German Unix User Group

Einführung in XEN 3.x

8. September 2008

B.Eng. Sascha Haupt
Fachschaft Elektrotechnik und Informationstechnik
Hochschule München

nulldevice@guug.de

German Unix User Group guug

Agenda

- Überblick Virtualisierung
  - Vor-/Nachteile
  - Virtualisierungskonzepte
- XEN 3.x
  - Aufbau
  - Netzwerk
  - Storage
  - Installation/ Konfiguration
- LowCost HA
  - Heartbeat
  - DRBD
- XEN Produkte u. Hersteller
- Informationsquellen

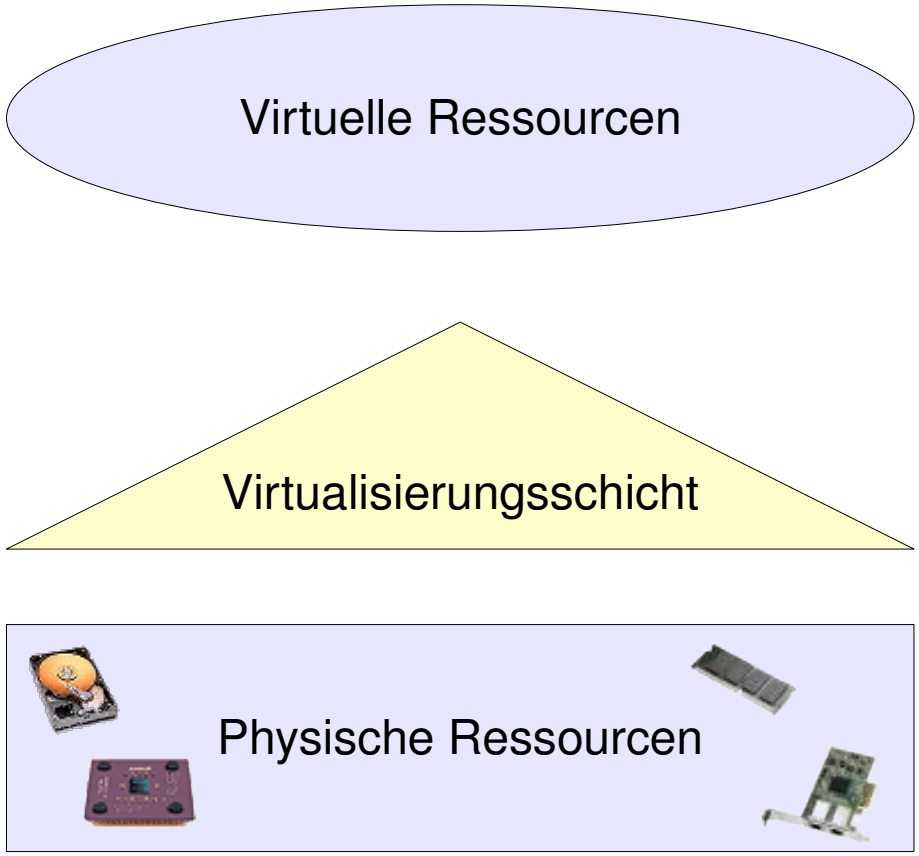
Überblick Virtualisierung

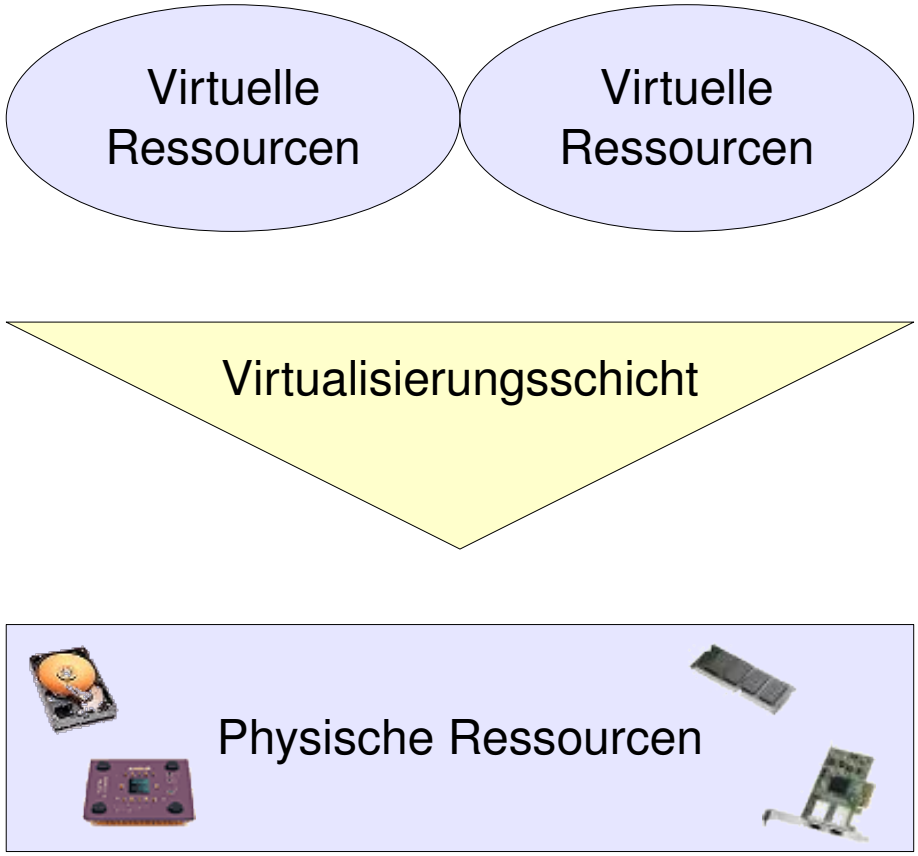
Was ist Virtualisierung?

Definition

Virtualisierung unterteilt einen Computer in logische Einheiten. Diese Einheiten repräsentieren virtuelle Rechner, welche eigenständig und parallel auf der gleichen Hardware laufen.

Was ist Virtualisierung?





Virtualisierung = Zusammenfassen und Aufteilen von Ressourcen

Vorteile der Virtualisierung

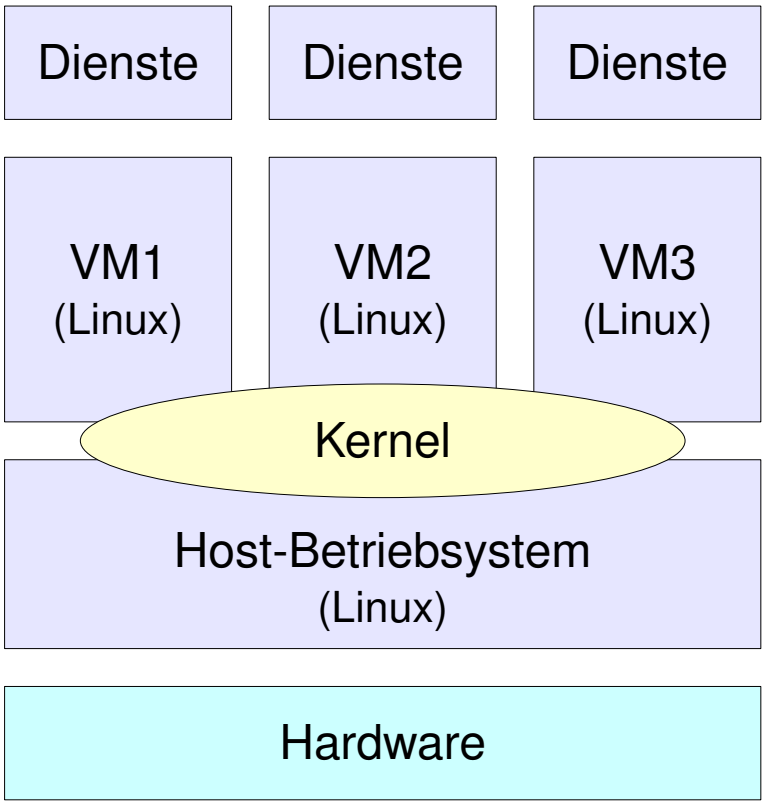
- Bessere Ausnutzung der physikalischen Ressourcen
  - Energieverbrauch
  - Hardware Kosten
  - Platz im Rechenzentrum
  - Cooling der Hardware
- Flexibilität
  - Leichtes Umziehen auf andere Hardware
  - Homogene Umgebung trotz unterschiedlicher Hardware
  - Mal „schnell“ einen (Test-)Rechner aufsetzen

Nachteile der Virtualisierung

- Mehr Know-How notwendig
- Evtl. Security Probleme
- Geringere Performance
  - Mehrere Rechner teilen die selbe Hardware
  - Bandbreiten (Netzwerk, Speicher) müssen geteilt werden
  - Kompliziertere Ressourcen-Verwaltung
- Höhere Anforderungen an Ausfallsicherheit
  - Ein Hardware-Ausfall führt zum Ausfall vieler virtueller Maschinen

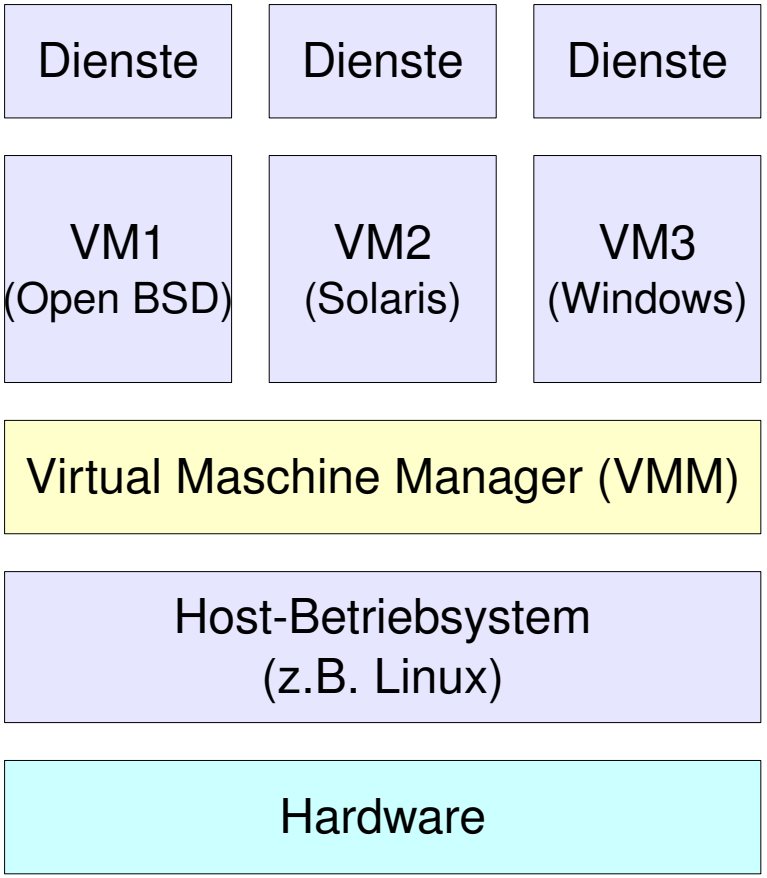
Virtualisierung auf OS-Ebene

- Es läuft nur ein Kernel „Single-Kernel-Image (SKI)“
- Host-Betriebsystem erzeugt weitere Instanzen seiner selbst: Ressourcen-Container
- Kernel schottet Container gegeneinander ab
- Bsp.: Solaris Container, BSD Jails, Linux Vserver, OpenVZ/Virtuozzo
- Pro: schnell, schlank
- Contra: nur ein OS/Kernel möglich



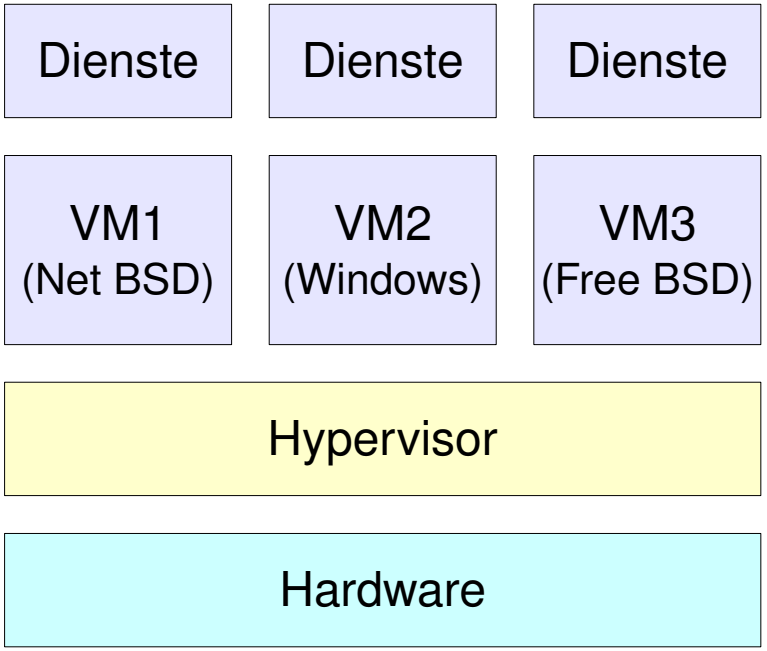
Vollständige Virtualisierung im User Space

- Software im User-Space simuliert Hardware
- Verschiedene Gast-Betriebsysteme möglich
- Beispiele:
  VMware GSX Server,
  VMware Workstation,
  Microsoft Virtual PC/Server
- Pro: keine Anpassung am Gast-Betriebssystem notwendig
- Contra: großer Overhead, schlechte Performance



Vollständige Virtualisierung mit Hypervisor / Paravirtualisierung

- Mit Anpassung (Paravirtualisierung) und/oder ohne Anpassung (Vollständige Virtualisierung) des Gast-Betriebssystems
- Hypervisor läuft direkt auf der Hardware
- Verschiedene Gast-Betriebssysteme sind möglich
- Bsp.: XEN, VMware ESX Server
- Pro: schneller als Virtual Machine Monitor im User-Space



XEN 3.x

German Unix User Group guug

About XEN

- Open Source: GNU Public Licence GPL
- Paravirtualisierung (Linux, OpenSolaris, NetBSD, FreeBSD, Plan 9)
- Vollständige Virtualisierung (beliebiges OS)
  (ab XEN 3.1 + Hardware Unterstützung: Intel-VT „Vanderpool“ oder AMD-V „Pacifica“)
- Plattformen: x86, x86-64, IA-64, PowerPC
- PAE (Physical Address Extention)
  für mehr als 4 GB Ram auf 32 Bit-Systemen
- Bootloader: Grub
- Kein ACPI, APM, Highend-Grafik, Sound in den Gästen
- Live-Migration auf andere Hardware möglich
- Aktuell: Xen 3.3

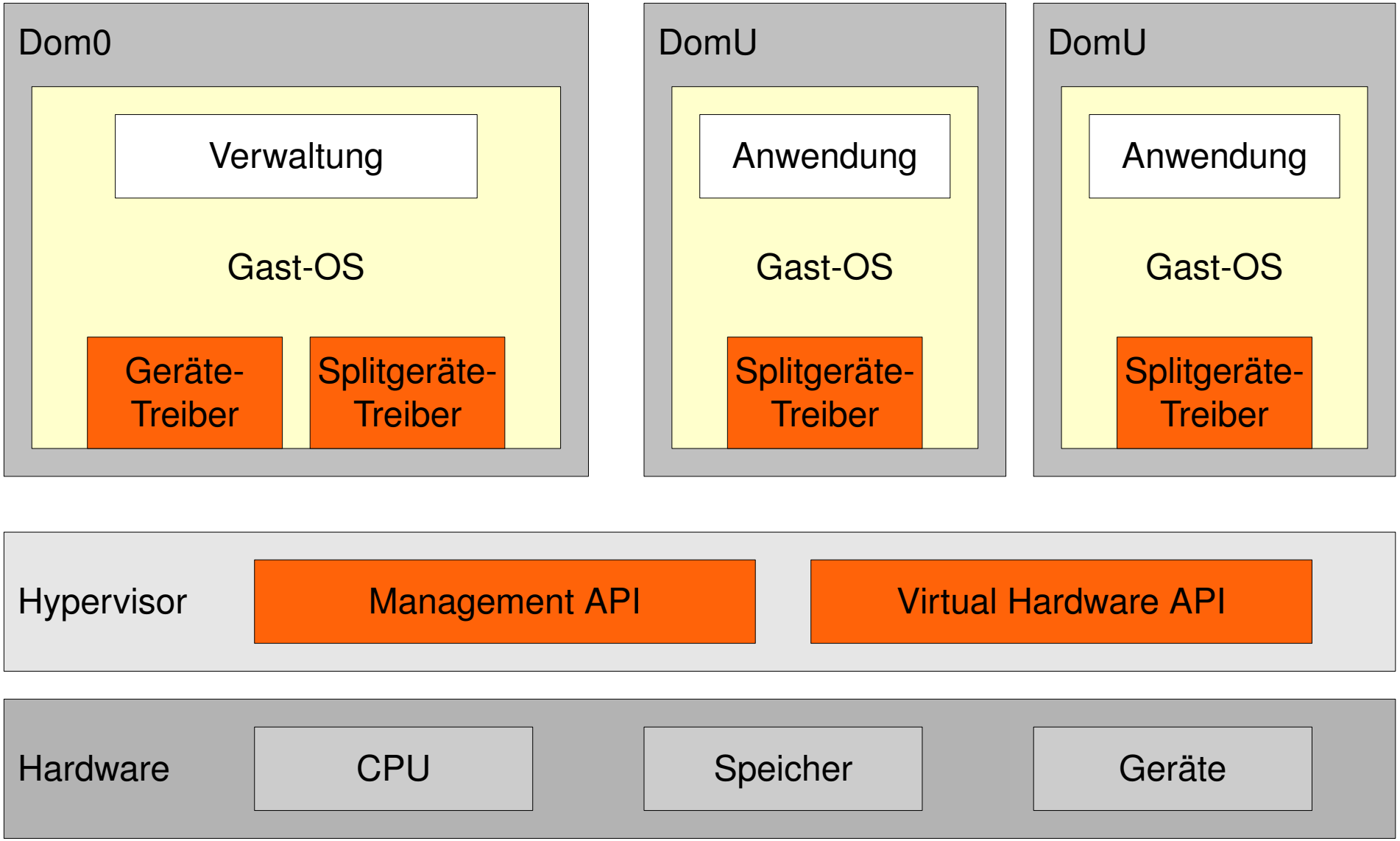
XEN History

- 2001: Ursprünge in der Systems Research Group am Computer Laboratory der Universität Cambridge (Ursprünglich Entwicklung von Grid Computing Komponenten)
- 2003: XEN wird das erste mal in „XEN and the Art of Virtualization“ öffentlich beschrieben
- Dezember 2004: Gründung der Firma XenSource
- August 2007: XenSource wird für 500 Millionen US-Dollar von Citrix aufgekauft

Architektur (1)

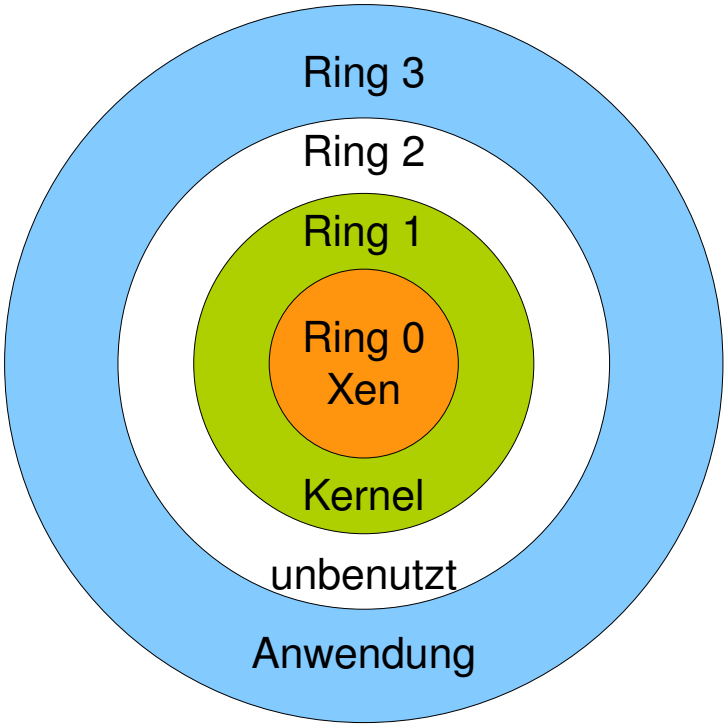
- Aufteilung in Hypervisor und Domänen
- Hypervisor zwischen Hardware und Domänen
  - Abstraktionsebene
  - Verwaltungs-API: Verwaltung des Gesamtsystems
  - Virtual Hardware API: Kapselung von Hardware in virtuelle Geräte
  - Prinzip: Möglichst viel Funktion in die Gäste verlagern
- Dom0: Verwaltungs Domäne
  - Wird nach Hypervisor gestartet
  - Verwaltung und Kontrolle der DomU-Gäste
  - Erhöhte Rechte
- DomU: Gast Domänen
  - „U“ für *unprivileged*
  - Eingeschränkte Rechte
  - Mehrere Instanzen
  - Die eigentlichen virtuellen Maschinen

Architektur (2)



Virtualisierung der CPU (x86)

- Xen in Ring 0, Gast in Ring 1
- Problem: Gast-OS läuft in Ring 1 => darf bestimmte privilegierte Instruktionen nicht mehr ausführen
- Lösung: Gast sieht keine x86-Architektur, sondern Xen/x86
- Xen/x86 = x86 ohne kritische Instr.
- Xen kann Gast über Hypercalls auf kritische Instr. zugreifen lassen
- Anpassung des Gastes notwendig!
- Dom0 hat mehr Rechte

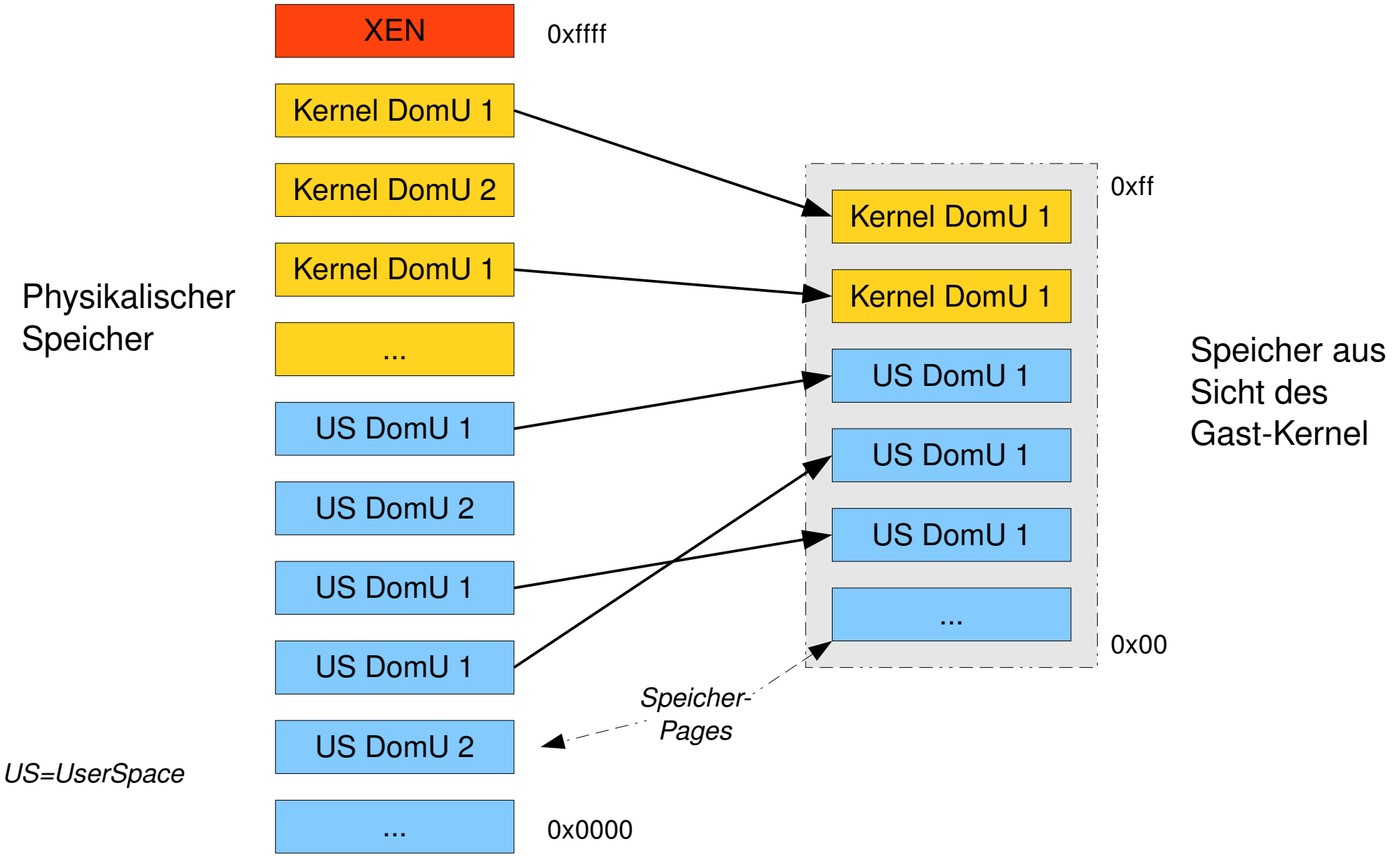


German Unix User Group guug

Scheduling

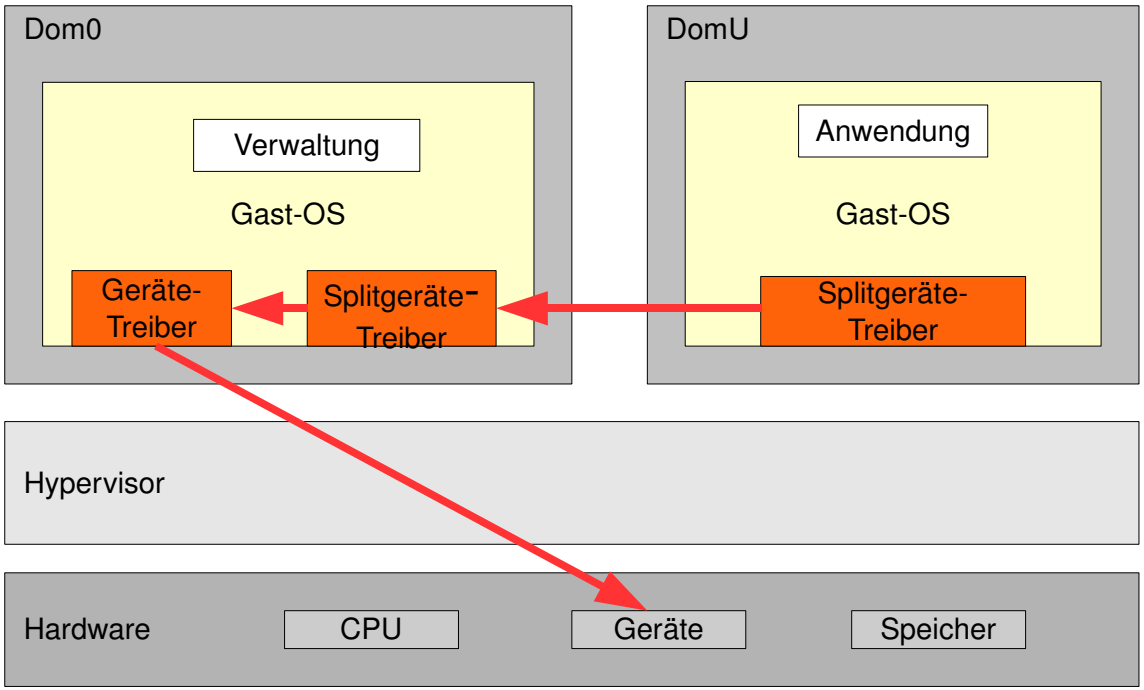
- Mehrere VM's auf einer Hardware => Zeitmultiplexing
- Xen-Scheduler ähnlich Schedulern in Betriebssystemen
- Mehrere Scheduling-Algorithmen in XEN
  (Auswahl über Parameter in Grub: menu.lst)
- Hat Auswirkung auf Systemzeiten in Gast-Domänen:
  - Kernel kann Uhrzeit nicht bestimmen, da seine Zähler durch den Scheduler unterbrochen werden
  - Xen stellt Zeit für Domänen zur Verfügung
  - Zeit in Dom0 einstellen
  - Kein NTP oder hwclock in DomU

Speichermanagement



Virtualisierung von I/O-Geräten

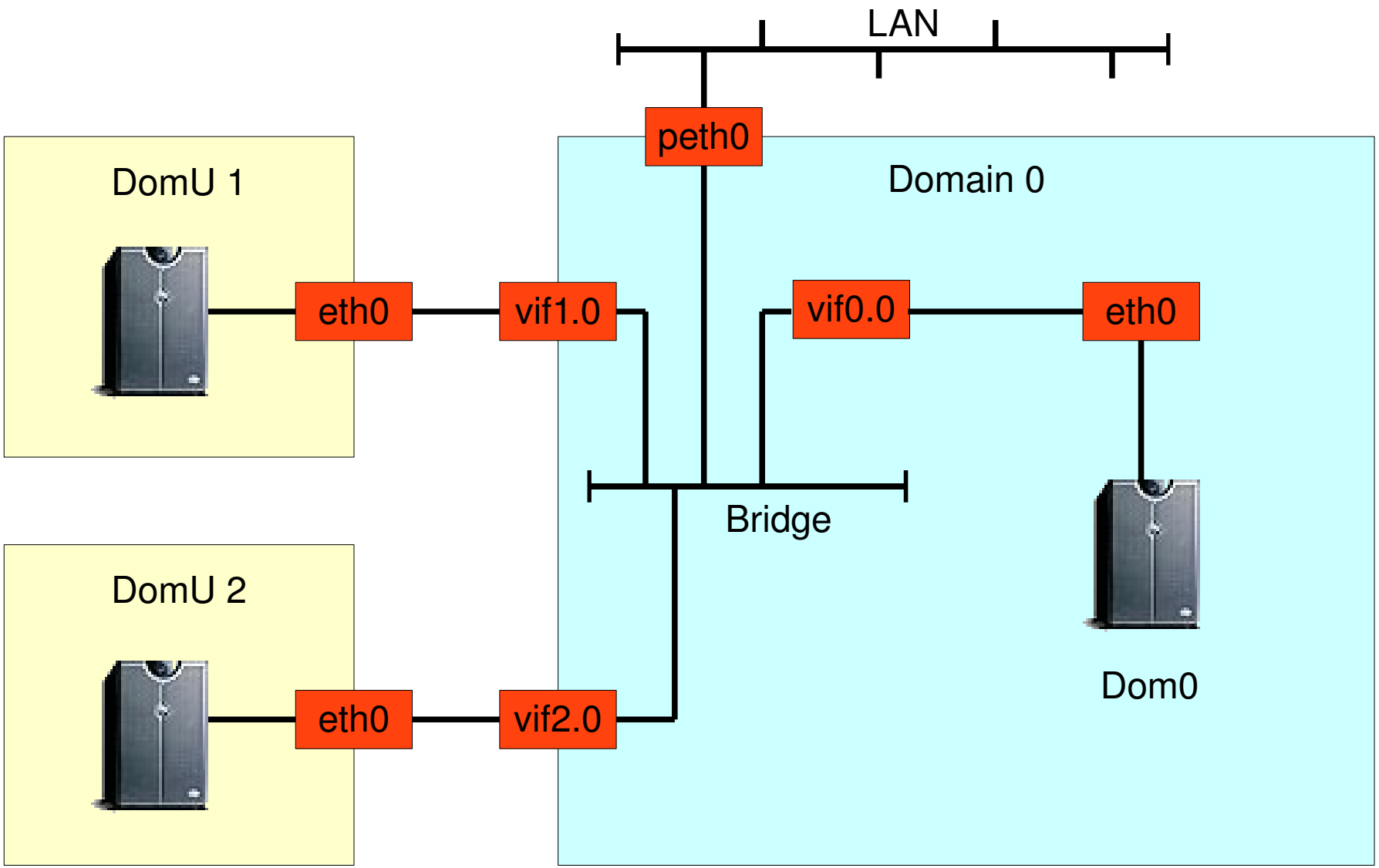
- I/O-Geräte sind nicht für simultanen Zugriff ausgelegt
- Lösung: Nur ein Gast hat Macht über ein I/O-Gerät (Default: Dom0)
- Andere Gäste werden über Split-Gerätetreiber angebunden
- Split-Gerätetreiber besteht aus Frontend- und Backend-Treiber



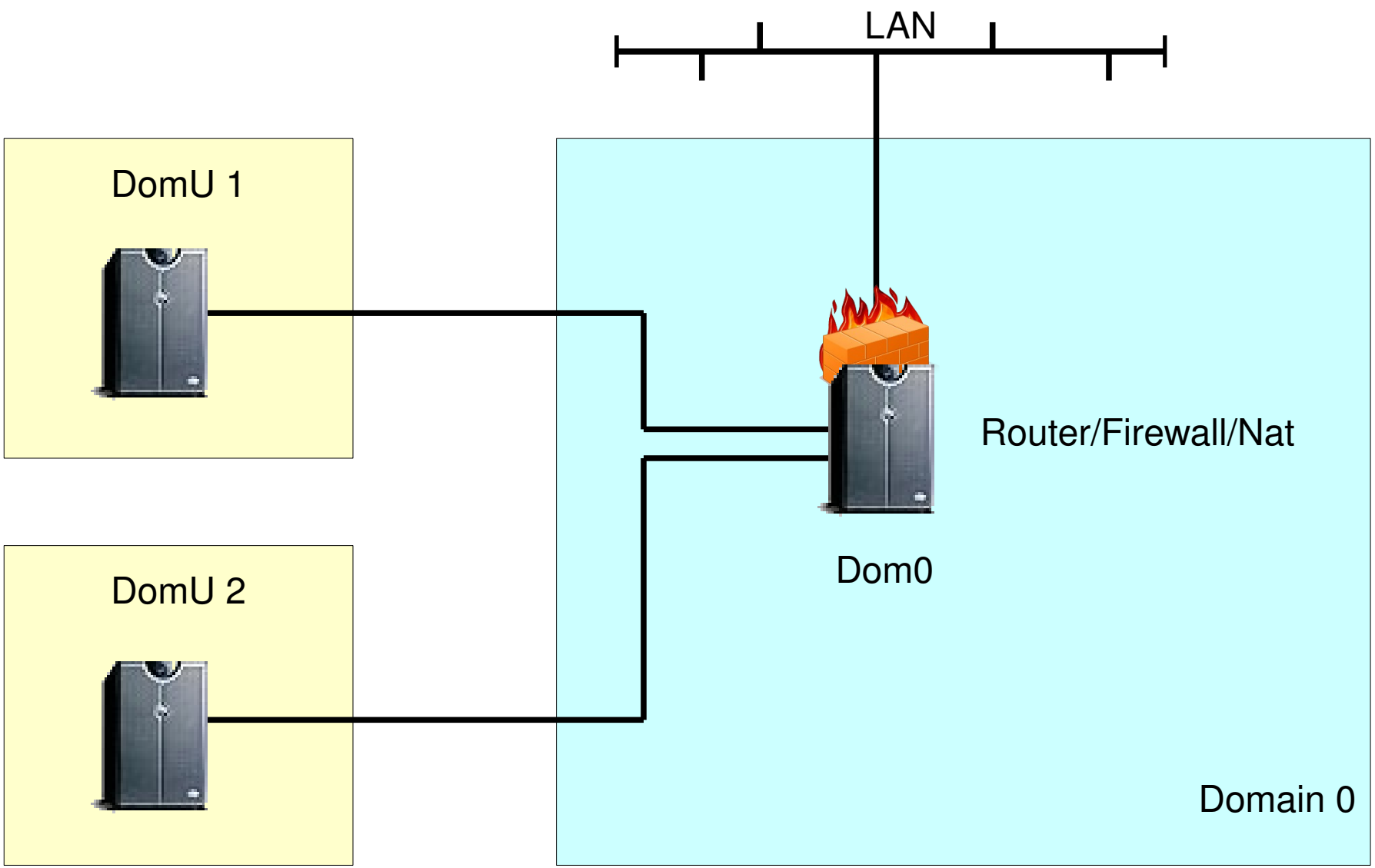
Netzwerk unter XEN (1)

- Dom0 fungiert als Bridge, Router oder sogar Firewall
- Basierend auf Betriebssystem-Tools
- Skripten zum konfigurieren der Netzwerk-Schnittstellen
  - Fertige Skripten für einfache Szenarien werden mitgeliefert
  - Mit eigenen Skripten komplexere Szenarien realisierbar
- Verschiedene Netztopologien möglich:
  - Bridge
  - Routing
  - NAT (Network Address Translation)
  - Firewalling/DMZ

Netzwerk unter XEN (2)



Netzwerk unter XEN (3)



Massenspeicher

- Fast alle Massenspeicher möglich:
  - Image-Datei
  - Partition
    (auch LVM, EVMS, DRBD, ...)
  - Netzwerk-Speicher
    (NFS, NAS, SAN, iSCSI, Infiniband, ...)
  - CD/DVD/ISO-Images
- Clustering mit GFS, OCSF2, ...
- Mit LVM Snapshots im Betrieb mögich
- NFS Server in DomU hat
  sehr schlechte Performance!

Xen-Dämonen und Xen-Tools

- Tools
  - *xm*: Kommandozeilen-Tool zur Steuerung von Xen
- Dämonen
  - *xend*: Hauptdämon für Steuerung der Domänen
  - *xenstored*: Hält verschiedene Verwaltungsinformationen bereit
  - *xenconsoled*: Dämon für Konsolensteuerung

Installation unter Debian - Dom0

Bsp.: Installation von Xen mit bridged-Network unter Debian 4.0 Etch

Hint: Xen 3.2 Pakete für Etch sind im Debian Backports-Repository zu finden!

1) Xen Pakete installieren:

```
apt-get install iproute python bridge-utils \
                xen-docs-3.0 xen-hypervisor-3.0.3-1-amd64 xen-ioemu-3.0.3-1 xen-tools \
                xen-utils-common linux-image-xen-amd64
```

2) Konfiguration anpassen: /etc/xen/xend-config.sxp

```
(network-script 'network-bridge bridge=my-bridge')
(vif-script vif-bridge)
(dom0-min-mem 196)
(dom0-cpus 0)
```

3) Rebooten um Hypervisor zu laden

Grub wird bei Paketinstallation angepasst!

Kurzer Abstecher: Grub

/boot/grub/menu.lst:

```
title          Xen 3.2-1-amd64 / Debian GNU/Linux, kernel 2.6.18-6-xen-amd64
root           (hd0,0)
kernel         /boot/xen-3.2-1-amd64.gz
module         /boot/vmlinuz-2.6.18-6-xen-amd64 root=/dev/sda1ro noacpi noapic console=tty0
module         /boot/initrd.img-2.6.18-6-xen-amd64
savedefault

title          Debian GNU/Linux, kernel 2.6.18-6-amd64
root           (hd0,0)
kernel         /boot/vmlinuz-2.6.18-6-amd64 root=/dev/sda1 ro noacpi noapic
initrd         /boot/initrd.img-2.6.18-6-amd64
savedefault
```

Installation unter Debian - DomU (1)

DomU in 6 Schritten:

Ein Debian Etch Gast soll installiert werden. Die Partition */dev/sdb2* als Root- und */dev/sdb3* als Swap-Partition genutzt werden. Der virtuelle Host soll in die Bridge '*my-bridge*' eingebunden werden.

1) Partitionen für Gast vorbereiten

```
mkfs -t ext3 /dev/sdb2
mkswap /dev/sdb3
mount -t ext3 /dev/sdb2 /mnt/sdb2
```

2) Basissystem per debootstrap installieren

```
apt-get install debootstrap
debootstrap --arch amd64 etch /mnt/sdb2 http://ftp2.de.debian.org/debian
```

Installation unter Debian - DomU (2)

3) Konfigurationsdatei für Xen DomU anlegen: /etc/xen/my-domu-1.cfg

```
kernel      = '/boot/vmlinuz-2.6.18-6-xen-amd64'
ramdisk     = '/boot/initrd.img-2.6.18-6-xen-amd64'
memory      = '1024'
vcpus       = '4'
cpus        = '2'
root        = '/dev/sda1 ro'
disk        = [ 'phy:sdb2,sda1,w',
                'phy:sdb3,sda2,w']
name        = 'my-domu-1'
hostname    = 'my-domu-1'
vif         = [ 'mac=AA:AA:AA:AA:AA:01, bridge=my-bridge' ]
```

Der Kernel liegt in dieser Konfiguration im Dateisystem der Dom0!
Alternativ kann mit *pygrub* bzw. *pvgrub* auch ein Kernel
aus der DomU-Partition gestartet werden.

Installation unter Debian - DomU (3)

4) Konfiguration der DomU anpassen

vi /mnt/sdb2/etc/network/interfaces
vi /mnt/sdb2/etc/fstab
vi /mnt/sdb2/etc/apt/sources.list
...

5) Gettys auskommentieren:

vi /mnt/sdb2/etc/inittab

```
1:2345:respawn:/sbin/getty 38400 tty1
#2:23:respawn:/sbin/getty 38400 tty2
#3:23:respawn:/sbin/getty 38400 tty3
#4:23:respawn:/sbin/getty 38400 tty4
#5:23:respawn:/sbin/getty 38400 tty5
#6:23:respawn:/sbin/getty 38400 tty6
```

Installation unter Debian - DomU (4)

5) Chroot um SSH-Paket zu installieren

```
chroot /dev/sdb2
apt-get install ssh ...
exit
```

6) DomU starten

```
umount /mnt/sdb2
xm create /etc/xen/my-domu-1.cfg
```

Die wichtigsten xm-Optionen

• *xm console*	Auf DomU Konsole zugreifen
• *xm create <configfile>*	DomU erstellen und starten
• *xm list*	DomU's auflisten
• *xm migrate*	DomU migrieren (für HA)
• *xm reboot*	DomU rebooten
• *xm shutdown*	DomU herunterfahren
• *xm top*	DomU monitoring
• *xm help*	Hilfe

Fallstricke (1)

- Das */lib/tls* Performance Problem
  - TLS = Thread Local Storage
  - Ab Linux 2.6: Segmentregister für „threadlokalen Speicher“
  - Xen-Hypervisor muss Legalität des Zugriffs prüfen
    => Performance Bremse
  - Umbennen von */lib/tls* in */lib/tls.disabled* nicht zu empfehlen
    - Manueller Eingriff nach jedem libc-Update notwendig
    - Statisch gelinkte Programme machen Probleme (z.B. Java)
    - Programme laufen langsamer, da */lib/tls* komplett fehlt
  - Lösung: *libc6-xen* Pakete installieren
    - Offizielle Unterstützung in Debian
    - Bei Updates bereits berücksichtigt

German Unix User Group

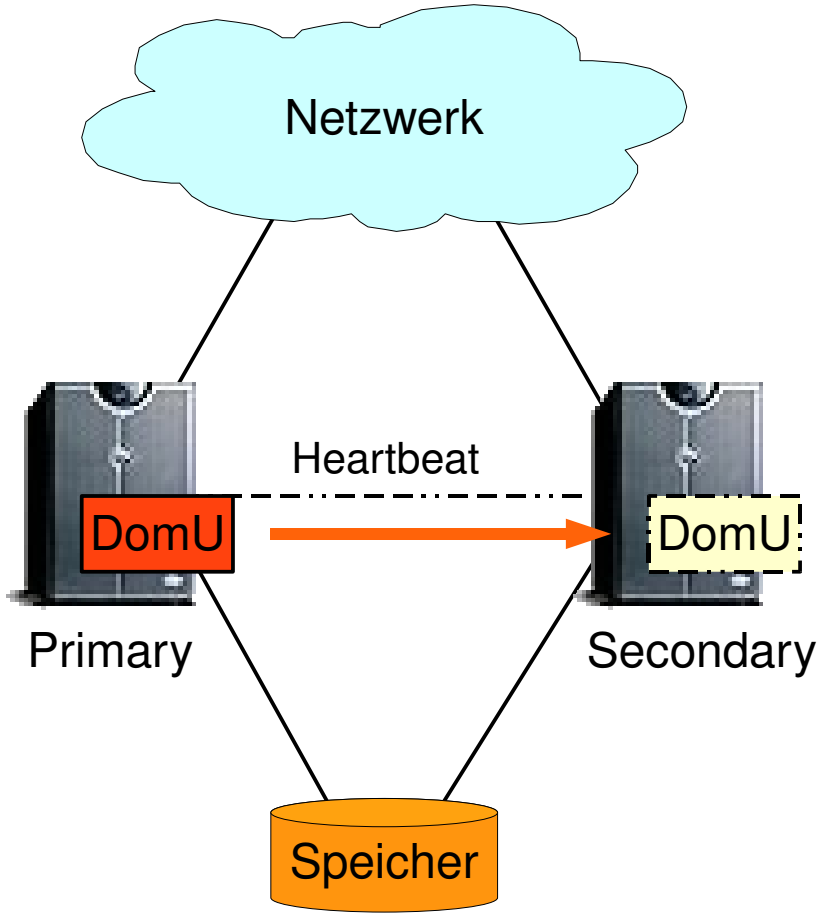
Fallstricke (2)

- ReiserFS und Image-Dateien = schlechte Kombination
  - Ausführen von fsck in Dom0
  - fsck sucht nach Reiser-Superblocks
  - fsck findet Reiser-Superblock in Image-Datei
  - fsck nimmt Dateisystemfehler an
  - fsck versucht Reparatur
  - Dateisystem wird vernichtet
  - Datenverlust kann die Folge sein

Hochverfügbarkeit
für den kleinen Geldbeutel

Typisches Failover-Cluster

- Zwei oder mehr Server bilden Failover-Cluster
- Daten liegen auf teurem Netzwerk-Speicher
  - iSCSI
  - Infiniband
  - SAN, NAS, NFS
  - Etc.
- Netzwerk-Speicher muss ebenfalls Hochverfügbar sein

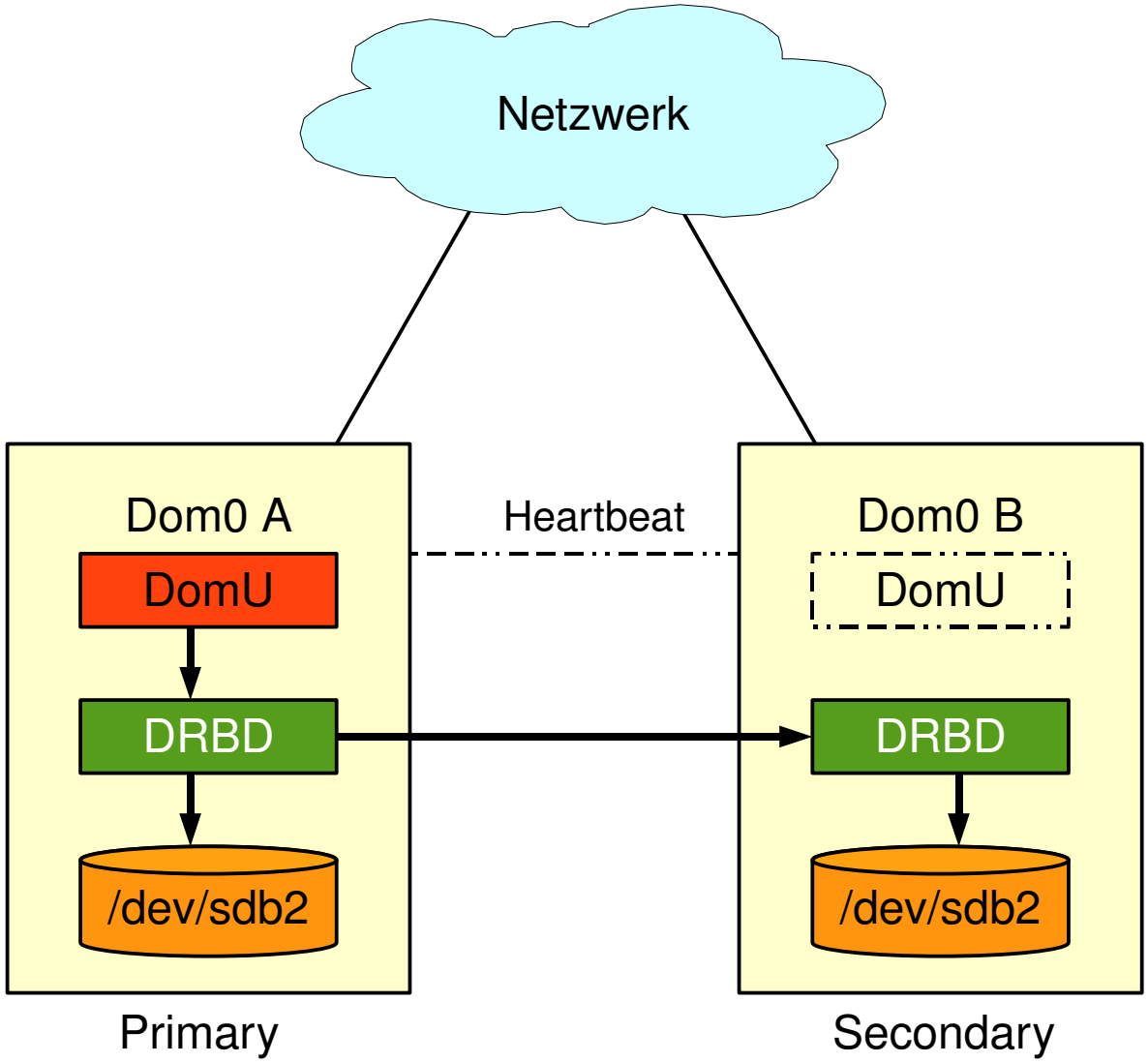


=> für kleine Projekte zu teuer!

Low-Cost Xen-Cluster

- Kein Netzwerk-Speicher notwendig
- OpenSource Komponenten
  - DRBD 7/8
  - Heartbeat
  - Xen 3
- Ab DRBD 8.0.6 ist Live-Migration möglich

Lösung für das Rechenzentrum und den privaten Keller!



XEN
Produkte und Hersteller

Xen Produkte und Hersteller

- Citrix XenServer
  - 2007 Übernahme von XenSource durch Citrix
  - Features wie OpenSource Xen
  - Kommerzielle Management-Tools
  - Kommerzieller Support
- Google Ganeti Cluster
  - Hochverfügbarkeits-Cluster
  - GPL v2
  - Basiert auf Xen, LVM und DRBD
  - Vorgeschlagene Cluster-Größe: 1-25 Nodes
  - Keine Live-Migration

German Unix User Group guug

Fazit

- Vorteile
  - Performanter durch Paravirtualisierung
  - Support durch große Firmen: Novel, Citrix, HP, IBM, Intel, ...
  - OpenSource
  - Live Migration
  - Einsparungen (Energie, Hardware, Cooling, Rack-Space)
  - Flexibilität
- Nachteile
  - An bestimmte Kernel-Versionen gebunden
  - Probleme können bei Cross-Distributions-Installationen auftreten
  - Probleme mit Closed-Source Treibern (Grafik, ...)
  - Schlechte Dokumentation
  - Virtualisierung eröffnet neue Angriffsvektoren

German Unix User Group

Fragen

Quellen, Literatur, Links

- Andrej Radonic, Frank Meyer; XEN 3, Franzis Verlag, Poing, 2006
- Bastian Neuburger, René Palige; Der XEN-Hypervisor; Online unter: http://www.sec.informatik.tu-darmstadt.de/pages/lehre/SS08/seminar_stumpf/doc/Topic02-Xen.pdf
- DRBD 8.0.6 brings full live migration for Xen on DRBD; Online unter: http://fghaas.wordpress.com/2007/09/03/drbd-806-brings-full-live-migration-for-xen-on-drbd/
- http://www.pro-linux.de/work/virtual-ha/index.html
- http://www.jailtime.org (fertige VM-Images)
- http://code.google.com/p/ganeti/ (Cluster Software)
- http://www.xen.org
- http://chaosradio.ccc.de/archive/chaosradio_express_092.mp3 (Potcast zu Virtualisierung mit Focus auf XEN)