

Virtualization level 1

1 : Introduction

Que signifie virtualisation ?

Application Mail Server	Application Web Server
Operating System Win Server 2008	Operating System Linux
Virtual Hardware CPU + RAM + ...	Virtual Hardware CPU + RAM + ...
Virtualization Layer (Type 1) = Hypervisor, ESX, Linux-KVM, ...	
Physical Hardware = CPU + RAM + ethernet + local disk	

Virtual Machine (VM)

Chaque machine virtuelle
croit posséder seule le
matériel disponible

- Compatibilité binaire (OS + applic)
- Transparence (performance)
- Cloisonnement (sécurité)

Arguments commerciaux

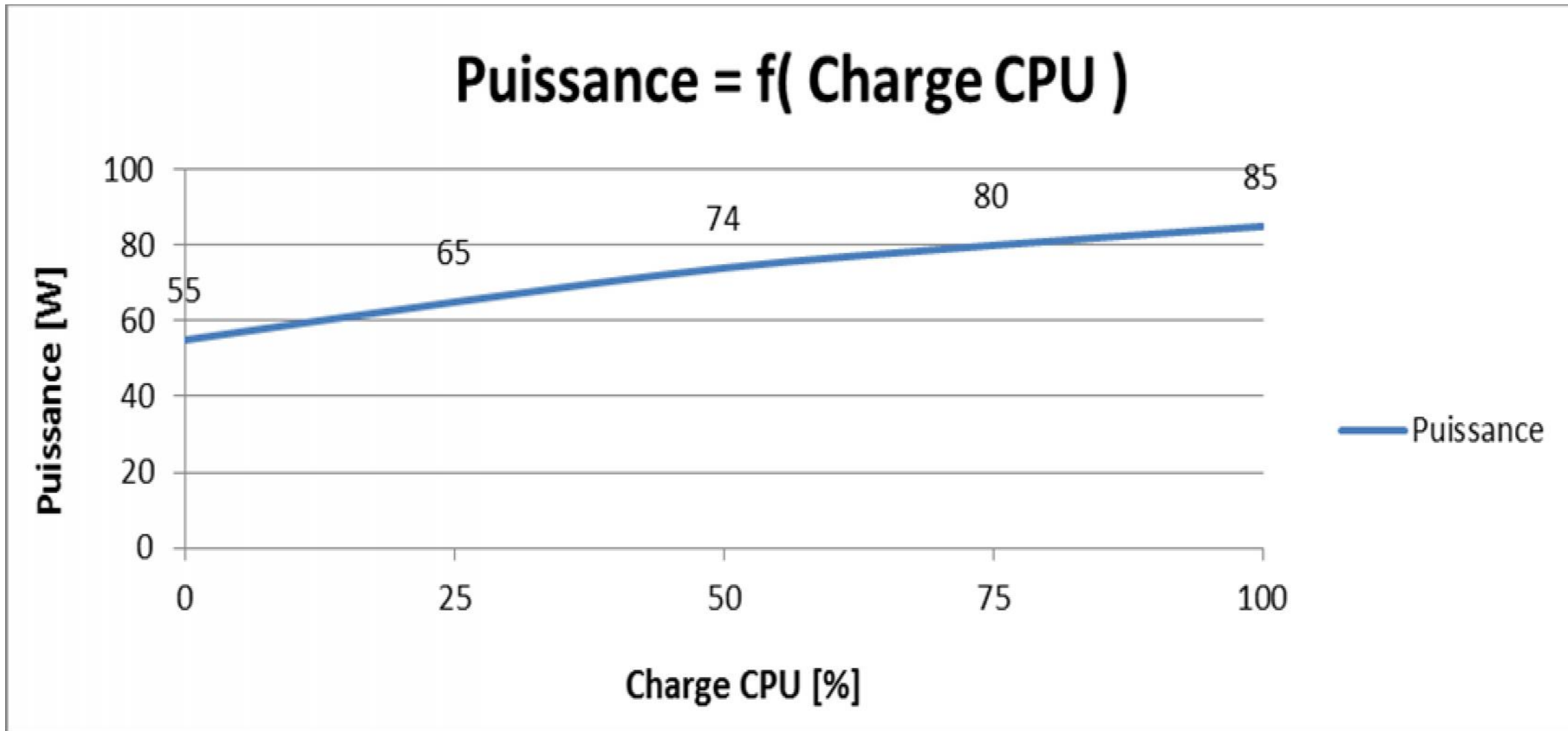
- Meilleure utilisation du matériel → économie

Consolider des serveurs

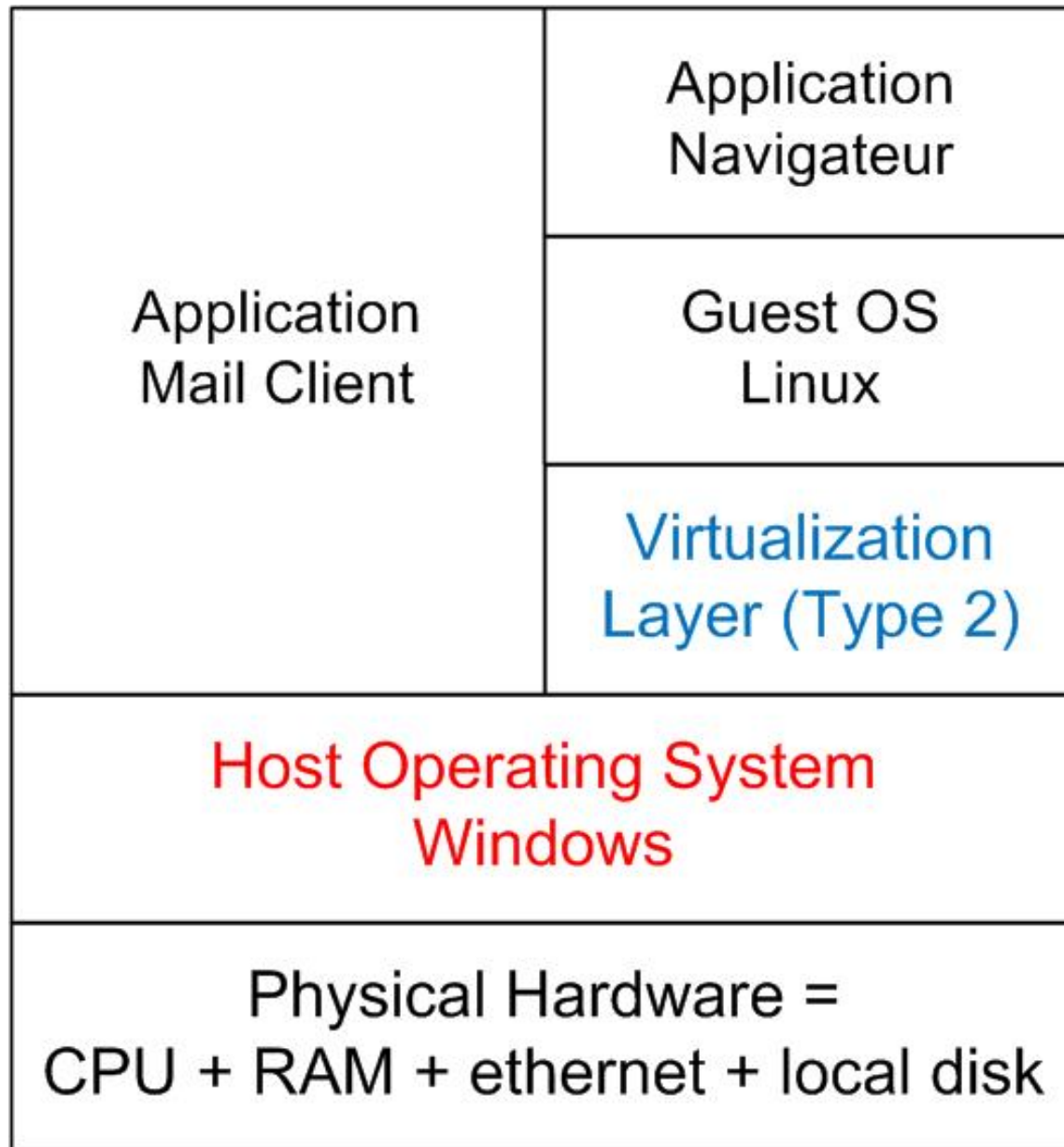
- Facilité d'ajouter une machine virtuelle supplémentaire → souplesse
- Idéal pour les tests (régression, avec plusieurs versions de l'OS)
- Sauvegarde simplifiée → disponibilité
- Gartner prédisait l'installation de 4 Mio de machines virtuelles en 2009, mais que 60% sera moins bien sécurisée qu'avec une architecture traditionnelle
- IBM proposait la virtualisation (VM/370) dès 1972

Consommation des PC du labo

- Carte mère Asus avec 2 CPU-64bit de 3 GHz, 4 GByte de RAM, un disque SATA de 320 GByte et 3 interfaces ethernet



Variante (Type 2) basée sur Host OS + VirtualBox, ...



- Avantages
Excellent support du matériel
- Inconvénients
Performances
- Compatibilité
Les machines virtuelles fonctionnent sur les 2 architectures Type 1 & Type 2
- Cours + lab des sem 1 & 3

Fonctions supportées par ESX (Virtualization Layer)

- Emuler le matériel

VMM (*Virtual Machine Monitor*) implémente la couche d'abstraction matérielle (CPU, RAM, ...) → **vNIC** (*virtual Network Interface Card*)

- Répartir le temps CPU

Scheduler qui alloue du temps aux **vCPU** (*virtual CPU*)

PC Gigabyte possède 1 pCPU Core2Duo → 2 vCPU disponibles

- Gérer la mémoire RAM

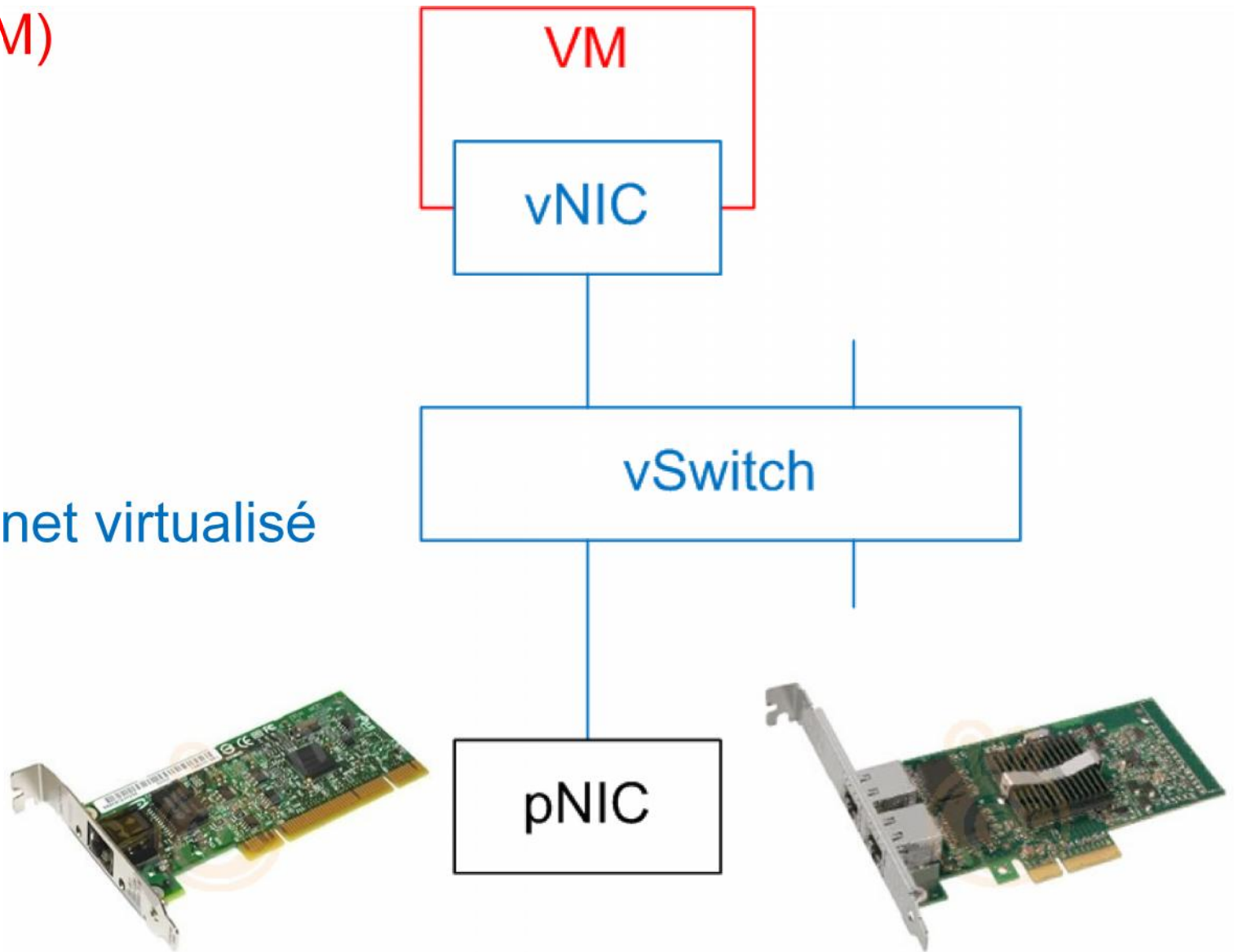
- Offrir des commutateurs ethernet virtuels → **vSwitch**

- Gérer le matériel (*physical*) → pNIC, pCPU, ...

- Choix des Guest OS → <http://www.vmware.com/guides>

VM, vNIC, vSwitch, pNIC

- *Virtual Machine (VM)*
- *virtual NIC (vNIC)*
interface virtualisé
- *virtual Switch*
commutateur ethernet virtualisé
- Interface physique
Intel PRO/1000

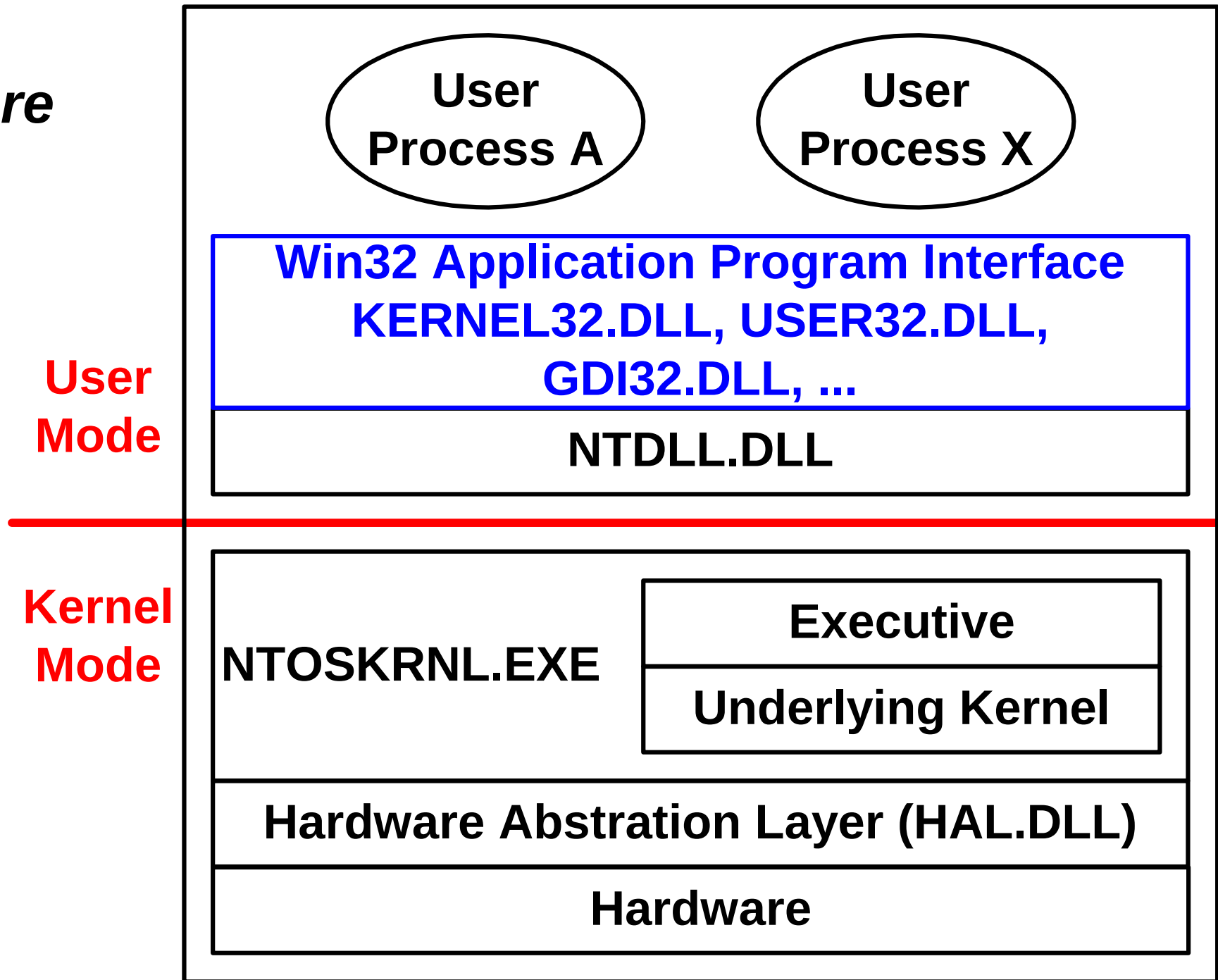


Device Manager (XP)

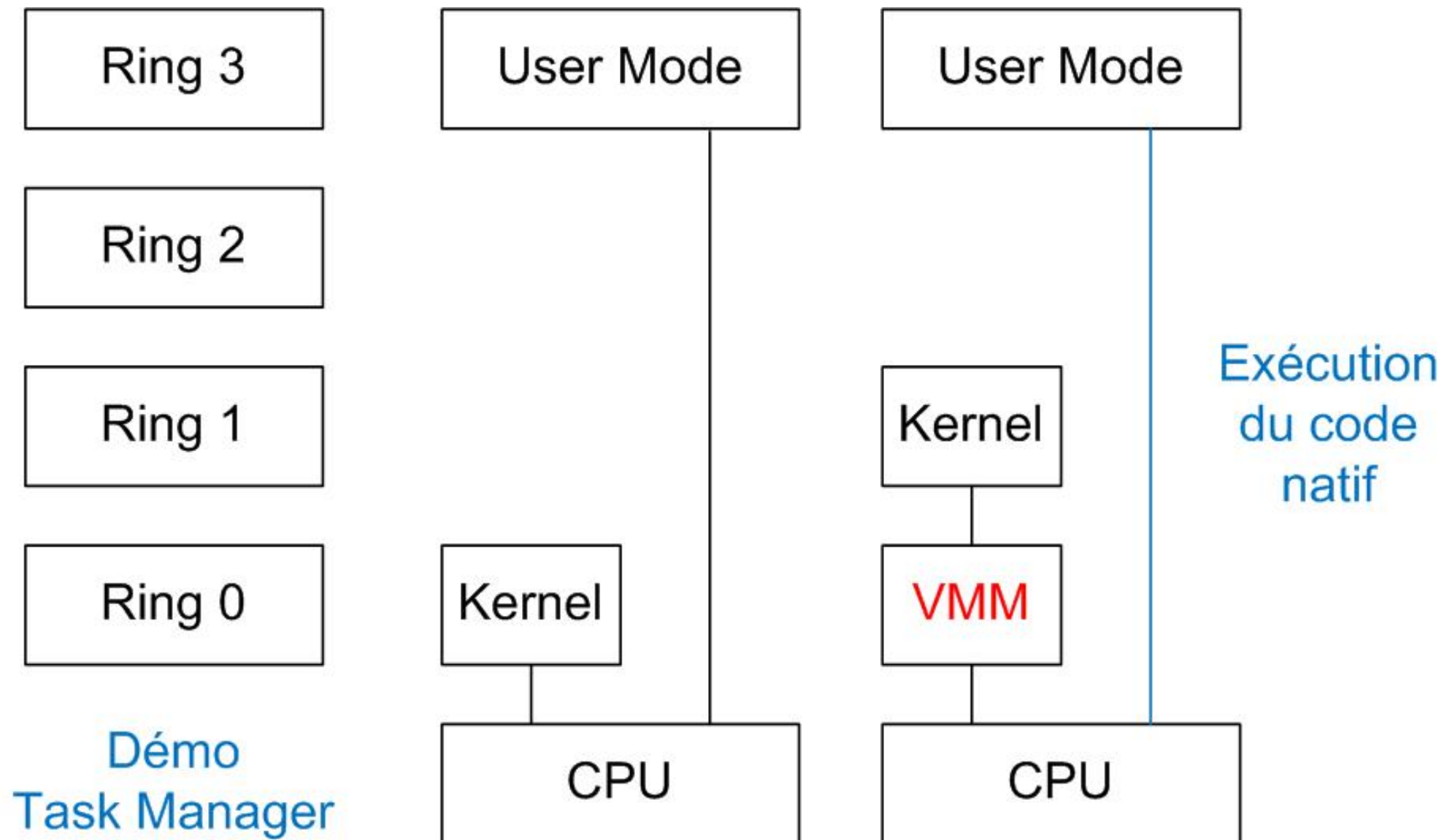
- ESXi remplace les pilotes (*drivers*)



Windows Architecture

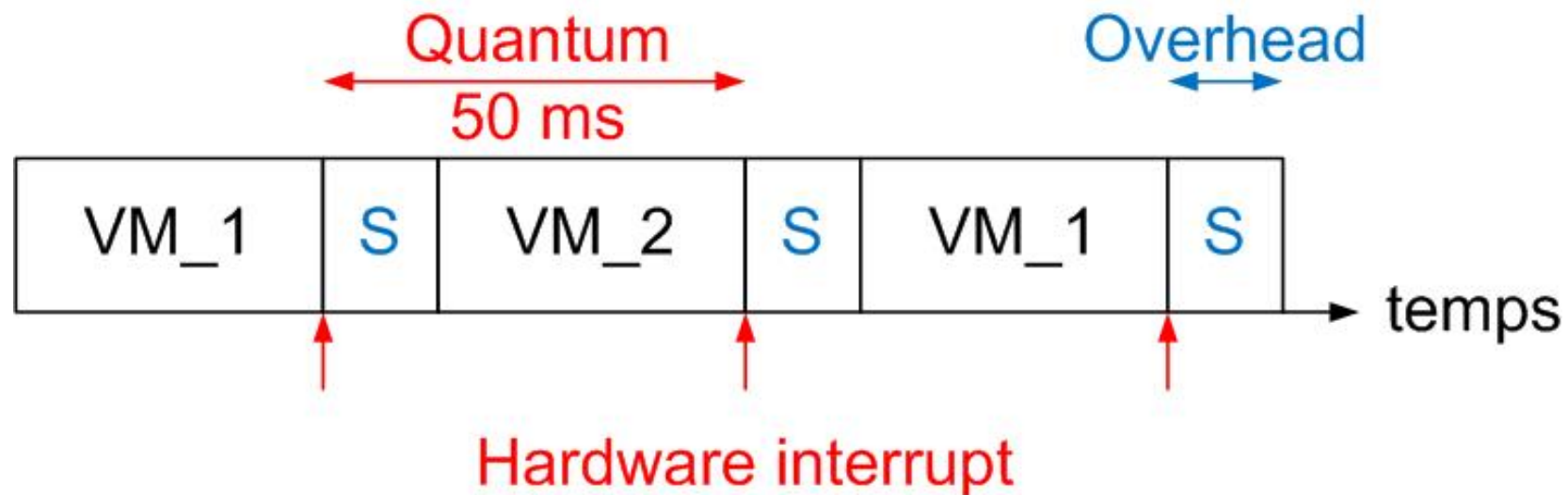


Exécution code x86 sans Hardware-assisted Virtualization



Scheduler (ordonnanceur)

- Périodiquement (quantum de temps = 50 ms), l'ordonnanceur doit allouer une ressource pCPU au processus (VM) qui le demande



- L'ordonnanceur consomme du temps **S** (**Système**); ce qui représente un coût (*overhead*)
- Etats de la VM (processus) = *Running / Waiting / Ready*

Virtual Machine (VM)

Getting Started Summary **Virtual Machines** Resource Allocation Performance Configuration Users & Groups Events

Name, State, Host or Guest OS contains:

Name	State	Status	Host CPU - MHz	Host Mem - MB	Guest Mem - %
MESA	Powered Off	○○●	0	0	
FW_ubuntu-server	Powered Off	○○●	0	0	
Nostalgia	Powered On	○○●	3025	26	54

- Power On Ctrl+B
- Power Off Ctrl+E
- Suspend Ctrl+Z
- Reset Ctrl+R
- Shut Down Guest
- Restart Guest
- Snapshot ▶
- Open Console

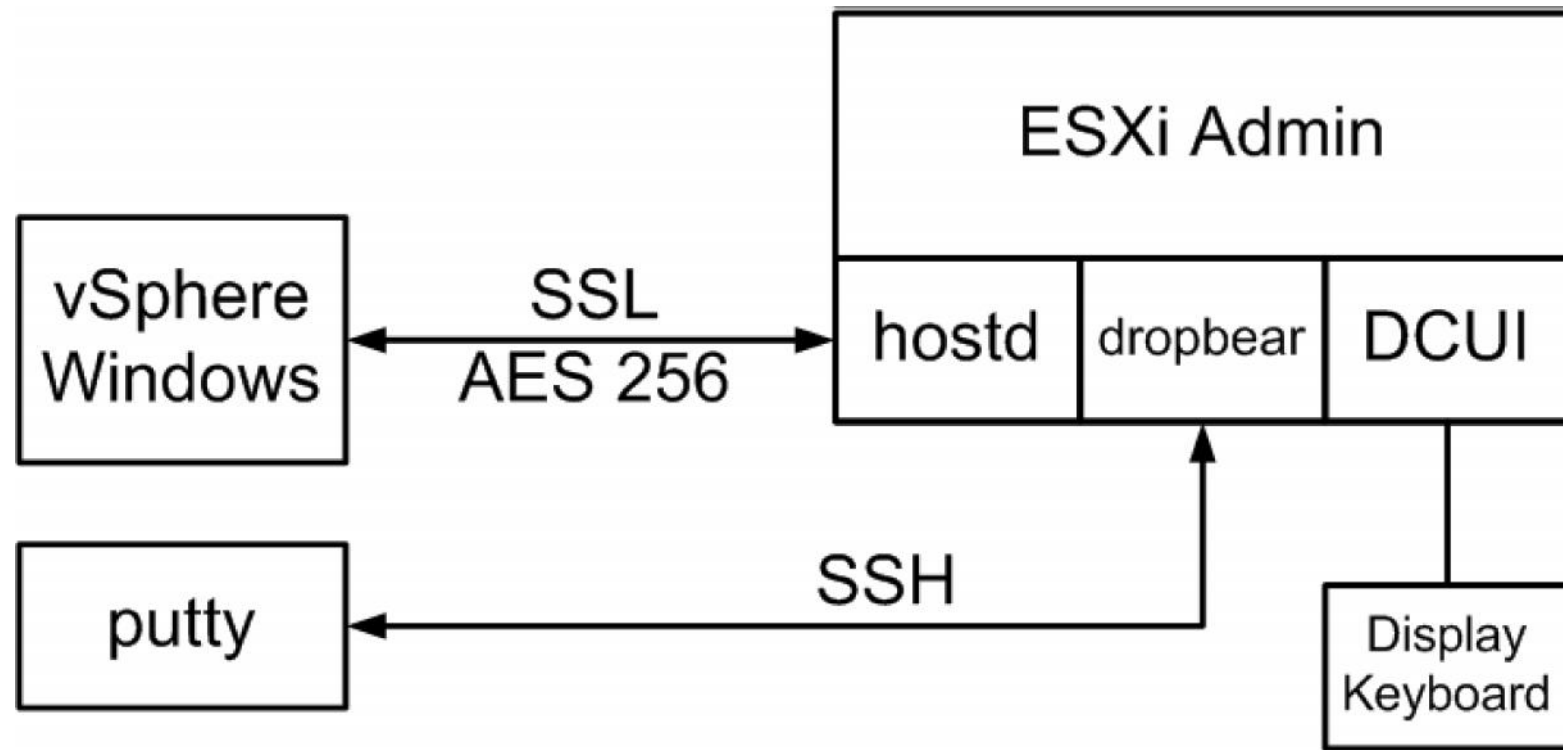
- 3 machines virtuelles

- 1 machine virtuelle active

- Mise sous tension / arrêt d'un serveur traditionnel
- Fonctions équivalentes (*Power ON / OFF*) pour une machine virtuelle

Administration avec vSphere

- vSphere compatible Windows & .NET [Démo](#)
Outil GUI (*Graphical User Interface*) recommandé pour l'admin.



Accès SSH pour
admin. en mode CLI

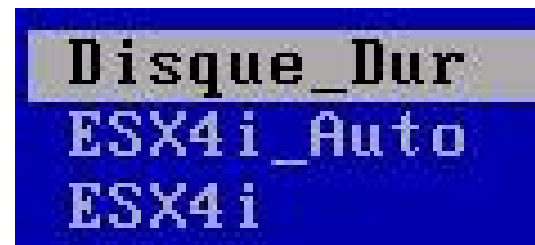
Direct Console UI
→ config. de base

Labo level 1 §1-3 : ESXi, vSphere, appliance (40 min)

- Matériel à disposition (CHF = 700) : 1 CPU dual Core-64bit de 3 GHz, RAM=8 GByte, disque=320 Gbyte SATA, 3 interfaces ethernet
- §1 Installer ESXi en mode PXE → next slide
ESXi = *Hypervisor, bare-metal architecture* de 32 Mbyte
Configurer (clavier, mot de passe et réseau) via la console physique DCUI (*Direct Console UI*)
- §2 Utiliser l'outil d'administration vSphere
- §3 Télécharger et utiliser une *appliance* = logiciel préinstallé

Installation via PXE (Pre-boot eXecution Environment)

- 1 PC_ESXi démarre en mode PXE par défaut (BIOS)
- 2 Serveur DHCP fournit les paramètres habituels +
Option 66 = adr_IP du serveur TFTP
Option 67 = pxelinux.0 (nom du fichier à télécharger)
- 3 Serveur TFTP propose le menu
Pour le détail → <\\10.2.1.1\tftpboot>



Chargement en 2-3 min sur LAN 1 Gbit/s

<http://www.syslinux.org/wiki/index.php/PXELINUX>

Virtualization level 1

2 Approfondissement

Synthèse du labo précédent

- Comment fonctionne l'installation via PXE ?

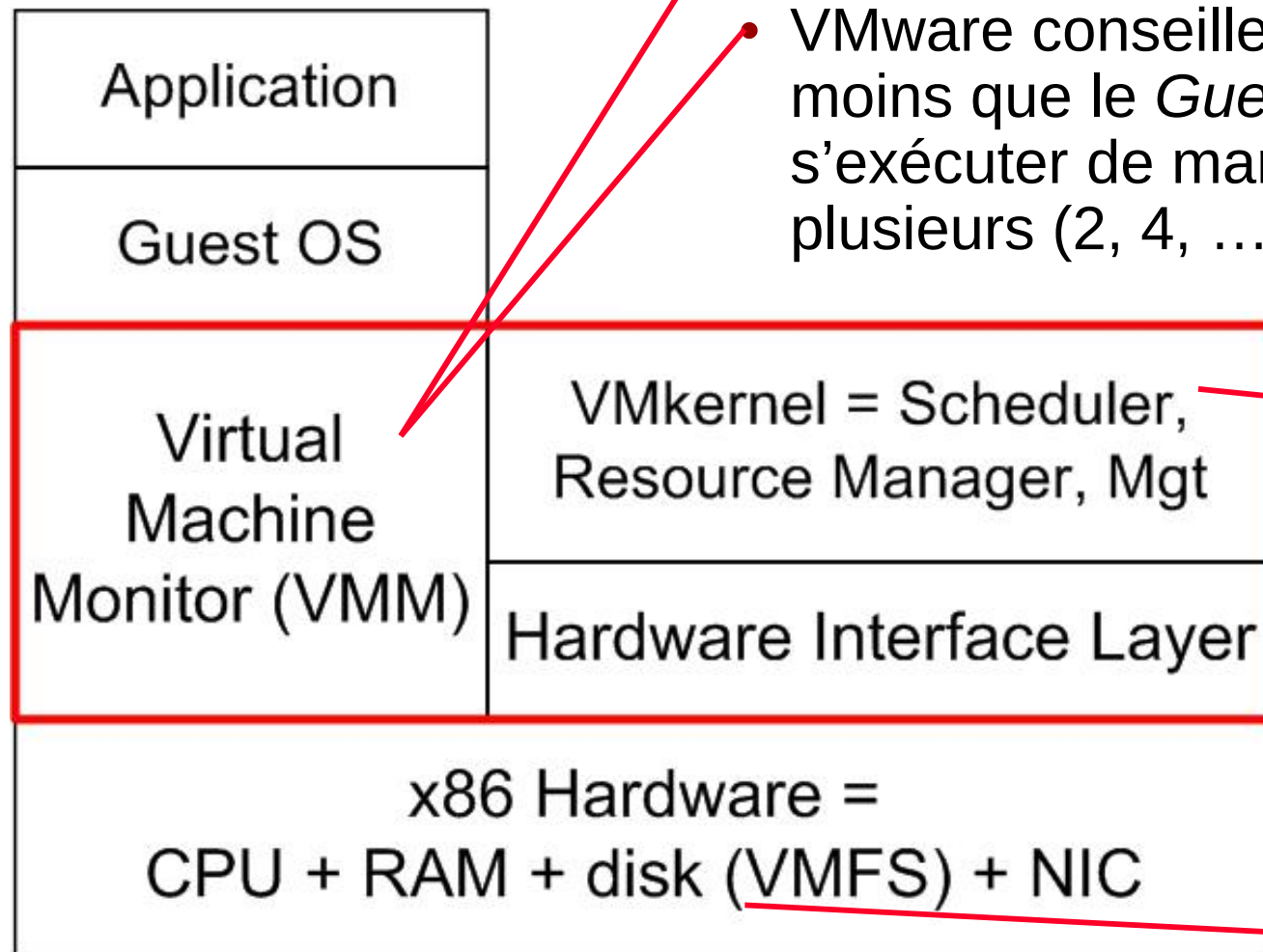
Voir slide 15

- Quels sont les principaux paramètres qui caractérisent le matériel utilisé ? Utiliser vSphere pour répondre

Voir slide 19

- Pourquoi la taille (105 MB) du fichier vmdk est différente de celle (6 MB) du fichier vmdk téléchargé ?
- Avec la VM Nostalgia, pourquoi la charge CPU = 3000 MHz ?

Architecture ESXi



Dans notre cas, VMM peut émuler **2 vCPU (virtual CPU)** → car 2 pCPU

VMware conseille de laisser ce paramètre à 1 à moins que le *Guest OS* soit capable de s'exécuter de manière concurrente sur plusieurs (2, 4, ...) CPU physiques

• VMkernel 64bit basé sur noyau Linux

• Management (web) via vSphere

• VMFS (*Virtual Machine File System*)

PC Gigabyte du labo

localhost.localdomain VMware ESX Server 3i, 3.5.0, 110271

Getting Started Summary Virtual Machines Resource Allocation Performance Configuration Users & Groups Events Permissions

General

Manufacturer: **Gigabyte Technology Co., Ltd.**
Model: G33M-S2L
Processors: **2 CPU x 2.999 GHz**
Processor Type: Intel(R) Core(TM)2 Duo CPU E8400 @ 3.0...
Hyperthreading: **Inactive**
Number of NICs: **2**

Resources

CPU usage: **57 MHz**
 2 x 2.999 GHz

Memory usage: **390.00 MB**
 3.99 GB

Datastore	Capacity	Free
 datastore1	293.25 GB	276.52 GB

- CPU = Core2Duo 3Ghz → 6 GHz à disposition
- RAM = 2x2GByte → 4 GByte à disposition
- NIC = 2 dans l'exemple (3 sur les PCs du labo)
- HD = env 290 GByte

Infrastructure matérielle SIACG utilisée par Korso

- Hyperviseurs → 4 x Cisco UCS B200 M3 Blade Server



80 pCPU – 2.8 = 223 GHz, 1 TB pRAM, 1 TB SSD en RAID5
200 VMs Win7

- SAN EMC-VNX = 2 systèmes RAID5

5 x SAS 600 GB = 3 TB = 2 x LUN (1.5 TB + 0.5 TB)

5 x SSD 200 GB = 1 TB

VMware Datastore

	vnx-lun160-Server	511.75 GB
	vnx-lun161-Lclone	1.53 TB
	vnx-lun170-Master	149.75 GB
	vnx-lun171-Profils	149.75 GB
	vnx-lun172-Replicas	299.75 GB

Resource Manager : Resources available

localhost.localdomain VMware ESX Server 3i, 3.5.0, 110271			
Getting Started		Summary	
Virtual Machines		Resource Allocation	
CPU Reservation:		4648 MHz	Memory Reservation:
CPU Reservation Used:		0 MHz	Memory Reservation Used:
CPU Unreserved:		4648 MHz	Memory Unreserved:
			2934 MB

- CPU Reservation = 4648 MHz Ressources disponibles = 77%
- CPU Reservation Used = 0 **Aucune machine virtuelle active**
- Memory Reservation = 2934 MByte Ressources disponibles = 73%
- Ces chiffres permettent d'estimer le coût de fonctionnement (overhead) de ESXi

Resource Manager : CPU

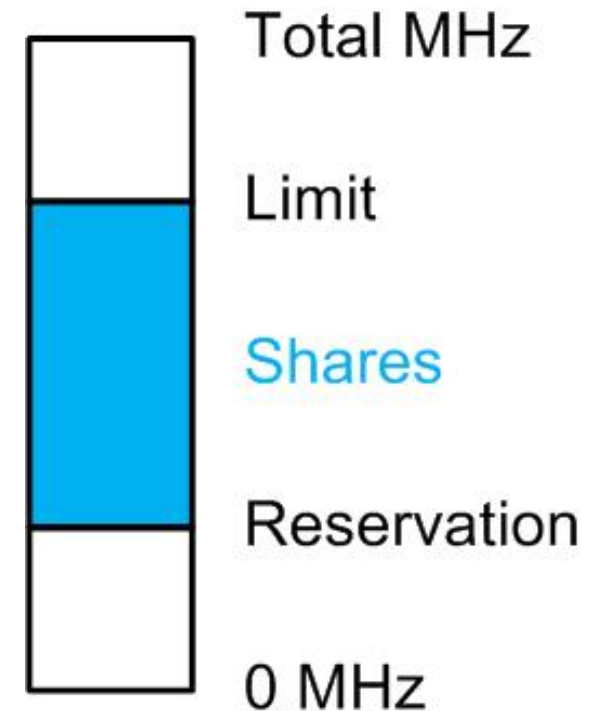
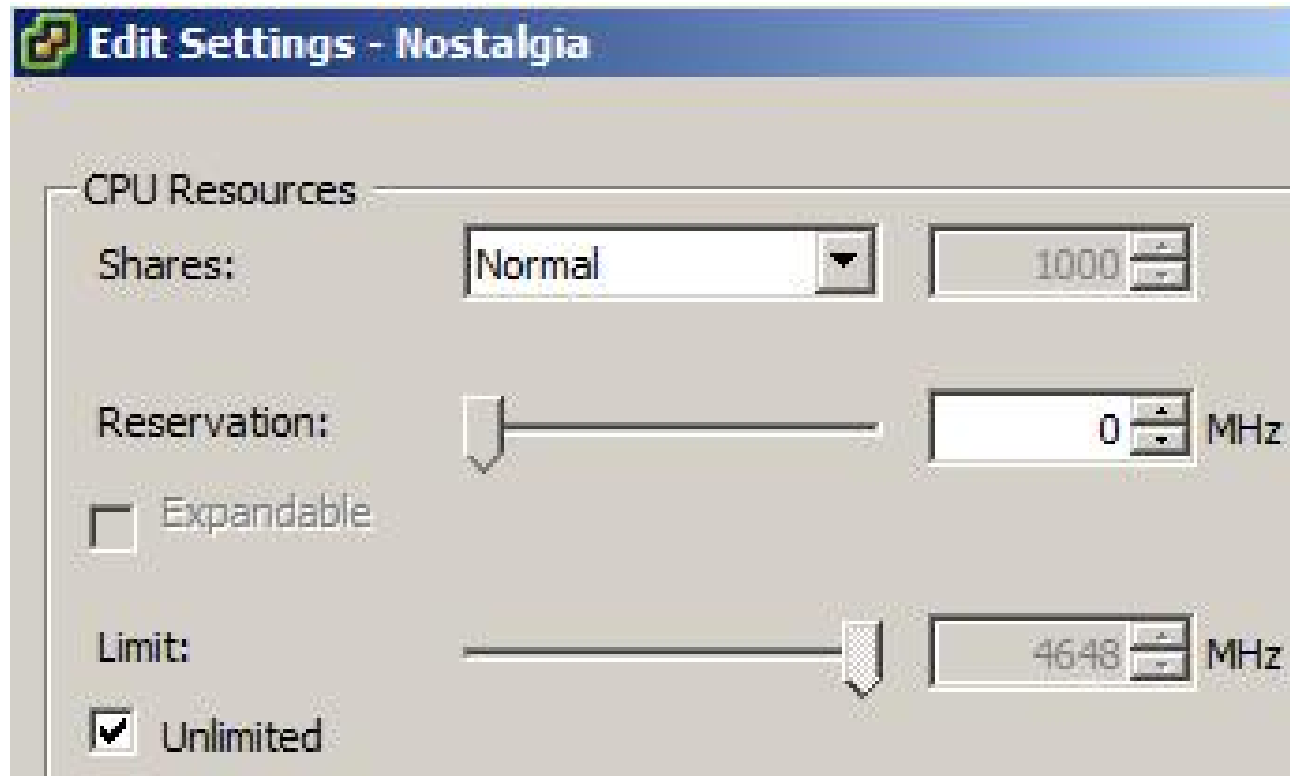
- Par défaut, **allocation équitable** (*fair*) du temps CPU **entre les N machines virtuelles (VM)**

View: ☒ CPU ☐ Memory

Name	Reservation - MHz	Limit - MHz	Shares	Shares Value	% Shares
 MESA	0	4648	Normal	1000	33
 FW_ubuntu-server	0	Unlimited	Normal	1000	33
 Nostalgia	0	Unlimited	Normal	1000	33

- 1 VM → 100%, 2 VM → 50%, 3 VM → 33%, ...
- Possibilité de définir des priorités relatives (*shares*) entre VM
High = 2000, *Normal* = 1000, *Low* = 500, *Custom* = X
- Mécanisme actif qu'en cas de **contention (no more IDLE)**

Resource Manager : CPU reservation & limit



- Possibilité de réserver (garantir) du temps CPU par VM
Redistribution du temps si la VM ne consomme pas l'intégralité
- Possibilité de limiter le temps CPU par VM
- Analogie avec USB → flux bulk (*fair*) & isochrone (*guaranteed*)

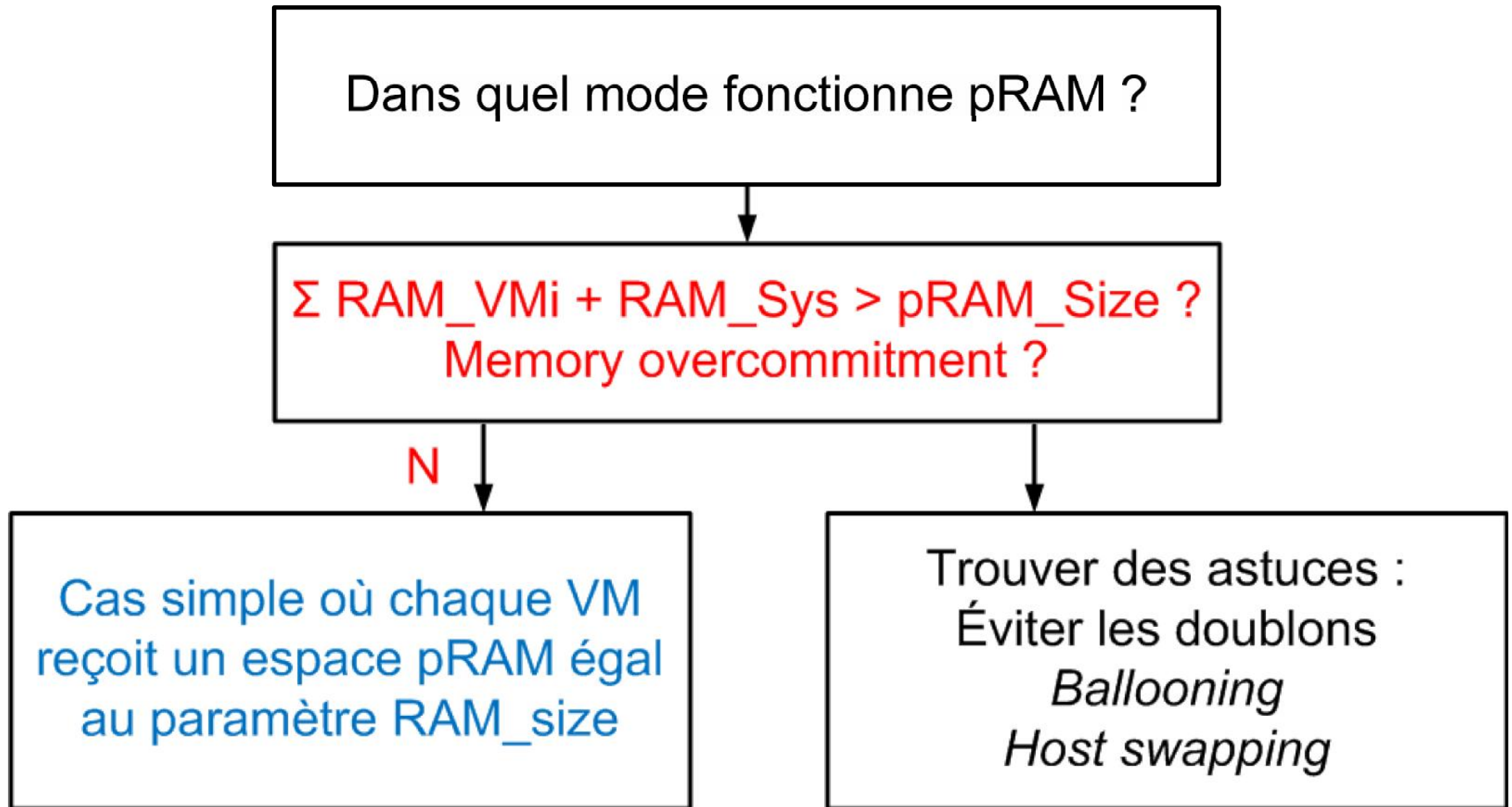
Resource Manager : CPU (Remarques)

- Le mode sans réservation est préférable
 - pour se familiariser avec ESXi
 - en cas de changements fréquents
- ESXi tente toujours d'allouer toutes les ressources disponibles
Le temps CPU réservé mais non consommé est disponible pour les autres VM
- Réserver une valeur minimale (*Best Practices*)
- Tenir compte du coût de ESXi → *overhead*
Ralentissement < 5% *for CPU-intensive applications* (SPECcpu2000)
http://www.vmware.com/pdf/hypervisor_performance.pdf

Virtualisation temporelle

- Les OS (*scheduler*) ont besoin de quantum (intervalle) de temps
Composant matériel qui génère 100 interruptions par seconde (100 Hz)
Lecture d'un compteur qui évolue en fonction du temps
- Ils ont aussi besoin de connaître le temps absolu (GMT, ...)
- Composants matériels sont très variés (CMOS, ACPI, ...)
- Gestion spécifique par chaque OS (Windows, Linux, Solaris, ...)
- Idem pour *Idle* (temps CPU non consommé) !!!
- Recommandation : synchroniser les horloges avec une base de temps externe NTP (*Network Time Protocol*) → Labo §2.5
- http://www.vmware.com/pdf/vmware_timekeeping.pdf

Modes de fonctionnement de la mémoire RAM



Fichiers des machines virtuelles

- **Virtual Machine File System (VMFS)**

Système de fichiers natif utilisé par VMkernel

- Pour VM = Nostalgia

1.2k	VM.vmx	Fichier principal de config.
373	VM.vmdk	Config. disque
102.4M	VM-flat.vmdk	Code machine (64 MByte) + <i>overhead</i>

- Autres extensions utilisés

VM.log, **VM.nvram** (BIOS), **VM.vmss** (suspend)

VM.vswp (swap), **VM.vmx**, **VM.vmsn** (snapshot), ...

- <http://searchvmware.techtarget.com/tip/Understanding-the-files-that-make-up-a-VMware-virtual-machine>

Affichage approximatif des fichiers par vSphere

[datastore1] Nostalgia4			
Name		Size	Type
 Nostalgia4.vmx		1.20 KB	Virtual Machine
 Nostalgia4.vmdk		105'472.00 KB	Virtual Disk
 Nostalgia4.vmx		0.26 KB	File
 Nostalgia4.vmsd		0.00 KB	File

```
ls -lah
```

```
102.4M Nostalgia4-flat.vmdk
```

```
373 Nostalgia4.vmdk
```

```
0 Nostalgia4.vmsd
```

```
1.2k Nostalgia4.vmx
```

```
265 Nostalgia4.vmx
```

Snapshot (instantané)

- Un *snapshot* est une **capture à chaud** à un moment donné de l'état du système virtualisé
- Il permet de **sauvegarder** l'état actuel du système (d'exploitation et des applications) pour pouvoir y revenir ultérieurement
- On peut par exemple créer un *snapshot* avant de mettre à jour le système et ainsi revenir à l'état initial en cas de problème !
- La sauvegarde d'un système à chaud rentre souvent en conflit avec des **fichiers réservés** par le système d'exploitation comme le fichier SAM de Vista mis en exclusion mutuelle par le processus LSASS
- La solution consiste à effectuer un *snapshot* du système, de copier cette image (les fichiers) vers un serveur de backup puis d'effacer ce *snapshot*

Snapshot (suite)

- Image créé **à chaud** avec vSphere (*snapshot manager*)
- Copie de VM-flat.vmdk selon un mode incrémental (bloc de 16 MByte)
- La taille du fichier créé reflète le niveau de changement (écriture) de la VM
- Créer plusieurs points de démarrage d'une VM et utiliser *Go to* pour **démarrer**
- **Effacer** tous les snapshots
→ nouvelle VM intègre les modifications contenues dans les *snapshots*
- **Revenir** (*revert*) au *snapshot* précédent en perdant les éventuelles modifications



Gestion des machines virtuelles

1	<i>Power ON</i>	copy VMFS → RAM	vCPU = 25%
2	<i>Suspend</i>		vCPU = 0
3	<i>Power ON</i>		vCPU = 25%
4	<i>Snapshot</i>	copy RAM → VMFS	vCPU = 25%
5.1	<i>Shutdown</i>	copy RAM → VMFS	vCPU = 0
5.2	<i>Power OFF</i>	contenu RAM perdu	vCPU = 0

- *Scheduler* du noyau Linux, processus dans l'état *Running*, *Waiting*, ... → *slide esx*top

Toolbar & Power Control

- Pour activer → *View – Toolbar*



power control

snapshot

console

- Sélectionner une VM puis clic-droit

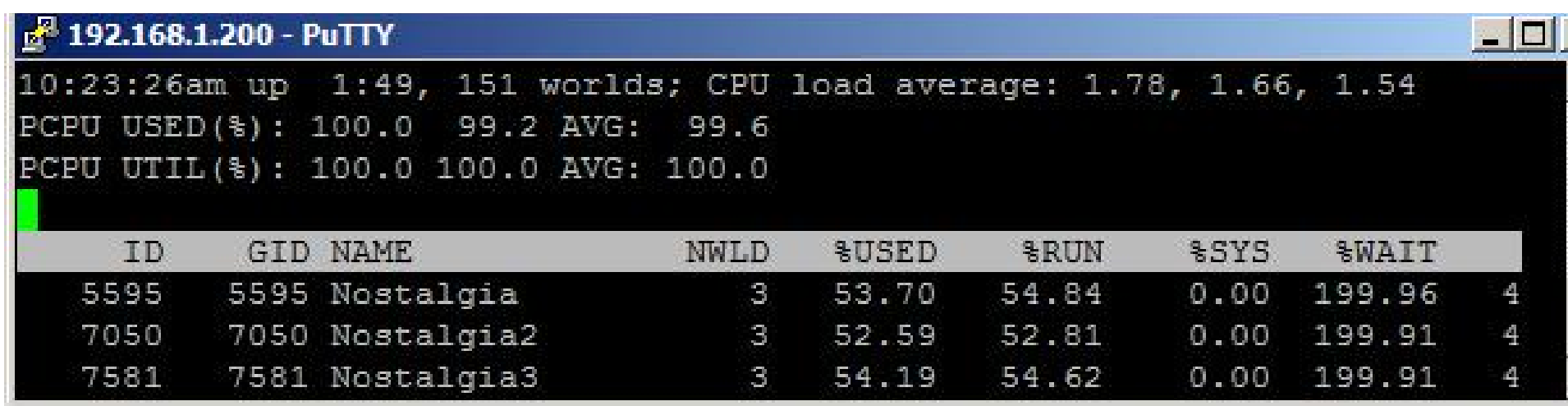
A screenshot showing the VMware context menu for a selected VM. The menu is open, displaying options like Power, Guest, Snapshot, Open Console, Edit Settings..., and Add Permission... The 'Power' option is highlighted, and a sub-menu is visible showing power-related actions and their keyboard shortcuts.

Power	▶	Power On	Ctrl+B
Guest	▶	Power Off	Ctrl+E
Snapshot	▶	Suspend	Ctrl+Z
Open Console		Reset	Ctrl+T
Edit Settings...		Shut Down Guest	Ctrl+D
Add Permission...	Ctrl+P	Restart Guest	Ctrl+R

Unsupported ESXi Console & esxtop

- ESXi, comme ESX, est basé sur un système Linux et offre une console **non officielle** et rudimentaire
- Exécutable = **/bin/busybox**
tiny versions of many UNIX utilities into a single small executable
<http://www.busybox.net/about.html>
- Démo depuis un client SSH après l'avoir autoriser dans /etc/inetd.conf
http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1003677
- esxtop = *performance monitoring tool*
*command-line tool provides a fine-grained look at how ESX Server uses **resources in real time***

esxtop



```
10:23:26am up 1:49, 151 worlds; CPU load average: 1.78, 1.66, 1.54
PCPU USED(%): 100.0 99.2 AVG: 99.6
PCPU UTIL(%): 100.0 100.0 AVG: 100.0
```

ID	GID	NAME	NWLD	%USED	%RUN	%SYS	%WAIT	
5595	5595	Nostalgia	3	53.70	54.84	0.00	199.96	4
7050	7050	Nostalgia2	3	52.59	52.81	0.00	199.91	4
7581	7581	Nostalgia3	3	54.19	54.62	0.00	199.91	4

- Heure système, *uptime*, *world* = VMkernel schedulable entity
- **PCPU USED(%)** (physical CPU) = charge en %
- **%USED** = % of physical CPU used by the resource (process, VM)
- **%IDLE** = % of time the resource was idle
- <http://communities.vmware.com/docs/DOC-11812>
- <http://www.linux-tutorial.info/modules.php?name=MContent&pageid=84>

Labo Level 1 §4-5 : Virtual Machine (VM) – 50 min

- §4 **Gérer des machines virtuelles Nostalgia**
 - §4.1 Créer une VM → comprendre les paramètres à entrer
 - §4.2 Installer un *guest OS* → principe pour gagner du temps
 - §4.4 *Unsupported ESXi Console & esxtop*
Mesurer les charges CPU
 - §4.5 Fonctionnement **sans réservation des ressources**
 - §4.6 Fonctionnement **avec réservation des ressources**
 - §4.7 Fonctionnement **avec niveaux de priorité des ressources**
 - §4.8 Fonctionnement **avec limitation des ressources**
- §5 **Gérer une machine virtuelle Ubuntu Server**
 - §5.1 Installer une *appliance* d'Ubuntu Server
 - §5.3 *Snapshot*

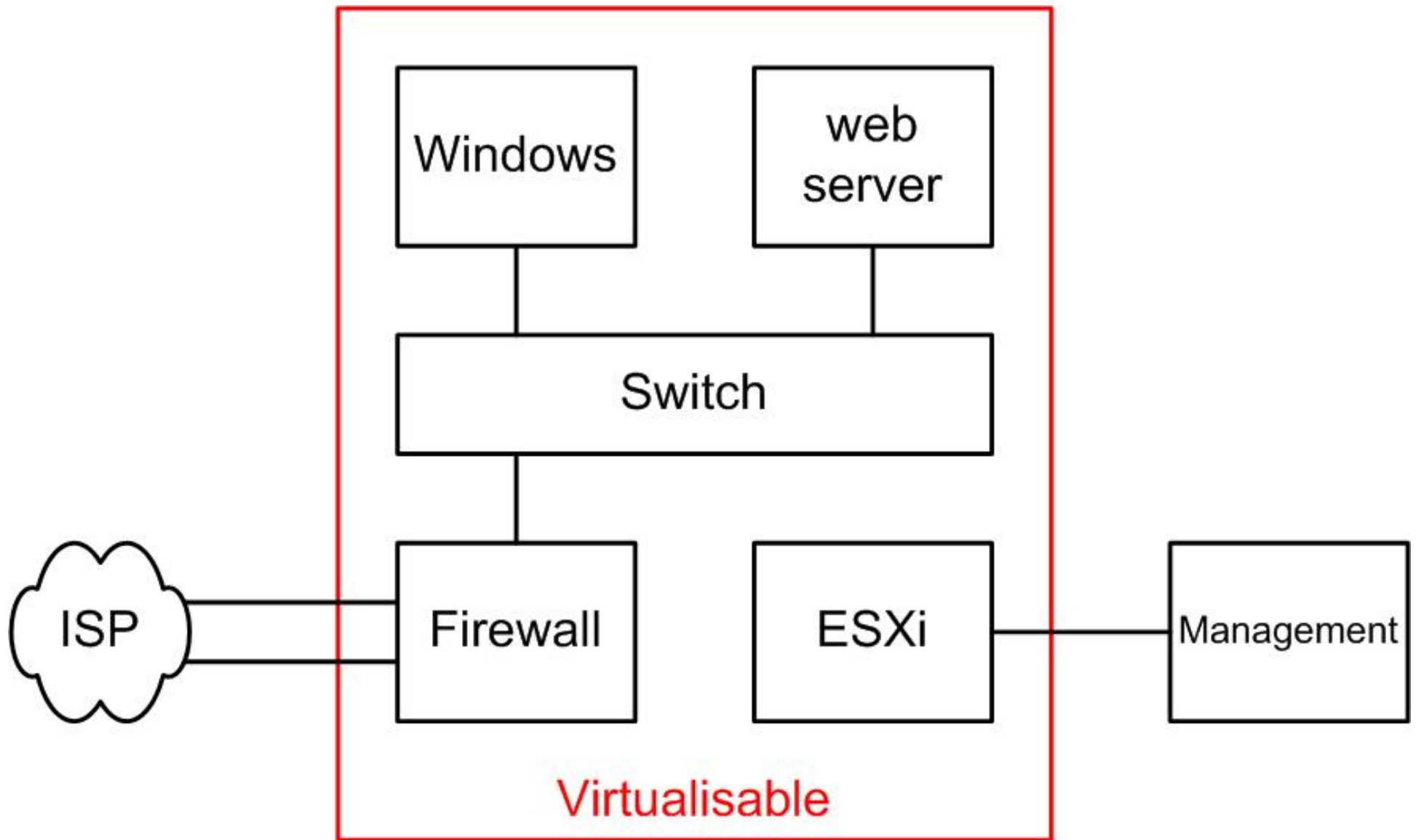
Virtualization level 1

3 Réseau & Compatibilité

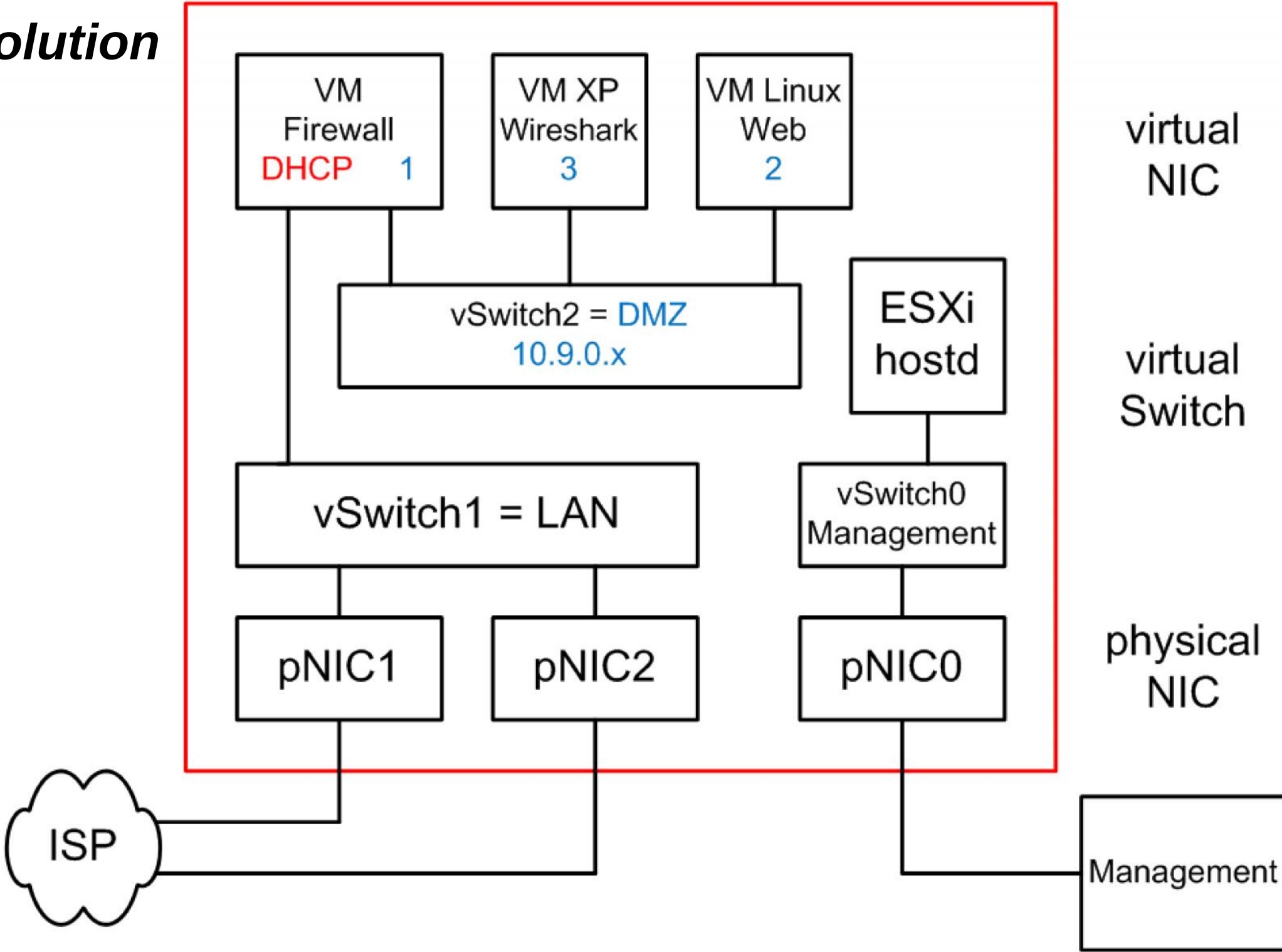
Synthèse du labo précédent

- Quels sont les principaux paramètres à contrôler lors de la création d'une VM ?
- Quelle est la charge CPU sans VM ?
 - avec 1 VM Nostalgia ?
 - avec 2 VM Nostalgia ?
 - avec 3 VM Nostalgia ?
- Identifier les cas de congestion dans les scénarios précédents

Scénario : virtualisation de la DMZ



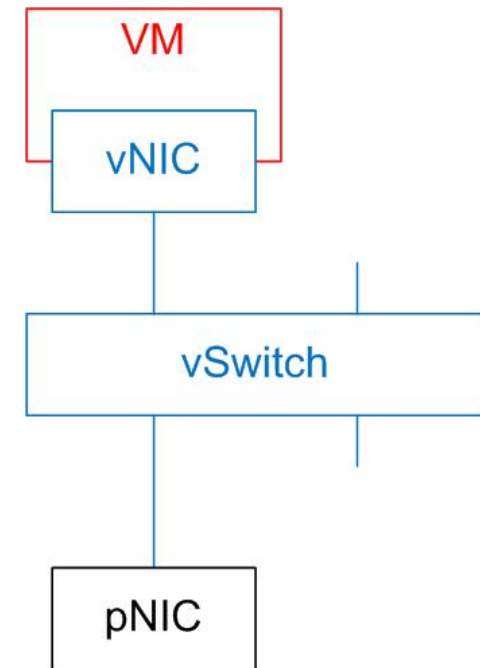
Solution



Illustration

- VM Linux Web = serveur web Ubuntu
- VM XP Wireshark = outil Wireshark sous XP (*troubleshooting*)
- VM Firewall = *firewall* sous Ubuntu
- vSwitch2 = commutateur ethernet de la DMZ
- vSwitch1 = commutateur LAN côté internet
- pNIC0 = interface physique Ethernet de management
- pNIC1 & pNIC2 = 2 interfaces physiques Ethernet pour les flux applicatifs
- Cartes Intel supportées par ESXi

<http://www.intel.com/assets/pdf/general/252454.pdf>



- pNIC (*physical NIC*)

Interface physique **compatible ESXi**

Nécessaire pour l'administration avec vSphere, SSH

Nécessaire pour les flux applicatifs

Offrir de la redondance

→ disposer d'un nombre suffisant (3) de ports physiques de 1 Gbit/s

- vNIC (*virtual NIC*)

Interface ethernet virtualisée utilisée par une VM

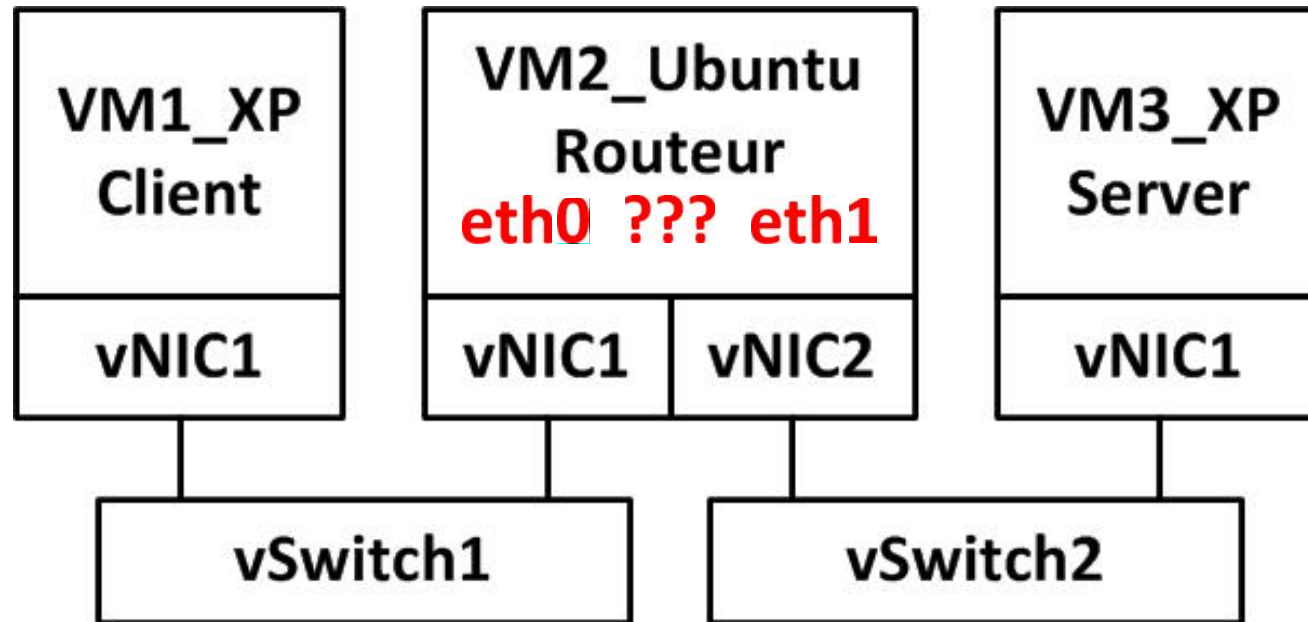
Les paramètres traditionnels (*speed, duplex*) n'ont plus de sens suite à la virtualisation (car l'échange n'est pas sériel)

Une VM (*firewall*) peut utiliser plusieurs vNIC → *next slide*

Difficulté liée à la virtualisation

- Problématique (projet de semestre 2011) :

Client (VM1) – Routeur (VM2) – Serveur (VM3)



- Difficulté liée à la virtualisation (idem dans **Labo §6.3**)

Ubuntu → eth0, eth1

ESXi → vNIC1-vSwitch1, vNIC2-vSwitch2

vSwitch

- Commutateur ethernet virtualisé (par VMkernel)
Consomme des ressources !
- Config. globale au niveau vSwitch (pas au niveau port)

- Paramètres



- *Security*

Promiscuous Mode

MAC Address Changes

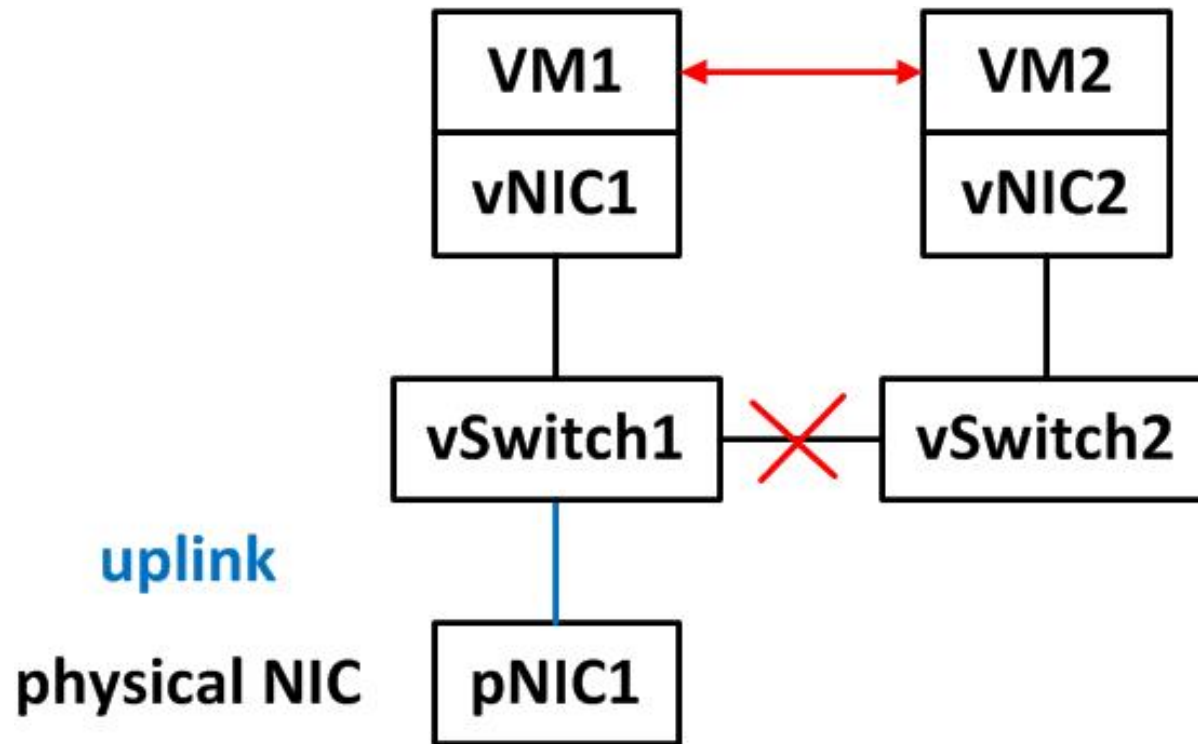
Forged Transmits

conserver les valeurs par défaut
mode switch / hub → [Lab](#)

autorise/**interdit** qu'une VM
modifie son adresse ethernet

autorise/**interdit** qu'une VM
émette une trame avec une
adresse ethernet source
différente de celle autorisée

Terminologie, limite et risque



- **Uplink** = vSwitch connecté à un port physique Ethernet
- Pas de connexion possible entre vSwitch
- Echange possible entre VM sans passer par vSwitch !!!

Contrôler **Disable VM to VM communication through VMCI**
(disabled by default)

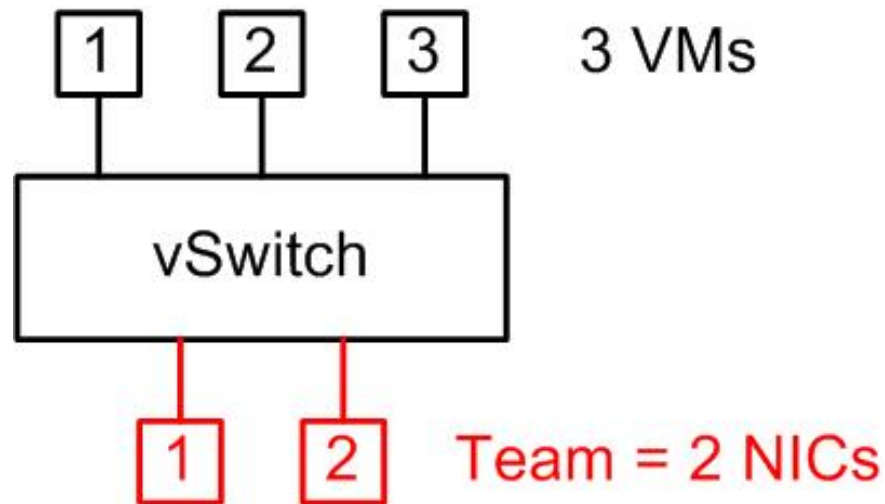
vSwitch

- *Traffic Shaping*

possibilité de lisser le flux en définissant
une valeur moyenne → *Average Bandwidth*
une valeur de crête → *Peak Bandwidth*
une taille → *Burst Size*

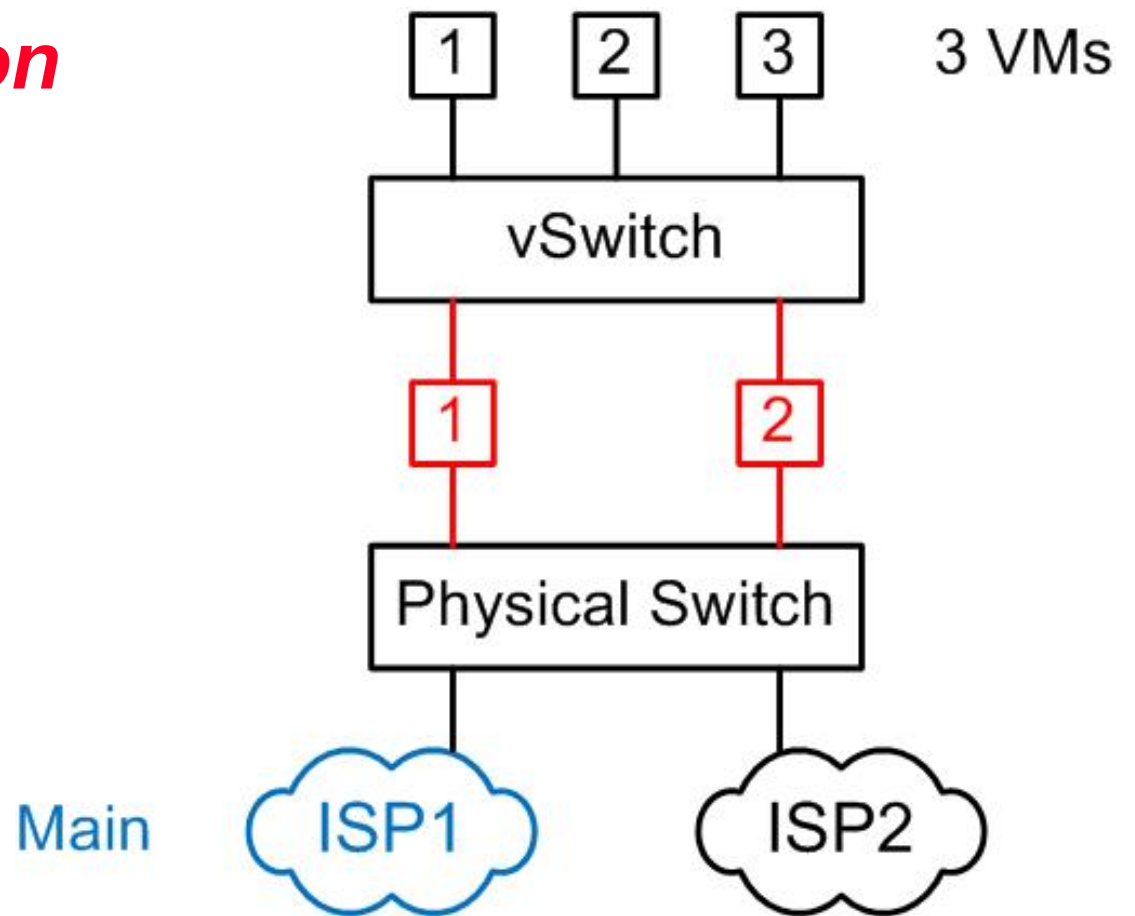
- *NIC Teaming*

réunion de plusieurs liens avec les ports
physiques



Network Failover Detection

- Un seul lien actif à la fois
- ISP1 principal et ISP2 de secours
- Config du labo

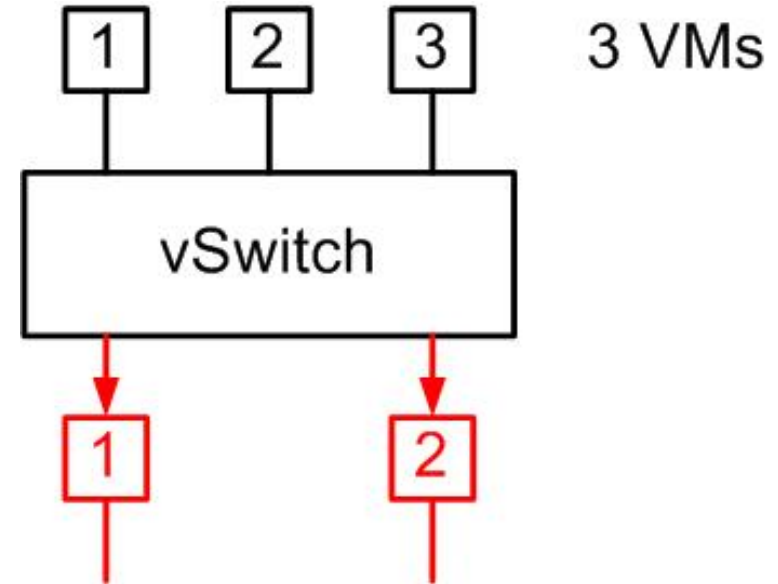


Load Balancing:	Use explicit failover order
Network Failover Detection:	Link Status only
Notify Switches:	Yes
Failback:	Yes

Détection
Signaler au switch
Retour lien principal

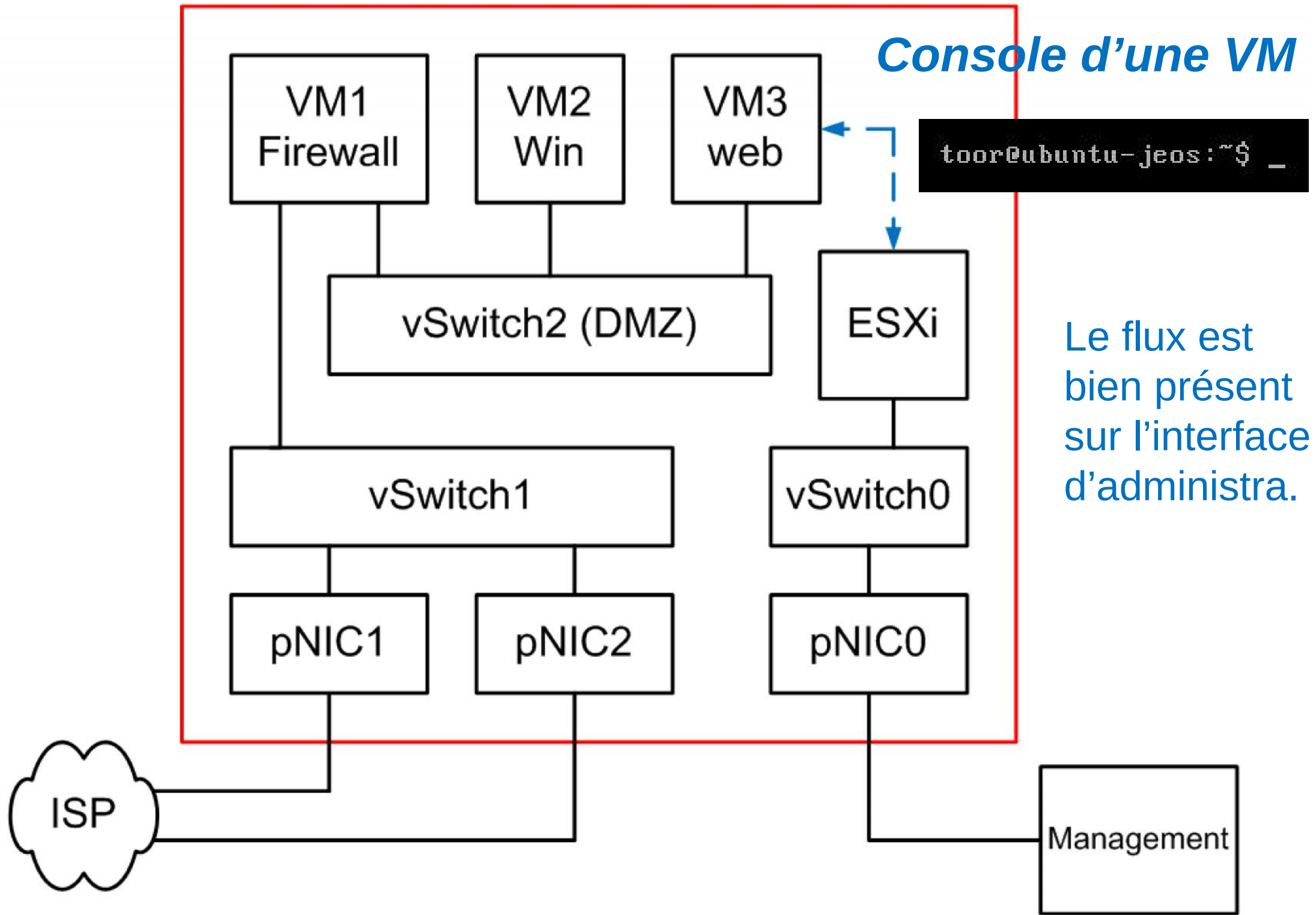
Load Balancing (agrégation)

- les N liens fonctionnent simultanément
pour le trafic sortant !



Variantes :

- *Route based on the originating virtual port ID*
Chemin en fonction du numéro de port → [mode par défaut](#)
- *Route based on source MAC hash*
Chemin en fonction de l'adresse Ethernet
- *Route based on ip hash*
Chemin en fonction des adresses IP → [Labo](#)



Outil Converter

- Outil gratuit compatible Windows
- Source = machine physique / format VMware / Ghost / Virtual PC / ...
- Destination = produits VMware (ESXi, Player, ...)

- Utilité

Copier (cloner) une VM de ESXi_1 vers ESXi_2

Produire une *appliance* (format ovf) à partir d'une VM (vmdk, vmx, ...)

Installer une *appliance* sur ESXi

Produire une VM depuis une machine physique Windows (*next slide*)

Hot cloning via snapshot / Cold cloning via Live CD

- <http://vmware-land.com/Converter.html>
- <http://www.vmware.com/fr/products/converter/>

Physical To Virtual (P2V)

- Opération facile (à chaud) sous Windows avec outil Converter → §7
- Linux → *Cold cloning* via *Live CD* → §2 (copie bit à bit avec dd)
Procédure spécifique pour chaque distribution et **config. matérielle**
§3 : Ubuntu Desktop 8.04 LTS 32 bit (www.tdeig.ch)
§6 : Debian 4.0 32 bit (serveur DNS de tdeig.ch)
- Eviter si possible P2V (problèmes de pilote, GUID, ...)
- <http://www.tdeig.ch/vmware/P2V.pdf> + impression

Labo level 1 §6 : Réseaux

- §6.1 Créer 2 *virtual Switches* LAN & DMZ
 - §6.2 Déplacer la VM Ubuntu (web server) sur le *switch* DMZ
 - §6.3 Installer une *appliance firewall*
 - §6.4 Configurer la nouvelle route
 - §6.5 Analyser (Wireshark) l'échange des paquets sur un *virtual switch* → *troubleshooting*
 - §6.6 Configurer le *virtual switch* DMZ en mode *hub*
-

Semaine prochaine

- §7.1 Configurer 2 interfaces en mode *failover*
- §7.2 Configurer 2 interfaces en mode *load balancing*

Virtualization level 1

4 Sécurité & *Best Practices*

Synthèse du labo précédent

- Principales difficultés liées à la virtualisation → risques !!!

§6.1 : Par défaut, les VMs sont connectés à vSwitch0 = management

§6.3 : Difficulté d'associer le port (eth0) de la VM avec son vNIC

§6.4 : Configurer une route

- Autres difficultés ?

Virtualisation & Sécurité

Cette partie présente les résultats du projet Visag (16 mois terminé en janvier 2012) financé par HESSO

- **Rapport** → <http://www.tdeig.ch/visag/Security.pdf>

Principaux thèmes d'étude :

- Comment aborder la sécurité d'une architecture virtualisée ?
- Quels sont les principaux risques liés à la virtualisation ?
- Peut-on réutiliser l'expérience du monde physique ?
- Quelles sont les bonnes pratiques ?
- La virtualisation peut-elle augmenter la sécurité d'un service ?
- Quels sont les avantages et les inconvénients du produit gratuit VMware ESXi 4.0 ?

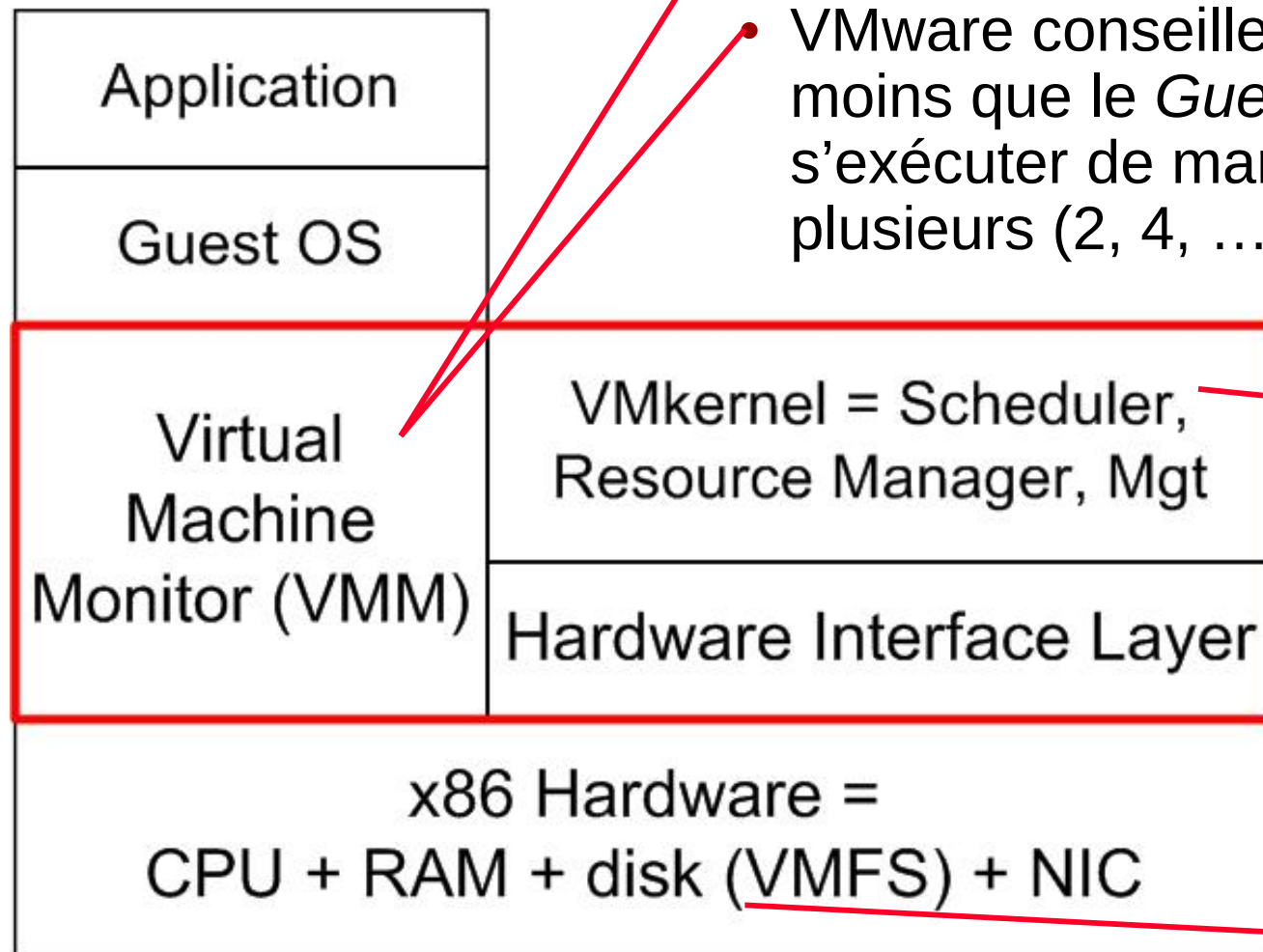
Risques théoriques & réels

- Héberger plusieurs VMs sur un matériel partagé
La couche VMM (hyperviseur) doit garantir l'isolation entre les VMs
- Souplesse d'exploitation : ajouter, tester, supprimer, sauvegarder, ..
Risques réels de laisser une VM de test (rogue VM)
- Gartner prédisait l'installation de 4 Mio de machines virtuelles en 2009, mais que 60% sera moins bien sécurisée qu'avec une architecture traditionnelle
- Les implémentations (IBM-1972, VMware-2001, KVM-2007) de la virtualisation ont une excellente réputation

Pourquoi les entreprises choisissent la virtualisation ?

- Source = Lancelot Institute – Advanced VMware Security – oct 2010
 - 80% Consolidation
 - 60% Disaster recovery (plan de continuité)**
 - 50% Agility of provisioning ressources to users
 - 50% Business agility
 - 10% Competitive advantage
- Cloud Computing
 - ... il y a plus de 5 ans que j'utilise un Cloud Computing sans que j'en susse rien ...
 - ... il y a plus de quarante ans que je dis de la prose sans que j'en susse rien ...*
 - M. Jourdain dans le Bourgeois gentilhomme de Molière*

Architecture ESXi



Dans notre cas, **VMM** peut émuler **2 vCPU** (*virtual CPU*) → car 2 pCPU

VMware conseille de laisser ce paramètre à 1 à moins que le *Guest OS* soit capable de s'exécuter de manière concurrente sur plusieurs (2, 4, ...) CPU physiques

VMkernel 64bit basé sur noyau Linux

Management (web) via vSphere

VMFS (*Virtual Machine File System*)

5 principaux domaines d'un système virtualisé (§2.2)

- Réseau

séparation, défense périmétrique

- Système

comptes actifs, services, logs

- Management

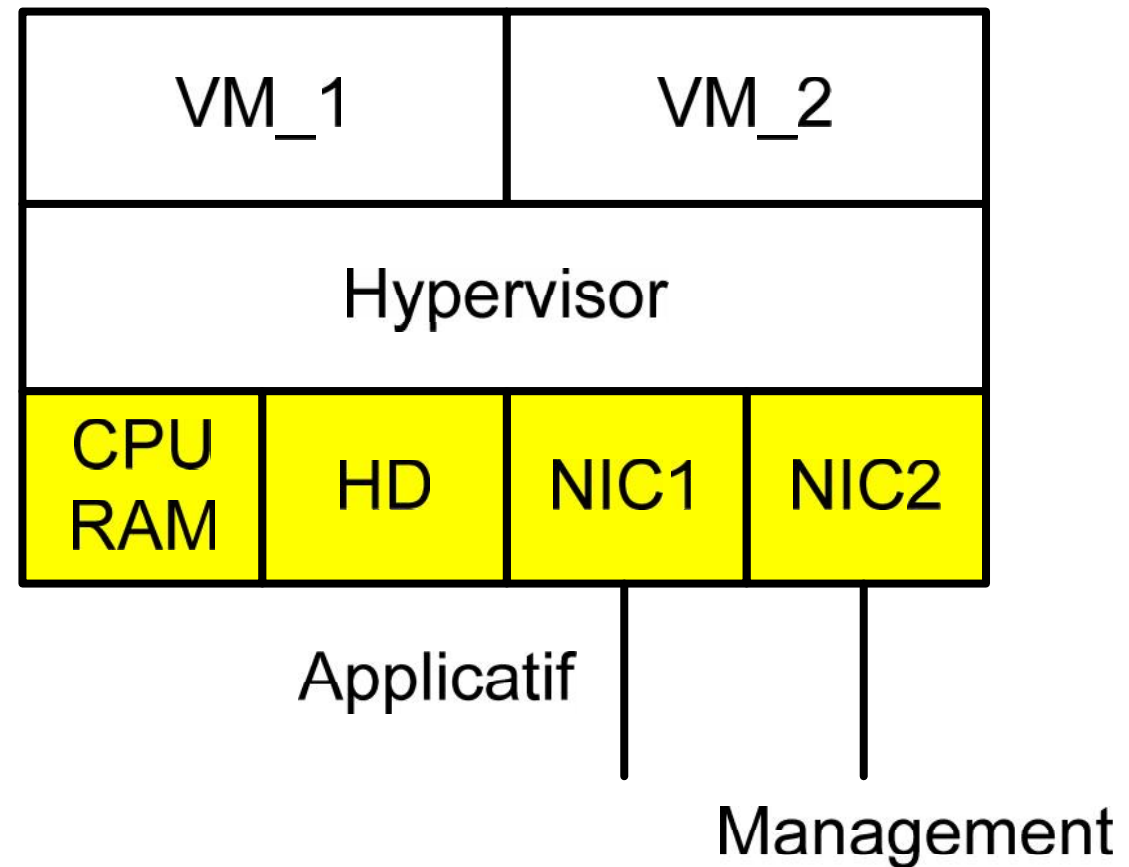
accès distant, backup, ...

- Virtualisation

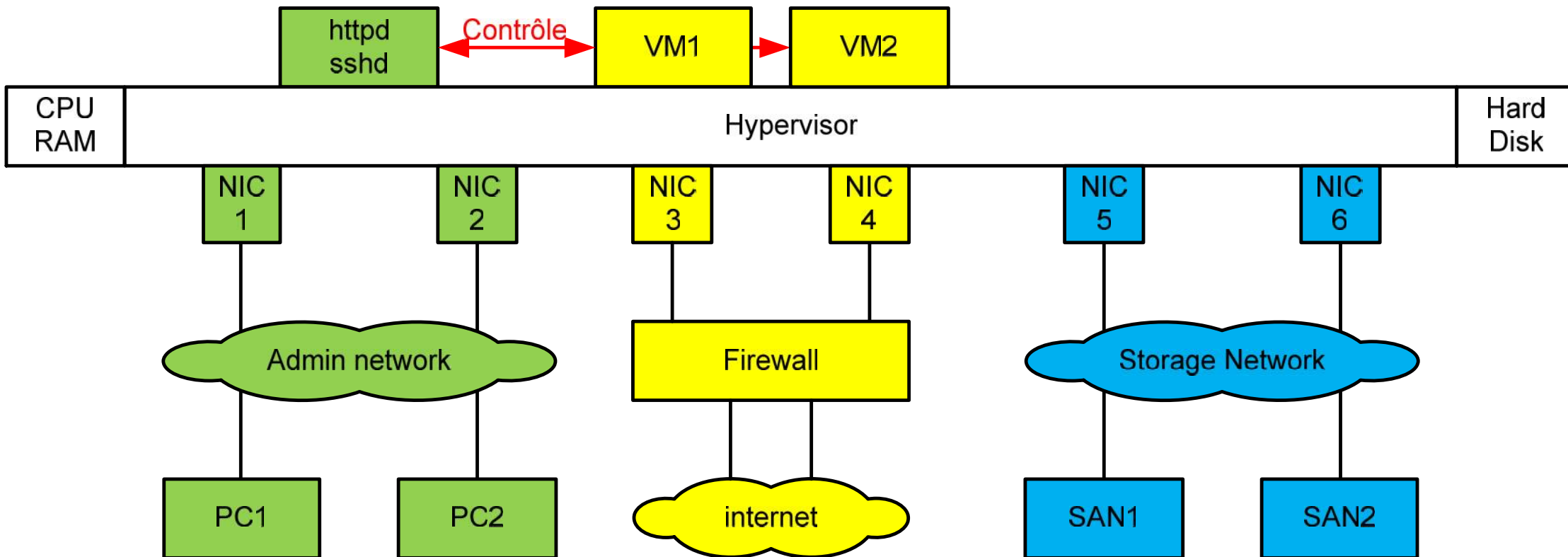
cloisonnement des VMs, ...

- Applicatif

flux, ...



Redondance et séparations physiques (§3)



- Administration via un réseau (des hommes) de confiance
 - *Storage Area Network (SAN)* digne de confiance
 - Risques au niveau des VMs (services) offerts sur internet
- Utiliser les bonnes pratiques sécuritaires du monde physique !

Performances & disponibilité

Pour le responsable technique :

- Connaître les besoins (cahier des charges, SLA, ...)
- Trouver le meilleur compromis entre coût et performances
- Surveiller certains indicateurs : charge CPUs, occupation RAM, accès disque, charge réseaux
- Choix du système de stockage, prise en compte des exigences de sécurité
- Redondance (haute disponibilité)

Pour l'utilisateur :

- Serveur virtualisé doit présenter des caractéristiques semblables au serveur physique
- L'utilisateur ne peut pas savoir s'il utilise un service virtualisé ou s'il accède à un serveur physique

Nom	CPU	RAM	Disque	Température CPU	Température MB
• Fedora14 - 2	14.32	38.48	0.00	0.00	0.00
VM-Fedora14-2	5.86	0/0			
VM-Fedora14-1	0.00	0/0			
• Fedora14	43.08	57.55	0.00	38.00	34.00
5-Fedora-14-64bits	0.00	2047/2047			
4-Fedora-14-64bits	0.00	512/512			
3-Fedora-14-64bits	11.25	512/512			
2-Fedora-14-64bits	11.68	1024/1024			
1-Fedora-14-64bits	0.00	1024/1024			

Lionel Schaub : [Enoncé](#) [Résumé](#) [Mémoire](#) [Présentation](#) [Documents](#)



Really Risks

- Trust mesh
 - Single point of failure
 - High value target
 - "New" layer (VMM)
 - **Misconfiguration**
-
- Source = Lancelot Institute – Advanced VMware Security – oct 2010

Risques

- **Disponibilité** - §2.3.1

Réservation des ressources CPU, RAM, ... Quality of Service

Denial of Service

Une VM peut monopoliser toutes les ressources

→ *Single point of failure, High value target*

Mêmes solutions que pour le monde physique

+ limiter le nb de vCPU, ajuster des limites de charge CPU, ...

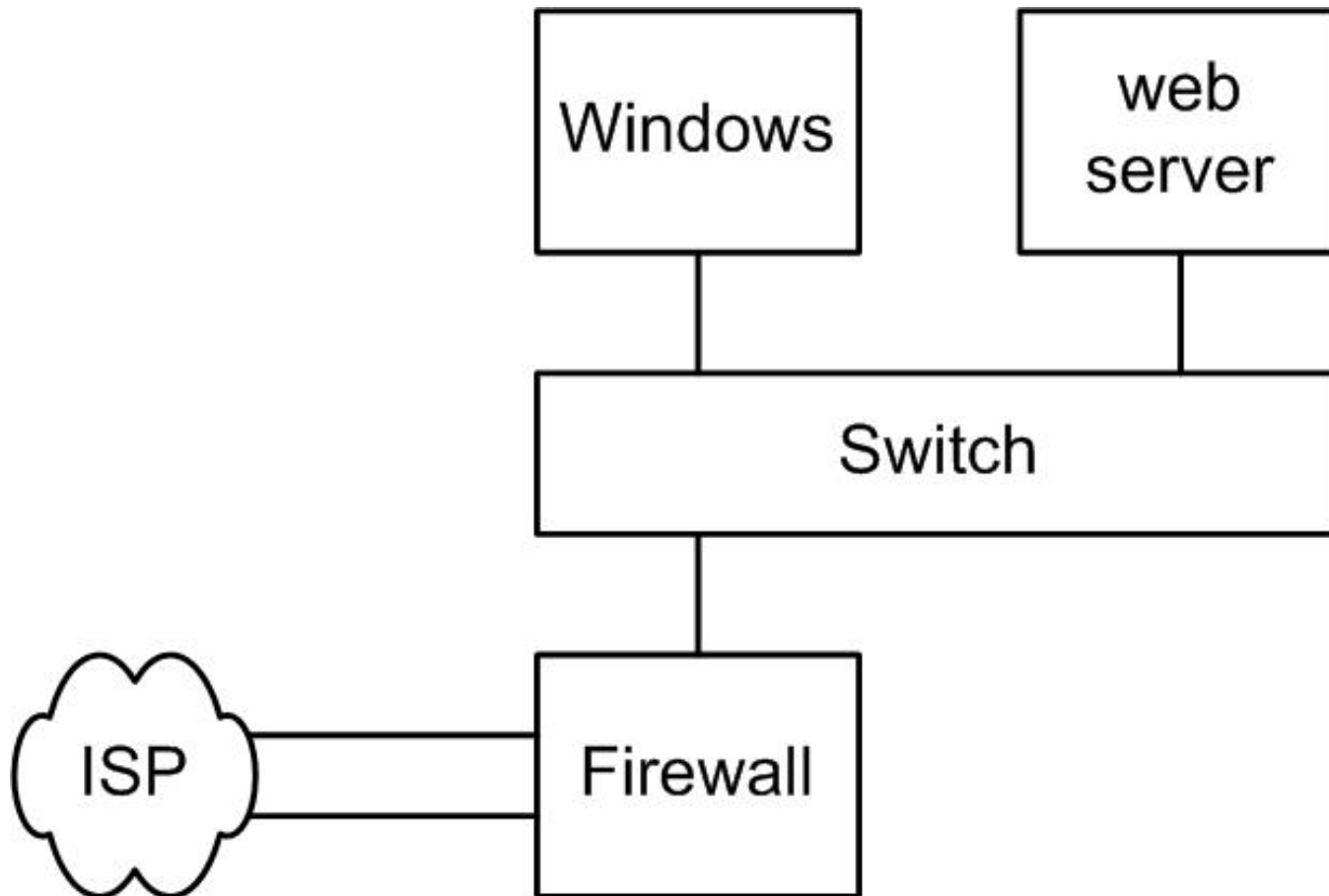
- **Intégrité** - §2.3.2

Confiance dans la couche de virtualisation, les OS, les applic., ...

- **Confidentialité** - §2.3.3

Les méthodes classiques (séparation physique, échanges sécurisés avec SSL/TLS, IPSec, ...) restent valables

Architecture physique



Spécialistes

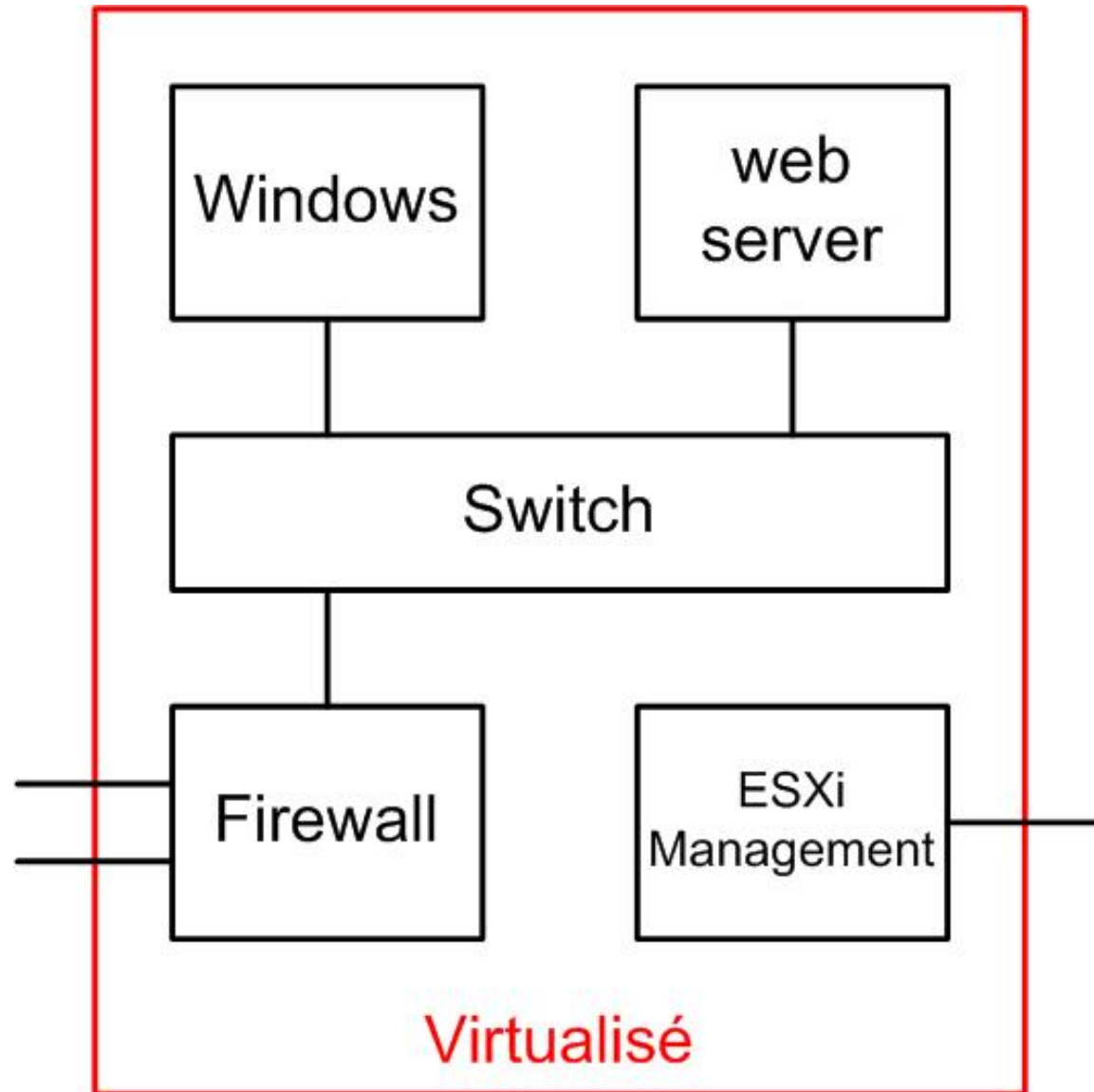
Serveurs,
Contexte applicatif

Réseau

Sécurité

Taille de l'entreprise

Architecture virtualisée



Risques

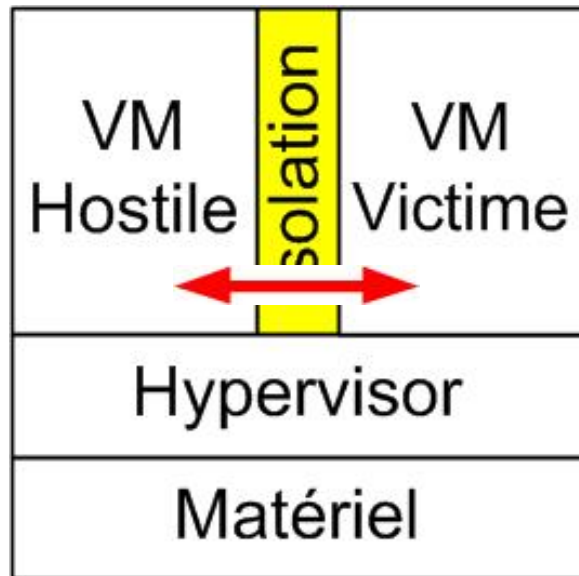
- *Misconfiguration*
- Eviter que chaque spécialiste s'occupe de problématique qu'il ne maîtrise pas
- Comprendre la virtualisation
- Très (trop) facile d'ajouter une VM de test, ...

Taille de l'entreprise

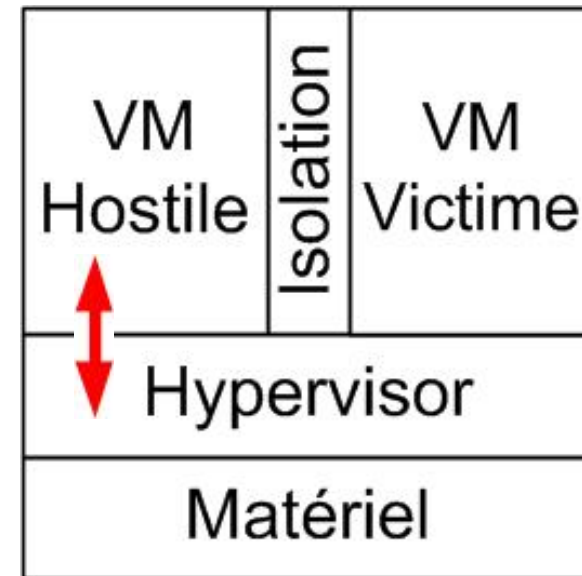
Risques au niveau de l'hyperviseur

- "New" layer (VMM) → Le cloisonnement entre VM est-il garanti ?
- Hyperviseurs (ESXi, Linux-KVM) sont des composants éprouvés

Pas d'attaque publiée sur **VM Hopping** ou **VM Escape**



- **VM Hopping** : VM Hostile tente d'accéder à l'espace RAM de la VM Victime

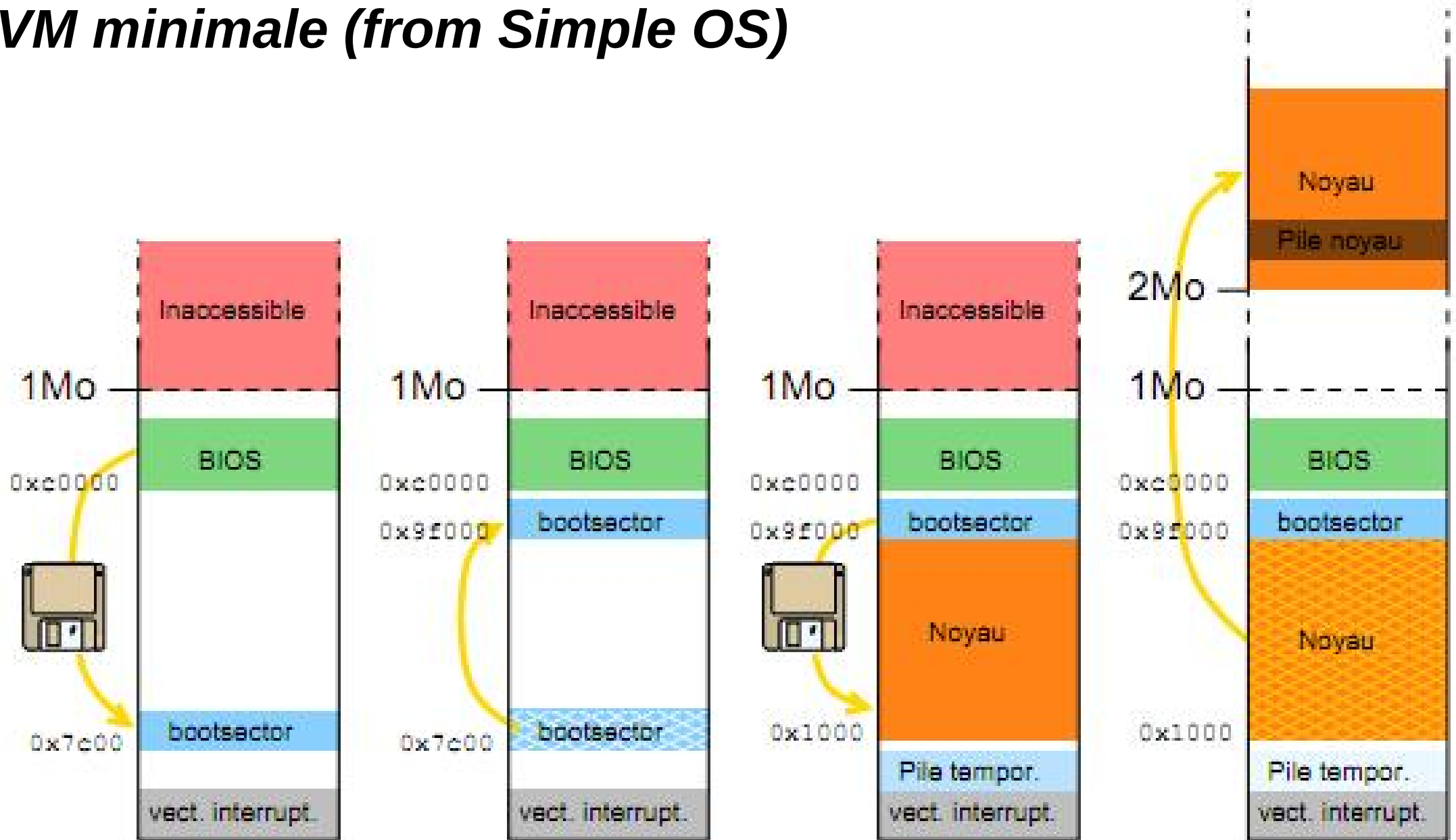


- **VM Escape** : VM Hostile tente de prendre le contrôle de l'hyperviseur

Méthodologie pour tester le cloisonnement mémoire

- VM minimale (sans système d'exploitation)
- Tentative d'accès à la mémoire RAM (no OS memory management)
VM accède en fait à vRAM et non directement à pRAM
Seul l'hyperviseur peut accéder au matériel
- Tentative de sortir de l'espace RAM alloué (réservé)
- Accès au niveau du CPU (ring)

VM minimale (from Simple OS)



- <http://sos.enix.org/wiki-fr/upload/SOSDownload/sos-texte-art1.pdf>



Nous sommes dans une machine virtuelle VMware
Fréquence CPU selon VMware: 3000 MHz
Mémoire RAM selon VMware: 64 Mo
CPUID: Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GHz

Lecture de 2Mo de la memoire au demarrage

Les meme 2Mo apres ecriture



Légende: _ 0-20%, ▤ 20-40%, ▥ 40-60%, ▧ 60-80%, ■ 80-100%

Lecture des bits 12 et 13 de EFLAGS pour connaitre notre niveau de privilèges.
Valeur hexadécimal de EFLAGS: 200206
Nous sommes en RING0

Test d'ecriture dans les 72 premiers Mo de la memoire:
Adresse: 2195456 - Motif écrit: A - Motif lu:
Adresse: 67076096 - Motif écrit: A - Motif lu: A
Adresse: 75464704 - Motif écrit: A - Motif lu:

Accès direct à la mémoire

- L'espace RAM est mis à zéro lors du démarrage de la VM

```
Lecture de 2Mo de la memoire au demarrage
```

- Il n'est pas possible de sortir de l'espace RAM alloué

```
Test d'ecriture dans les 72 premiers Mo de la memoire:  
Adresse: 2195456 - Motif écrit: A - Motif lu: A  
Adresse: 67076096 - Motif écrit: A - Motif lu: A  
Adresse: 75464704 - Motif écrit: A - Motif lu: A
```

- Dans quel niveau de protection (ring) se trouve le CPU ?

```
Lecture des bits 12 et 13 de EFLAGS pour connaitre notre  
Valeur hexadécimal de EFLAGS: 200206  
Nous sommes en RING0
```

Can we trust ESXi ?

- Peut-on valider scientifiquement la bonne sécurité de l'hyperviseur ESXi ?

Non, sans code source, l'utilisateur doit faire confiance à VMware (comme à MS) ou tenter une analyse de type reverse engineering

- Peut-on considérer le cloisonnement entre VMs comme sûr ?

Oui, aucun exploit disponible sur internet confirmé par nos tentatives de sortie de l'espace alloué

- Qui gère ce cloisonnement ?

VMM logiciel dans les systèmes 32 bit (Binary Translation)

Extension VT-Intel pour les systèmes 64 bit

http://www.vmware.com/files/pdf/perf-vsphere-monitor_modes.pdf

Conclusion : principaux risques

- Principaux risques sont **humains**

Mauvaise configuration, VMs fantômes ?, ...

Documentation à jour

Connaître le profil (min – moy – max) des charges → **références !**

- Difficulté d'agir lors d'un problème (affirmation VMware)

La couche de virtualisation complexifie le système

Outils de supervision, formation, training (mise en condition, ...)

- Peut-on auditer une architecture virtualisée selon les principes utilisés pour le monde physique ?

Oui l'hyperviseur n'est qu'un composant du système d'information

Posséder un **référentiel** basé sur les bonnes pratiques

Conclusion : Bonnes pratiques

1. Documentation & simplicité
2. Gestion du changement
3. Sécurité physique
4. Administration
5. Systèmes : profils
6. Redondance
7. Réseau
8. Applicatifs
9. Stockage & sauvegarde
10. Supervision

Conclusion : impact positif sur la sécurité ?

- La virtualisation peut-elle augmenter la sécurité ?
- Sous Windows, un processus hérite des droits (autorisations NTFS) du compte actif (Système, Service réseau, ..., Admin, User)
- Tous les processus démarrés avec le même compte X peuvent interférer entre eux.
- Attaques de type injection DLL
- L'isolation entre VMs offre un cloisonnement sûr

Synthèse du labo précédent

- Quelle est la principale différence entre Failover et Load balancing ?
- Vous avez contrôlé l'effet du Failover avec esxtop
Proposer une autre méthode de mesure externe au PC_ESXi
- Dans quel cas ces modes peuvent donner des résultats différents ?

Load Balancing:

Route based on the originating virtual port ID

Load Balancing:

Route based on source MAC hash

Load Balancing:

Route based on ip hash

General

Manufacturer:	Gigabyte Technology...
Model:	G33M-S2L
CPU Cores:	2 CPUs x 2.999 GHz
Processor Type:	Intel(R) Core(TM)2 D...
License:	ESXi 4 Single ServerLi...
Processor Sockets:	1
Cores per Socket:	2
Logical Processors:	2
Hyperthreading:	Inactive
Number of NICs:	2
State:	Connected
Virtual Machines and VMotion Enabled:	5
VMware EVC Mode:	N/A
FaultTolerance Enabled:	N/A
Active Tasks:	
Host Profile:	N/A
Profile Compliance:	 N/A

Resources

CPU usage 209 MHz	Capacity
	2 x 2.999 GHz
Memory usage 3594.00 MB	Capacity
	4081.62 MB

Datastore	Capacity	Free	Last Upd
 datastore1	293.00 G	213.30 G	09.10.20

Network	Type	Nu
 VM Network	Standard switch network	1
 DMZ	Standard switch network	4

Host Management

[Manage this host through VMware vCenter.](#)

Name	State	Provisioned ...	Used Space	Host CPU - MHz	Host Mem - MB	Guest Mem - ...
 www.tdeig.ch	Powered On	200.50 GB	22.34 GB	3 	533 	10 
 Windows Server 2008 (CA)	Powered On	44.01 GB	23.02 GB	68 	1366 	6 
 ubuntu-apache-ssl	Powered On	17.63 GB	9.30 GB	62 	467 	3 
 DNS	Powered On	8.51 GB	8.01 GB	55 	461 	1 
 vSphere Management Assi...	Powered Off	5.50 GB	5.50 GB	0	0	

- 10.2.1.10
- DNS**
- ubuntu-apache-ssl
- vSphere Management
- Windows Server 2008
- www.tdeig.ch

DNS

Getting Started

Summary

Resource Allocation

Performance

Events

Console

Permissions

General

Guest OS: Other 2.6x Linux (32-bit)
 VM Version: 7
 CPU: 1 vCPU
 Memory: 512 MB
 Memory Overhead: 444.18 MB
 VMware Tools: OK
 IP Addresses: 10.10.10.20 [View all](#)
 DNS Name: dns-publique
 State: Powered On
 Host: [localhost.localdomain](#)
 Active Tasks:

Commands

Resources

Consumed Host **38 MHz**
 Consumed Host **461.00 MB**
 Active Guest **7.00 MB**

[Refresh Storage Usage](#)

Provisioned **8.51 GB**
 Not-shared Storage: **8.01 GB**
 Used Storage: **8.01 GB**

Datastore	Capacity	Free	Last Upd
 datastore1	293.00 G	213.30 G	09.10.20

Network	Type
 DMZ	Standard switch network