L'ESSENTIEL

- **Les experts en informatique judiciaire recherchent des éléments significatifs dans les mémoires de dispositifs électroniques variés : ordinateurs, téléphones portables, routeurs de réseaux, etc.**
- **Pour ce faire, ils ont développé des techniques élaborées, visant par exemple à reconstituer des fichiers partiellement effacés ou à détecter des photos retouchées.**
- **Les défis auxquels ils sont confrontés s'accroissent avec l'augmentation de la quantité de données et la diversification des matériels et des logiciels.**

Pour la Science, © Shutterstock/Igor Latsi

Depuis les années 1980, les ordinateurs sont de plus en plus présents dans la vie de tous les jours. Le monde du crime n'y échappe pas. Les criminels, comme d'autres personnes, stockent des données et communiquent avec leurs complices. Ces activités laissent des traces dans les appareils utilisés. Cela a conduit au développement de l'informatique judiciaire, qui vise à recueillir et à analyser les indices présents sur tout dispositif doté d'une mémoire numérique : ordinateur, téléphone portable, réseau...

Les experts de ce domaine sont confrontés à certains des problèmes les plus difficiles de l'informatique, tels que l'analyse de données massives et variées, le traitement automatique des langues naturelles et la cybersécurité. Ils sont impliqués dans l'élucidation de deux types de forfaits : ceux perpétrés dans le monde physique, mais dont des preuves sont enregistrées sur un dispositif électronique, et ceux qui sont intrinsèquement liés aux systèmes informatiques, tel le piratage. Nous examinerons ici la façon dont sont stockées les données sur un dispositif électronique, ainsi que les techniques développées par l'informatique judiciaire pour les extraire, les analyser, voire les reconstruire quand l'utilisateur a tenté de les effacer.

L'efficacité de l'informatique judiciaire vient de ce que nombre de systèmes numériques conservent de grandes quantités d'informations. Il s'agit d'examiner aussi bien les ordinateurs et les téléphones portables des suspects de crime que les données électroniques des entreprises lors de litiges. Les enquêteurs récupèrent souvent d'anciens courriers électroniques, des historiques de discussion en ligne, des termes de recherche sur *Google* ou d'autres types de données, créés des semaines, des mois, voire des années plus tôt. Ces enregistrements peuvent révéler l'état d'esprit d'un individu ou ses intentions au moment où un crime a été commis. Les techniques de l'informatique judiciaire peuvent aussi renseigner sur les victimes : un téléphone retrouvé sur un cadavre non identifié serait étudié de la même façon qu'un appareil perdu pendant un cambriolage.

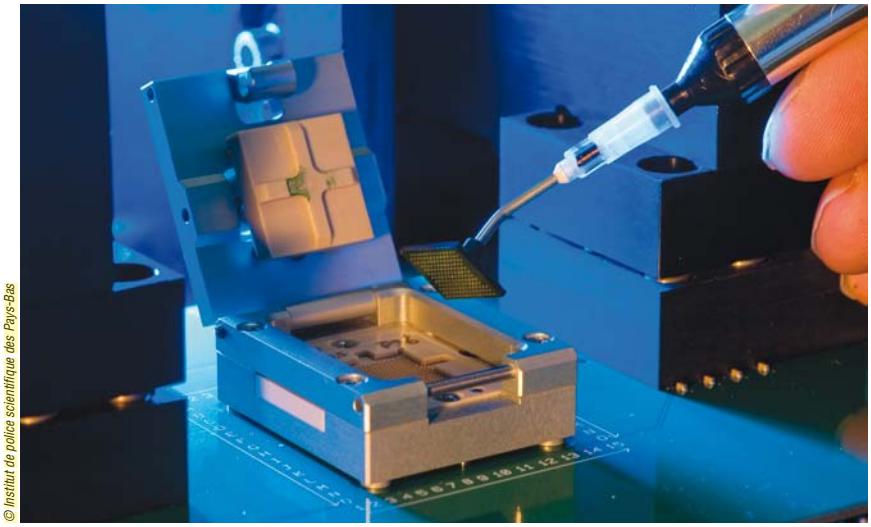
L'informatisation rend l'enquête plus compliquée que si les données étaient inscrites sur du papier. Par exemple, Bernard Madoff, arrêté en 2008 pour des escroqueries financières, suivait les comptes de ses victimes sur un ordinateur IBM datant des années 1980. L'ancienneté de cette machine explique en partie la durée exceptionnellement longue de la fraude – qui a couru sur plus de deux décennies –, car peu de personnes à Wall Street maîtrisaient encore ce type de matériel. Elle a aussi compliqué l'interprétation des données après l'arrestation de B. Madoff.

Les indices numériques peuvent aussi révéler une absence de manipulation. En mai 2006, un ordinateur portable et un disque dur externe ont été dérobés au Département américain des anciens combattants. Ils contenaient des données sensibles concernant plus de 26 millions de vétérans et d'employés de l'armée. Quand l'ordinateur a été retrouvé en juin 2006, les enquêteurs ont montré que les dossiers sensibles n'avaient probablement pas été consultés. Pour ce faire, ils ont notamment examiné les heures d'accès et de modification associées à chaque fichier. Mais en utilisant les mêmes techniques judiciaires qu'eux, les coupables auraient pu visionner les fichiers sans modifier ces horodatages ; en fait, les enquêteurs ont juste déterminé que les fichiers n'avaient pas été ouverts par des moyens classiques.

Des données volatiles et invisibles

En pratique, les informations obtenues ne sont pas tant limitées par la technologie que par les coûts et les besoins. Le désir intellectuel d'examiner jusqu'au dernier octet est rarement justifié et l'on s'arrête souvent quand on en sait assez pour incriminer un suspect. Les contraintes légales conditionnent aussi la façon dont l'analyse est effectuée.

Les données électroniques nécessitent une attention particulière. Une mauvaise manipulation peut vite les modifier, les endommager ou les effacer. Le simple fait d'allumer un GPS du commerce risque d'entraîner la suppression de preuves cruciales. En outre, les ordinateurs contiennent souvent des informations cachées, pour lesquelles ont doit utiliser des outils spécifiques. Par exemple, un appareil photo qui semble contenir 30 photos peut en réalité en renfermer plusieurs centaines, censées être effacées mais en réalité récupérables. En effet, quand un appareil « efface » un fichier, il ne nettoie pas la mémoire, il note juste que l'espace est disponible ; le fichier n'est supprimé que lorsqu'un autre



© Institut de police scientifique des Pays-Bas

1. LES OUTILS DE L'INFORMATIQUE JUDICIAIRE permettent d'accéder directement aux puces de mémoire de dispositifs électroniques variés : téléphones portables, appareils GPS, systèmes automobiles, clés USB... On peut ainsi récupérer des données sur des appareils endommagés ou protégés par des mots de passe.

est inscrit par-dessus. Selon un rapport de 2009 du Département de la défense américain, de nombreux disques durs ne sont pas correctement nettoyés avant d'être mis au rebut par le gouvernement.

De nombreuses enquêtes judiciaires commencent par la recherche de fichiers appartenant à l'utilisateur de l'ordinateur. Avant d'examiner les données, on doit les extraire de leur support et les enregistrer de façon fiable et durable. Une bonne compréhension des mécanismes de stockage informatique est nécessaire. Bien que l'informatique soit fondée sur la manipulation de chiffres binaires (0 et 1), les « bits », les ordinateurs modernes font la plupart de leurs calculs sur des groupes de huit bits, les octets. Il y a 256 (2^8) octets possibles. Les octets sont souvent utilisés pour stocker du texte, chaque lettre étant représentée par une chaîne binaire spécifique. Le système de codage UTF-8, très répandu, représente la lettre A par la séquence 01000001, la lettre B par 01000010, et ainsi de suite. Les bits sont aussi souvent traduits en codes hexadécimaux (comportant 16 éléments de base, à savoir les chiffres de 0 à 9 et les lettres de A à F), notamment pour stocker les images : on associe à chaque pixel un code caractéristique de sa couleur.

Quand ils sont enregistrés sur un disque dur ou sur une carte mémoire, les octets sont groupés en « secteurs », qui font typiquement 512 ou 4 096 octets de longueur.

Deux enquêtes numériques

■ Le 1er janvier 2002, Scott Tyree kidnappe Alicia Kozakiewich, âgée de 13 ans. Il envoie une photo de l'adolescente à un homme sur la messagerie instantanée de *Yahoo*. Le destinataire prévient le FBI et lui fournit le nom d'utilisateur de l'expéditeur. Les enquêteurs contactent *Yahoo*, qui leur donne l'adresse IP correspondante, puis obtiennent les coordonnées du titulaire de cette adresse auprès du fournisseur d'accès à Internet.

■ En 2005, Dennis Rader, tueur en série du Kansas, fait une nouvelle victime avant d'envoyer à la police une disquette contenant une lettre. Les enquêteurs y trouvent un fichier *Word* mal effacé. Les métadonnées de ce fichier indiquent qu'un certain Dennis est la dernière personne à l'avoir modifié et contiennent un lien vers un temple luthérien où D. Rader est diacre, ce qui permet de le retrouver.

Chaque secteur a un numéro d'identification unique, que l'on nomme l'adresse du bloc logique du secteur. Il faut en général 10 à 20 secteurs pour stocker un courriel, et des centaines de milliers pour un film. Un téléphone portable doté d'une mémoire de huit gigaoctets comporte environ 15 millions de secteurs.

Les secteurs de la mémoire peuvent être stockés sous forme d'une seule longue séquence ou être fragmentés et enregistrés à de nombreux endroits différents. D'autres secteurs contiennent les métadonnées (littéralement, les données sur les données) du système de fichier, c'est-à-dire les informations nécessaires à l'ordinateur pour qu'il retrouve les données.

Des copies intégrales des mémoires

Pour préserver les données d'un ordinateur ou d'un téléphone, chacun des secteurs doit être copié et stocké sur un autre ordinateur, dans un fichier unique nommé image disque ou image physique. Ce fichier contient tous les octets de l'appareil examiné, y compris ceux correspondant aux données « effacées » mais non encore écrasées par le système d'exploitation. En pratique, l'image disque est copiée en plusieurs exemplaires, dont on vérifie la fidélité à l'original en leur associant une signature numérique par la technique dite du hachage. Cette technique a de nombreuses autres applications et son apparition a été un événement majeur de l'informatique judiciaire et de la cybersécurité *(voir l'encadré page ci-contre)*.

On s'intéresse également à la mémoire vive, qui joue le rôle d'espace de travail et de stockage temporaire pour le système d'exploitation et les programmes de l'ordinateur. Cette mémoire est aussi nommée RAM, pour *Random access memory* (mémoire à accès aléatoire), car on peut accéder directement et rapidement à n'importe quelle donnée (tandis qu'un accès séquentiel suivrait un chemin défini). La mémoire vive d'un ordinateur ou d'un téléphone portable est une mosaïque de blocs de 4 096 octets qui contiennent des informations variées : code de programmes en cours d'exécution ou qui ont tourné récemment, instructions du système d'exploitation, fragments d'éléments envoyés et reçus par le réseau, morceaux de fenêtres affichées à l'écran, presse-papier (où sont stockées les

informations quand on utilise la fonction « copier » ou « couper »)…

Pour dupliquer la mémoire vive, on se sert d'un programme spécifique, nommé imageur de mémoire. Tous les types d'appareils électroniques (ordinateurs, téléphones, équipements du réseau tels les routeurs, etc.) ont une RAM, mais ils utilisent des logiciels différents, si bien que les programmes conçus pour analyser certains systèmes ne fonctionnent pas toujours sur les autres. Aujourd'hui, aucune méthode universelle n'existe.

Dans un nombre croissant d'enquêtes, les données à analyser ne sont pas celles enregistrées sur une machine individuelle, mais celles échangées *via* une connexion réseau. Cette connexion est alors mise sur écoute.

La sauvegarde des données n'est que la première étape. On doit ensuite identifier les informations pertinentes pour l'enquête. Certains outils peuvent extraire les fichiers de l'image disque, rechercher ceux qui contiennent un mot ou une phrase particuliers (dans des langues variées), détecter des données cryptées...

Sur un ordinateur, les fichiers dits alloués peuvent être visualisés grâce au système de gestion des fichiers (par exemple dans des répertoires arborescents). Le terme alloué se réfère aux secteurs du disque où le contenu du fichier est stocké : ils sont dédiés à ce fichier particulier et ne peuvent pas être affectés à d'autres. De nombreux outils d'informatique judiciaire permettent de voir les fichiers alloués d'une image disque sans utiliser le système d'exploitation de l'ordinateur, ce qui préserve l'intégrité judiciaire des preuves.

L'AUTEUR

Simson GARFINKEL est maître de conférences à la *Naval Postgraduate School*, à Monterey, en Californie.

Article publié avec l'aimable autorisation de la revue *American Scientist*.

À la recherche des fichiers perdus

Quand on supprime un fichier, c'est-à-dire quand on le place dans la poubelle de l'ordinateur et qu'on vide cette dernière, l'espace qu'il occupait est dit désalloué. Son contenu peut rester en mémoire, mais le fichier n'est plus visible et les métadonnées permettant de le localiser sont perdues. Un des principaux progrès de l'informatique judiciaire au cours des 15 dernières années concerne la récupération de ces fichiers « supprimés ». On utilise une technique nommée *file carving* (littéralement « sculpture de fichier »), inventée vers 1999 par Dan Farmer, un chercheur indépendant en sécurité.

De nombreux types de fichiers commencent et se terminent par des séquences caractéristiques d'octets, respectivement nommées en-tête et pied de fichiers. Ainsi, les fichiers JPEG créés par les appareils photo numériques commencent en général par la séquence FF D8 FF E0 et se terminent par FF D9. Les premiers dispositifs de *carving* passaient en revue l'image disque à la recherche de ces séquences

Le hachage, technique clef de la cybersécurité et de l'informatique judiciaire

Le hachage a été inventé au début des années 1950 par l'informaticien allemand Hans Peter Luhn, chez IBM. Une fonction de hachage de x bits associe un nombre binaire de taille x, c'est-à-dire comportant x bits, à une chaîne de caractères, tel un extrait de texte. Le terme hachage vient de ce que, en pratique, les fonctions utilisées segmentent les données avant de les mélanger. Une fonction de hachage de 32 bits peut ainsi produire $2^{32} = 4\,294\,967\,296$ valeurs. Ces fonctions sont conçues de façon à obtenir un résultat complètement différent quand on change un seul caractère de la chaîne d'entrée. Certaines chaînes ont la même valeur de hachage (on parle de collision de hachage), mais plus celle-ci est longue, moins cela risque de se produire.

En 1979, Ralph Merkle, alors doctorant à l'Université de Stanford, aux États-Unis, a appliqué le hachage au domaine de la sécurité informatique. Il a utilisé une fonction de hachage produisant plus de 100 bits en sortie et dite à sens unique : il est facile de calculer la valeur de hachage d'une chaîne, mais presque impossible, à partir d'une valeur de hachage, de trouver la chaîne correspondante. L'idée de R. Merkle était d'utiliser la valeur de hachage d'un document comme signature numérique. Le nombre considérable de valeurs de hachage possibles (2^{100}, soit environ 10^{30}) devait empêcher les pirates de trouver celle associée à un document cible. Aujourd'hui, les signatures électroniques obtenues par hachage sont à la base de nombreux systèmes de cybersécurité, protégeant par exemple les numéros de cartes bancaires envoyés par Internet.

Dans le domaine de l'informatique judiciaire, le hachage sert notamment à vérifier que les copies des données sont fidèles à l'original. Pour un même système informatique, on réalise souvent deux images disques, auxquelles on applique une fonction de hachage. Si les images ont la même valeur de hachage, on suppose qu'elles sont toutes deux une copie fidèle des données d'origine. Tout enquêteur en possession d'une copie ultérieure des données peut calculer la valeur de hachage et vérifier qu'elle correspond à la valeur initiale. De nombreux outils d'informatique judiciaire font automatiquement cette comparaison.

Le hachage est aussi utilisé pour étiqueter des fichiers ou des segments de fichiers connus comme étant « bons », tels les programmes fournis avec les systèmes d'exploitation, ou « mauvais », tels des virus informatiques, des documents volés ou de la pédopornographie. Aujourd'hui, les experts judiciaires distribuent les valeurs de hachage de ces fichiers, afin de faciliter leur détection. Grâce à l'identification de fichiers complets ou partiels par hachage, on peut également déterminer automatiquement si l'un des millions de fichiers que contient un disque dur apparaît dans une base de données renfermant des centaines de millions de fichiers, et ce en un temps assez court, de l'ordre de quelques heures.