FORMATION INITIALE Récupération de données diskcard 5/08/2011 – Licence CC-BY-SA

I- PROTÉGER

1/ Analyse du contexte de la perte de données

Le support est-il allumé ou éteint ?
Existe-t-il des spécificités structurelles (raid, serveur, lvm) ?
Quel est le type de sinistre rencontré ?

- Suppression
- Défaillance logique
- Défaillance structurelle
- Défaillance matérielle

2/ Élaborer un plan d'action

En général, le plan d'action est le suivant, à la variation près des spécificités rencontrées :

- Si le support est allumé, il faut effectuer les sauvegardes d'urgence vitale, puis éteindre
- Une fois le support éteint, on procède à l'examen initial.

Sauvegardes d'urgence vitale :

1. Spécificités structurelles (config raid, config lvm, config serveur, mbr).
2. Fichiers vitaux (mots de passes, clé de chiffrage, puis fichiers personnels essentiels en commençant par les plus petits).
3. Fichiers secondaires importants si l'on soupçonne une panne définitive au reboot.

II- EXAMINER

1/ Examen matériel

Examen physique :

1. Le disque a-t-il déjà été ouvert ?
2. Présente-t-il des traces de chocs ?
3. Observation de l'état du PCB (connectiques et composants).

Examen auditif : Lors d'un branchement rapide sur un connecteur externe (sans connectique data), le disque montre-t-il des symptômes spécifiques ?

1. Rotation régulière ou non ?
2. Frottements ou claquements ?
3. Rotation forcée ou alternée ?

Note au sujet de l'examen auditif : si il n'y a pas de rotation, il faut vérifier l'état des composants du pcb en priorité. Sinon, c'est probablement le moteur qui est HS. Tout autre symptôme auditif provoquant une rotation qui n'est pas régulière est le signe d'une défaillance matérielle lourde ou imminente. Lorsque la rotation est régulière mais que le disque n'est pas reconnu (dans le bios ou dans les logs du noyau) il faut en priorité penser à une défaillance du firmware du disque.

2/ Préparation à l'examen avancé

Préparer l'OS support :

1. Désactiver l'auto-mount et l'auto-open
2. Désactiver les miniatures
   → sous ubuntu, Alt+F2 puis gconf-editor / nautilus

Préparer le matériel nécessaire :

1. Support de sauvegarde avec taille suffisante
2. Connectique externe éventuelle
3. Sécurité électrique

3/ Examen logique

1. Prise d'informations sur l'état matériel
   Le disque est-il visible dans le bios (si branchement interne) ?
   Le disque est-il visible dans lsusb et dmesg (si branchement externe) ?
   Sa dénomination et sa taille sont-elles correctement détectées ?

2. Prise d'informations sur l'état structurel
   Y a-t-il un raid et quel est son état (dmraid, mdadm) ?
   Ex : « sudo mdadm --examine /dev/sda2 »
   Ex : « sudo mdadm –details /dev/md0 »
   Y a-t-il un lvm et quel est son état (outils lvm2) ?
   Ex : « sudo [vg|pv|lv]display »
   Y a-t-il un conteneur chiffré (outils cryptsetup) ?

3. Prise d'informations sur l'état des partitions et des systèmes de fichiers
   Vérification graphique (gparted et/ou utilitaire de disque)
   Vérification terminal (fdisk, sfdisk, blkid, mount, parted)
   Ex : « sudo sfdisk -luS », « sudo blkid » etc…
   Vérifications complexes (hexdump et/ou textdisk)
   Ex : « sudo dd if=/dev/sda bs=512 count=1 | hexdump -C »
   Ex : « sudo testdisk /dev/sda »

III- ALERTER : Le cas de la panne physique importante ou imminente

Lorsque l'on constate une panne physique importante ou imminente, il faut garder à l'esprit que nous ne serons pas capables de faire de miracles. Cela nécessite un traitement en salle blanche par un personnel qualifié, expérimenté, et doté de matériel spécifique. De plus, il faut se souvenir à tout instant que lors d'une panne physique la première tentative est bien souvent déterminante : c'est celle qui donne les meilleurs résultats lorsqu'elle est correctement réalisée, c'est aussi celle qui aggrave le plus la situation au point que parfois les professionnels ne pourront rien récupérer si l'insistance et la méconnaissance entraînent des dégâts irréversibles.

IV- SECOURIR :

1/ Sauvegarder l'état initial

Outils utiles : dd, dcfldd, dd_rescue, dd_rhelp, ddrescue
Ma préférence : ddrescue

Ex : « sudo ddrescue /dev/sda /media/mondisk/image.dd /media/mondisk/image.log -n »
Options très utiles :

-n : éviter les secteurs défecteux
-d : accès direct-to-disk (bypass firmware, parfois utile, plus lent)
-i : déplacement position début copie (lorsqu'on connait une zone HS ou inutile)
-T : reprendre après déconnexion de la source
-r : seconde passe avec retry des zones HS (préciser le nombre d'essais)
-c : préciser un nombre de blocs à traiter en une fois (128 par défaut)

2/ Mener l'enquête

Lorsqu'on ne connaît pas exactement la cause de la perte de données (et même lorsqu'un ami/client vous la donne en pensant la connaître), il est impératif de prendre le temps d'analyser la situation.

Il faut mettre en corrélation les observations effectuées sur la situation actuelle, sur les éléments qui sont intervenus préalablement et dont on a la certitude, et réaliser une démarche de cheminement inverse pour tenter de déduire quelle était la situation première (avant la perte de données). C'est cet état, ou un état le plus proche de celui-ci, qu'il faut viser à reproduire.

3/ Passer à l'action

En cas de suppression volontaire ou accidentelle, utiliser testdisk (pour de la FAT32 ou du NTFS) ou ext3grep/extundelete pour de l'ext 2/3/4. On peut aussi se tourner vers des softs propriétaires et payants (Rstdudio, ufsexplorer) qui donnent parfois des résultats différents. Dans certains cas, il peut être utile de compléter par une récupération brute (photorec).

Si la table de partition a été altérée, rechercher les partitions préalables avec testdisk ou gpart. Il peut parfois être utile de les monter en lecture seule grâce à un loopdevice avec un ofset. Par exemple, si dans testdisk on trouve la trace d'une partition commençant à C:H:S 1201/27/31, on pourra procéder de la manière suivante (note, ces données sont peu probables, elles ne sont pas alignées sur une frontière de cylindre, cf. remarque ci-dessous) :

déterminer le loopdevice disponible : « sudo losetup -f » → /dev/loop**0**
calculer l'ofset (attention, ici taille de secteur de 512, géométrie X/255/63) :
ofset = (Cx255x63 + Hx63 + S – 1)x512
= (1201x255x63 + 27x63 + 31 -1)x512
= 9879447552
affecter le loopdevice avec l'ofset :
« sudo losetup /dev/loop0 /media/mondisk/image.dd -o 9879447552 »
monter en read only
« mkdir test && sudo mount /dev/loop0 test/ -o ro »

Note au sujet de testdisk : la configuration par défaut recherche en priorité les partitions qui débutent « correctement » sur des frontières de cylindre. Il est parfois utile de ne pas spécifier de type de table (intel → none) et de changer les options en « cylinder boudary : no ». Pour avoir des résultats plus lents, mais plus pertinents.

En cas d'écrasement trop important et de corruption rédhibitoire de la table de partitions, il est finalement souvent nécessaire de procéder à une récupération brute. Celle-ci peut aisément se faire avec photorec. On peut agir directement sur l'image :

« sudo photorec /media/mondisk/image.dd »

Il faudra veiller à spécifier uniquement les types de fichiers ciblés pour ne pas avoir un trop grand nombre de résultats, et SURTOUT, donner en sortie un autre support que le disque original pour ne pas écrire sur celui-ci (d'où l'intérêt de travailler sur l'image)

Les résultats sont triés par 500 éléments dans des dossiers de type recupdir.1, recupdir.2 etc…

Enfin, il faudra passer au post-traitement qui se compose du tri des fichiers, de la suppression des doublons, de la suppression des faux-positifs, et quand c'est possible de renommer et classer les données en fonction d'informations pertinentes pour l'utilisateur final. Cette partie, fastidieuse, longue, est souvent celle qui va déterminer la qualité de la prestation et la satisfaction ou non du client final. Il ne faut pas hésiter à créer ses propres scripts, spécifiquement à chaque situation. On ne peut faire de généralité que pour certaines parties, mais il est réellement nécessaire d'adapter.

V- LE POST-TRAITEMENT

1/ Suppression des doublons

Peut être effectué graphiquement avec fslint, ou en ligne de commande à l'aide de fdupes. Ou bien l'on peut faire des scripts plus complexes (comparaison de trois dossiers avec priorisation etc…) en s'appuyant sur des comparaisons de sommes de contrôle md5 (md5sum) ou tout autre système de calcul de hachage.

2/ Tri des résultats photorec

Il faut extraire des dossiers les fichiers cible et les placer dans des dossiers portant le nom de l'extension dans un premier temps. C'est plus facile à traiter. On peut utiliser un script du type :

```
for i in
        {jpg,png,doc,xls,ppt,pdf, (etc… autant d'xtension souhaitées) };
                do
                mkdir $i;
                for j in {1..47}; note : ici on suppose qu'il y a 47 dossiers créés par photorec
                        do
                        mv -v recup_dir.$j/*.$i $i/;
                        done;
                done;
rmdir *;
```

3/ Supprimer les faux-positifs

La plupart du temps cela ne peut se faire que par l'observation du contenu des dossiers et des tailles attendues pour ce type de fichier (exemple, un jpg de 100Mo…). On peut parfois faire un script sur la base de « identify » ou de « file » si cela donne des résultats pertinents.

4/ Renommer

Graphiquement, on peut utiliser « pyrenamer » qui donne d'assez bons résultats pour les images et la musique, mais cela n'est généralement pas suffisant. Il faut alors compléter par des scripts qui prennent les informations de métadonnées en entrée et qui renomment en sortie.

« jhead » peut parfois donner des résultats intéressants en une seule commande pour les images.

Pour les métadonnées, on peut utiliser jhead, exiv2, hachoir-metadata, file, identify…

Pour renommer, il faut faire un script à partir de sed.

Exemple sur des images :

```
for i in `ls`; do j=`hachoir-metadata $i | grep "Camera manufacturer:" |
                    sed 's,.*Camera manufacturer: \(.*\),\1,g'`; k=`echo $i |
                    sed 's,.*\([f|t|b][0-9]*\).*,\1,g'`; echo $i"  "$j"_"$k.jpg;done
```

Note : ici le script ne renomme pas, il permet juste d'afficher le résultat qui serait utilisé par exemple par « mv » pour renommer… cela permet de se faire une première idée…