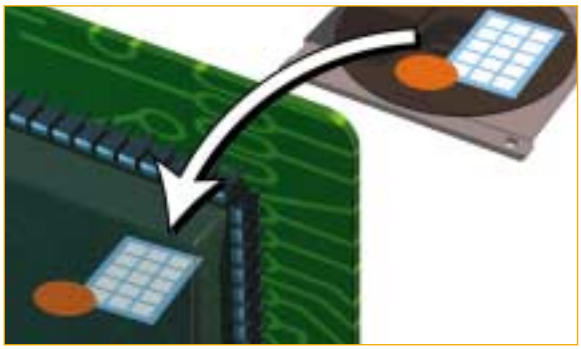
Le cycle d'un virus qui infecte les programmes



Lorsqu'un utilisateur lance un programme infecté, l'ordinateur copie le programme du disque, où il est stocké et inactif, dans la mémoire vive, où il peut être exécuté.

Le programme viral s'exécute en premier, tandis que le programme infecté reste au repos.

Le virus se recopie dans une zone de la mémoire vive séparée du programme, de sorte qu'il peut rester actif même après que l'utilisateur a lancé d'autres logiciels.

virus de macros, responsables de la moitié de toutes les attaques virales.

En plus des séquences qui assurent la recopie, les virus contiennent des séquences variées : certains virus impriment un message ou affichent une image ; d'autres effacent des programmes ou des données. Même ceux qui ne sont potentiellement pas nuisibles peuvent créer des dégâts quand ils se placent dans des configurations différentes de celle où ils avaient été créés. Par exemple, le virus *Form*, qui ne produit habituellement qu'un léger cliquetis chaque mois, écrit sur une partie du disque qui est inoccupée dans les vieux ordinateurs, mais qui contient des données cruciales dans les nouvelles machines, où les répertoires sont agencés différemment.

Techniques immunitaires

Les logiciels contre les virus sont apparus peu après les virus. Des programmes généraux de détection de virus surveillent si le comportement d'un ordinateur est modifié par la présence d'un virus (par exemple, si le virus modifie certains fichiers importants ou des parties de la mémoire) et cherchent périodiquement des modifications suspectes des programmes. Ils détectent les virus sans les connaître, mais ils peuvent aussi déclencher de fausses alarmes, car des activités normales de l'ordinateur ressemblent parfois à celles de virus en action.

En revanche, des programmes d'analyse recherchent dans les fichiers, dans les secteurs de démarrage ou dans la mémoire de l'ordinateur, des morceaux de programmes (nommés aussi signatures virales) caractéristiques de virus déjà connus. Pour rester efficaces, ils doivent être mis à jour régulièrement, mais ils ne déclenchent que rarement de fausses alarmes. Ces programmes ont la même efficacité que les récepteurs biologiques du système immunitaire : ces derniers reconnaissent les protéines virales en se liant à de courtes séquences (8 à 15 acides aminés seulement, contre des milliers pour la protéine virale complète) ; de même, les signatures virales reconnues par les programmes d'analyse sont très courtes, de 16 à 30 octets (un octet représente huit bits), parmi les quelques milliers qui constituent un virus informatique complet. La recherche de petits segments, au lieu du virus complet, accélère l'analyse, d'autant que des virus différents partagent parfois la même signature. La plupart des programmes d'analyse utilisent des algorithmes issus des techniques d'appariement en biologie, qui cherchent plusieurs signatures en même temps. Le plus rapide de ces programmes vérifie la présence de 10 000 signatures dans 10 000 programmes en moins de dix minutes.

Comment éliminer un virus qui a été détecté? Une technique brutale, mais efficace, consiste à effacer le programme contaminé. La technique est identique à celle qui est utilisée par certains types de cellules immunitaires, qui détruisent les cellules infectées ; mais si les cellules du corps sont aisément remplacées, il n'en est pas de même des programmes et des documents informatiques. Les programmes antiviraux perfectionnés essaient plutôt de réparer les fichiers infectés. Cette tâche est facilitée par le fait que les virus informatiques ne se transmettent que s'ils n'éveillent pas l'attention de l'utilisateur, de sorte qu'ils laissent leur programme hôte intact.

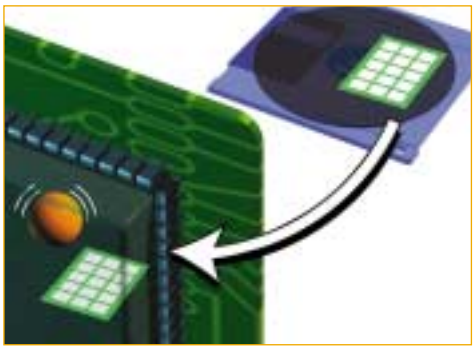
Lorsqu'un programme d'analyse détecte un fichier infecté, il exécute habituellement une ordonnance détaillée, fournie par ses concepteurs, afin d'ôter le code viral et de reconstruire une copie opérationnelle du programme original. D'autres techniques de désinfection générale fonctionnent aussi bien sur des virus connus que sur des virus inconnus. Nous avons mis au point un programme qui élabore un résumé mathématique (nommé somme de contrôle) de chacun des programmes du système, de sorte que l'on peut reconstituer une copie de l'original lorsqu'un programme est infecté.

Les techniques de détection et d'éli-

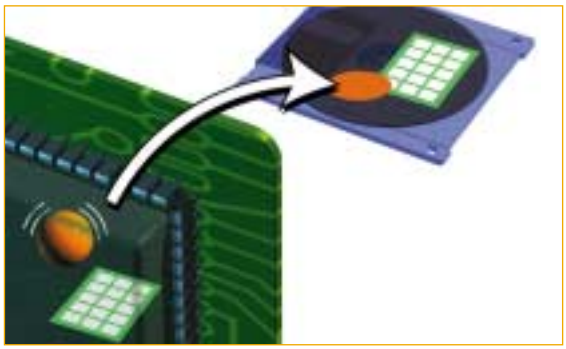
James Gary

2. SUR LE CHAMP DE BATAILLE NUMÉRIQUE, des programmes antiviraux recherchent les programmes malfaisants (virus, vers ou chevaux de Troie) et réparent les dégâts qu'ils causent. Les nombreux types de virus informatiques qui existent peuvent être reconnus à leur action et à leur structure caractéristique. Les programmes antiviraux protègent aussi les programmes ordinaires en les analysant et en prévenant les utilisateurs de tout changement.

Son travail initial effectué, le virus redonne le contrôle au programme infecté.



Lorsque l'utilisateur lance un autre programme, le virus dormant se réveille.



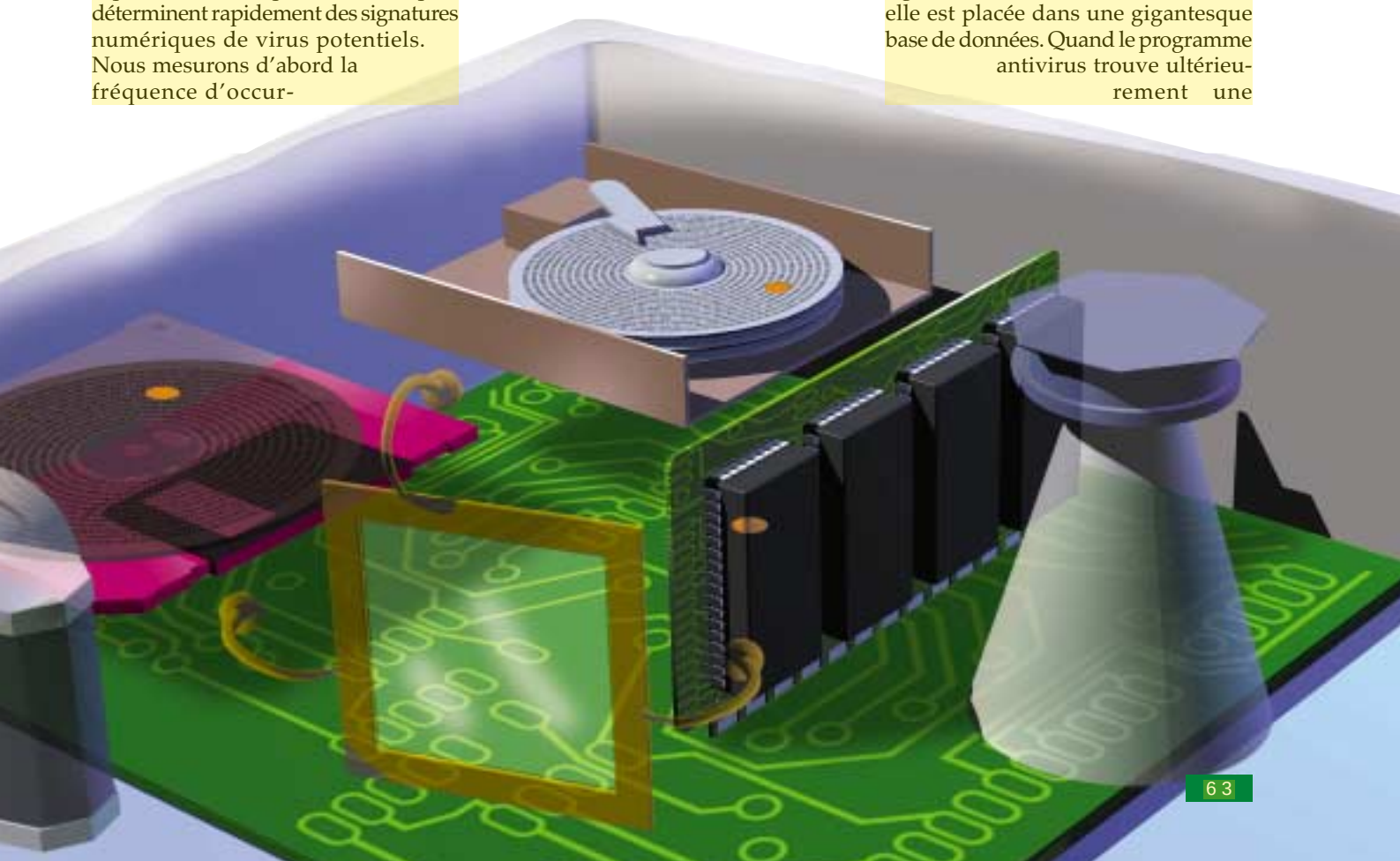
Il se copie dans le logiciel sain et peut aller infecter un autre ordinateur ou une disquette.

James Gary

mination de virus spécifiques nécessitent une analyse détaillée de chaque nouveau virus. Des spécialistes doivent identifier des séquences inhabituelles d'instructions qui apparaissent dans le programme viral, mais pas dans les programmes normaux. Ce processus est fondé sur une bonne connaissance des virus, mais aussi sur l'intuition. Ces spécialistes élaborent également les prescriptions pour enlever le virus de n'importe quel ordinateur hôte. Comme six nouveaux virus sont créés chaque jour, nous avons mis au point des outils et des procédures automatiques, afin d'aider les spécialistes humains, voire de les remplacer.

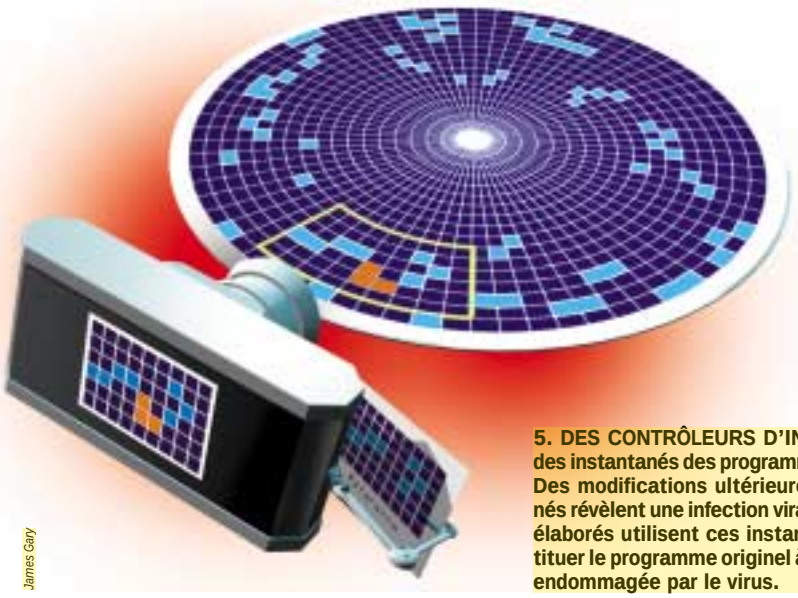
Fondées sur des méthodes statistiques, ces techniques automatiques déterminent rapidement des signatures numériques de virus potentiels. Nous mesurons d'abord la fréquence d'occurrence de certaines courtes séquences d'octets dans un grand nombre de programmes normaux. Quand on nous envoie un nouveau virus, notre logiciel le détecte en trouvant la séquence qu'il est statistiquement le moins probable de rencontrer dans les programmes légitimes. Cette méthode est beaucoup plus rapide qu'une analyse manuelle, et les signatures identifiées conduisent à moins de fausses alarmes que les signatures choisies par des experts humains. Cette méthode heuristique, initialement mise au point par Fridrik Skulason et Franz Veldman, ressemble à un mécanisme qui avait semblé être celui du système immunitaire : on croyait naguère que les anticorps se modelaient d'après les envahisseurs ; de même, nos signatures sont réalisées spécifiquement pour chacun des nouveaux virus rencontrés.

Avec ses collègues, Stephanie Forrest a mis au point une autre méthode, copiée sur la «sélection clonale», qui semble être le mécanisme du système immunitaire : le corps produit d'emblée une grande variété de cellules immunitaires, mais seules prolifèrent les cellules qui reconnaissent l'agent pathogène. De même, dans le programme fondé sur cette méthode, des signatures sont engendrées au hasard, et chaque signature est ensuite comparée à tous les programmes de l'ordinateur non infecté ; quand une signature ne ressemble à rien de connu, elle est placée dans une gigantesque base de données. Quand le programme antivirus trouve ultérieurement une



3. UN PROGRAMME ANTIVIRAL GÉNÉRAL surveille les comportements qui peuvent être dus à un virus en action, telle la modification de certaines zones de la mémoire vive ou de fichiers particuliers du disque dur. En bloquant ces comportements illicites, il n'élimine pas le virus, mais peut l'empêcher d'infecter d'autres programmes ou de perturber le fonctionnement de l'ordinateur.

4. UN ANALYSEUR DE SIGNATURES recherche sur les disques des fragments de programme qui apparaissent dans les virus connus.



5. DES CONTRÔLEURS D'INTÉGRITÉ prennent des instantanés des programmes et des données. Des modifications ultérieures de ces instantanés révèlent une infection virale. Des algorithmes élaborés utilisent ces instantanés pour reconstituer le programme originel à partir de la version endommagée par le virus.

James Gary

telle signature dans un programme, il sait que le programme a été modifié, bien qu'une analyse ultérieure soit requise pour déterminer si un virus en est la cause.

La biologie est également à la base d'une autre technique antivirale : on recherche des parentés entre les nouveaux et les anciens virus. En effet, les programmeurs de virus utilisent souvent des parties de virus existants pour en créer de nouveaux. Ces parties, ou «gènes» viraux, permettent de retracer l'évolution historique des virus informatiques, comme le font les biologistes lorsqu'ils établissent les relations entre espèces vivantes. En étudiant de nombreux virus, nous construisons automatiquement un ensemble de signatures valables pour tous les membres de la famille virale, incluant des variantes précédemment inconnues. Cette technique réduit les problèmes de stockage des signatures : une seule signature de 20 octets peut reconnaître des dizaines de virus différents.

Nous avons également mis au point des techniques pour reconnaître des virus en analysant la présence de très courts segments de programme, chacun de trois à cinq octets de longueur. Ces segments minuscules représentent des instructions dont l'exécution est spécifique d'une infection virale. Bien que des logiciels classiques puissent en contenir certains, leur répétition témoigne d'une infection virale. Les logiciels antiviraux en détectent la présence et, ce qui est plus important, reconnaissent de nombreux virus sans les avoir rencontrés auparavant.

La propagation des virus en milieu naturel

Depuis 1990, nous avons étudié les virus sur une population de plusieurs centaines de milliers d'ordinateurs qui nous avaient été confiés par des clients. Nous enregistrons le lieu et la date de chaque incident, le nombre d'ordinateurs ou de disquettes infectés et l'identité des virus. Ces statistiques nous ont permis de comprendre la propagation des infections virales dans le milieu naturel. Seulement cinq pour cent des virus connus ont été observés dans notre échantillon, et certains n'ont été vus qu'une seule fois. Les dix virus les plus fréquents déclenchent deux tiers des incidents, et leur prévalence est semblable : au cours du temps, le nombre de ces virus à succès augmente

linéairement jusqu'à atteindre un plateau ; ensuite, ils continuent d'apparaître à un rythme constant dans les ordinateurs, et quelquefois ils disparaissent complètement.

Afin de comprendre ces caractéristiques, nous avons utilisé des modèles mathématiques élaborés par les épidémiologistes. Les modèles les plus simples prédisent l'évolution d'une infection à partir d'un petit nombre de paramètres : le «taux de naissance», c'est-à-dire la vitesse à laquelle les individus malades infectent les autres, et le «taux de mortalité», la vitesse à laquelle un individu malade meurt ou est guéri. Lorsque le quotient de ces deux quantités est inférieur à un seuil, l'infection disparaît rapidement. Plus ce quotient est élevé et plus, en l'absence de mesures prophylactiques ou de réactions immunitaires, le nombre de personnes infectées augmente rapidement.

Toutefois, cette théorie simple ne rend pas bien compte de l'évolution des virus informatiques, car elle stipule que, sauf quand le quotient des naissances et des morts est proche de la valeur critique, les virus disparaissent ou se propagent à toute la population. Or, de nombreux virus infectent une très faible proportion de la population et restent à ce niveau. Pourquoi cette évolution? Parce que la théorie suppose qu'un individu a la même chance de contaminer n'importe quel autre individu de la population à risque. En informatique, ce n'est pas le cas, et l'on doit compliquer la théorie pour décrire la structure en groupes fermés de la population qui utilise les ordinateurs : en moyenne, chaque personne ne partage des logiciels et des données qu'avec un petit nombre de personnes, et tous les échanges prennent place à l'intérieur de groupes. Si Philippe échange des logiciels avec Pascale et Pascale avec Françoise, alors, il est raisonnable de penser que Philippe et Françoise peuvent échanger des virus.

Des simulations informatiques ont montré que la proximité des contacts ralentit la croissance des épidémies conformément à nos observations. Des échanges peu fréquents réduisent la probabilité d'une épidémie et abaissent le niveau du plateau, mais n'expliquent pas complètement les données.

L'évolution en action

De même que les disettes, les campagnes de santé publique ou les migrations modifient le cours des épidémies biologiques, des changements de l'environnement informatique sont responsables de plusieurs époques distinctes dans l'histoire de la virologie informatique. Jusqu'en 1992, les virus infectant les programmes ou les secteurs de démarrage se trouvaient

Les virus de messagerie électronique

La croissance rapide du réseau Internet accroît proportionnellement les menaces liées au contenu des messageries électroniques, et plus particulièrement, aux virus informatiques. Un message électronique qui transite sur le réseau peut être séparé en trois parties : l'en-tête, le corps du message et les fichiers attachés. La principale cible des virus informatiques est le fichier attaché. Ces fichiers sont soit des programmes exécutables, soit des documents. Aujourd'hui, on se prémunit contre ce type d'infection informatique par deux approches principales : la protection répartie et la protection centralisée.

Dans la protection répartie, chaque poste susceptible de recevoir un message est protégé. La plupart des programmes antiviraux du marché disposent de modules de surveillance. Ces derniers détectent la présence du virus lorsque les éléments attachés sont exécutés ou ouverts. Malheureusement, cette méthode n'empêche pas la propagation du virus sur l'ensemble des postes des destinataires de messages : si, sur certains postes infectés, le virus n'est pas éradiqué, l'épidémie peut s'étendre. De surcroît, la détection n'intervient qu'à l'ouverture du fichier. Dès lors, un fichier attaché infecté pourra être transmis d'un utilisateur à l'autre sans qu'aucune alerte soit donnée.

Certains programmes antiviraux possèdent des modules spécifiques pour la messagerie électronique. Ces programmes décomposent le message et l'analysent sans ouvrir ni exécuter le fichier attaché. Si cette détection précoce évite l'épidémie, elle dépend du type de messagerie électronique et n'empêche pas la propagation d'un virus infecté à de nombreux utilisateurs. Ainsi, une note de service envoyée à l'ensemble du personnel d'une société entraînera de nombreuses alertes.

L'autre approche, nommée protection centralisée, consiste à utiliser une machine de contrôle. Tous les messages qui transitent passent par cette machine. À leur arrivée, les messages sont décomposés et décompressés (les fichiers attachés sont souvent compressés). Sur la machine centrale, des logiciels antivirus éradiquent alors les éventuels virus avant de faire suivre le message. Cette solution a l'avantage d'être préventive et centralisée : la note de service infectée envoyée à l'ensemble du personnel sera bloquée avant d'atteindre ses destinataires. Malheureusement, ces solutions ne fonctionnent qu'avec certains types de messageries.

La dernière approche qui s'apparente également à de la protection centralisée, n'est pas préventive : elle consiste à analyser tous les messages présents sur une machine centrale. Cette machine centrale, ou serveur de messagerie, vérifie l'innocuité des messages avant qu'ils ne parviennent à l'utilisateur. Cette solution est encore peu répandue, mais elle présente l'intérêt d'être économique en terme de ressource mémoire et de système.

Quelle que soit la solution mise en œuvre, l'utilisation de l'Internet et des messageries électroniques doit se faire avec une extrême prudence. Sans protection adéquate, quelques règles élémentaires de conduite peuvent éviter le pire : ne pas ouvrir un message en provenance d'un expéditeur inconnu, ne pas exécuter de programmes ni ouvrir de document attaché sans en connaître l'origine et le but, sauvegarder sur disquette les zones et les fichiers sensibles.

Sans ces précautions, le *surf and buy* («se promener sur l'Internet et acheter») peut rapidement se transformer par un *surf and die* («se promener» sur l'Internet et mourir) pour les disques durs et les informations des internautes non avertis.

Stéphane BOUCHÉ, Société *OMNISET*