Effacement sécurisé de disques durs

En 1996, Peter Gutmann a publié une étude (cf. [1]) sur la difficulté d'effacer de façon sécurisée un disque dur. Cette étude est considérée comme une référence majeure sur ce sujet, et de nombreuses personnes sont convaincues sur cette base qu'il est possible de relire des données effacées sur un disque dur si l'on n'a pas effectué auparavant de nombreuses passes d'écrasement de données (Gutmann proposait en 1996 de faire **35 passes**). Cette hypothèse est cependant controversée (cf. [2]) car personne n'a jamais montré qu'il avait pu récupérer d'un disque dur des données effacées par écrasement.

Dix ans après les travaux de Gutmann, nous faisons un point sur ce domaine.

La persistance des données effacées

Un disque dur ne dispose pas de fonction d'effacement; une fois une donnée écrite, la seule façon (logicielle) pour l'effacer est donc d'écrire d'autres données par-dessus les données existantes (écrasement). Cette technique est qualifiée par la DCSSI ("Direction Centrale de la Sécurité des Systèmes d'Information") d'effacement par "surcharge".

Cependant, si l'on observe au microscope électronique la surface du disque (par la technique de MFM – "Magnetic Force Microscopy"), on voit qu'il reste des résidus des données précédemment écrasées. L'illustration [3] montre un exemple flagrant de ces résidus. Dans cette image (parfois appelée "traces de tracteur"), on voit au centre de l'image la dernière donnée écrite sur le disque, et autour de cette trace principale, les traces d'anciennes données précédemment stockées au même endroit sur le disque. L'une des raisons de ces traces résiduelles est le fait que la tête d'écriture du disque n'est pas toujours parfaitement alignée sur sa piste; elle écrit donc souvent un peu à gauche et à droite de la piste. Une autre raison est le vieillissement du disque (vieillissement du support magnétisable, des têtes d'écriture, etc…). Ce second cas est intéressant parce que, du fait de ce vieillissement, il se peut que des données anciennes ne puissent plus être effacées correctement, même si l'on effectue de multiples ré-écritures.

Méthodes de récupération de données effacées

Plusieurs méthodes pour récupérer les données effacées sont mentionnées dans la littérature.

L'imagerie MFM :

Cette méthode est évoquée précédemment. Elle est effectuée par analyse microscopique de la surface des plateaux du disque (typiquement par zones élémentaires de 100x100 microns). Elle est très efficace dans sa capacité à identifier des résidus de données effacées. Par contre son "utilisabilité" réelle est questionnable. L'étude [4] estime ainsi que le temps nécessaire pour obtenir une image MFM de la surface complète du plateau d'un disque de 3,5 pouces (disque dur conventionnel d'un PC de bureau) est de l'ordre de 60 semaines (c'est le temps qu'il faut mécaniquement pour "balayer" avec la tête MFM la surface du plateau, en fonctionnement 24 heures sur 24, indépendamment de la densité du disque). Cette technique ne semble donc utilisable que sur des très petits volumes de données. Il n'existe pas, à notre connaissance, de démonstration de données qui auraient été reconstruites à partir d'images MFM. Cependant la NSA ("National Security Agency") est réputée maîtriser cette technique, probablement parce qu'elle a déposé en 1996 des brevets sur cette technologie d'analyse de disques.

Analyse du signal électronique lu sur un banc de test ("Spin-stand") :

Cette fois le disque est lu de la même façon qu'une tête de lecture d'un disque dur, mais sur des équipements plus

performants comme les "spin-stands". Un "spin-stand" est un banc de test qui est utilisé par les laboratoires pour mettre au point ou tester les technologies d'écriture sur disque. Il reproduit le fonctionnement d'un disque dur, mais avec une précision bien supérieure.

L'analyse repose sur un principe simple : lorsqu'un bit "1" est écrit par-dessus un bit "0" sur un disque dur, la valeur qui sera relue ensuite sera 0,95 (plutôt que 1,0). Par contre si un bit "1" est écrit par-dessus un bit "1", la valeur relue ensuite sera 1,05 (plutôt que 1.0). Au moment de la lecture les deux valeurs 0,95 et 1.05 sont bien sûr toutes les deux considérées comme représentant la valeur "1".

Cette explication est une large simplification de la réalité. Le signal électronique lu sur un disque est un signal analogique (une courbe sur un oscilloscope) qui est ensuite interprété par le pilote du disque dur. Ce signal analogique contient effectivement des résidus des signaux précédents enregistrés au même endroit sur le disque. Si l'on calcule l'onde théorique parfaite qui correspond aux données lues sur disque, et que l'on soustrait cette onde parfaite à l'onde réelle présente, on peut extraire ainsi une onde résiduelle qui est fonction des données précédemment stockées sur le disque.

Il existe au moins une démonstration publique de ce type d'analyse (cf. [5]). Dans ce cas, l'exercice (réussi) consistait à retrouver sur un disque effacé (par surcharge au moyen de données aléatoires) la trace d'un signal connu à l'avance (chaine de caractères "HELPHELP"). On le voit, cette démonstration est loin d'un cas réel de récupération de données (car dans la réalité, on ne connaît pas d'avance le contenu originel du disque).

Lecture d'un disque après désalignement des têtes de lecture :

Cette dernière technique n'est pas explicitement mentionnée dans la littérature (hormis éventuellement dans [4] sous le nom de "Spin-Stand MFM"), mais elle découle intuitivement des deux méthodes précédentes. Il s'agit ici de tenter de lire le bord des pistes d'un disque dur (qui contient les résidus des écritures précédentes, cf. MFM) en désalignant volontairement la tête de lecture du disque (ou en utilisant un banc "Spin-stand").

Qui sait réellement extraire des données effacées sur un disque ?

Il est certain qu'il subsiste une trace des données d'un disque dur, même après les avoir écrasées. Par contre, les méthodes (que nous venons de voir) pour exploiter ces traces sont très complexes à mettre en œuvre. On peut donc résumer la situation en disant qu'il est théoriquement possible d'extraire des données effacées d'un disque, mais que pratiquement c'est presque impossible.

On est alors en droit de se demander qui serait potentiellement capable d'effectuer ce type d'opération.

Il existe des sociétés spécialisées dans la récupération de données sur disque (comme par exemple les sociétés "**Ontrack.com**" ou "**DataClinic.co.uk**"), mais il nous parait peu probable que ces sociétés soient capables de réaliser de la récupération de données effacées par écrasement. Les techniques mises en jeu par ces sociétés vont en effet apparemment de l'analyse logicielle d'un disque (récupération de données effacées au niveau logique mais pas effacées physiquement par écrasement), jusqu'au remplacement de pièces sur un disque dur endommagé. Il s'agit donc de travaux très différents par leurs natures des techniques de récupération que nous avons évoquées.

En fait, la récupération de données effacées par écrasement demande des travaux importants de recherches préalables (pour mettre au point les procédures opérationnelles de récupération), et implique des outils d'analyse puissants et couteux. Il nous semble donc que seul un état (ou une puissance financière comparable) pourrait financer ces travaux de recherches. Il est probable que ces travaux seraient réalisés en collaboration avec des laboratoires de recherches spécialisés (tels que le CMRR de l'université de San Diego) ou les laboratoires des constructeurs de disques.

La menace "un disque effacé est récupéré par un tiers, et des données effacées sont extraites de ce disque par ce tiers" est donc, de notre point de vue, une menace concernant un niveau de secret de type "secret d'état". Il est clair cependant que ce risque existe, et que la question "quelqu'un est-il vraiment capable de le faire ?" n'est pas forcement pertinente : certains risques sont inacceptables, quelque soit la probabilité de l'occurrence du risque.

Comment effacer un disque dur ?

Destruction physique

La destruction physique d'un disque dur est sans doute la méthode la plus sûre pour neutraliser un disque qui a contenu des données sensibles : c'est un moyen rapide, et le bon déroulement de l'opération est facilement vérifiable.

Il existe des équipements spécialement vendus pour cette tâche. L'exemple que nous avons vu se présente sous forme d'une masse qui plie en deux le disque dur inséré dans la machine. On notera qu'avec ce type d'appareil les plateaux du disque dur sont cassés, mais que les fragments restent cependant très probablement analysables par MFM (cette méthode d'analyse se fait au moyen d'une sonde qui se déplace sur la surface du support analysé). Seule la destruction totale des surfaces magnétiques (abrasion ou fonte des plateaux) nous parait donc une solution absolument sûre.

Démagnétisation du disque

Il existe des matériels vendus pour supprimer les données d'un disque en les passant dans un champ magnétique intense (démagnétiseurs, ou "degaussers" en anglais). Cette méthode d'effacement est réputée extrêmement efficace, **si le champ magnétique est suffisamment fort**. D'après la littérature (cf [1]), ce champ doit être de 5 fois la coercivité des plateaux du disque pour que l'effacement soit "de sécurité" (la coercivité mesure le pouvoir de magnétisation des supports. Elle est couramment mesurée en "Oersted", unité de mesure dont l'abréviation est "Oe"). Les disques actuels ont couramment une coercivité comprise entre 2500 et 4500 Oe. Peu de démagnétiseurs commerciaux sont aujourd'hui capables de produire un champ magnétique assurant un effacement de sécurité pour les fortes coercivités.

On notera enfin que la démagnétisation d'un disque dur rend celui-ci inutilisable car elle détruit les informations de gestion contenus sur le disque (signal "servo", tables de paramétrages, etc…). Le résultat est donc équivalent à une destruction physique.

La destruction physique et la démagnétisation sont (dans cet ordre) les deux mesures les plus sûres pour neutraliser un disque. Ce sont les deux méthodes recommandées par la DCSSI (cf. [7]) pour neutraliser les disques qui quittent une organisation (cas dit de "l'exportation d'un disque") lorsque l'on craint que les disques puissent être analysés par des techniques "bas niveaux" (i.e. par les techniques que nous couvrons dans cet article).

Effacement par ré-écriture (surcharge)

La dernière méthode d'effacement (la moins robuste) est l'écriture (logicielle) de données sur l'ensemble du disque pour écraser les données sensibles qui y étaient stockées.

Le nombre de ré-écritures successives qu'il faut faire varie beaucoup selon les sources (et le niveau de sécurité souhaité).

Le cas le plus simple est l'écrasement en une seule passe avec des zéros. Cette solution nous parait insuffisante. La principale difficulté de l'effacement par ré-écriture (au moyen d'un logiciel) est qu'il existe de nombreuses couches logicielles à traverser avant que les données ne soient physiquement écrites sur le disque. En particulier le pilote qui existe à l'intérieur du disque (sur le circuit imprimé serti sur le boitier) effectue sa propre "bufférisation" (pour augmenter les performances). Il est possible aussi (pour gagner en fiabilité) que les demandes d'écriture disque soient journalisées (i.e. que l'on écrive sur des pistes réservées du disque un ordre qui devra être effectué plus tard) plutôt que d'effectuer réellement les opérations demandées sur les pistes du disque. Dans ce cas l'ordre "écrire partout sur le disque des zéros" pourrait ne pas être réellement effectué sur le disque, et un écrasement de ce type ne nous parait donc pas sûr. Le niveau minimal pour un effacement sécurisé nous parait donc d'effectuer sur le disque une seule passe d'écriture, en écrivant des

données les plus aléatoires possibles (par exemple via "/dev/urandom" sur Linux).

Le cas le plus extrême que nous avons rencontré est la recommandation donnée par Gutmann en 1996 d'effectuer 35 passes successives d'écriture pour écarter tout risque de récupération de données (cf. [1]). Ces 35 passes ont pour objet de prendre en compte toutes les techniques d'encodage des disques durs qui ont existées durant les 3 dernières décennies (en date de rédaction de l'étude de Gutmann), et Gutmann reconnaît dans une épilogue ajoutée postérieurement à son étude que pour les technologies contemporaines (utilisant la reconnaissance du signal magnétique par la technique "PRML" – "Partial Response Maximum Likelihood"), quelques passes d'écriture de données aléatoires sont probablement suffisantes.

La dernière approche qu'il nous parait intéressant de mentionner est la fonctionnalité de "**Secure Erase**" qui existe dans les versions récentes de la spécification de l'interface ATA (et qui existe donc en particulier dans les disques "Serial-ATA"). Dans ce type de disque, le pilote du disque dispose d'une commande "Secure Erase" qui, lorsqu'elle est activée, provoque un effacement (par surcharge) de l'ensemble des blocs du disque. L'intérêt majeur de cette solution est qu'elle est a priori plus fiable qu'une solution logicielle de plus haut niveau : plus l'ordre d'effacement est donné à un niveau proche de la couche matérielle, plus il y a de chances que cet ordre soit exécuté sans erreur. De notre point de vue cependant, cette solution "Secure Erase" ne nous parait pas sûre (sans tests préalables d'efficacité effective) si l'on envisage le cas où il existerait des commandes non documentées permettant d'accéder aux données prétendument effacées. Cette éventualité nous semble devoir être envisagée, lorsque l'on a auparavant envisagé la possibilité d'une reconstitution de données écrasées par technique de MFM (puisque cette technologie MFM est extrêmement complexe à réaliser). On notera cependant que le NIST ("National Institute of Standards and Technology") américain recommande dans plusieurs cas cette méthode d'effacement (cf. [8]).

En conclusion, les personnes pour lesquelles le risque que des données effacées soient ensuite retrouvées est inacceptable, détruiront physiquement le disque (ou le démagnétiseront, ce qui le rend de toutes façons inutilisable), même si ce risque est hypothétique. Les autres utiliseront sans doute un outil logiciel de surcharge. Le logiciel PC "open source" de référence dans ce domaine est le logiciel DBAN (**http://dban.sourceforge.net/**). Il se présente sous forme d'une disquette "bootable", et propose plusieurs options d'effacement par surcharge, dont :

- une passe de surcharge avec des zéros (mode "Quick Erase")
- une passe de surcharge avec des données aléatoires (mode "PRNG Wipe")
- un effacement conforme à au standard américain "DoD 5220-22.M" (trois passes : la première avec un caractère fixe, une seconde avec son complément, puis une troisième avec des données aléatoires)
- ou les 35 passes recommandées par Gutmann

Sur un PC contemporain, une passe d'effacement sur un disque de 60 Go dure typiquement 50 minutes.

Pour plus d'informations :

- [1] : Peter Gutmann (1996)
  http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html
- [2] : Daniel Feenberg (2003)
  "Can Intelligence Agencies Read Overwritten Data? A response to Gutmann."
  http://www.nber.org/sys-admin/overwritten-data-guttman.html
- [3] : Image MFM de données effacées sur un disque dur
  http://www.veeco.com/nanotheatre/nano_view_detail.asp?ImageID=78

- [4] : ActionFront (2004)
"Recovering Unrecoverable Data"
http://www.actionfront.com/whitepaper/Drive-Independent%20Data%20Recovery%20Ver14AIrs.pdf

- [5] : Secure Erase of Disk Drive Data (2002)
http://www.idema.org/_smartsite/modules/local/data_file/show_file.php?cmd=download&data_file_id=1093

- [6] Coercivité des disques magnétiques
http://www.hitachigst.com/hdd/technolo/overview/chart11.html
- [7] DCSSI (2003)
Guide technique pour la confidentialité des informations enregistrées sur les disques durs à recycler ou exporter
http://www.ssi.gouv.fr/fr/documentation/Guide_effaceur_V1.12du040517.pdf
- [8] NIST (2006)
Guidelines for media sanitization (DRAFT)
http://csrc.nist.gov/publications/drafts/DRAFT-sp800-88-Feb3_2006.pdf