# La virologie informatique

Eric FILIOL et Jean-Yves MARION

**L'émergence des virus informatiques échappe à tout contrôle. Seules des mesures « sanitaires » de prévention et de surveillance nous permettront de juguler les épidémies virales sur les réseaux.**

Dans le film Terminator 3, sorti en 2003, le système Skynet prend le contrôle des réseaux de défense par le biais d'un virus. La réalité rejoint la fiction. En effet, un tel scénario est désormais possible tant notre dépendance à l'informatique est grande et l'interconnexion illimitée de nos systèmes, incontrôlable. Le 2 novembre 1988, le virus Morris a attaqué avec succès environ 6 000 ordinateurs connectés au réseau Internet. Quinze ans plus tard, le 25 janvier 2003 à 5 h 30, temps universel, le virus Slammer a paralysé le réseau Internet en se servant d'une faille informatique découverte six mois plus tôt. En 10 minutes, ce virus s'est dupliqué et a contaminé 90 pour cent des machines qui possédaient cette faille. Aujourd'hui, un virus correctement programmé serait capable de paralyser Internet en quelques secondes.

Les épidémies virales informatiques coûtent des milliards d'euros de dégâts et représentent une réelle menace pour notre société. Les infrastructures actuelles comme les communications, le transport d'énergie ou la distribution de biens dépendent des réseaux informatiques. Une défaillance, due en particulier à un acte malveillant, nous priverait de ces services. Une attaque virale, comme celle réussie par le virus Slammer, bloque l'accès aux ressources qui sont contenues dans des systèmes d'information répartis sur Internet. À titre d'illustration, le réseau informatique de surveillance de la centrale nucléaire Besse-Davis dans l'Ohio, aux États-Unis, a été paralysé pendant près de 24 heures suite à l'infection par le virus Slammer.

D'autres types d'attaques sont capables de défigurer un site Internet, de modifier certaines informations, de nous en dérober ou pire, d'utiliser nos propres systèmes pour commettre crimes et délits, à notre insu. On imagine les conséquences possibles de tels détournements. Pour mener à bien leurs méfaits, les codes malveillants sont devenus plus discrets, comme des taupes qui espionnent un système. Ces nouveaux intrus se révèlent, en fonction de qui les contrôle, des armes de guerre économique, mais aussi politique.

Les virus sont-ils l'équivalent moderne et technologique des infections biologiques ? Les épidémies de virus informatiques représentent-elles une nouvelle menace, comme la guerre bactériologique ? Jusqu'où va la comparaison avec la biologique ? Et si l'expérience acquise en luttant contre les virus biologiques nous permettait de lutter contre leurs homologues informatiques... Les réponses

Shutterstock/Mikhail Levin

à ces questions demandent, tout d'abord, une connaissance des virus informatiques.

## Quand un programme viral prend vie

Les premiers pas de la virologie informatique datent de 1986, lorsque Fred Cohen présenta sa thèse de doctorat de l'Université de Californie du Sud, aux États-Unis. Il y introduisit la notion de « virus informatique » par ces mots : « On pourrait définir un virus approximativement comme une séquence de symboles qui, selon une interprétation et dans un environnement donné, provoque la modification d'autres séquences de symboles dans ce même environnement, de telle sorte qu'elles contiennent des virus (eux-mêmes éventuellement modifiés) ».

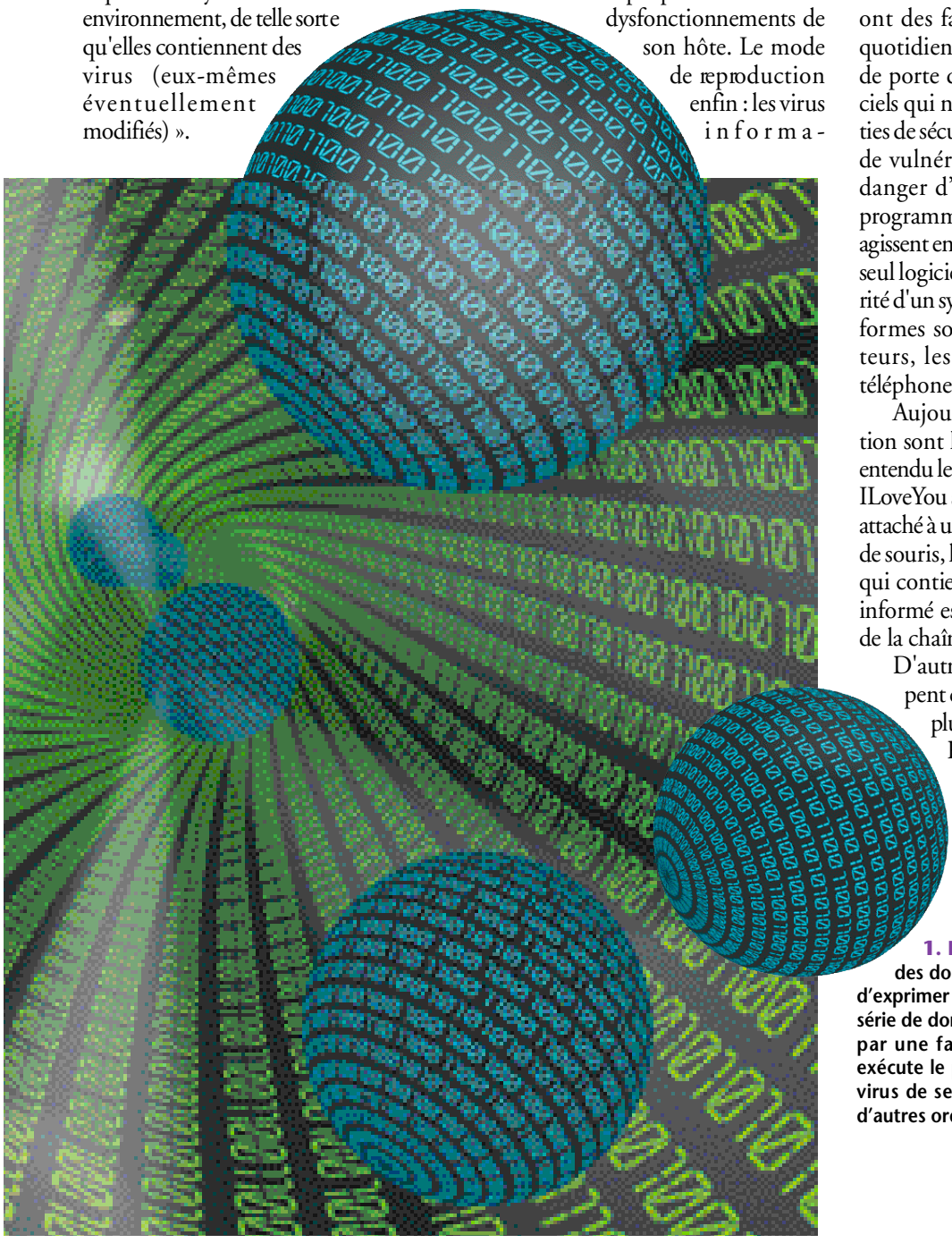
Le choix du terme « virus » suggère une analogie entre le monde de l'informatique et celui de la biologie. En effet, plusieurs correspondances existent entre ces deux mondes. La structure d'abord : un virus biologique est composé d'un filament d'acide nucléique qui correspond à une séquence de lettres qui codent des fonctions précises, comme un programme. Or, un virus informatique est un programme qui s'exécute une fois qu'il a pénétré dans un système. La fonction des deux types de virus ensuite : ils provoquent une infection. Le virus informatique force et détourne les ressources logicielles des systèmes informatiques pour son propre compte, et il engendre dans la plupart des cas des dysfonctionnements de son hôte. Le mode de reproduction enfin : les virus informatiques semblent avoir en commun avec leurs « semblables » biologiques, de se recopier indéfiniment et de muter.

La notion d'exécution de programmes convient également au monde cellulaire. En effet, le cycle de reproduction des virus aux seins des cellules s'exécute selon un schéma, ou programme, unique et codé par le génome viral. Toutefois, en informatique, la notion de programme et de donnée n'existe pas à proprement parler, car un ordinateur manipule des symboles. Toute donnée est un programme potentiel, et inversement, tout programme est une donnée. Le revers de la médaille est l'exploitation de cette ambivalence pour prendre le contrôle d'un système.

Les logiciels que nous utilisons ont des failles : nous les éprouvons quotidiennement. Ces failles servent de porte d'entrée aux virus. Les logiciels qui n'offrent pas toutes les garanties de sécurité sont une source majeure de vulnérabilité. Ils représentent un danger d'autant plus grand que les programmes communiquent et interagissent entre eux de telle manière qu'un seul logiciel vulnérable menace la sécurité d'un système entier. Toutes les plateformes sont concernées : les ordinateurs, les assistants personnels, les téléphones portables, etc.

Aujourd'hui, les vecteurs d'infection sont les dvd, les clés USB et bien entendu le réseau Internet. Ainsi, le virus ILoveYou se présente comme un fichier attaché à un courriel. Par un simple click de souris, l'utilisateur exécute ce fichier qui contient le virus. L'utilisateur mal informé est souvent un maillon faible de la chaîne infectieuse.

D'autres modes d'infections échappent cependant aux utilisateurs les plus avertis. Ainsi, les virus Code Red, Slammer ou Blaster emploient une méthode dite par « débordement de mémoire ». Lorsqu'un utilisateur se connecte à un système, ce dernier lui



**1. Les virus informatiques** sont des données numériques susceptibles d'exprimer un code malveillant. Lorsqu'une série de données pénètre dans un système par une faille de sécurité, l'ordinateur exécute le programme viral et permet au virus de se reproduire et de contaminer d'autres ordinateurs.

demande un mot de passe, constitué de huit lettres, par exemple. Pour livrer son attaque, le virus se fait passer pour un utilisateur et propose un mot de passe plus long, de plus de huit caractères. L'effet recherché est un débordement de mémoire. Si le système est vulnérable à ce type d'attaque, il est probable qu'il commette une faute. Afin de prendre le contrôle du système, les programmeurs de virus codent un mot de passe qui est en lui-même un programme exécutable. Lorsque le système lit le mot de passe, trop long, il est possible qu'il commette une erreur qui entraîne l'exécution du code malveillant. Le virus pénètre alors dans le système.

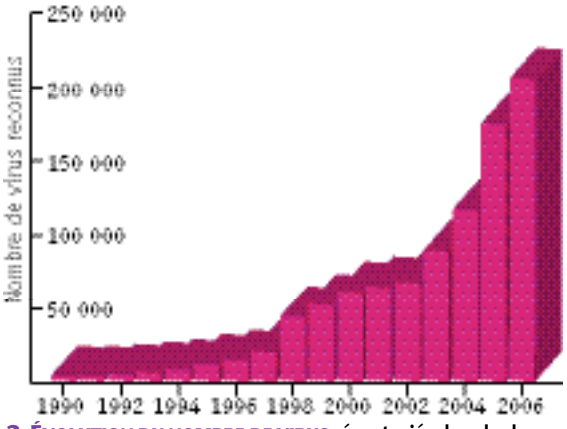
## Des failles à usage immédiat

La majorité des failles de ce type sont connues et corrigées par les éditeurs informatiques. Mais d'autres ne le sont pas : ce sont les failles logicielles dites 0-Days. Connues des seuls pirates, ces derniers les exploitent pour perpétrer leurs attaques. L'éditeur, qui ne les découvre que plus tard, n'est pas en mesure de publier un correctif immédiatement. Ce fut le cas lors de l'attaque qui a frappé, en janvier 2006, les installations du parlement britannique, notamment. Le code était dissimulé au sein de simples images, et le correctif n'a été disponible que trois semaines plus tard. Pendant ce temps, aucune solution technique ne permettait de colmater la faille.

Cette situation est devenue préoccupante, car le nombre des failles 0-Days augmente constamment et les pirates en ont fait une nouvelle monnaie d'échange : une faille se négocie actuellement entre 5 000 et 50 000 dollars. Les éditeurs de logiciels sont le plus souvent impuissants et attendent qu'une attaque se produise pour l'analyser et trouver la faille. Pire, une fois les correctifs publiés, les administrateurs et les utilisateurs ne les appliquent pas systématiquement. Ce défaut d'hygiène informatique laisse des millions d'ordinateurs à la merci des attaques. Le cas du virus Slammer est à ce titre éloquent : alors que la faille était connue et son correctif publié depuis 6 mois, près de 200 000 serveurs ont succombé à l'attaque, faute d'avoir « vacciné » leur système.

Les logiciels de conception médiocre sont courants et concernent tous les systèmes d'exploitation, à des degrés divers. Il arrive souvent que les produits soient rapidement élaborés pour faire face à la concurrence et les vérifications de qualité bâclées. Toutefois, la réalisation de logiciels sûrs est une activité difficile. En outre, on peut démontrer mathématiquement qu'il est impossible de garantir la qualité d'un programme dans l'absolu. L'indécidabilité est un mot-clé de l'informatique ! Par conséquent, une méthode de défense efficace contre les virus serait de mettre à jour régulièrement nos ordinateurs.

Un virus se sert d'une faille du système pour y pénétrer : il détourne à son avantage une fonction du système hôte. Les virus biologiques agissent de même en détournant les fonctions des récepteurs de la membrane cellulaire pour s'introduire dans la cellule hôte. Ces deux types de virus vont ensuite asservir les fonctions de leur hôte pour se multiplier. Mais, au contraire d'un virus biologique, le virus informatique a besoin d'une cible afin de se reproduire. Une fois la cible trouvée, il se duplique et change parfois son code. Le virus autonome est donc doué d'autoreproduction. Stephen Kleene, mathématicien de l'Université du Wisconsin, aux États-Unis, a certainement conçu le premier programme autoreproducteur.




**2. ÉVOLUTION DU NOMBRE DE VIRUS répertoriés dans les bases de données des logiciels antivirus. Une signature numérique identifie chaque virus et permet ainsi de les repérer lorsque le logiciel filtre les processus contenus dans l'ordinateur. La barre des 200 000 virus a été dépassée en 2006.**

Cependant, le danger d'une analogie est de réduire deux objets différents à un seul, sans tenir compte des particularités de chacun. La différence la plus notable concerne la mutation de code : alors que les virus biologiques mutent rarement, leurs homologues informatiques mutent très souvent, et les codes produits sont toujours viables. Les ordres de grandeur de la propagation des épidémies sont heureusement fort différents. Que donnerait une maladie virale comme le Syndrome respiratoire aigu sévère (SRAS), responsable de la mort de plus de 800 personnes en Asie entre 2002 et 2003, si elle se propageait à la vitesse d'un virus informatique !

## L'efficacité relative des antivirus

Une vingtaine d'années après l'apparition du premier virus informatique, les protections apportées par les antivirus actuels ne semblent manifestement pas à la hauteur. Il y a plusieurs raisons à cela. La première est mathématique : Fred Cohen a démontré qu'il n'est pas possible de construire un « antivirus universel » qui serait capable de détecter tous les virus. Il est néanmoins possible de concevoir des antivirus efficaces.

Les antivirus actuels reconnaissent les virus par une analyse de leur forme, ce qui requiert au préalable de connaître un élément caractéristique du virus, sa signature. Ces logiciels filtrent alors les programmes pour les analyser et retenir tous ceux dont la signature correspond à celle du virus. Les programmes désignés sont alors considérés comme des virus potentiels. En maintenant un logiciel antivirus à jour, on complète régulièrement la base de données des nouvelles signatures virales, condition *sine qua non* d'une lutte efficace contre les virus les plus courants.

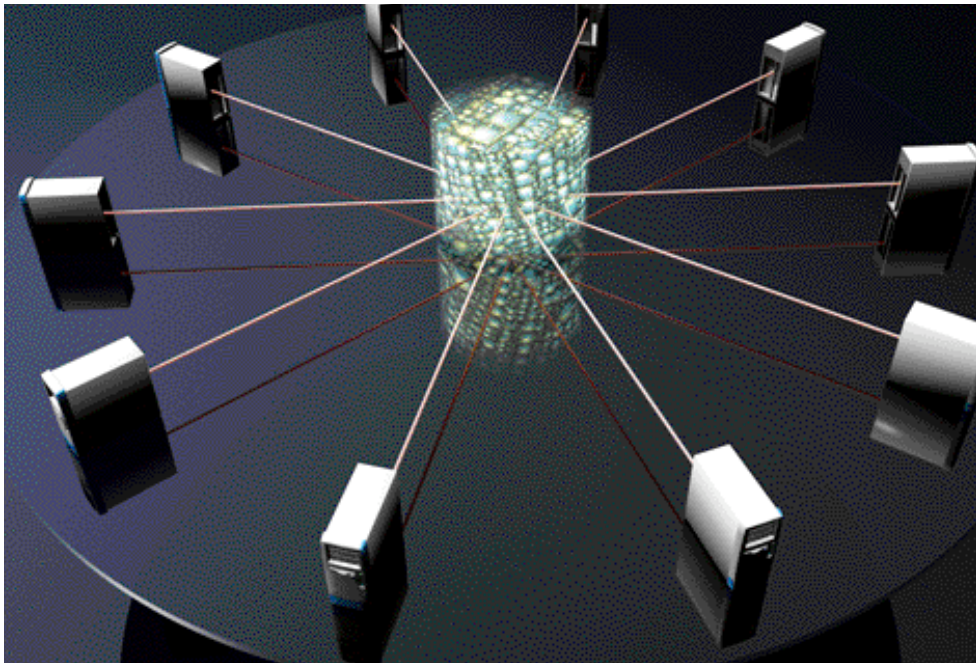
Cette méthode possède pourtant deux faiblesses conceptuelles. Tout d'abord, un nouveau virus passera au travers du tamis si aucune signature connue ne lui correspond. Enfin, le temps de réaction de la défense risque d'être

trop long face à une attaque fulgurante ! D'autres approches sont alors nécessaires. L'une d'elles repose sur l'analyse comportementale des programmes : il consiste à reconnaître une action malveillante dans l'ensemble de processus exécutés par le système. Autant chercher une aiguille dans une botte de foin...

Tous ces efforts sont-ils vains ? Sur le plan technique, il sera toujours possible de contourner tout système de protection. En effet, le pirate n'a nul besoin d'écrire un code élaboré : quelque 600 à 1 200 nouveaux codes apparaissent par mois, et cela suffit à compliquer le travail des éditeurs d'antivirus. En outre, les codes viraux les plus évolués sont écrits pour que leur analyse soit la plus difficile possible, et ils sont capables de muter constamment, ou de se cacher dans le système. Des expériences de laboratoires montrent que ces codes seront toujours en mesure de faire échouer n'importe quel antivirus, et ce pour plusieurs raisons.

Les codes malveillants et les antivirus ne sont pas soumis aux mêmes contraintes : un virus a la liberté d'agir sur plusieurs minutes, voire quelques dizaines de minutes. L'utilisateur n'acceptera jamais que son antivirus monopolise ce même temps pour déterminer si le code est malveillant ou non. Rappelons que si une analyse humaine est, sauf cas particuliers, toujours possible, une analyse antivirale faite par un programme est d'une efficacité limitée. D'autres problèmes sont par nature indécidables et aucun antivirus n'est capable de les résoudre. Ainsi certaines classes de mutations de code sont imparables.

L'expérience de la lutte antivirale en biologie peut alors venir au secours de la virologie informatique sur au moins deux plans. Premièrement, les organismes biologiques les plus évolués possèdent des systèmes immunitaires complexes et efficaces. Ces derniers distinguent ce qui est propre à l'organisme, de ce qui lui est étranger, notamment les agents infectieux. Un système immunitaire reconnaît une substance étrangère et met en mémoire cette rencontre. Ne faudrait-il pas s'en inspirer pour inventer des défenses antivirales plus efficaces, capables de s'adapter aux nouvelles menaces grâce à un système immunitaire formel ? L'analogie avec l'immunologie suggère de protéger le réseau informatique par un système de défense qui serait constitué d'une multitude de processus autonomes.



Shutterstock/P. Hermans

**3. LA RÉPLICATION DES VIRUS INFORMATIQUES** passe par l'infection d'autres ordinateurs qui sont connectés au réseau. Dans le foyer infectieux central, le virus est copié en autant d'exemplaires qu'il existe de cibles. Ce principe est totalement différent de la réplication biologique où les virus se multiplient au sein d'une cellule hôte en des centaines voire des milliers d'exemplaires, puis s'en échappent à la recherche d'autres cellules saines.

Deuxièmement, la médecine moderne a instauré un ensemble de mesures et de précautions pour prévenir les maladies. Une telle prophylaxie informatique, dont les deux axes seraient l'entretien des logiciels, similaire aux politiques de vaccination, et les comportements « sains », comparables à l'hygiène de vie, ne serait-elle pas à notre portée ?

## Une question d'hygiène

Le parallèle entre les virus informatiques et les virus biologiques s'impose avec force. De même qu'il est impossible de détecter par avance l'apparition des nouveaux virus biologiques, nous sommes condamnés à constater les foyers infectieux sur les réseaux informatiques, et en réaction, d'appliquer les correctifs adéquats. Seules la prévention et une politique de santé adaptées nous permettront une lutte efficace contre les virus informatiques… en espérant qu'elles soient respectées et appliquées par tous.

Si la biologie est un domaine d'inspiration légitime pour l'informatique, en revanche, elle n'apporte pas de solutions miracles aux problèmes du monde informatique. Aussi, ayons un regard critique quand certains éditeurs de logiciels n'hésitent pas à nous proposer des « vaccins » pour nos ordinateurs, et des antivirus qui affichent un stéthoscope, une gélule ou un caducée comme argument publicitaire…

Enfin, il reste à créer au niveau national un équivalent du réseau Sentinelle, qui collecte les informations épidémiologiques auprès des médecins sur tout le territoire. Sur le plan international, pourquoi ne pas imaginer une « OMS informatique », apte à organiser la défense face à toutes les alertes ? Pour garantir un fonctionnement efficace, ces organismes devront agir en toute indépendance, notamment envers les intérêts commerciaux et stratégiques, et se reposer sur une recherche autonome, théorique et appliquée.

**Eric FILIOL et Jean-Yves MARION sont membres du Laboratoire de Virologie et de Cryptologie de l'École Supérieure et d'Applications des Transmissions de l'Université de Nancy**

Eric FILIOL. Techniques virales avancées, Collection IRIS, Springer Verlag France, janvier 2007.

Eric Filiol. Les virus informatiques : théorie, pratique et applications. Collection IRIS, éditions Springer France, 2004.

G. Bonfante, Mathieu Kaczamarek, Jean-Yves Marion. On abstract computer virology from a recursion theoretic perspective. Journal in Computer Virology, vol. 1, numéro 3-4, 2005.

Fred Cohen. Computer Viruses, Thèse de doctorat, Université de Californie du Sud, 1986.