AC - Virtualisation

Semestre Automne 2008

Antoine Benkemoun
Romain Hinfray

Introduction

- Cadre du projet
- Choix de Xen
- Quatre parties
  - Tour d'horizon de la virtualisation
  - Pré-requis à la compréhension de Xen
  - Présentation de Xen
  - Architecture de Xen

Sommaire 1ère partie

- I Tour d'horizon
  - Historique
  - Pourquoi virtualiser ?
  - Solutions techniques
  - Solutions commerciales
- II Pré-requis à la compréhension de Xen
  - La notion d'anneau
  - Les problèmes liés à l'architecture x86
  - Gestion de la mémoire

Sommaire 2ème partie

- III Présentation de Xen
  - Interactions entre l'hyperviseur, le système d'exploitation et les applications
  - Les domaines
- IV Architecture de Xen
  - La gestion des informations système
  - La communication inter-domaines
  - La gestion du temps
  - La gestion de la mémoire
  - La gestion du temps
  - La réseau sous Xen
- Conclusion

Introduction à la virtualisation

- Objectif
  - Faire fonctionner plusieurs environnements logiques indépendants séparément
- Extension du principe d'émulation
  - Substitution de composants informatiques par une application
- Multiplexage de systèmes d'exploitation

Tour d'horizon : Historique

- 1960 : Composant VM/CMS du System/360 d'IBM
  - Cloisonnement d'environnements logiques
- 1990 : Amiga de Commodore International
  - Exécution des PC X386, Macintosh 6800 et X11
- 1990 : Architecture NUMA (Non Uniform Memory Access)
  - Cloisonnement et partitionnement de la mémoire via des bus
- 1999 : VMWare
  - Virtualisation de systèmes x86
- 2000 à aujourd'hui : Développements de projet libres
  - QEMU, Bochs, Xen (2003), KVM (2007)

Tour d'horizon : Pourquoi virtualiser ?

- Trois types d'avantages
  - Sécurité
  - Coût
  - Criticité et performances

Tour d'horizon : Pourquoi virtualiser ? (Sécurité)

- Isolation
  - Ignorance de la présence d'autres environnements
  - Utilisation des protocoles conventionnels
- Cloisonnement
  - Allocation de ressources physiques (exclusive ou temporelle)
- Etude de sécurité
  - Contrôle et étude d'environnements infectés
  - Répétition de scénarios

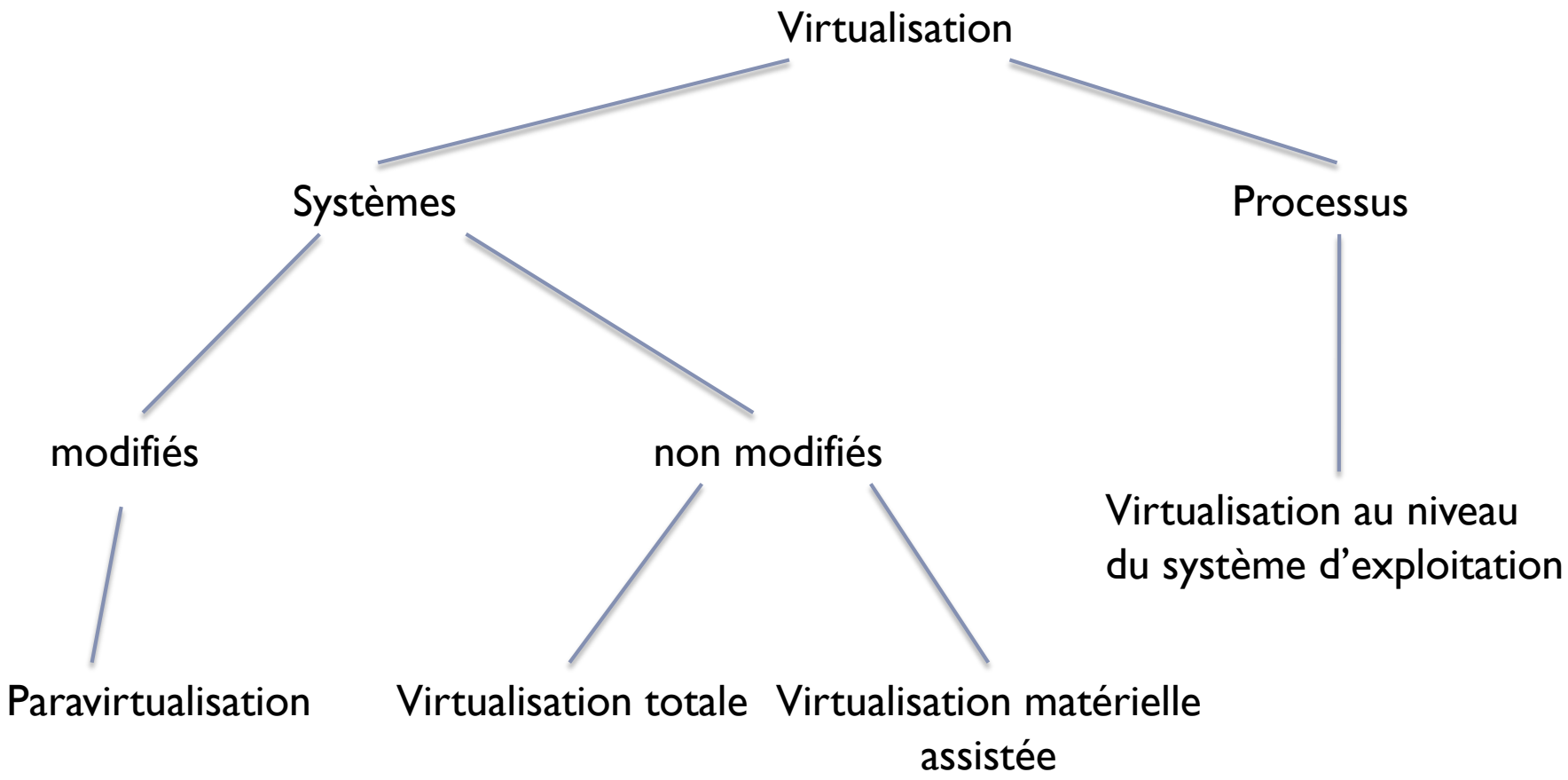
Pourquoi virtualiser ? (Coût)

- Sous utilisation actuelle des serveurs
  - Coût de l'énergie électrique
  - Coût de l'espace en centre de données
- Principe du cloisonnement des services
- Mutualisation de ressources physiques tout en maintenant un cloisonnement des services

Pourquoi virtualiser ? (Criticité et Performance)

- Possibilité de mettre en pause et de copier un environnement logiciel complet
  - Sauvegarde
  - Clonage
- Migration d'environnements logiciels
  - Transfert d'un environnement logiciel vers une autre machine physique
- Allocation dynamique de ressources
  - Flexibilité de l'offre
  - Adaptabilité en cas de montée en charge

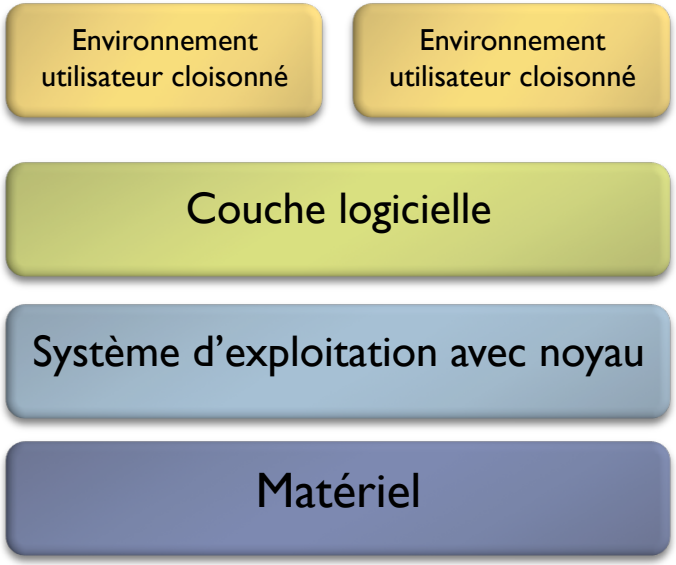
Tour d'horizon : Les solutions techniques



Tour d'horizon : Les solutions techniques

La virtualisation au niveau du système d'exploitation

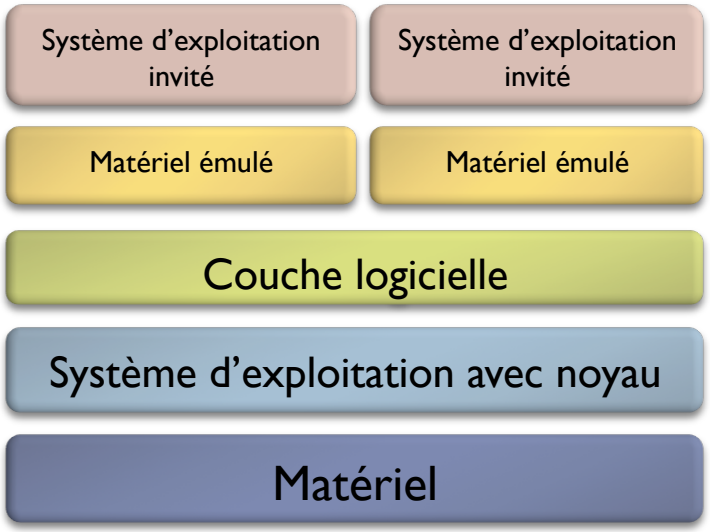
- Un seul système d'exploitation, un seul noyau
- Environnements utilisateurs entièrement cloisonnés
- Allocation de ressources entre les différents environnements



Tour d'horizon : Les solutions techniques

La virtualisation totale

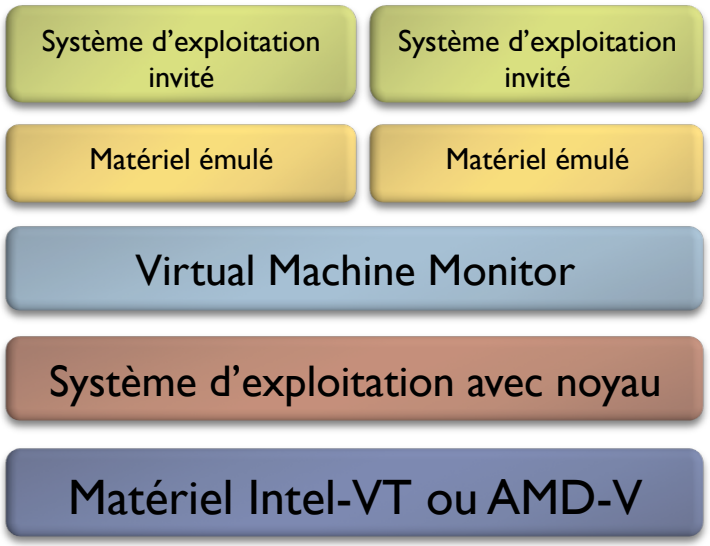
- Emulation de machines physiques
- Emulation de toutes les architectures possibles
- Les systèmes invités croient être sur des machines physiques
- Perte de performances importante



Tour d'horizon : Les solutions techniques

La virtualisation materielle assistée

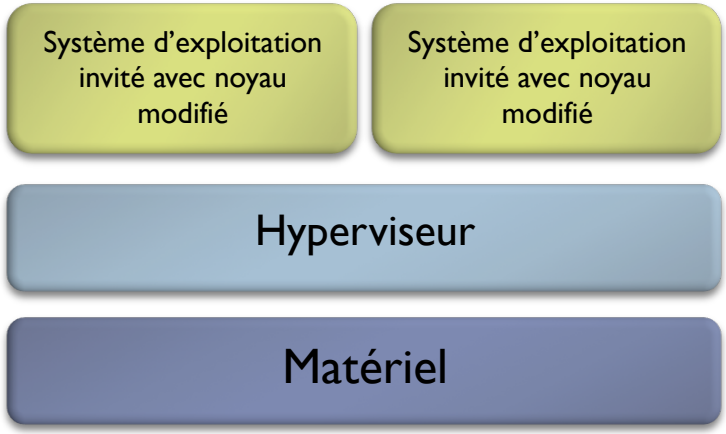
- Architecture processeur conçue pour la virtualisation
- Emulation du matériel mais moins de pertes de performances
- Systèmes d'exploitations invités non modifiés



Tour d'horizon : Les solutions techniques

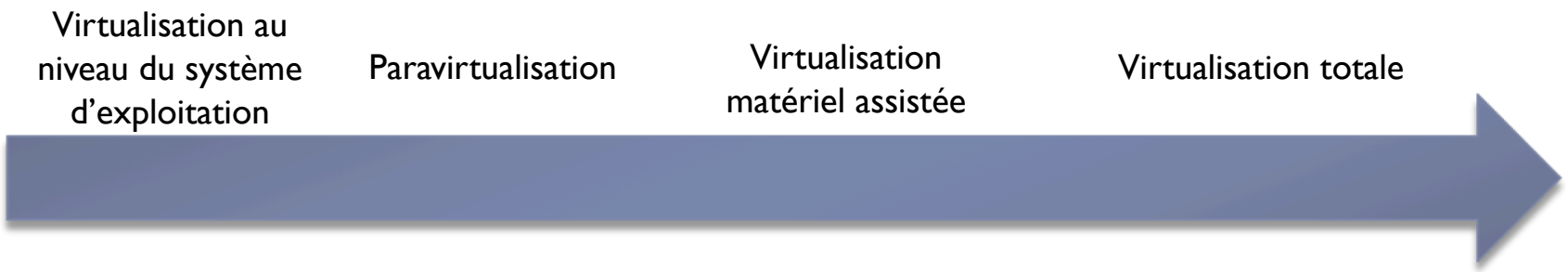
La paravirtulisation

- Systèmes d'exploitations invités modifiés
- Systèmes conscients d'être virtualisés
- Pas de systèmes d'exploitation entre les systèmes invités et le matériel



Tour d'horizon : Les solutions techniques

Bilan en terme de niveau d'émulation du matériel



Tour d'horizon : Les solutions commerciales

Les solutions propriétaires

- 1er éditeur VMWare
  - 13 logiciels
  - Virtualisation totale et matérielle assistée
  - VMWare Server ESX : Hyperviseur ?
- Virtual PC (Microsoft)
  - Gratuit
  - Virtualisation totale

Tour d'horizon : Les solutions commerciales

Les solutions Libres

- Xen
  - Paravirtualisation
- QEMU
  - Virtualisation totale
- OpenVZ
  - Virtualisation au niveau du système d'exploitation
- KVM
  - 2 parties :
    - Un module dans le noyau Linux (communication avec le processeur)
    - Un module dérivé de QEMU
  - Gestion des instructions processeur liées à la virtualisation

III Présentation de Xen

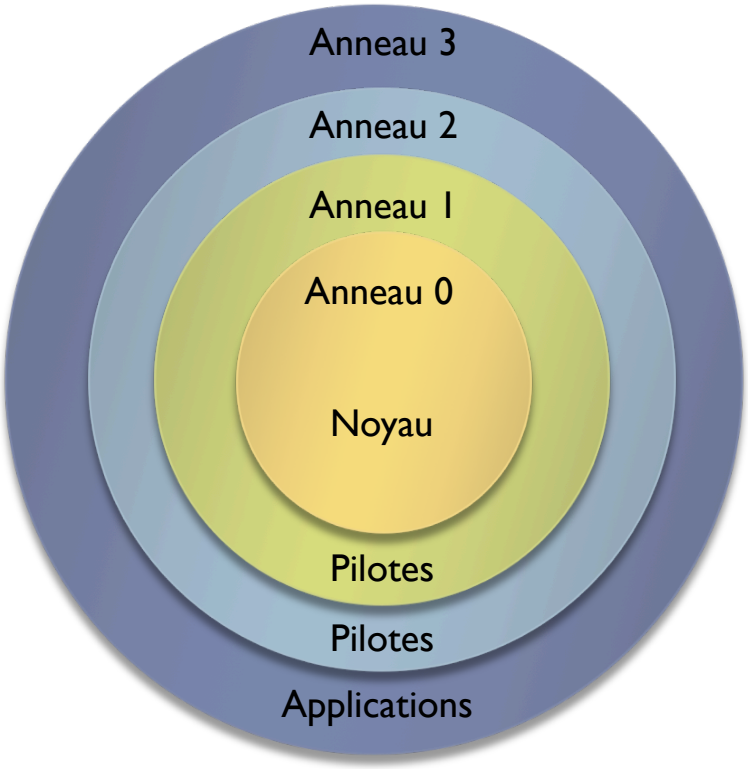
- Les problèmes liés à l'architecture x86 et ces solutions
- La notion d'anneau
- Interactions entre l'hyperviseur, le système d'exploitation et les applications
- Les domaines
  - Domaines privilégiés
  - Domaines non privilégiés
  - Domaines matériel assistés

Les problèmes liés à l'architecture x86 et ces solutions

- Selon Popek et Goldberg il existe 3 types d'instructions processeur problématiques pour la virtualisation :
  - Instructions privilégiées
  - Instructions sensibles à la configuration
  - Instructions sensibles de comportement
- Nécessité de capter ces instructions en virtualisation
- Certaines dérogent à la règle
- Nous avons donc un problème pour mettre en place une architecture de virtualisation sur x86
- Solutions :
  - VMWare : utilisation de la virtualisation totale
  - Xen : remplacer les instructions privilégiées par d'autres

La notion d'anneau

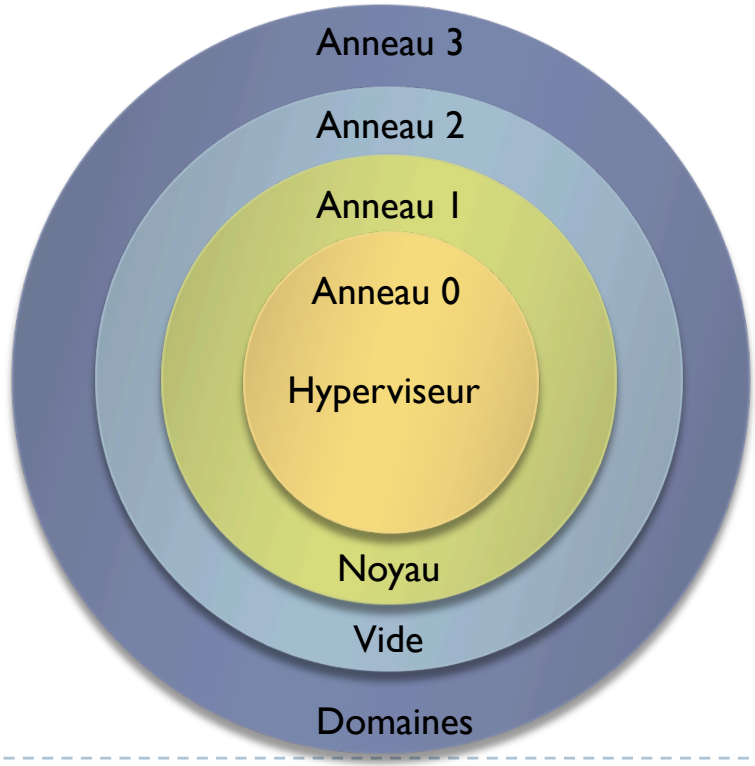
- Les anneaux identifient des niveaux de privilèges
- Limitation des accès aux informations du système
- Tous les systèmes actuels n'utilisent que 2 des 4 anneaux sauf quelques architectures spécifiques



Architecture x86 32 bits avec un système d'exploitation classique

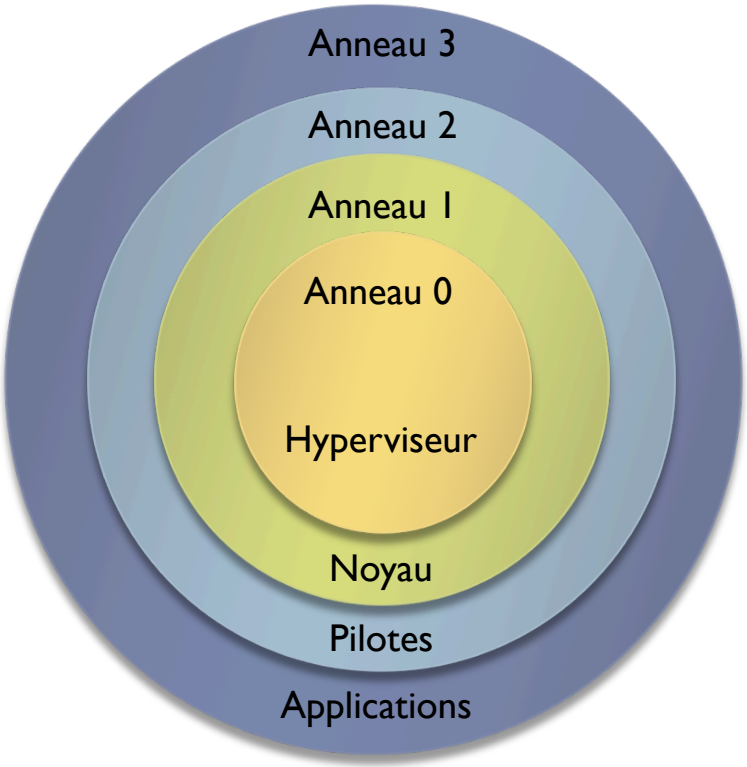
Interactions hyperviseur / applications / OS

- Qu'est ce que Xen ?
- Mise en place de la paravirtualisation
  - Motivation principale : remplacement des instructions
- Positionnement des différents composants

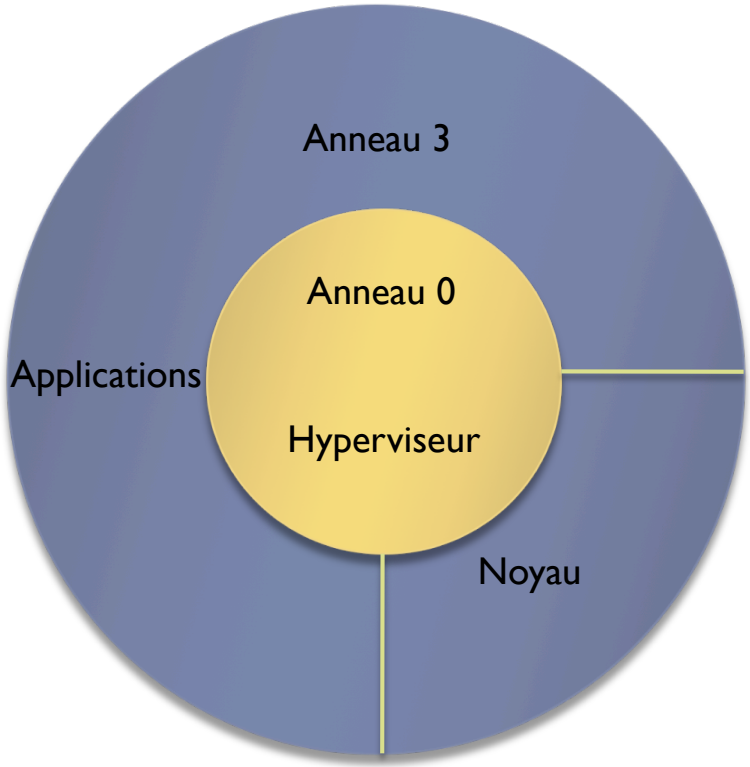


Interactions hyperviseur / applications / OS

Adaptation d'un hyperviseur dans les architectures non conçues pour la virtualisation



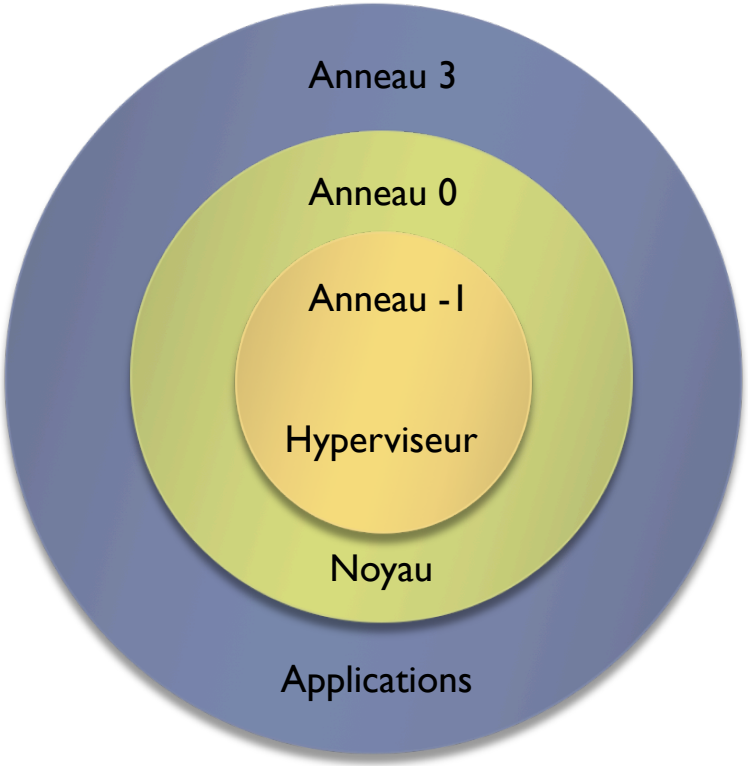
Architecture x86 32 bits avec un hyperviseur Xen



Architecture x86 64 bits avec un hyperviseur Xen

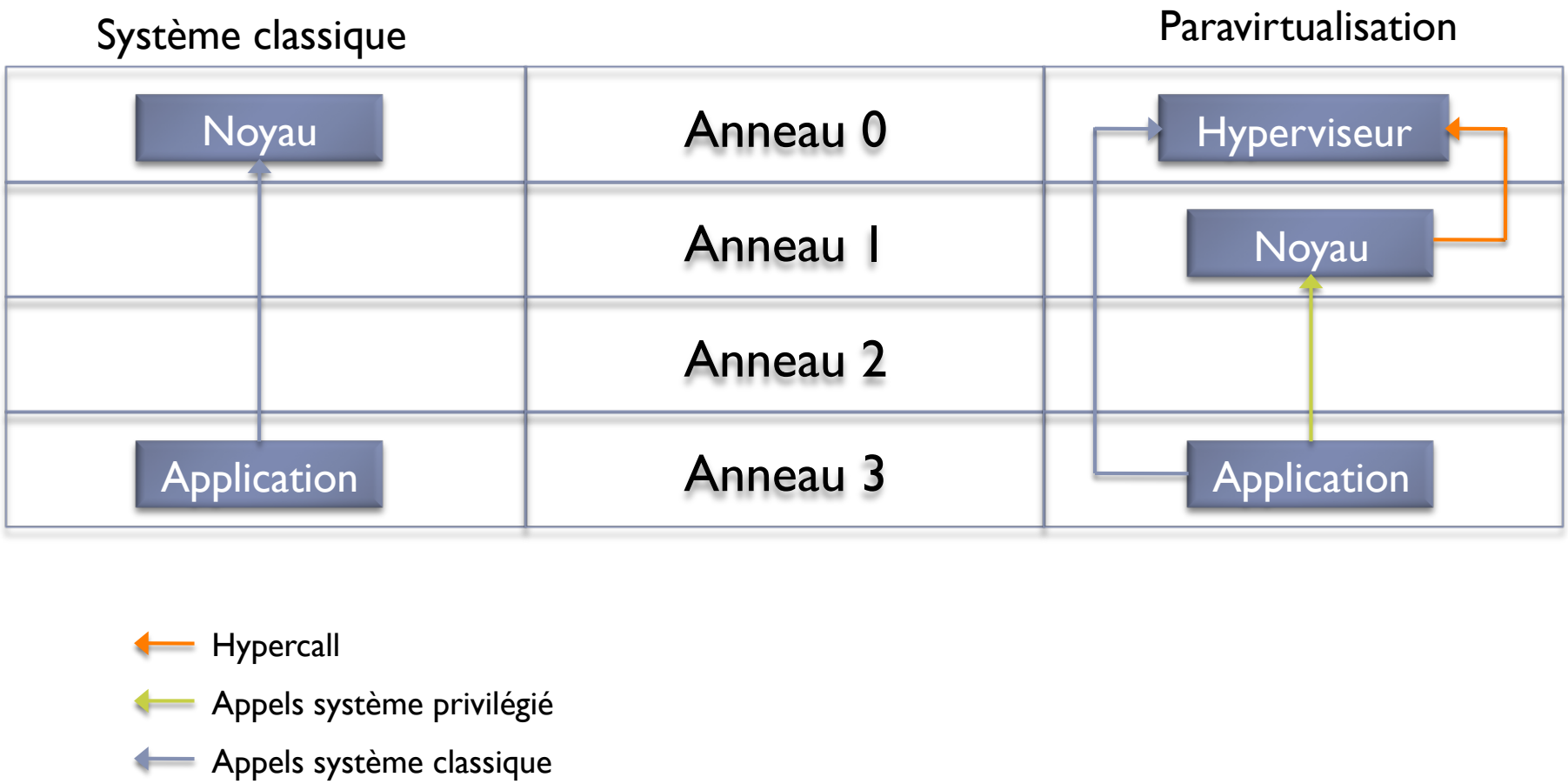
Interactions hyperviseur / applications / OS

- Apport des extensions de virtualisation



Architecture x86 64 bits et extensions processeurs liées à la virtualisation avec un hyperviseur Xen

Interactions hyperviseur / applications / OS



Les domaines

- Système d'exploitation invité

- Différents types de domaines pour différentes utilités
  - Domaine privilégié
  - Domaine non privilégié
  - Domaine matériel assisté

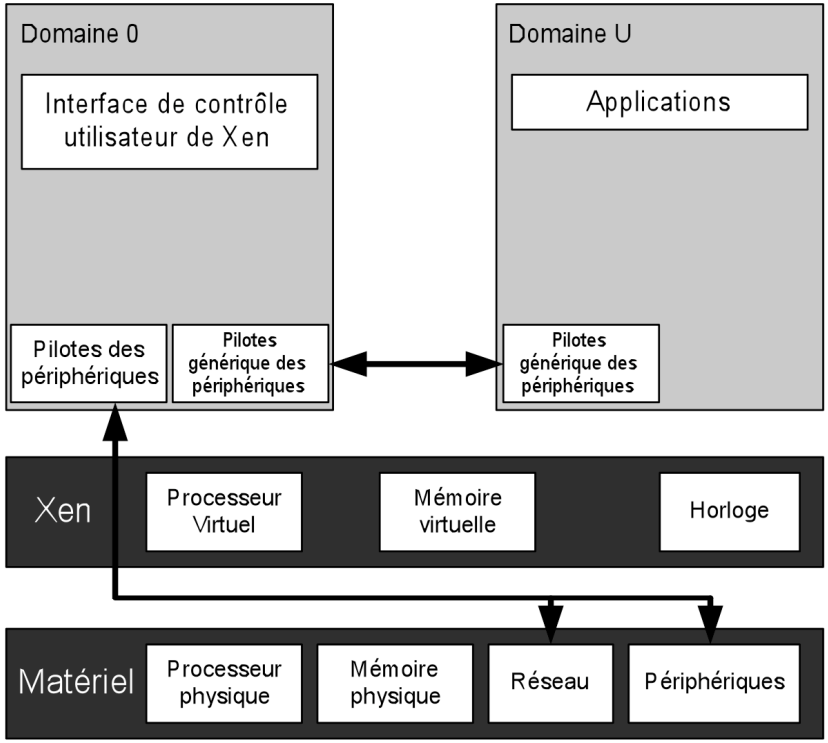
Domaines privilégiés

- Premier domaine exécuté au démarrage : Domaine 0
- Accès aux fonctionnalités de l'hyperviseur
- Gestion des autres domaines
- Accès direct à la machine physique et à ses périphériques
- Gestion de l'interface entre la machine physique et les domaines non privilégiés

Domaines non privilégiés

- Exécutés par le dom0
  - Limite de la capacité mémoire
- Domaines de paravirtualisation
- Accès aux périphériques à travers le dom0
- Système d'exploitation invité « classique »

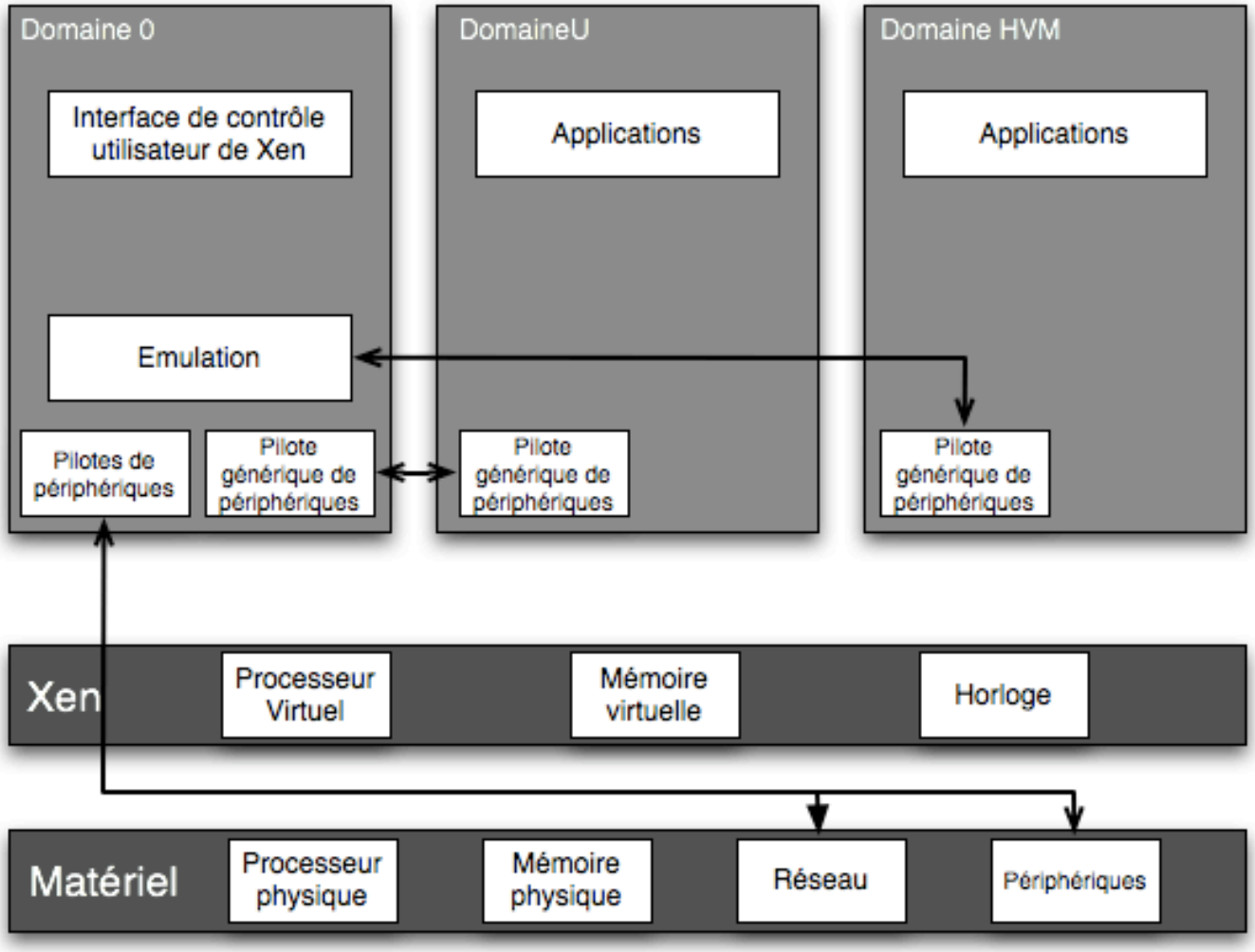
Schéma résumé



Domaine matériel assisté (HVM)

- Type de domaine non privilégié
- Virtualisation matérielle assistée
- Utilisation du module QEMU exécuté dans le dom0

Schéma résumé



IV Architecture de Xen en paravirtualisation

- La gestion des informations système
- La communication inter-domaines
- La gestion du temps
- La gestion de l'ordonnancement
- La gestion de la mémoire
- La réseau sous Xen

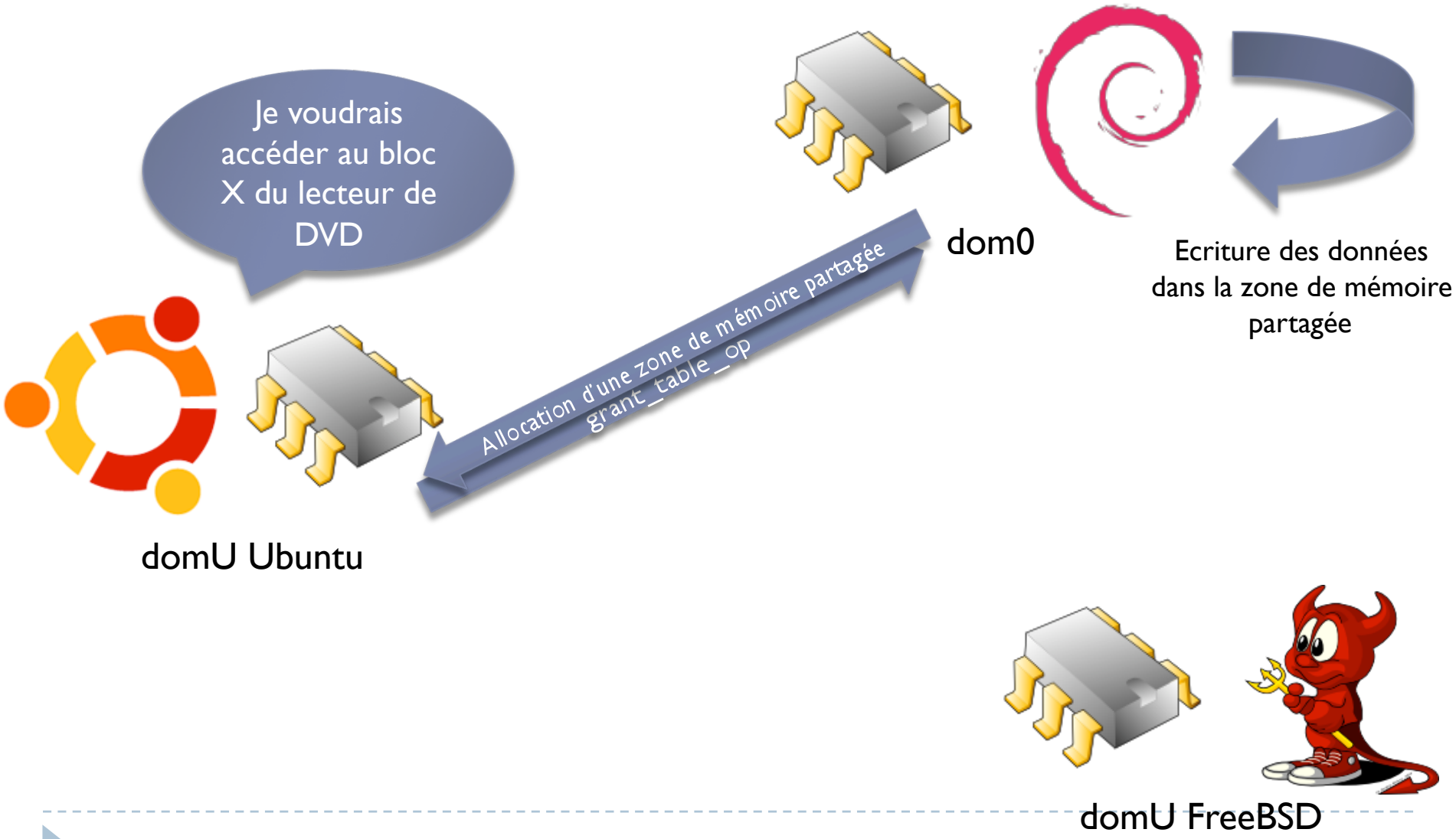
Gestion des informations système

- Nécessité d'avoir des informations sur le matériel sous-jacent
- Modification du mécanisme de récupération de ces informations système
- *Shared info pages* – pages d'information partagée
- Page d'information de démarrage
- Page *shared_info*

Communication interdomaines

- Isolation totale pas envisageable
- Nécessite d'accéder aux périphériques de la machine physique
  - Pilotes spécifiques dans le dom0
- Cas de processus normaux
- Zones de mémoire partagées
- Hypercall *grant_table_op*
  - Allocation de zones
  - Transfert de zones

Communication interdomaines



Gestion du temps

- Nécessité de la gestion du temps
- Ecoulement du temps
  - Lors de l'exécution
  - En dehors de l'exécution
- Trois valeurs contenues dans *shared_info*
  - Heure initiale du lancement du domaine
  - Temps écoulé depuis le lancement du domaine
  - Time Stamp Counter (TSC)

Gestion de l'ordonnancement

- Gestion de l'allocation de ressources de calcul aux domaines
- Ordonnanceurs
- Deux types
  - SEDF (Simple Earliest Deadline First)
  - Credit Scheduler

Gestion de l'ordonnancement

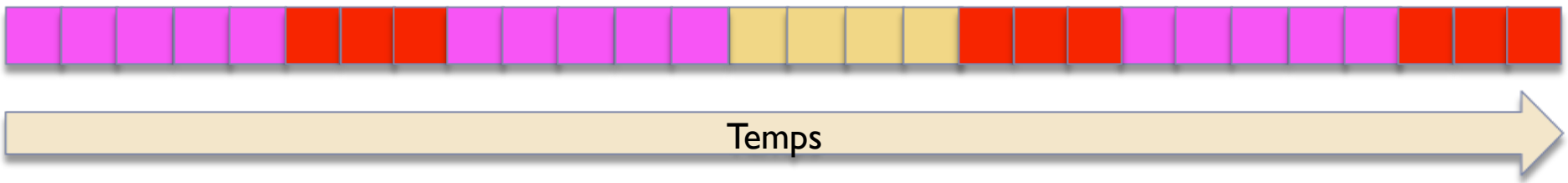
- SEDF
  - Temps d'exécution défini
  - Périodicité définie

Gestion de l'ordonnancement

Temps : 4 ms
Périodicité : 20ms

Temps : 3 ms
Périodicité : 10ms

Temps : 5 ms
Périodicité : 10ms



Allocation sur un VCPU

Gestion de l'ordonnancement

- *Credit scheduler*
- Deux attributs :
  - Poids
  - Limite
- Allocation de crédits
- Exécution jusqu'à l'épuisement des crédits

Gestion de l'ordonnancement

Poids : 2
Limite : 25%

Poids : 2
Limite : 0

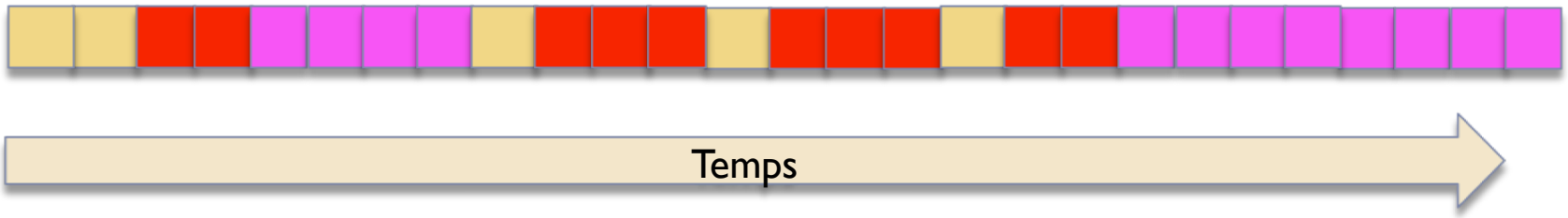
Poids : 4
Limite : 0

U : 2
F : 2
D : 4

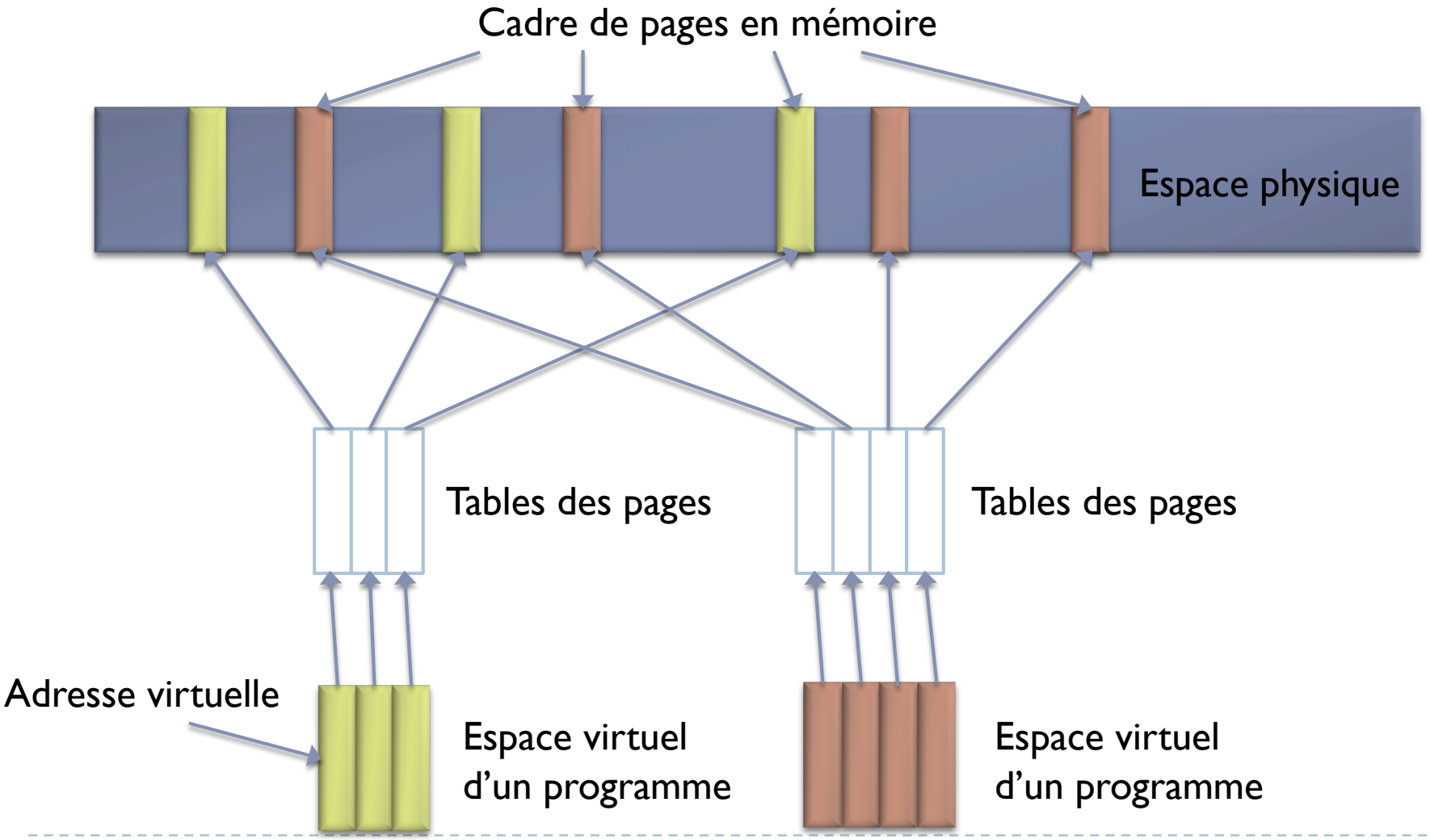
U : 1
F : 3
D : 4

U : 1
F : 2
D : 9

U : 1
F : 3
D : 12

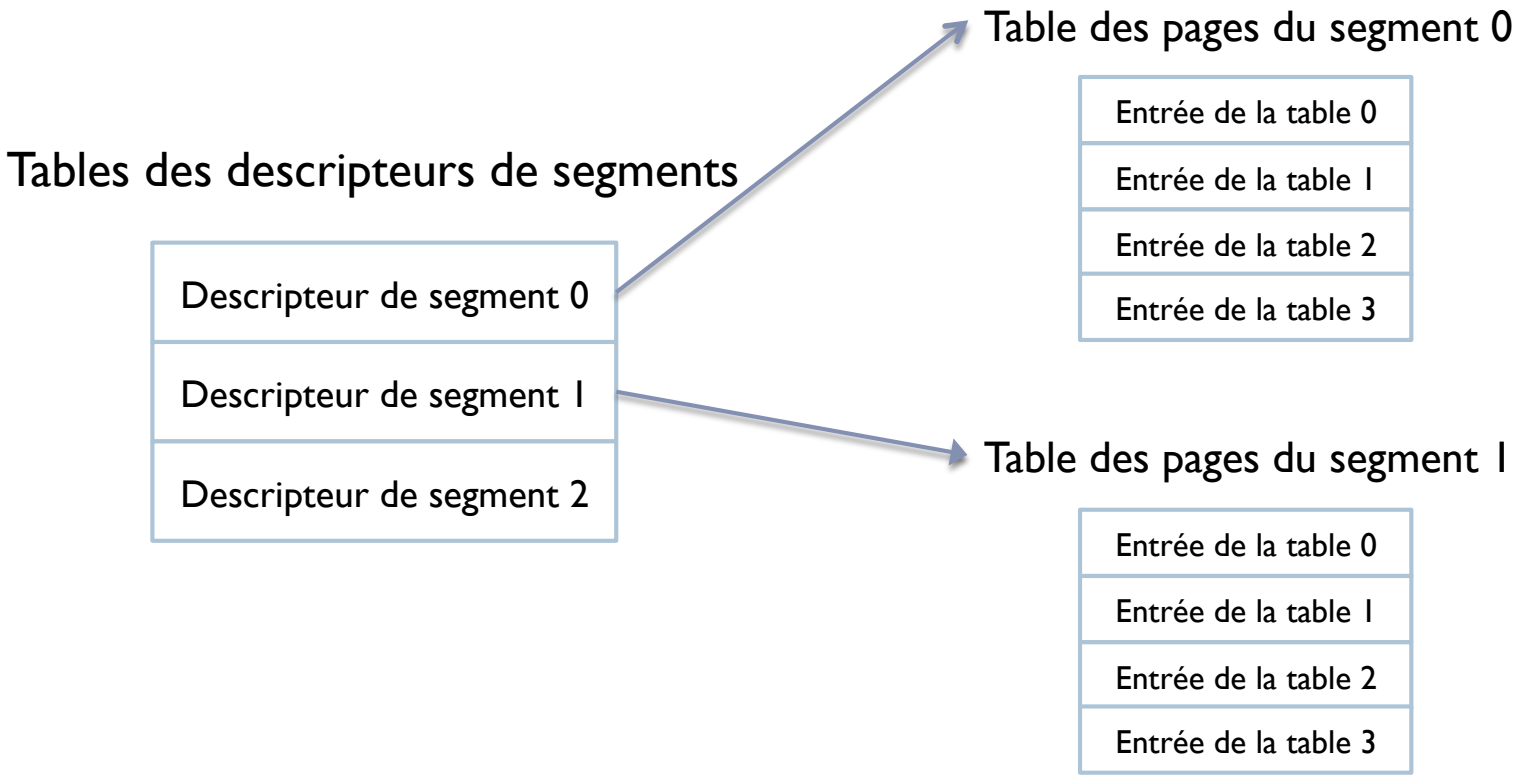


La pagination



La gestion de la mémoire dans un système classique

La segmentation



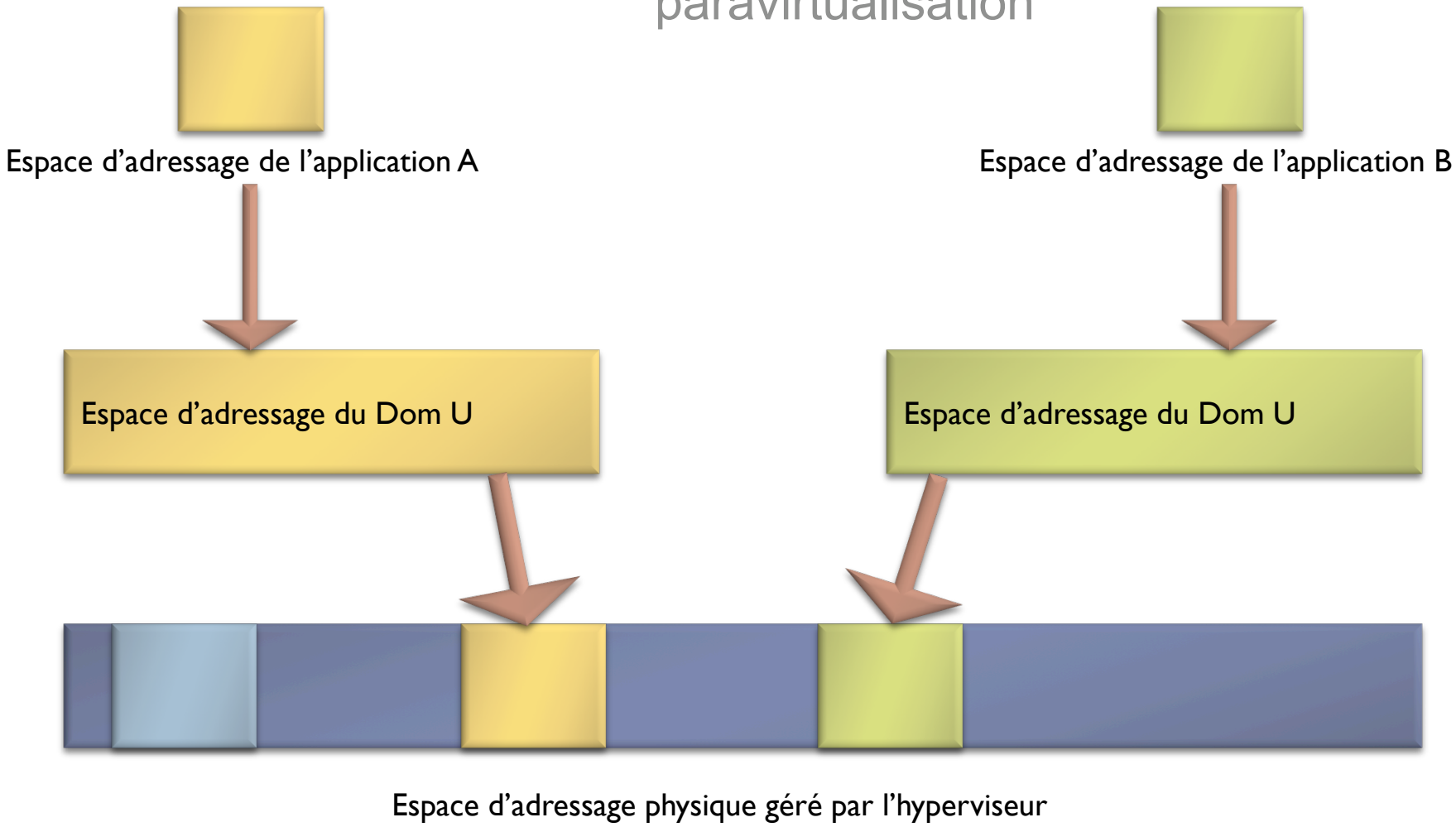
La segmentation

- En architecture x86, il existe 2 types de tables de descripteurs de segments :
  - Global Description Table
  - Local Description Table
- Nécessité d'un sélecteur de segment pour charger un descripteur dans un registre :

Numéro d'entrée dans la table	GDT/LDT	Niveau de privilèges 0-3
13 bits	1 bit	2 bits

La gestion de la mémoire dans Xen

Modèle de gestion de mémoire en paravirtualisation



La gestion de la mémoire dans Xen

Modèle de gestion de mémoire en paravirtualisation

- Pourquoi 3 niveaux d'indirections
  - Problème lié au fait qu'un système croit que son espace d'adressage est continu et qu'il est seul sur la machine physique
  - Problème de la mise en pause d'une machine et d'un redémarrage sur une autre machine physique
- Accès mémoire et mise à jour de la table des pages
  - Lecture et Ecriture uniquement via des hypercalls

Intérêt : Aucun système invité ne modifie l'espace mémoire sans en avertir l'Hyperviseur

La gestion du réseau

- Toutes les communications passent par le Dom 0
- 2 tunnels entre un Dom U et le Dom 0
  - Un tunnel de service et un tunnel de données
  - Chaque tunnel composé de 2 zones mémoire distinctes
  - Tunnel de service permet de localiser les pages mémoires de données brutes
- Communication réseau interdomaines gourmande en ressources processeur

Conclusion sur Xen

- Solution innovante et performante
- Solution libre permettant l'accès à la connaissance
- Support de l'industrie informatique
  - "IBM is a strong supporter of the open source community and has a long history with virtualization, having invented it for mainframe over 40 years ago." – Rich Lechner, Vice Président d'IBM Entreprise Systems
- Composante de base du « Cloud Computing »

Conclusion du projet

- Acquisition de connaissances sur un sujet d'actualité
- Approfondissement des connaissances en système d'exploitation
- Assimilation d'une méthodologie de rédaction
- Clarification et formalisation de notions
- Apport professionnel dans le domaine des systèmes d'information