

# Backuppc, retour d'expérience

Sylvain Ferrand

Ecole Polytechnique

10 octobre 2012, Journées Mathrice, Orléans



- Un laboratoire d'environ 150 personnes
- Parc en majorité sous linux, des machines windows et des macs. Fichiers centralisés et sauvegardés (Time Navigator, rsnapshot)
- Un parc de portables qui explose !
- De nombreux incidents avec les portables en 2010/2011: disque HS, vols, pertes
- Jusqu'à 2011, sauvegarde des portables à la charge de l'utilisateur (Time Machine surtout)



- Solution de sauvegarde supportant au minimum des clients Mac, Linux et Windows
- Limiter les interventions humaines pour déclencher la sauvegarde (tout auto si possible)
- Compatible avec les portables (qui ne sont donc pas toujours connectés au réseau)
- Restauration rapide, si possible directement par l'utilisateur
- Si possible déduplication pour éviter que la volumétrie explose
- Sauvegarde sur disque



# La solution choisie : BackupPC

Pourquoi ?

- Répond à notre cahier des charges ci-dessus
- multiplateforme : serveur UNIX, client : toutes les plateformes supportant les protocoles rsync/SMB/SSH/NFS
- Un outil “rustique”, déjà ancien (V.1 en 2001), écrit majoritairement en Perl. Pas vraiment à *la pointe*, mais robuste et éprouvé
- Format de donnée de stockage non propriétaire (hard links + compression zlib)

Les petits plus :

- Bien packagé sous debian (rapide à installer, facile à maintenir)
- Interface graphique web permettant d'effectuer pratiquement toutes les opérations d'administration (apprentissage facile)
- Pas de client nécessaire (en théorie), si on utilise SMB sous windows, ou ssh/rsync sous linux



- Backuppc controle régulièrement la présence des machines sur le réseau (ping ou nmblookup)
- Le serveur lance la sauvegarde de chaque machine pour laquelle il n'a pas de sauvegarde depuis 24H ou plus.
- Calcul de hash MD5 par fichier  $\Rightarrow$  déduplication (fait un lien hard si un fichiers identique existe, enregistre les metas-data à part)
- Chaque nuit nettoyage du pool
- Fonctionnement classique (sauvegarde complète/incrémentale) malgré la déduplication



# Installation : Facile !

Coté serveur : `apt-get install backuppc ...` c'est fini !

Fichiers de configuration :

- `config.pl` Configuration générale
- `hosts` : Les machines à sauver et les utilisateurs
- `ORDI_CLIENT.pl` : la configuration spécifique à chaque client

Données :

- `__TOPDIR__/pc/$host` log et états des sauvegardes et les backups (liens vers (c)pool)
- `__TOPDIR__/pool/` Pool non compressé
- `__TOPDIR__/cpool/` Pool compressé
- `__TOPDIR__/log/` Log
- `__TOPDIR__/trash` Fichier à effacer

Tous ces répertoires doivent être dans le même FS ! (hard link)



Coté client... suivant l'OS et le protocole choisi :

- Windows : Partage samba ou rsyncd (utiliser le rsyncd de cygwin avec ou sans ssh - un peu lourd mais quasi indispensable)
- Linux/Unix : Copier la clé public du serveur sur les clients
- Mac : Démarrer le serveur SSH sur le mac (Menu pomme/paramètres/partage puis cocher session à distance, penser aux conséquences !), copier la clé du serveur dans le repertoire de l'utilisateur.



Permet d'oublier la ligne de commande !

- Interface WEB
- Visualisation des logs, des états des sauvegardes, des files d'attente
- Configuration de l'outil (générale et machine par machine)
- Naviguer dans les sauvegardes - restaurer, forcer une sauvegarde
- Login utilisateur pour voir ses propres sauvegardes





- Les portables sont fortement ralentis lors de la sauvegarde (sollicitation CPU mais surtout accès disque)
- Les données **NE SONT PAS CHIFFRÉES** sur le serveur  $\Rightarrow$  il est impératif d'en informer les utilisateurs
- Problème avec les dual-boots



# Dimensionner le stockage

Quelle stockage prévoir ? Pas vraiment de formule magique tenant compte de la taille des données, des incrémentales, de la périodicité, de la durée de rétention, de la compression, de la déduplication...

Quelques estimations très largement pifométriques !

- Taille moyenne des données à sauvegarder par portable (2012) ~ 50Go
- Coté volumétrie la déduplication revient un peu à ne faire que des incrémentales après une totale  $\Rightarrow$  partons sur incrémentale = 5% d'une totale
- Compression - 30-40% en moyenne

Nous avons 10To (Linux ext4, RAID 5, 7 disque SATA 2To) pour 50 portables.



## Sécurité

- Blinder le serveur de sauvegarde, la clé privé permet d'accéder aux clients !
- Autorisez les connexions sur les clients uniquement depuis le serveur ( ajouter `from="IP du serveur"` au début des clés SSH, éventuellement configurer un firewall pour SMB et RSYNC sous windows)
- Eventuellement, créer un utilisateur non privilégié sur les clients et utilisez sudo : fausse bonne idée

## Divers :

- Fichiers verrouillés sous windows  $\Rightarrow$  utiliser Volume Shadow Copy Service (VSS). On prend un snapshot que l'on sauve et l'on détruit ensuite.

<http://serverfault.com/questions/119120/>

[how-to-use-a-volume-shadow-copy-to-make-backups](http://serverfault.com/questions/119120/how-to-use-a-volume-shadow-copy-to-make-backups)



# Pour être à la pointe du progrès !

D'autres approches de la sauvegarde (pas forcément au point)

- bup : Logiciel s'appuyant sur Git et en particulier sur les *packfile*. Redondance *PAR2*, *algo* à la *rsync*, *bup-fuse* pour monter un dépôt comme un FS...
- Obnam : Gère à la fois la déduplication des morceaux de fichiers sur le client et le chiffrement. Utilise des Copy-on-Write B-Tree pour stocker les différences.

