans la mesure de nos capacités, nous avons été contraints de ne pas tenir compte de certains paramètres/facteurs. Ces facteurs sont aussi bien internes au réseau qu'externes.

Le principal facteur externe que nous n'avons pas pris en compte dans notre modélisation est le fait qu'un virus informatique peut être en permanente évolution. Dans ce cas, il ne cesserait de se modifier au cours du temps. De ce fait, un unique correctif ne pourrait éradiquer ce virus qu'à moyen terme du réseau . Le virus ayant été modifié, il ne disparaîtrait pas totalement du réseau. Or, dans notre modèle, nous prenons comme hypothèse que le virus n'évolue pas au cours du temps, et donc, qu'un correctif de sécurité mis en place permet de définitivement protéger le poste de travail contre le virus.

De plus, notre virus se propage au sein d'un petit réseau et n'est donc pas à la vraie échelle. Notre modélisation du réseau ne peut être parfaitement réaliste car nous l'avons créé de manière aléatoire donc sans les vraies bases de données Internet.

Par ailleurs, notre modélisation insiste sur la manière de propagation du virus dans le réseau mais ne se concentre peut être pas assez sur le comportement du virus lui-même c'est-à-dire la manière dont il affecte un utilisateur: détecté ou non ,cheval de Troie ,etc.

auteur : Antoine Bailly
relecteurs: Edouard Matheu , Florian Guily, Achref Ouertatani

Conclusion:

Ainsi grâce à ce projet nous avons pu modéliser grâce à chaque membre de l'équipe la propagation d'un virus informatique à travers un modèle dont nous sommes plutôt fier car il correspond aux attentes que nous avions initialement puisque nous pouvons voir comment se propage un virus informatique en fonction de divers facteurs (temps , type de virus , infectiosité, …). Nous avons apprécié de travailler en équipe notamment grâce aux 2 groupes de travail formés (Florian GUILY/Antoine BAILLY et Edouard MATHEU/Achref OUERTATANI) car cela nous a permis de partager les tâches et d'observer des résultats plus rapidement (quand le groupe de conception avait terminé quelque chose le groupe affichage pouvait leur montrer visuellement le résultat obtenu par leur code). Les deux groupes de travail s'entraidaient pour trouver des nouvelles idées et contourner des problèmes tous ensemble. Ce projet nous a aussi permis de nous donner une idée de la façon dont nous serions amené à travailler plus tard : en équipe et en collaboration avec d'autres personnes.

Summary:

Throught this project, we tried to modelize the spread of a computer virus within a set of computers and servers connected to an internal and public network. Therefore, after some research sometimes unsuccessful on the propagation phenomenon of a virus ,we set up a propagation model of the virus. A network of users and servers generated randomly has been created. The development team split in 2 in order to share the work. One team created the network and the virus propagation and the other worked on the display but the two team worked together to share ideas for example. Thus we can present a network showing the propagation of a virus over time.

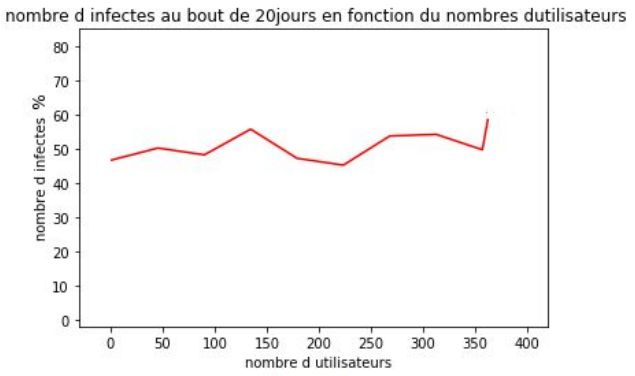
auteur : Achref Ouertatani
relecteurs: Edouard Matheu , Florian Guiiy, Antoine Bailly

Annexes:

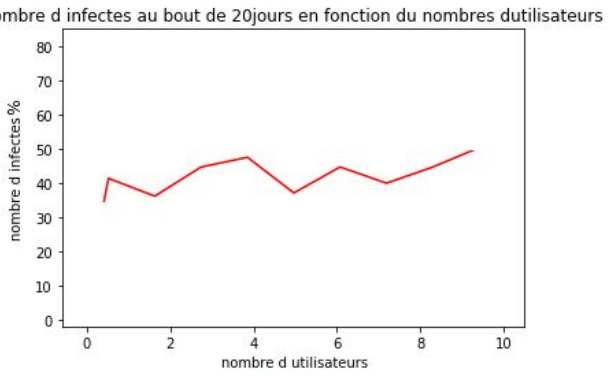
Tableau ARE - Feuille 1 (2).pdf

Attribut des agents				
Utilisateurs : les utilisateurs connectes a un serveur local et seulement ce serveur				
Nom	Type	Intervalle	Valeur initiale	Fixe
Infection	Bool	/	"False"	non
Possède un antivirus	Bool	/	Aléatoire (76% de Vrai)	oui
Protection	Float	[0,1]	Aléatoire	oui
Taux de désinfection	Float	[0,1]	depend de la protection	oui
Personnes connectées	set[int]	/	Aléatoire	oui
Serveur local connecté	str	/	Aléatoire	oui
Serveurs globaux connectés	set[str]	/	Aléatoire	oui
Serveurs globaux : (tout les serveurs globales sont liee entre eux)				
Nom	Type	Intervalle	Valeur initiale	Fixe
Infection	Bool	/	"False"	non
Possède un antivirus	Bool	/	"True"	oui
Protection	Float	[0,1]	Aléatoire	oui
Taux de désinfection	Float	[0,1]	Aléatoire	oui
Personnes connectées	set[int]	/	Aléatoire	oui
Serveur local connecté	str	/	""	oui
Serveurs globaux connectés	set[str]	/	Ensemble des serveurs globaux	oui
Antivirus :				
Nom	Type	Intervalle	Valeur initiale	Fixe
Connaissance du virus	Bool	/	"False"	non
Personnes protégées	set[alpha]	/	Aléatoire (76% des personnes et serveurs locaux, plus les serveurs globaux)	oui
efficacités	int	[0, + infinie[	variable	
Serveurs locaux : connectes a au moins un serveur local				
Nom	Type	Intervalle	Valeur initiale	Fixe
Infection	Bool	/	"False"	non
Possède un antivirus	Bool	/	Aléatoire (76% de Vrai)	oui
Protection	Float	[0,1]	Aléatoire	oui
Taux de désinfection	Float	[0,1]	Aléatoire	oui
Personnes connectées	set[int]	/	Aléatoire	oui
Serveur local connecté	str	/	""	oui
Serveurs globaux connectés	set[str]	/	Aléatoire	oui
Virus :				
Nom	Type	Intervalle	Valeur initiale	Fixe
Infectiosité	Float	[0,1]	variable	oui
Type	str	"vers"/"virus"	"False"	oui
patient 0	alpha (int ou str)	adresse	variable	oui
Paramètres du modèle				
Nom	Type	Intervalle	Valeur initiale	Fixe
Nombre d'utilisateurs	int	[1,N]	N	oui
Nombre serveur local	int	[1,NSL]	NSL	oui
				oui
Environnement : Environnement informatique				
Exprerience:	viralité du virus	type de logiciel malveillant	l'effet du patient 0	
Effet parametre Y sur Indicateur X:	nb infectés en fonction de l'infectiosité du virus	compararaison entre la viralité d'un virus et d'un vers	comparaison entre la viralité du meme virus mais des patients 0 diffrents	
Experience:	efficacité de l'antivirus	mode de propagation		
Effet parametre Y sur Indicateur X:	nb infectés en fonction de l'efficacité de l'antivirus	comparaison entre un modele de propagation par liaison et un modele de propagation par a la fois liaison et un pear to pear		

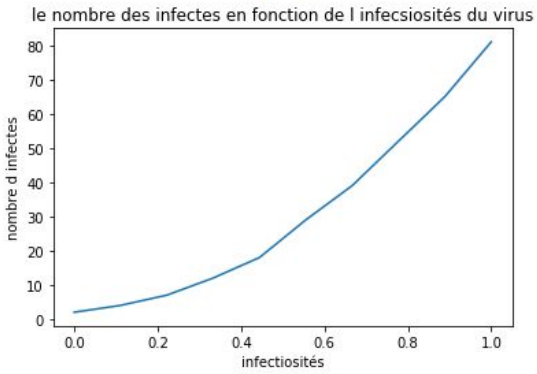
Indicateurs				
Nombre d'infectés	int	[0,108]		
Nombre de désinfectés	int	[0,108]		



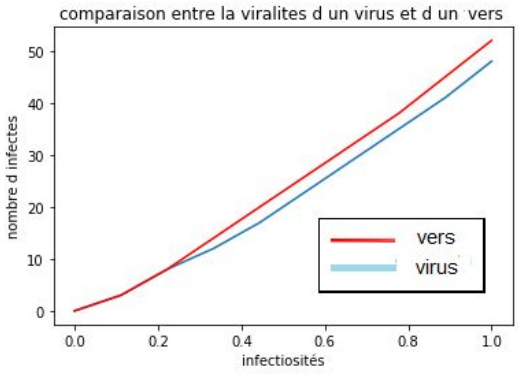
Annexe 1



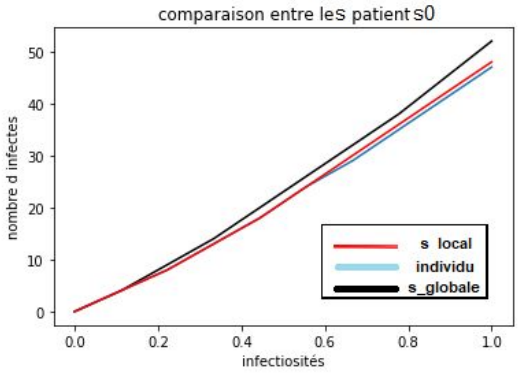
Annexe 2



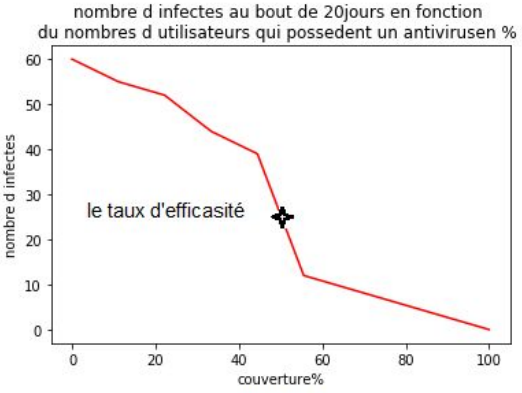
Annexe 3



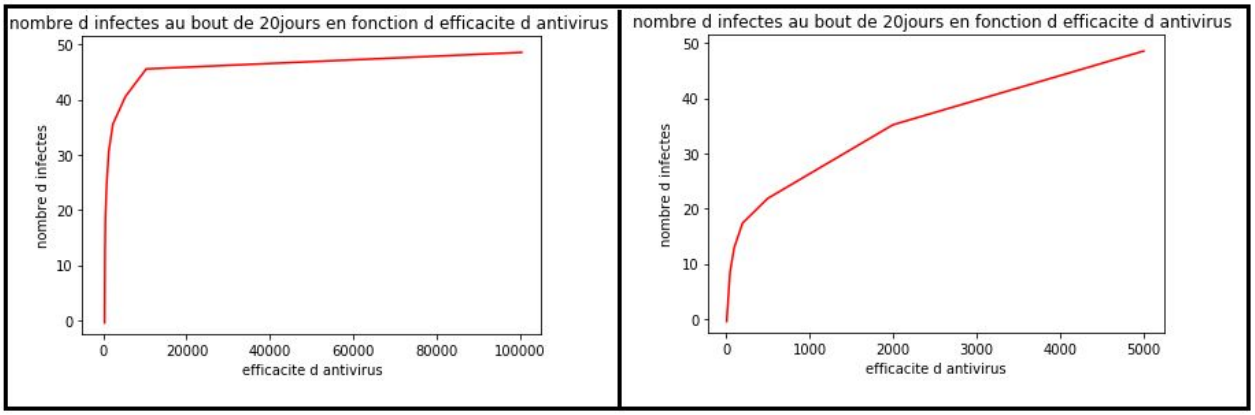
Annexe 4



Annexe 5



Annexe 7



Annexe 6

Source :

https://blogs.technet.microsoft.com/microsoft_blog/2013/04/17/latest-security-intelligence-report-shows-24-percent-of-pcs-are-unprotected/