La recherche de la preuve

Stanislas Krawezyk

Professeur associé université de Reims
Expert près la cour d'appel de Reims
skej@club-internet.fr

Résumé La recherche de la preuve informatique apporte au demandeur tous les éléments matériels pour infirmer ou confirmer une situation le plus souvent délictueuse ou délictuelle. Le demandeur est le plus souvent un magistrat dans le cadre d'une procédure pénale, civile, prud'homale. C'est l'employeur dans le cadre d'une procédure privée, rarement l'employé. Ces éléments matériels servant, ainsi de preuves, au magistrat, pour fonder en partie son jugement, poursuivre des investigations ou à l'employeur dans le cadre soit d'un licenciement ou dans l'utilisation abusive des technologies informatiques. Pour l'employé, il s'agit de se défendre essentiellement contre des allégations à son encontre.

1 Principes de l'irréfutabilité de la preuve

Le droit d'un inculpé dans le cadre d'une procédure pénale, est d'avoir un procès juste et équitable. Les principes de la recherche de la preuve informatique doivent viser cet objectif. Il en est de même dans tous les autres cas.

Dès lors, le rôle de l'expert judiciaire est primordial. Il doit, ainsi, démontrer en permanence que les preuves qu'il apporte sont sans faille.

Cependant, c'est après une intense réflexion, la mise à plat de toutes les hypothèses, que la preuve peut être apportée. Si un doute subsiste, il devra en faire part et le consigner dans son rapport.

2 Le constat et origine du besoin

Sans être exhaustif, les « affaires classiques » que l'on rencontre sont essentiellement :

Pour le pénal : faux et usage de faux, falsification de documents, utilisation frauduleuses de cartes de paiement, génération de codes de connexion à des à services payants (télévision, ...), pénétration dans des sites informatiques, commande par internet de produits en lieu et place de..., substitution d'identité, diffusion d'information à caractère pornographique, certains actes prémédités, courriers anonymes, séditieux, dénonciations, et bien d'autres.

Pour le civil : utilisation de logiciels piratés, logiciels inadaptés aux besoins, copie et utilisation illicite de tout ou partie d'un logiciel, non couverture fonctionnelle, non- respect des règles de l'art, performance, non-respect d'éléments contractuels dans la mise en œuvre, et bien d'autres.

Pour les prud'hommes : ce sont tous les problèmes relatifs au droit du travail aboutissant à une procédure de licenciement : données erronées ou fausses sur le temps de travail, le détournement de ressources informatiques pour son propre compte, la consultation de sites pornographiques pendant les heures de travail,... et bien d'autres.

L'expert agit dans le cadre d'une commission rogatoire ou d'une mission. Elle définit l'objet, les éléments de recherche ainsi que l'exploration de tout élément de preuves contributif à la vérité.

3 La démarche de mise en œuvre

Les différentes étapes (dans leurs grands traits) :
- Prise de connaissance du dossier et de la mission
- Vérification et enregistrement des scellés
- Clonage du support
- Recherches des éléments de preuves
- Rédaction du rapport

3.1 Prise de connaissance du dossier, de la mission

Il s'agit de retracer le besoin et/ou les constats effectués.

3.2 Vérification et enregistrement de scellés

Pour une procédure pénale, l'ordinateur est sous scellé. Nous devons respecter la procédure d'ouverture et de fermeture des scellés. Pour les expertises civiles et privées, l'expert applique les mêmes principes.

3.3 Clonage du support

L'irréfutabilité de la preuve s'obtient en clonant le disque dur ou le support d'information. Le calcul du hash code répond à cette vérification. Le hash code monodirectionnel, similaire à une somme de contrôle, est utilisé pour la vérification de l'authenticité de données. Contrairement aux systèmes avec CRC, le hash code ne laisse passer aucune erreur d'intégrité. Cependant différentes collisions ont été répertoriées essentiellement dans l'algorithme du Md5 et du Sha1. Il est préférable d'utiliser les algorithmes SHA-256 et PSCHF (Pukall Stream Cipher Hash Function)[1],[2],[3].

L'utilisation d'outils d'analyse de disque impose d'être en mode lecture seule, pour éviter les modifications de dates, de clusters, la génération de fichiers temporaires dus à l'utilisation du système d'exploitation.

3.4 Recherche de preuves

- Les différentes pistes de preuves se situent essentiellement dans la récupération de fichiers (documents, images, séquences vidéo, messages reçus, messages émis) présents, effacés, modifiés, compressés, cryptés.

3.5 Les pistes de preuves

- Récupérations de fichiers non altérés ou partiellement altérés : la recherche se réalise soit par nom, extension ou mieux encore à partir de son header. Par exemple un fichier -arj archive- dont l'extension est « arj » possède comme header la valeur « 0x60EA ». Certains fichiers nécessitent le positionnement de l'offset et pour d'autres, la précision du footer pour être certain de correctement identifier un type de fichier. Une base de données privée référence actuellement plus de 30 000 extensions de fichiers différentes. Il est évident que devant ce nombre important de fichiers, il est parfois difficile de lire le contenu précis d'un document.
- Recherche sur chaîne de caractères : ce sont les recherches classiques sur chaîne de caractères avec l'insertion de jokers facilitant le groupement d'informations. L'inconvénient d'une telle fonction réside dans le temps de recherche sur un disque de taille conséquente.
- Les formats de données : l'utilisation d'un éditeur puissant s'avère être précieux pour décoder les multiples formats de données. Nous retrouvons essentiellement 3 types : format date, nombre entier, donnée flottante.

- Recherche sur dates[4] : Les principaux formats de dates sont les suivants :

- MS-DOS Date & Heure (4 octets) : le mot le plus bas détermine l'heure et le mot le plus haut la date. Bit 0-4 secondes divisées par 2 ; bit 5-10 minutes (0-59) ; bit 11-15 heures (0-23 sur 24 heures) ; bits 16-20 jour du mois (1-31) ; bit 21-24 mois (1 = Janvier, 2 = Février, etc.) ; bit 25-31 décalage d'années depuis 1980

- Win32 FILETIME (8 octets) : la structure FILETIME est une valeur entière de 64 bits représentant le nombre d'intervalles de 100 nanosecondes depuis le 1^{er} janvier 1601.

- OLE 2.0 Date & Heure (8 octets) : Une double valeur flottante détermine pour la partie entière le nombre de jours passés depuis le 30 décembre 1899. La partie fractionnaire est interprétée comme l'heure du jour.

- ANSI SQL Date & Heure (8 octets) : constitué de 2 valeurs entières consécutives de 32 bits. La première détermine le nombre de jours depuis le 17 Novembre 1858. La seconde est le nombre d'intervalles de 100 microsecondes depuis minuit.

- UNIX/C Date & Heure (4 octets) : Une valeur entière de 32 bits détermine le nombre de secondes depuis le 1^{er} janvier 1970. Ce type de données est utilisé par UNIX, par C et C++ ("*time_t*"), et par les programmes FORTRAN depuis les années 80. Ce format est parfois utilisé pour définir le nombre de minutes écoulé depuis le 1^{er} janvier 1970.

- Java Date & Heure (8 octets) : Une valeur entière de 64 bits représentant les millisecondes depuis le 1^{er} janvier 1970. Essentiellement stockée selon le format "*poids fort en tête*", correspondant à l'ordre des octets caractéristique de Java.

– Donnée entière :

Signé, 8 bits ; Non signé, 8 bits ; Signé, 16 bits ; Non signé, 16 bits ; Signé, 24 bits ; Non signé, 24 bits ; Signé, 32 bits ; Non signé, 32 bits ; Signé, 64 Bits ;Les nombres multi octets sont stockés au format "poids faible en tête" (little endian). C'est le format commun sous Windows. Si l'on respecte le modèle "poids faible en

tête" (little endian), la valeur hexadécimale 10 27 peut être interprétée comme le nombre hexadécimal 2710 (en décimal : 10 000).

– Donnée de type flottant :
Float (simple) 4 octets ; Real 6 octets, Double (Double) 8 octets ; Long Double (étendu) 10 octets.

D'autres espaces de données se situent dans trois zones particulières que sont les espaces chutes, les espaces libres, les espaces inter partitions. On retrouve souvent ici, des preuves oubliées parce que difficilement atteignables. Les espaces chutes concernent les données présentes sur le disque entre un enregistrement et la fin du cluster. Les espaces libres correspondent à l'ensemble des clusters non utilisés mais contenant le plus souvent de l'information antérieure, les espaces inter partitions sont analysés lorsque un disque comporte plusieurs espaces logiques ou si un disque a été reformaté pour tenter d'effacer des données.

L'accès au secteur de Boot, répertoire racine ainsi qu'à la table d'allocation des fichiers fournit de précieuses indications sur le contenu du disque. À partir de la MFT miroir, il est possible de reconstituer, en partie une partition.

Un comparateur de fichier montre les différences de contenu entre 2 dates. Un comparateur d'images permet d'automatiser la recherche d'images présentant des couleurs de peaux et donc la sélection d'images contenant des couleurs proches de la peau humaine.

Les deux derniers points concernent la base de registre ainsi que le contenu de la mémoire.

Le rapport final doit immanquablement reprendre l'ensemble des traces trouvées ainsi que le journal des recherches effectuées.

La recherche de la preuve est un enjeu important. Elle doit respecter un ensemble de formalismes. Compte tenu de la multiplicité des formats de données, une attention particulière est à apporter dans l'interprétation des différents éléments trouvés.

Références

1. Menaces liées à l'application du MD5 – Philippe Schwada, Réné Heinzl –hahin9 2-2005
2. Revue MISC n ° 18 – les fonctions de hachages sont-elles condamnées ?
3. www.eprint.iacr.org/2004/199 - Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD *Xiaoyun Wang and Dengguo Feng and Xuejia Lai and Hongbo Yu*
4. Stefan Fleischmann X-Ways Software