COURS DE MOBILITÉ
Sujet n°20

Comment effacer vraiment un disque dur, symétriquement comment retrouver des informations?

Université Paris-Diderot
M2 IMPAIRS

Joaquim LEFRANC <lefrancjoaquim@gmail.com>
Jérôme SKODA <contact@jeromeskoda.fr>

18/03/2018

1. Comment fonctionne un disque dur

1.1. Deux type de disque dur

Un disque dur est une mémoire physique permettant de stocker et de garder des informations même hors tension.

Il faut cependant distinguer deux type :

Les HDD : qui fonctionnent sur le principe électromagnétique, ils sont constitués d'un ou plusieurs plateaux tournants qui sont magnétisés par une tête de lecture et d'écriture.

HDD ouvert

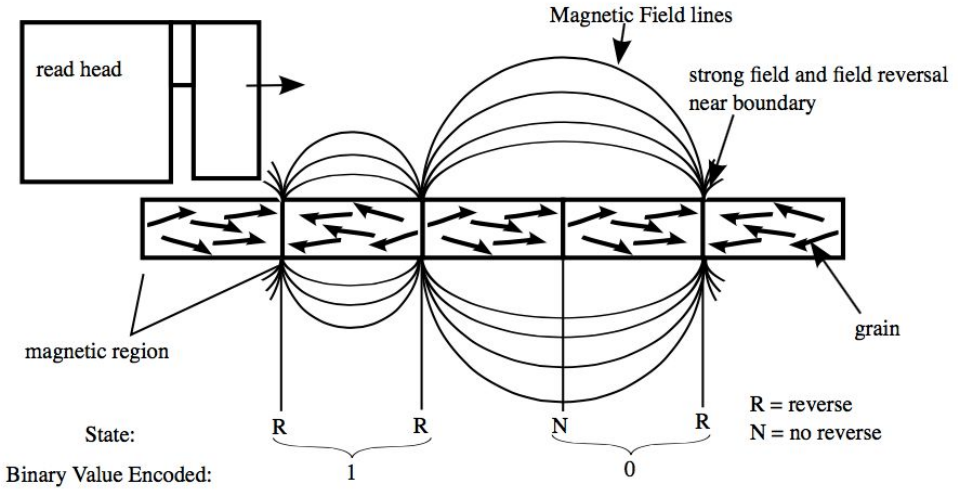
Les SSD : qui fonctionnent sur des cellules de mémoire électronique, ils sont constitués de plusieurs blocs de mémoires statiques.

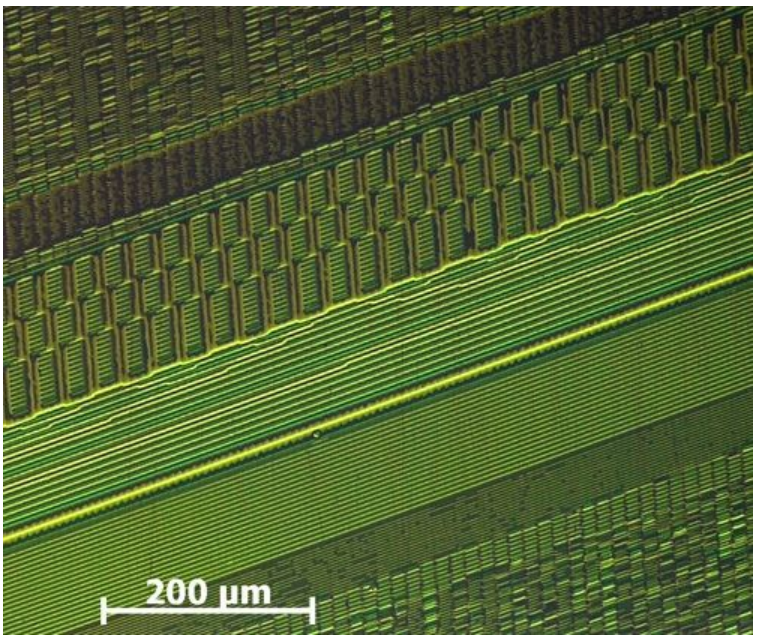
SSD ouvert

1.2. Représentation des données

1.2.1. HDD

Les HDD ont donc un plateau divisé en milliers de petits blocs contenant des particules sensibles aux champs magnétiques. Ces particules peuvent donc être orientées vers le NORD ou le SUD magnétique, ce qui correspond aux états 1 et 0.
Pour la suite il faut noter que cette direction NORD ou SUD n'est pas parfaitement exacte, par conséquent le contrôleur du disque s'arrange pour arrondir les informations provenant de la tête de lecture.
De plus l'écriture d'un bit influe légèrement sur les bits situés juste à côté, cet effet est également à prendre en compte lors de l'effacement du disque.



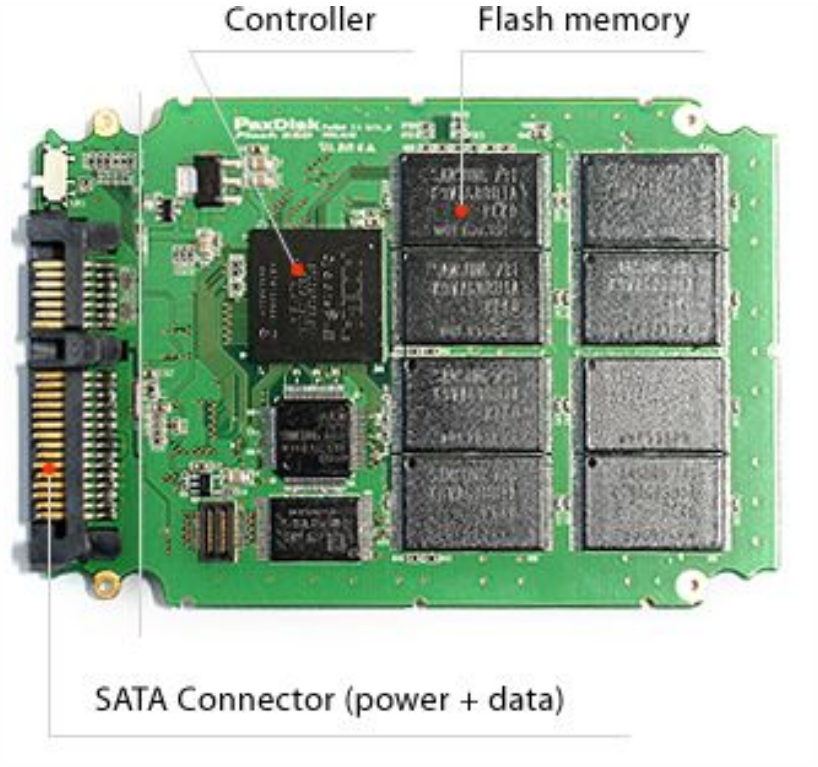


1.2.2. SSD

Les SSD utilisent plusieurs mémoires flash non volatiles capable de conserver les données sans alimentation constante et en garantissant la persistance des données après une coupure de courant soudaine.

Ils sont composés de plusieur éléments clefs:

- Le contrôleur : processeur exécutant le microprogramme d'exploitation du SSD, permettant à l'ordinateur hôte d'interagir avec le SSD.
- La mémoire : composant permettant de stocker les données.
- L'Interface hôte : connecteur physique.

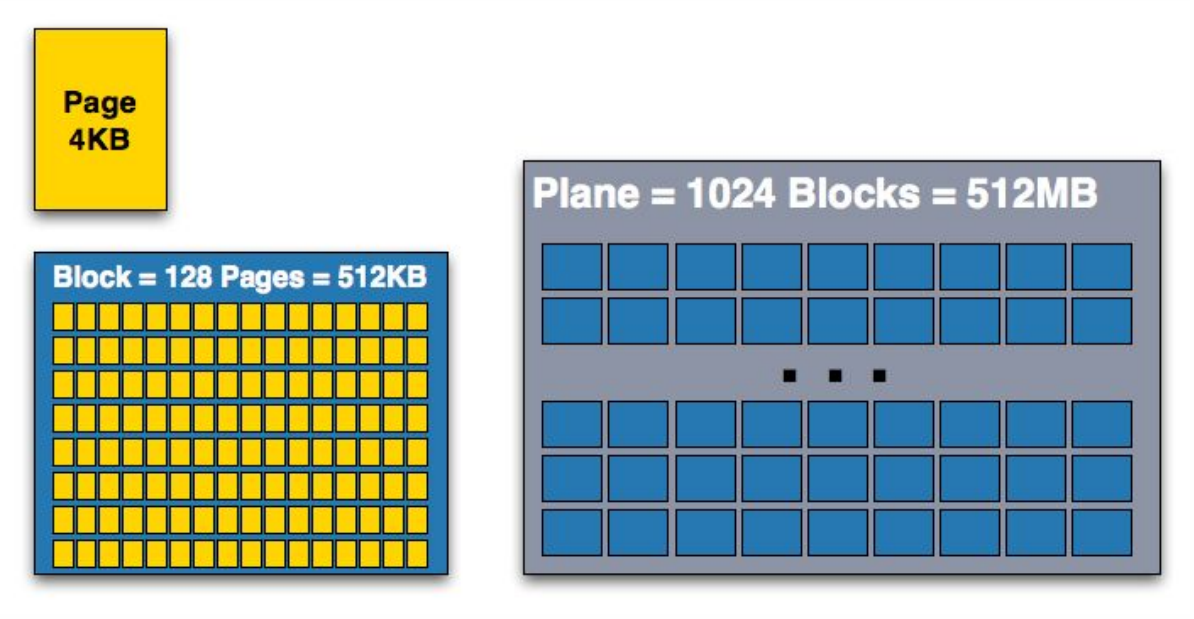


Composition d'un SSD

Pour garantir des performances en lecture et écriture, les puces mémoire sont utilisées en parallèle et sont conçues de telle manière que les données sont répartie uniformément entre chacune des puces.

Les mémoire flash sont structurées de la manière suivante:

- Page : Généralement de 4Kb
- Block : équivalent à 128 pages
- Plane: contenant 1024 blocks



Répartition en Page / Block / Plane

Une mémoire flash est généralement constituée de plusieur planes.

L'une des spécificité d'un SSD est que l'unité la plus petite effaçable est une page complète (4KB), il n'est pas possible d'effacer un octet uniquement.

2. Rémanence des données

2.1. Les disques durs

2.1.1. Provenance

Les disques durs ont deux fonctionnalités : l'écriture et la lecture. Un disque dur ne dispose pas de la fonctionnalité d'effacement.
Ainsi l'opération d'effacement des données consiste à « écraser » les données c'est à dire écrire d'autres données sur l'emplacement des données à effacer.

2.1.1.1. Positionnement de la tête de lecture

Cependant cette opération n'est pas parfaite : la tête de lecture d'un disque dur n'est pas toujours alignée parfaitement sur la piste, soit légèrement à gauche ou soit légèrement à droite, il reste donc des « traces de l'ancienne donnée » plus ou moins exploitable après chaque opération d'écrasement.
Avec le vieillissement, le décalage rendent certaines « traces » ineffaçable même après un grand nombre d'écrasement.

2.1.1.2. Orientation imparfaite des bits

De plus, il faut rappeler que les défauts physiques font que l'orientation des bits n'est pas exactement NORD ni SUD.

Exemple simplifié :

NORD parfait: 1
SUD parfait: 0

NORD réel: entre 0.7 et 1.0
SUD réel: entre 0.0 et 0.3

ERREUR lecture: 0.4 - 0.6

Donnée à effacer : **D = 1** 0 **1** 0
Donnée effacée : **E = 0 0 0 0**

En réalité les inclinaisons NORD / SUD ressembles à ça :

D = 0.8 0.2 **0.8** 0.2
E = 0.3 0.0 **0.3** 0.1

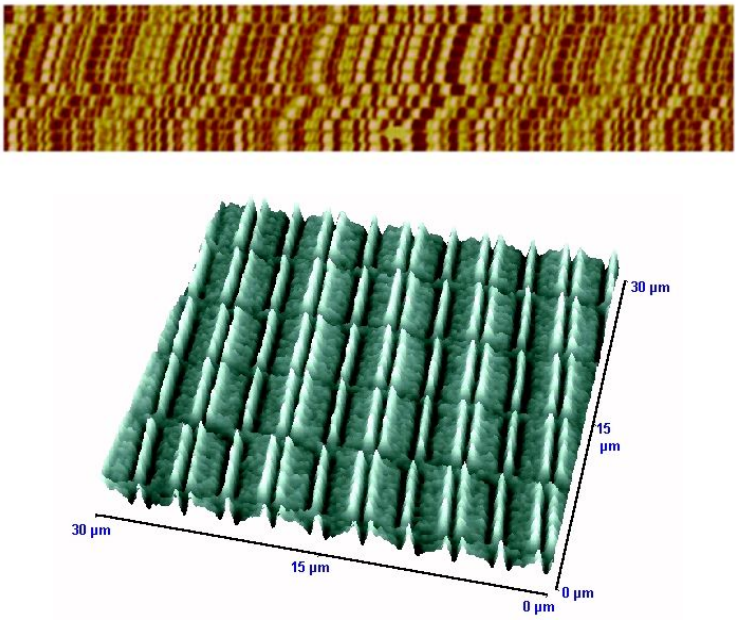
On peut voir ici que l'écriture à 0 des bits 1 et 3 a eu pour conséquence de "renforcer" les bits 2 et 4.

2.1.2. Outils d’analyse

L’écrasement des données n’est pas parfait et laisse des traces. Nous allons maintenant explorer les différentes manières de récupérer ces traces.

2.1.2.1. La microscopie à force magnétique (MFM)

Un microscope à force magnétique (MFM) permet de mesurer les variations de l’aimantation de la surface de l’échantillon.

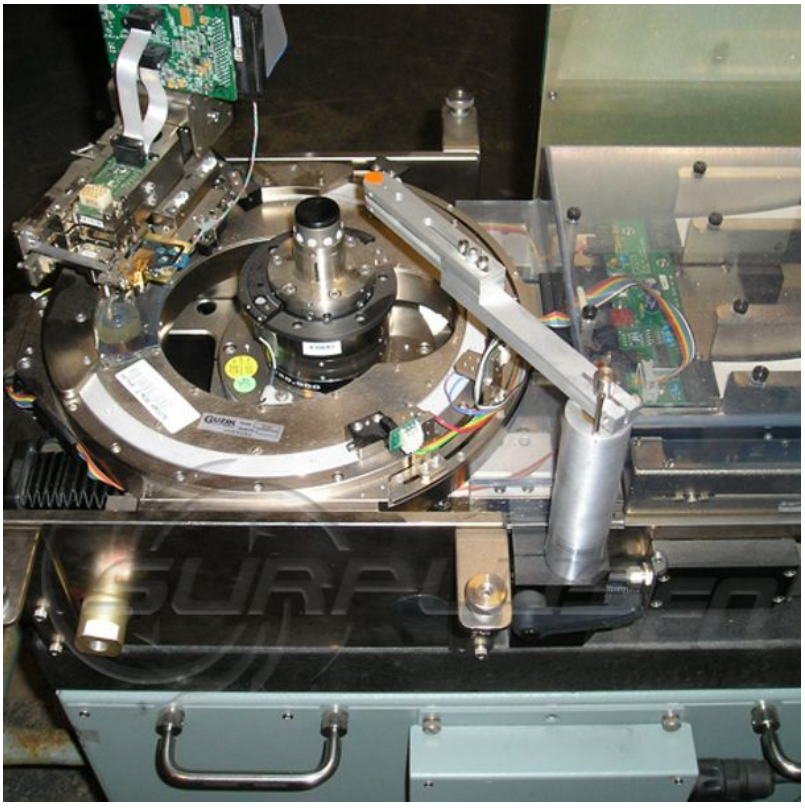


Modélisation d'imagerie MFM sur un disque dur

Les images obtenues par MFM sont très efficaces pour identifier les traces, cependant cette technique n’est à ce jour pas envisageable pour récupérer des données effacées car elle demande énormément de temps de l’ordre d’un an pour faire une imagerie complète d’un seul plateau de disque dur standard (3.5 pouces) et n’a pas eu de démonstration de récupération de donnée connu à ce jour.

2.1.2.2. Analyse sur un Spin-stand

En utilisant une tête de lecture plus performante comme celle des « spin-stand », équipement utilisé dans les laboratoires pour concevoir les disques dur il est possible d'obtenir plus de précision sur la lecture qu'en utilisant une tête de lecture classique.



Head test spin stand

Avec la précision supplémentaire du « spin-stand » et en partant du constat :

- Lorsqu'un bit « 1 » écrase un bit « 0 » sa valeur sera lue à environ à 0.95 sur un « spin-stand »
- Lorsqu'un bit « 1 » écrase un bit « 1 » sa valeur sera lue à environ à 1.05 sur un « spin-stand »

Cette technique a été expérimentée, la publication « Secure Erase of Disk Drive Data » [6], démontre qu'il est possible de récupérer des données écrasées.
Cependant, dans cette expérimentation les données recherchées étaient déjà connues.

Cette technique reste donc expérimentale et réservée aux laboratoires.

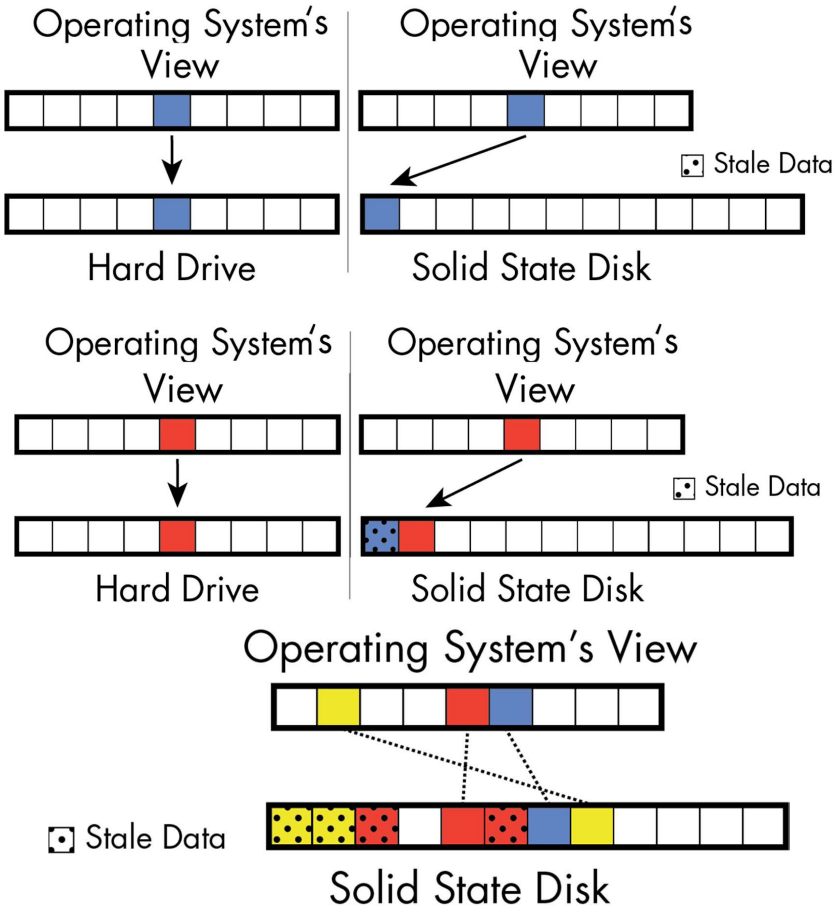
2.1.3. Conclusion

Actuellement, sur un disque dur, l'effacement par écrasement des données bien qu'imparfaite reste sûr. Il n'existe aucun outil connu ni d'application capable de récupérer des données écrasées.

2.2. Les SSD

Les SSD possèdent des contraintes d'effacement et des contrôleurs sophistiqués capable de réallouer des blocs et copier des fichiers partiellement ou entièrement.

Ces opérations permettent de garantir des performances en lecture et écriture et d'augmenter la durée de vie des cellules de mémoire flash en répartissant les données. Cependant, en effectuant ces opérations, beaucoup d'anciennes données peuvent être laissées sur le lecteur.



Répartition des anciennes données conservé sur SSD

Avec ce fonctionnement, il est impossible de s'assurer qu'un fichier est définitivement supprimé sur un SSD sans procéder à un effacement complet.
D'après l'article « Reliably Erasing Data From Flash-Based Solid State Drives » [x], la commande d'effacement complet du SSD "Secure Erase" fournie par les fabricants est sûre et vérifié.

Alors que selon le même article, il n'y a aucune méthode logicielle vérifiée permettant la suppression sécurisée d'un unique fichier.

3. Les systèmes d’exploitation

3.1. Supprime t-il vraiment les fichiers ?

Dans la majorité des systèmes d’exploitation la possibilité de récupérer des fichiers supprimés accidentellement est un standard pour l’expérience utilisateur.
Les utilisateurs ne s’attendent pas à la suppression définitive de leur fichier mais sont habitués à la métaphore de la mise en corbeille.

Avec le mécanisme de corbeille, lorsque que l’on supprime un fichier, il est transféré dans la corbeille avec l’ajout de métadonnées contenant:

- Son nom et son emplacement d’origine (avant suppression)
- Sa date de suppression
- L’id de l’utilisateur (pour certains environnements multi-utilisateurs)

Avec ces métadonnées, le système d’exploitation est capable de restaurer le fichier.

Si l’utilisateur décide de supprimer un fichier de la corbeille plusieur opération se produise:

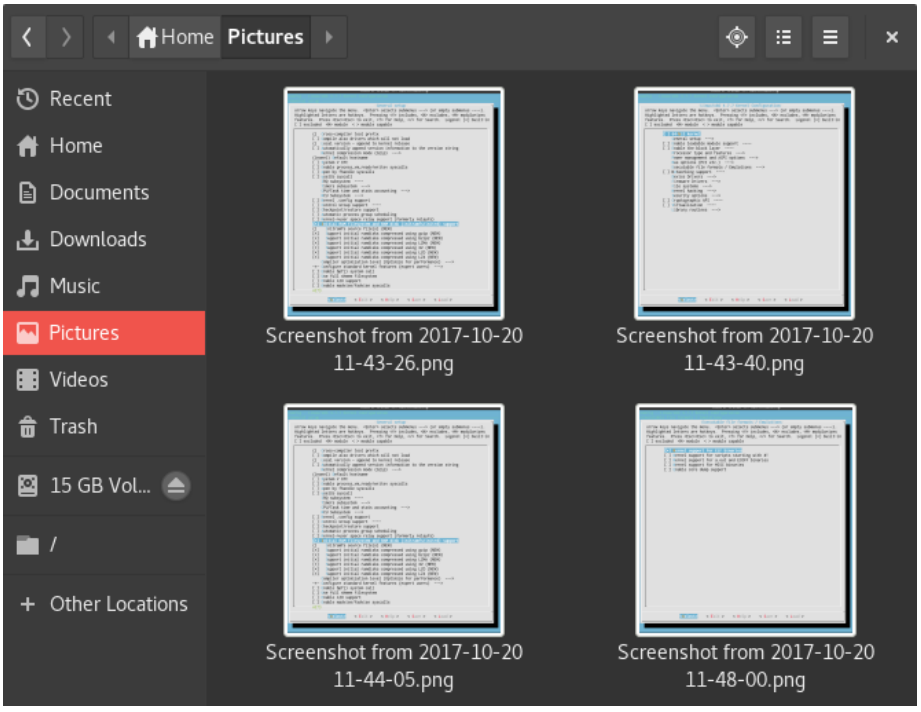
- Les fichiers sont dé-répertoriés de la corbeille, ils ne seront plus visible dans la corbeille.
- L’espace occupé par ces fichier alors est libéré sur le disque, ce qui signifie que l’espace devient de nouveau disponible en écriture pour d’autre fichiers

Après l'opération de suppression dans la corbeille, l’espace libéré par le fichier contient toujours les données car il n’y a pas de réécriture par dessus automatiquement. Le fichier n’est donc pas vraiment supprimé.

3.2. Les métadonnées persistantes

Après avoir supprimé un fichier, le système d'exploitation peut laisser des traces de celui-ci sous différentes formes de métadonnées comme par exemple: l'indexation pour la recherche, fichier temporaire ou les miniatures. Nous allons nous intéresser à la dernière qui est la plus répandue.

Les miniatures sont des versions d'une image ou vidéo dont la taille est réduite, utilisé pour les reconnaître plus facilement dans un environnement de bureau. Les miniatures sont présentes dans la plupart des environnement de bureau tels que Windows, MacOS, GNOME, KDE etc.



Miniature sous GNOME

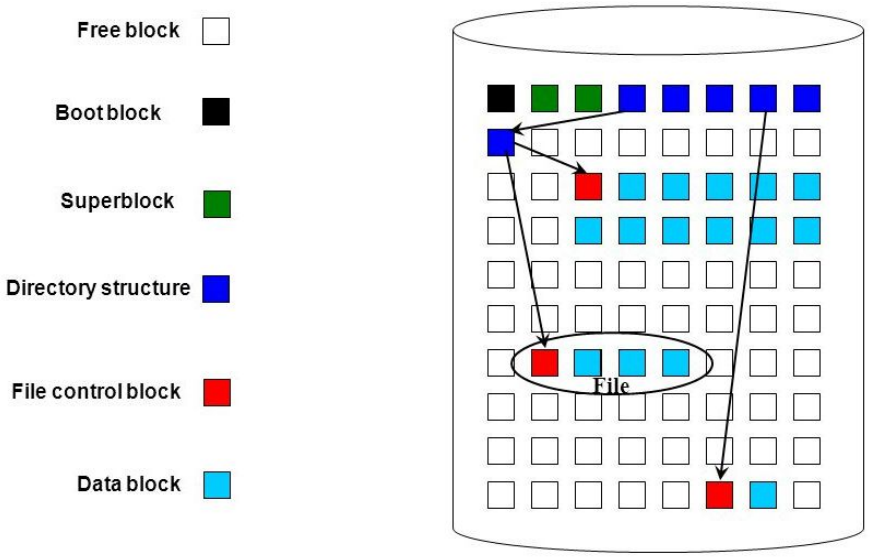
Les fichiers supprimés peuvent avoir leurs miniatures conservées dans des fichiers cachés générés automatiquement par le système comme par exemple les fichiers ".DS_Store" sous MacOs ou "Thumbs.db" sous Windows.

Ces informations peuvent constituer un indice sur les fichiers susceptibles d'avoir été supprimées ou visionnées. Par exemple dans une affaire de pornogaphie juvénile: "UNITED STATES vs VOSBURGH", les enquêteurs ont récupéré deux miniatures (dans un fichier "Thumbs.db") qui on pu servir de preuve sans avoir trouvé les deux images taille réel dans le disque dur.

4. Comment récupérer nos données effacé

Comment nous l'avons vu précédemment, la plupart des systèmes d'exploitation et des systèmes de fichiers n'effacent pas toujour directement les données.
Il est donc possible de récupérer un fichier supprimé tant qu'il n'a pas été partiellement ou totalement remplacé.

Les systèmes de fichier comme par exemple FAT, NTFS, ext2, ext3 et ext4 stockent les fichiers dans des blocs de données dont la taille est fixée à l'initialisation du système de fichiers. De plus, la plupart des systèmes d'exploitation essaient de stocker les données des fichiers de manière contiguë.



Structure d'un système de fichiers

Les logiciels de récupération des données utilisent cette spécificité des systèmes de fichiers pour retrouver les fichiers supprimé.

Dans un premier temps, le logiciel recherche la taille des blocs. Cette information est contenue dans les superbloc (dans ext2, ext3 et ext4) ou dans le bootblock (FAT ou NTFS).

Une fois la taille des blocs connue, le logiciel parcourt chacun des blocs en recherchant une signature (nombre magique) de type de fichiers connu.

Par exemple, identifier un fichier JPEG lorsqu'un bloc commence par:

0xff 0xd8 0xff 0xe0 ou **0xff 0xd8 0xff 0xe1** ou **0xff 0xd8 0xff 0xfe**

C'est sur ce principe que la plupart des logiciels de récupération fonctionnent. De plus, ils utilisent aussi des informations générés par les systèmes d'exploitation (métadonnées, journaux, configuration etc) pour affiner leurs recherches.

5. Comment effacer les données définitivement

Les enjeux d'une suppression efficace sont multiples : suppression d'informations personnelles ou professionnelles avant de jeter ou vendre un HDD, pour éviter la vente d'informations personnelles ou la diffusion d'informations industrielles.

Pour contrer la rémanence des données et les données libérées mais non supprimées par le système d'exploitation, il existe plusieurs méthodes dont le niveau de sécurité et de vitesse peut varier. Les plus connu sont la destruction du support physique, effacement réitérées (overwriting) et effacement des clefs de déchiffrement.

Dans le graphique ci-dessous recouvre des méthodes d'effacement classé selon leur niveau de sécurité et leur vitesse.

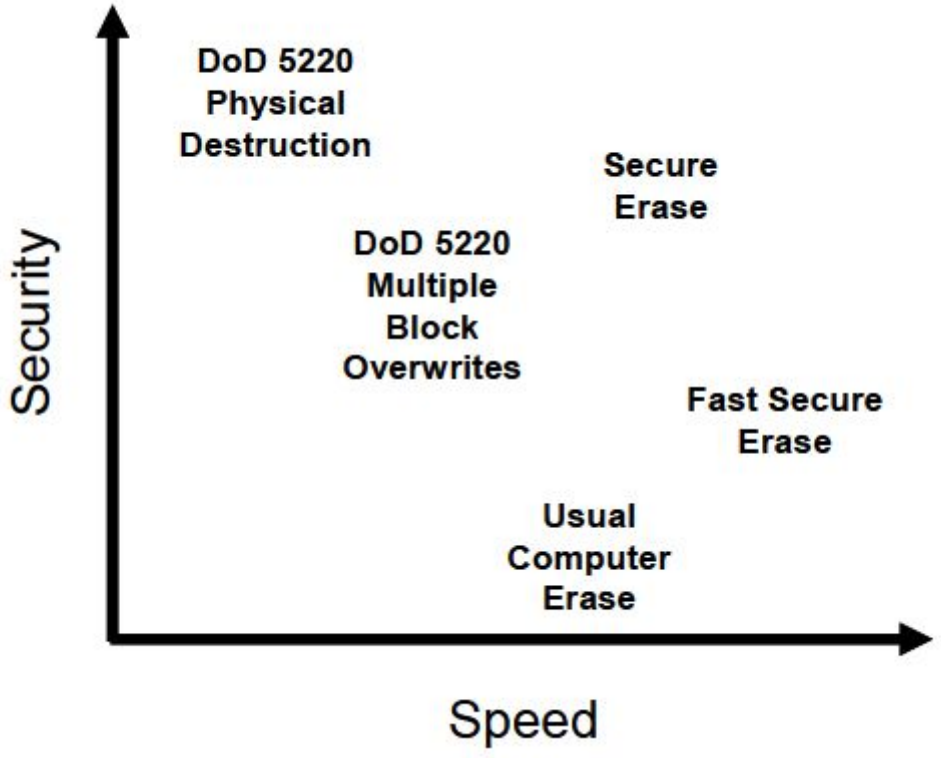


Tableau comparatif Sécurité vs Vitesse

5.1. Destruction du lecteur physique

Cette méthode qui consiste à détruire le support physique des données mais rend le disque dur inutilisable. Le résultat peut être variable, toutes les méthodes ne garantissent pas l'impossibilité de récupérer toute ou une partie des données.

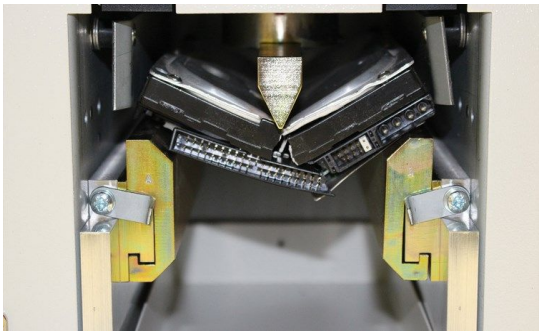
Voici une liste de méthodes possible que chacun peut utiliser:

- Les dégât des eaux, feu ou surtension sont à éviter. Ce sont des causes très répandues de pannes dans les entreprises, les centre de récupération sont habitués à récupérer les données après ce genre de dégât. Il n'est pas rare dans le cadre d'enquête judiciaire, de pouvoir récupérer les données d'un disque dur brûlé après un incendie ou pêché après un naufrage.

- Le micro-onde n'est pas suffisamment puissant pour démagnétiser un disque dur. Démagnétiser un disque dur demande du matériel professionnel (dégausseur) et un certain savoir faire car l'efficacité de cette méthode varie en fonction des modèle de disque dur.

- Le découpage, perçage ou pliage des plateaux sont sans doute parmi les méthodes les plus sûr de tout un chacun. Il est extrêmement difficile pour un centre de récupération de récupérer des données sur un support ainsi détruit.

Perçage d'un disque dur



Pliage d'un disque dur

5.3. Effacement par réécriture

5.3.1. Effacement total d'un disque dur

Une solution naïve consiste à écrire des '0' partout sur l'ensemble des enregistrements. A première vue, ceci écrase toutes les données.

```
$ sudo dd if=/dev/zero of=/dev/sdX
```

Mais comme nous avons pu voir précédemment en 2. Rémanence des données (page 7), cette méthode n'est pas suffisante. L'orientation imparfaite des bits après écriture et le mauvais positionnement de la tête de lecture peuvent suffir à laisser des traces exploitable pour la récupération des données présentes avant l'écrasement.

La façon de contrer ce problème est d'effectuer plusieur passes et d'utiliser des valeurs aléatoires de bit pour l'écriture.

Par exemple écrire des '0' puis un bit aléatoire.

```
$ sudo dd if=/dev/zero of=/dev/sdX
$ sudo dd if=/dev/urandom of=/dev/sdX
```

C'est sur ce principe que fonctionne la plupart des algorithmes d'effacement sécurisé.

Voici la liste des algorithmes les plus connus:

Algorithme	Passe 1	Passe 2	Passe 3
DoD 5220.22-M	'0'	'1'	aléatoire
CSEC ITSG-06	'1' or '0'	not(passe1)	aléatoire
AR 380-19	aléatoire	spécifié	not(spécifié)
AFSSI-5020	1	0	aléatoire

Pour effectuer un effacement de fichier sécurisé, des outils sont disponibles comme par exemple Darik's Boot and Nuke (DBAN) un logiciel libre pouvant être lancé depuis un clef USB bootable.

5.3.2. Effacement total d'un SSD

Un SSD n'est pas constitué de la même manière qu'un disque dur, l'effacement s'effectue uniquement par page de 4kB et le systéme ne contrôle pas où il écrit dans le SSD.
La seule manière d'effacer totalement un SSD est d'utiliser la commande "secure erase" du fabricant. Cette commande consiste à vider tous les électrons stockés dans les cellule de mémoire.

Voici comment procéder:

1. Identifier le SSD à effacer
   $ hdparm -I /dev/sdX - Affiche les informations sur un lecteur.
   $ hdparm -I /dev/sdX | grep Model - Affiche le modèle du lecteur
2. Vérifier que le disque est ni bloqué ni gelée
   $ hdparm -I /dev/sdX | grep locked - Affiche si le disque est bloqué
   $ hdparm -I /dev/sdX | grep frozen - Affiche si le disk est gelée
3. Effacement du SSD
   $ hdparm –security-erase PASS /dev/sdX - Efface le SSD

5.3.3. Effacement d'un fichier sur disque dur

Il existe la possibilité d'effacer définitivement les fichiers d'un disque dur avec la commande "shred" ou "srm".

Ces commande fonctionnent de la même manière que les algorithme présentés en 5.3.1. Effacement total d'un disque dur (page 16) en effectuant plusieur passes d'écrasement des données.

5.3.4. Effacement d'un fichier sur SSD

Il n'existe pas de possibilité de supprimer définitivement un fichier d'un SDD avec certitude sans passer par un effacement complet.

Ce problème est dû à la structure des SSD, qui laisse les données des ancienne version du fichier présentes en mémoire (voir 2.2. Les SSD page 10)

5.4. Effacement sécurisé de donnée chiffré

Dans un disque dur chiffré, il est possible de supprimer les clés de déchiffrement ce qui rend toutes les données stockées inaccessible. Cette méthode est simple, rapide et a une grande fiabilité.

6. Conclusion

Effacer vraiment les informations d'un disque dur n'est pas une opération aussi simple qu'il n'y paraît et ceci pour plusieurs raisons:

- Le matériel n'est pas parfait, de fait de la rémanence des données laissées malgré des réécritures.

- Les systèmes d'exploitation ne suppriment pas vraiment les informations et conserve des métadonnées exploitables pour retrouver des données pourtant censées être effacées.

- Il existe de nombreux outils de récupération capables de retrouver des données ou leurs traces.

Cependant, il existe des méthodes effaçant définitivement des données dont l'efficacité et la sécurité ont été prouvé.

7. Ressources

- “Secure Deletion of Data from Magnetic and Solid-State Memory”, Peter Gutmann, https://www.cs.auckland.ac.nz/%7Epgut001/pubs/secure_del.html
- “Image MFM” : http://electron.mit.edu/~gsteele/mirrors/elchem.kaist.ac.kr/jhkwak/TopometrixWeb/images/hdbt3d3.jpg
- “Data Reconstruction from a Hard Disk Drive using Magnetic Force Microscopy”, Kanekal, Vasu : https://escholarship.org/uc/item/26g4p84b
- Image spin stand : https://www.surpluseq.com/image/cache/catalog/product/guzik/rwa-1632-analyzer/guzik-rwa-1632-analyzer_c-1000x1000-w2-57-98-31-0.jpg
- “Can Intelligence Agencies Read Overwritten Data?”, https://www.nber.org/sys-admin/overwritten-data-guttman.html
- “Secure Erase of Disk Drive Data”, Gordon Hughes, Tom Coughlin : https://www.researchgate.net/publication/228740643_Secure_erase_of_disk_drive_data
- Image SS:D http://cseweb.ucsd.edu/~m3wei/assets/pdf/LISA2011-sanitize.pdf
- “Reliably Erasing Data From Flash-Based Solid State Drives”, Michael Wei, Laura M. Grupp, Frederick E. Spada, Steven Swanson: https://cseweb.ucsd.edu/~swanson/papers/Fast2011SecErase.pdf
- “Tutorial on Disk Drive Data Sanitization”, Gordon Hughes, Tom Coughlin: cmrr.ucsd.edu/_files/data-sanitization-tutorial.pdf
- “The FreeDesktop.org Trash specification”: https://standards.freedesktop.org/trash-spec/trashspec-latest.html
- “UNITED STATES v. VOSBURGH”: http://caselaw.findlaw.com/us-3rd-circuit/1522221.html
- “Exemple de la récupération de données perdues”: https://www.nodisknorisk.com/fr/exemple-de-la-recuperation-de-donnees-perdues
- “A Practical Guide to Computer Forensics Investigations” Dr Darren Hayes
- “Shred : détruire efficacement un fichier“: https://doc.ubuntu-fr.org/shred
- “PhotoRec, Digital Picture and File Recovery”: https://www.cgsecurity.org/wiki/PhotoRec