Présentation de Xen

Centre de Physique Théorique UMR 6207
Vincent BAYLE

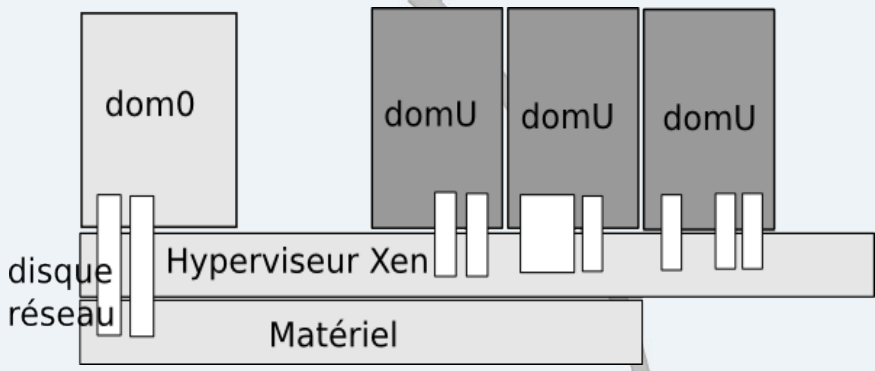
Plan de la présentation

- Introduction : historique de l'hyperviseur Xen
- L'architecture de Xen : l'hyperviseur, les domaines privilégiés et invités.
- Les deux modes de fonctionnement : paravirtualisation et virtualisation matérielle.
- Installation sous la distribution linux debian squeeze.
- L'utilisation de Xen :
  - Création d'une machine virtuelle
  - Le fichier de configuration d'une machine virtuelle
  - La création des domaines invités
  - La sauvegarde des domaines invités
  - Le partage / la protection des ressources
- Autour de xen
- Retour d'expérience au Centre de Physique Théorique

L'hyperviseur Xen : historique

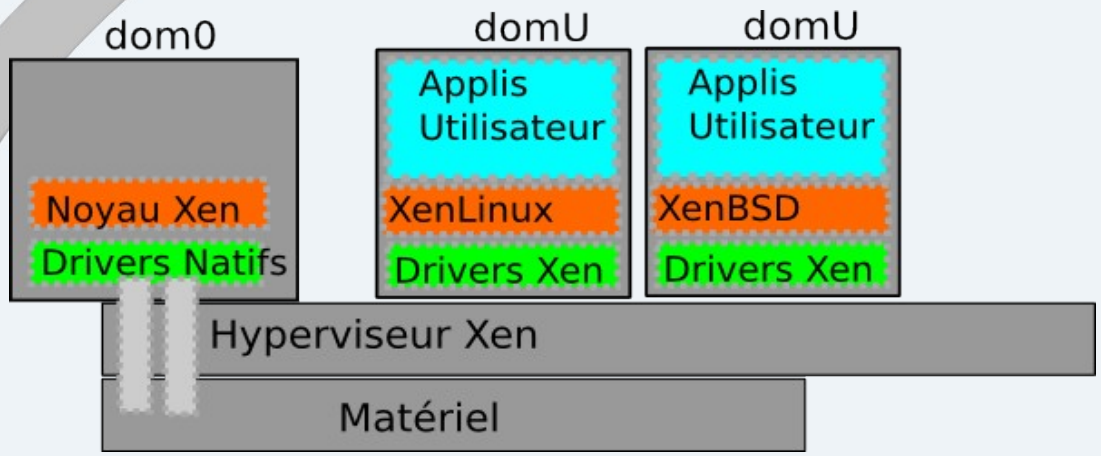
- Xen né à Cambridge, en 2003. Produit par Xensource.
- Logiciel libre.
- Rachat de Xensource par Citrix, en 2007.
- Projet xen.org :
  - Développement Xen continue : depuis la version 1.0 en 2003 jusqu'à la version 4.1 en mars 2011
  - Autres projet : Xen Cloud Platform...
- Gamme professionnelle chez Citrix, dérivée du code Xen.

La virtualisation selon Xen



- Hyperviseur : c'est l'ordonnanceur du système, qui fait le lien entre la machine privilégiée, le(s) processeur(s), la mémoire.
- Domaine privilégié ou dom0 : c'est lui qui réalise, pour l'hyperviseur, les accès disques, les accès au réseau, qui crée le réseau virtuel, etc.
- Domaine invité ou domU. Ce sont les machines virtuelles à proprement parler. Elles disposent de périphériques fournis par l'hyperviseur et le domaine privilégié. Elles disposent de leur propre noyau, distinct de celui du domaine privilégié. Avec une distinction supplémentaire :
  - Machine virtuelle : entité permanente qui réside sur le disque (comme un programme)
  - Domaine : c'est une machine virtuelle chargée pour éxécution (comme un processus).

Le mode para-virtualisé



Historiquement, fonctionne sur le modèle de la paravirtualisation.

- Les domaines hébergés sont para-virtualisés : ils ont connaissance du fait qu'ils tournent dans un hyperviseur. Leur accès disques, réseau, etc. se font par des drivers spécifiques.
- Ces appels sont retranscrits facilement dans des appels au niveau du dom0 : en ce sens, la paravirtualisation n'entraine que peu de traitement supplémentaire.
- Dans ce mode, ne peut faire tourner que des systèmes “xenifiés” : distributions linux, OpenSolaris, NetBSD, FreeBSD.
- Le système hôte est aussi un système modifié. Les OS disponibles pour ce rôle sont : les distributions linux, Opensolaris, NetBSD, , avec un noyau modifié.

Le mode virtualisation matérielle (HVM)

Depuis la version 3.0, Xen peut utiliser les technologies de virtualisation embarquées dans les processeurs (HVM : soit Vt-x ou i pour intel et pacifica pour amd) :

- Il peut alors faire tourner des Os qui n'ont pas été “xénifiés”.
- Emule complètement une machine : un bios, une carte réseau, une carte graphique...
- Les performances sont moins bonnes. Les appels ont besoin d'être traités différemment : il simule une carte réseau, la gestion de la mémoire impose de faire croire à l'Os embarqué qu'il dispose d'un espace contiguë, etc.
- En contrepartie, il peut faire tourner une plus grande variété de système d'exploitation (Windows notamment, sans aucune modification) dans les domaines invités.
- Les deux modes concernent les domaines invités : ils peuvent coexister au sein d'un domaine privilégié

Installation du domaine privilégié dom0 sous linux debian

- Procédure décrite ici : http://wiki.debian.org/Xen
- Installation des paquets :

```
aptitude -P install xen-linux-system
```

- Par defaut, le premier noyau n'est pas le bon :

```
mv -i /etc/grub.d/10_linux /etc/grub.d/50_linux
```

- Modification du fichier /etc/default/grub,

```
GRUB_DISABLE_OS_PROBER=true
```

- Reconstitution de la liste :

```
update-grub
```

- On a alors une entrée (dans le fichier /etc/grub/grub.cfg) de la forme :

```
menuentry 'Debian GNU/Linux, with Linux 2.6.32-5-xen-amd64 and XEN 4.0-amd64' --class debian --class
gnu-linux --class gnu --class os --class xen {
      insmod part_msdos
      set root='(hd0,msdos3)'
      multiboot       /xen-4.0-amd64.gz placeholder
      module /vmlinuz-2.6.32-5-xen-amd64 placeholder root=/dev/mapper/cpt-disk ro quiet
      module /initrd.img-2.6.32-5-xen-amd64
}
```

- Après un redémarrage, on dispose d'un dom0 fonctionnel :

```
root@cpt18:~# xm list
Name                                        ID   Mem VCPUs      State   Time(s)
Domain-0                                     0 46702    16     r-----   5750.2
```

Le domaine privilégié dom0

Il accède à l'hyperviseur par l'intermédiaire de 2 services : xend et xendomains, qui démarrent aussi les démons xenstored et xenconsoled.

Fichier de configuration de xend : /etc/xen/xend-config.sxp.

- la configuration du service lui-même → accessible comme service Unix, comme serveur http, comme serveur de migration, et les règles pour accéder au service.

```
(xend-http-server yes)
#(xend-unix-server no)
#(xend-tcp-xmlrpc-server no)
#(xend-unix-xmlrpc-server yes)
#(xend-relocation-server no)
#(xend-relocation-ssl-server no)
#(xend-udev-event-server no)
```

- Configuration du comportement par rapport aux ressources.

```
(dom0-cpus 0)
(dom0-min-mem 196)
```

- Configuration du réseau

```
(network-script 'network-bridge antispoof=yes' )
(vif-script vif-bridge)
```

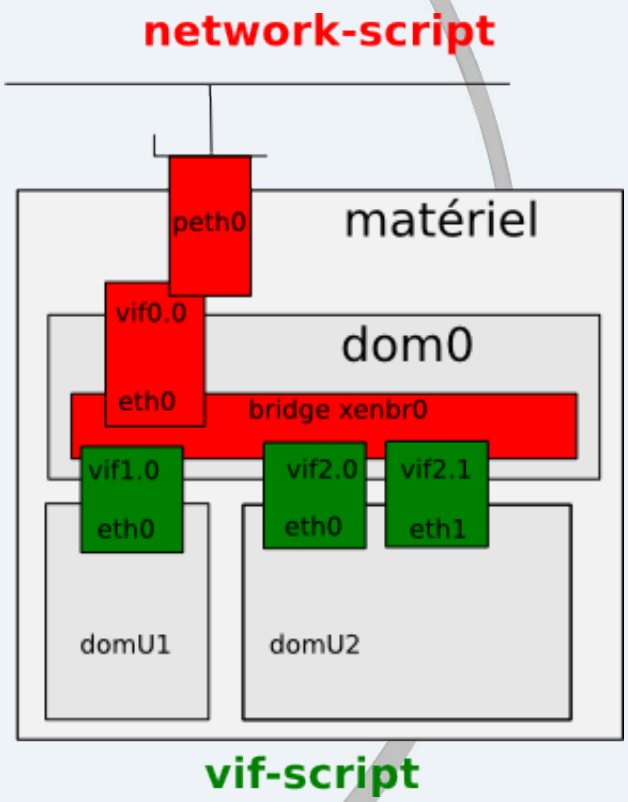
Le domaine privilégié dom0 (2)

- C'est le système le plus sensible, il faut le protéger :
  - Ne pas installer de services dessus.
  - Limiter les accès au maximum : accès ssh, firewall, interface réseau dédiée à l'administration.
- Possibilité, à partir de ce domaine d'administrer les autres domaines, les ressources, l'hyperviseur, via notamment la commande xm
  - `xm create domu.cfg` pour créer les domaines
  - `xm list` pour lister les domaines présents sur la machine.
  - `xm console domU` pour accéder à un domaine en mode console.
- Traces de fonctionnement du demon dans les fichiers de log :
  - `/var/log/xend.log`
  - `/var/log/xen-debug.log`
  - Possibilité de voir les messages de l'hyperviseur : `xm dmesg`
- C'est dans ce domaine là que l'on peut créer des machines virtuelles et instancier des domaines.

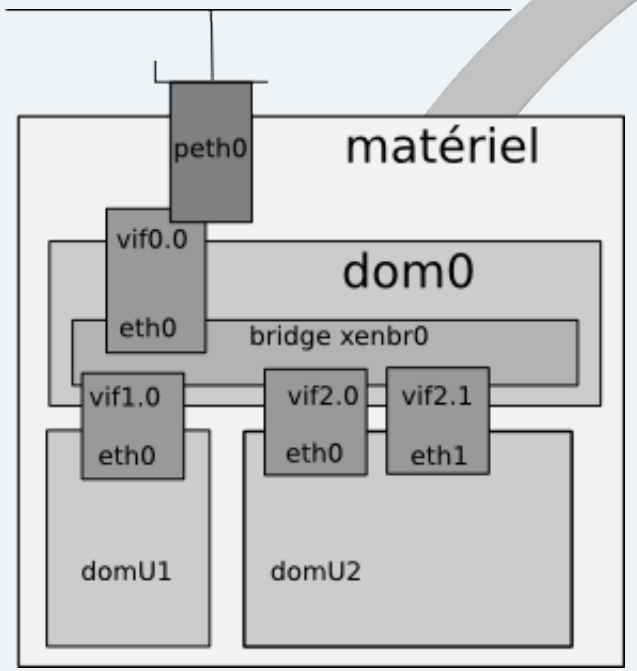
Le réseau virtuel

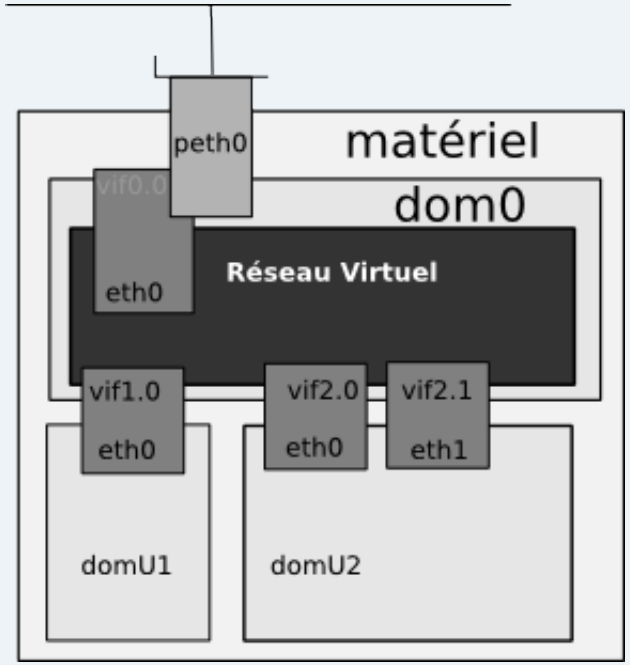
Le réseau est défini par Xen via 2 scripts.

- Ces scripts permettent de :
  - générer le réseau virtuel (`network-script -`)
  - créer les cartes réseau lors de l'instanciation des machines virtuelles (`vif-script --`)
- Les interfaces des domaines invités sont créées avec le domaine. Elles ont deux noms : un dans le domaine privilégié, et un dans le domaine invité.



Le réseau virtuel (2)





- Schéma simple basé sur la programme bridge : suffisant dans la majorité des cas.
- Autres possibilités disponibles avec xen : route, nat.
- Possibilité de constituer le réseau virtuel à partir de scripts, en fonction de ses besoins : plusieurs bridges rattachés chacun à son interface physique, etc.

Exemple de création d'un domaine invité

- Utilisation du paquet debian xen-tools :

```
cpt76:~# apt-get install xen-tools
```

- Ensuite, simplement, création d'une machine virtuelle

```
cpt76:~# xen-create-image --hostname origami.cpt.univ-mrs.fr --ip 139.124.7.164 --netmask 255.255.255.0
--gateway 139.124.7.250  --pygrub --dist squeeze --role ssh --lvm=cpt -size=5G
…
Installation Summary
--------------------
Hostname        :  origami.cpt.univ-mrs.fr
Distribution    :  squeeze
IP-Address(es)  :  139.124.7.164
RSA Fingerprint :  be:94:37:1d:a3:d7:8d:68:9e:2e:b7:e2:cc:08:7d:c3
Root Password   :  pJvXjxEm
```

- Et pour booter (instancier le domaine invité) :

```
cpt76:~# xm create origami.cpt.univ-mrs.fr.cfg
Using config file "/etc/xen/origami.cpt.univ-mrs.fr.cfg".
Started domain origami.cpt.univ-mrs.fr (id=1)
```

- On peut désomais se connecter dessus.

```
cpt76:~# xm console origami.cpt.univ-mrs.fr
Debian GNU/Linux 6.0 origami hvc0
origami login: root
Password:
```

La création des machines virtuelles

- De nombreux scénarios sont possibles :
  - Depuis la création à partir du disque d'installation du système d'exploitation (machine qui disposent de capacités hvm)
  - copie d'images du système
  - Création avec des outils tiers → xen-tools pour debian, mais aussi virt-manager pour les machines redHat.
- Permettent de générer une image disque et un fichier de configuration
- Exemple de fichier de configuration : `/etc/xen/origami.cfg`

```
bootloader = '/usr/lib/xen-default/bin/pygrub'
vcpus       = '1'
memory      = '128'
root        = '/dev/xvda2 ro'
disk        = [
                  'phy:/dev/cpt/origami.cpt.univ-mrs.fr-disk,xvda2,w',
                  'phy:/dev/cpt/origami.cpt.univ-mrs.fr-swap,xvda1,w',
              ]
vif         = [ 'ip=139.124.7.164,mac=00:16:3E:78:20:21' ]
```

- Caractéristiques du système : mémoire (memory), processeur (vcpus), réseau (vif)
- Caractéristiques des disques du domaine invité. Xen permet de stocker ces images disques sur un grand nombre de support :
  - un système de fichier embarqué dans un fichier, au format raw ou CoW (via les drivers blktap2)
  - périphérique en mode bloc, comme une partition système ou une partition lvm (via le driver phy)
  - NFS, NBD, iSCSI, GFS.

L'instanciation des domaines invités : différentes méthodes de boot

Plusieurs méthodes possible, configurée dans le fichier de création des domU, domU.cfg

- Chargement à partir d'un noyau existant dans le domaine privilégié :

```
kernel      = '/boot/vmlinuz-2.6.32-5-xen-amd64'
ramdisk     = '/boot/initrd.img-2.6.32-5-xen-amd64'
```

- Pygrub : émulation de grub pour les machines virtuelles. Le programme pygrub est éxécuté dans le dom0. Cette méthode permet de donner plus de possibilités à l'administrateur du domaine invité, p. ex. d'installer un noyau de son choix, sans intervention de l'administrateur du dom0.

```
bootloader = '/usr/lib/xen-default/bin/pygrub'
```

- Hvm-loader : ici, le domaine est démarré à partir du programme hvmloader, qui fabrique la machine virtuelle, attache les éléments pour donner à l'Os embarqué l'impression qu'il est sur une machine physique. Il attache un bios bochs.

```
kernel = "/usr/lib/xen/boot/hvmloader" builder='hvm'
```

La sauvegarde des domaines invités

- Question ouverte : elle n'est absolument pas réglée par Xen. Il n'existe de solutions de sauvegarde intégrée que dans des logiciels commerciaux, et dans la version Xenserver.
- Sauvegarde des domU eux même :
  - `xm save domaine fichier` : permet de faire une sauvegarde de la mémoire de la machine à un instant donné (avec toutes ses connexions etc.). C'est l'équivalent d'un mode hibernation → il libère les ressources, mémoire et cpu.
  - `xm restore fichier` : permet de le réveiller.
- Sauvegarde des machines virtuelles :
  - Possibilité de sauvegarder les disques, p.ex. après un shutdown propre
  - Script, pour avoir un état complet de la machine (voir présentation suivante).
  - virtualisation du stockage dans un san p.ex., auquel on délègue la responsabilité de sauvegarder les machines virtuelles.

Nécessité de concevoir la solution de sauvegarde.

La migration “live” des domaines

Xen propose une fonctionnalité de migration “live” des machines virtuelles.

- La migration, effectuée par la commande :
  xm migrate --live domaine hôteDestination
  permet de migrer le domaine vers la machine hôteDestination.
- Le domaine invité est migré avec l'ensemble de ses connexions réseaux qui ne sont pas interrompues (entre 60 et 300 ms de latence à l'instant précis de la migration seulement). L'interface réseau est migrée sur la nouvelle machine.
- Nécessite que les domaines privilégiés :
  - Disposent de la même version de Xen
  - Soient correctement configurés (options du fichier xend-config.sxp).
  - Accèdent tous les deux au disque du domaine invité (disque SAN, NFS, disque répliqué par DRBD, etc).
  - Que les configurations réseau soient compatibles

Partage des ressources

- Xen permet de partager, de limiter l'accès à certaines ressources :
  - Cpu (nombre) : à froid dans le fichier de configuration, ou à chaud via les commandes

    `xm vcpu-set domId #vcpu`
  - Mémoire (quantité) : possibilité d'attribuer une quantité de mémoire, de modifier cette quantité, dans la limite de la mémoire du système :

    `xm mem-set dom-id tailleMem`
  - Disque : possibilité de modifier la taille du disque (outil de redimensionnement du système privilégié, à froid, ou invité, à chaud).
  - Les interfaces réseau : possibilité d'attribuer ou de supprimer des interfaces réseaux depuis le dom0.
- D'autres éléments permettent de limiter l'accès à certaines ressources, notamment le réseau :
  - pour des fonctions de firewall. Exemple : http://www.shorewall.net/Xen.html
  - Pour une gestion de bande passante en utilisant les fonctionnalités de gestionnaire de bande passante du domaine privilégié : http://book.xen.prgmr.com/mediawiki/index.php/Chapter_7:_Hosting_Untrusted_Users_Under_Xen:_Lessons_from_the_Trenches#Controlling_Network_Resources

Autour de xen : rajouter des fonctionnalités

- Utilitaires en mode texte :
  - xen-tools de debian. Il contient la commande xen-create-image que l'on a utilisé plus haut, et aussi xen-list-images, xen-delete-image, xen-create-nfs et xen-update-image.
  - virsh qui fait partie de la suite libvirt de redHat
- Utilitaires graphiques :
  - virt-manager, et toute la suite libvirt de RedHat : rajoute une interface graphique.
  - convirt (ex Xenman) : rajoute aussi une interface graphique.
- XenServer : incompatible avec les versions xen open-source. Linux profilé pour la virtualisation, auquel sont ajoutés des utilitaires professionnels.

Retour d'expérience

- Laboratoire de recherche : environ 55 chercheurs, 4 tutelles, une cinquantaine de machines Linux, postes nomades, un petit réseau windows pour l'administration.
- Domaine : Physique Théorique.
- Service informatique : 1 poste ETP.
- Service réseau: réseau filaire, wifi, serveurs d'authentification, DNS, Web (labo + conférences), git, gestionnaire de bibliothèque, messagerie.
- Utilisation de Xen → réduction du nombre de petits serveurs (en conservant la philosophie : un service, un serveur).
- Aujourd'hui, une quinzaine de machines en production : DNS, LDAP, www, git, proxy, impression, nagios, ocs...
- Systèmes debian sur les dom0s comme sur les domUs
- Consolidation sur deux machines serveurs récentes.
- Avantages principal : deux machines physique seulement : coût (notamment de la garantie), encombrement...
- Autres avantages : moins de consommation électrique, de dégagement de chaleur. Schéma électrique simplifié : 2 onduleurs vers 2 doubles alims.
- Au quotidien, création de machines de test, pour le service informatique, quelquefois pour des utilisateurs (site web de conférence, serveur git pour une équipe de recherche), pour de l'enseignement : création/destruction de machines linux par script.
- Création de serveurs pour des besoins précis : rapide, efficace.

Dans le contexte de notre laboratoire : Avantages/Inconvénients de Xen

Avantages :

- Simplicité
- Stabilité
- Maturité du projet
- Intégration dans les systèmes debian

Inconvénients :

- Documentation pas évidente à trouver, et concerne des versions parfois anciennes

En guise de conclusion, l'hyperviseur xen, sur le matériel dont nous disposons satisfait la totalité de nos besoins.

Documentation

- Site de l'éditeur du logiciel

  http://www.xen.org

  http://wiki.xensource.com/xenwiki/FrontPage

- Différentes présentations disponibles en français dans la communauté :

  http://raisin.u-bordeaux.fr/IMG/pdf/cours_xen.pdf

  http://cesar.com.univ-mrs.fr/IMG/pdf/Virtualisation_Xen_avec_commentaires.pdf

  http://laser.cbs.cnrs.fr/IMG/pdf/Xen-TP.pdf

- Site web du livre : The book of XEN. Intéressant pour un éclairage transversal sur Xen : utilisation dans un environnement privé d'hébergement de machines virtuelles Xen.

  http://book.xen.prgmr.com/mediawiki/index.php/Introduction