

Xen and the Art of Consolidation

Tom Eastep
Linuxfest NW
April 29, 2007

<http://www.shorewall.net/Linuxfest-2007.pdf>

Outline

- Introduction
- Why I use Xen
- Xen Overview
- Xen Installation
- Xen Networking Options
 - Bridged
 - Routed
 - NAT
- Q&A

Introduction

- About me
 - Developing or Supporting Software for 38 years
 - Software Architect for HP
 - Developer of Shorewall (Linux firewall tool)
 - Disclaimers
 - I am here as an individual and not as an HP employee
 - What I tell you represents my own views and not those of HP
 - I am not affiliated with the Xen project or with XenSource
 - I am not affiliated with Novell/SuSE

Why I use Xen - Before I used Xen

- Home office crowded with 5 systems, three monitors a scanner and a printer
- Systems:
 - Firewall
 - Public shorewall.net Server (Web, FTP, Mail)
 - Private File/Print Server
 - Personal Desktop (Linux)
 - Work System (Docked Laptop running Windows™ XP)

Why I use Xen - Problems with my old setup

- Crowding
- Noise
- Heat
- High power consumption
- Lots of systems to administer
- Systems underutilized

Why I use Xen - Before I used Xen

- Why so many systems?
 - Isolate Internet-accessible systems from other systems
 - Isolate functions
 - Strong isolation preferred
 - Reproduce Shorewall user configurations
- I attended a presentation by Ian Pratt and became interested (2004 - pre-Xen 2 timeframe)

Why I use Xen - After I Adopted Xen

- Three Systems with Two Monitors and an HP AllInOne™ Printer/Copier/FAX
- Systems:
 - Combination Firewall/Public Server/Private Server/Wireless Gateway using Xen.
 - Desktop
 - Work System
- KVM switch and LCD Monitor with Dual Video Inputs (although the Xen system is essentially headless)

Why I use Xen - Benefits

- Less clutter
- Less noise
- Less power consumed
- Fewer systems to administer
- Fuller utilization of hardware assets
- There are other benefits that I don't take advantage of
 - Dynamic re-balancing to achieve SLAs
 - Avoid downtime with VM relocation

Xen Overview

- Developed at Cambridge University
- Allows multiple virtual machines within one physical machine
- Open Source version available as option with Linux Distributions
 - Text-mode tools (but that is improving)
- Now also provided by XenSource (“Commercial Open Source”)
 - Free XenServer Express (complete distribution – currently 32-bit only)
 - Nice graphical tools
 - I haven't used it yet

Xen Overview – VM Terminology

- Common Terminology
 - Host: The system hosting one or more virtual machines
 - Host OS: The operating system running on the Host
 - Guest: A virtual machine environment
 - Guest OS: The OS running on a particular guest

Xen Overview - Virtualization

- Single OS Image – Virtuozzo, Vservers, OpenVZ, Zones.
 - “chroot on steroids”
 - Hard to establish protection zones
 - Confuse Administrators
- Full Virtualization – Vmware, VirtualPC, QEMU
 - Run multiple unmodified guest Oses
 - Hard to Virtualize X86 efficiently
- Para-virtualization – Xen
 - Run multiple guest Oses ported to a special architecture (Xen/X86)
 - Xen/X86 is very close to true X86
 - Also Xen/X86_64

Xen Overview - Xen and Para-virtualization

- Initially supported only *para*-virtualization
- Now supports full virtualization
 - Requires hardware support available in current server-class CPUs
 - AMD & Intel's *Pacifica* and *Vanderpool* extensions (think of them as ring “-1”).
- I use paravirtualization

Xen Overview – Pros/Cons of Para-virtualization

- **PROS:**
 - The main advantage of para-virtualization over other approaches is performance
 - Varies with benchmark
 - www.cl.cam.ac.uk/netos/papers/2003-xensosp.pdf
- **CONS:**
 - With para-virtualization, you can't just drop an unmodified installation DVD into the drive and install the Guest OS

Xen Overview - Domains

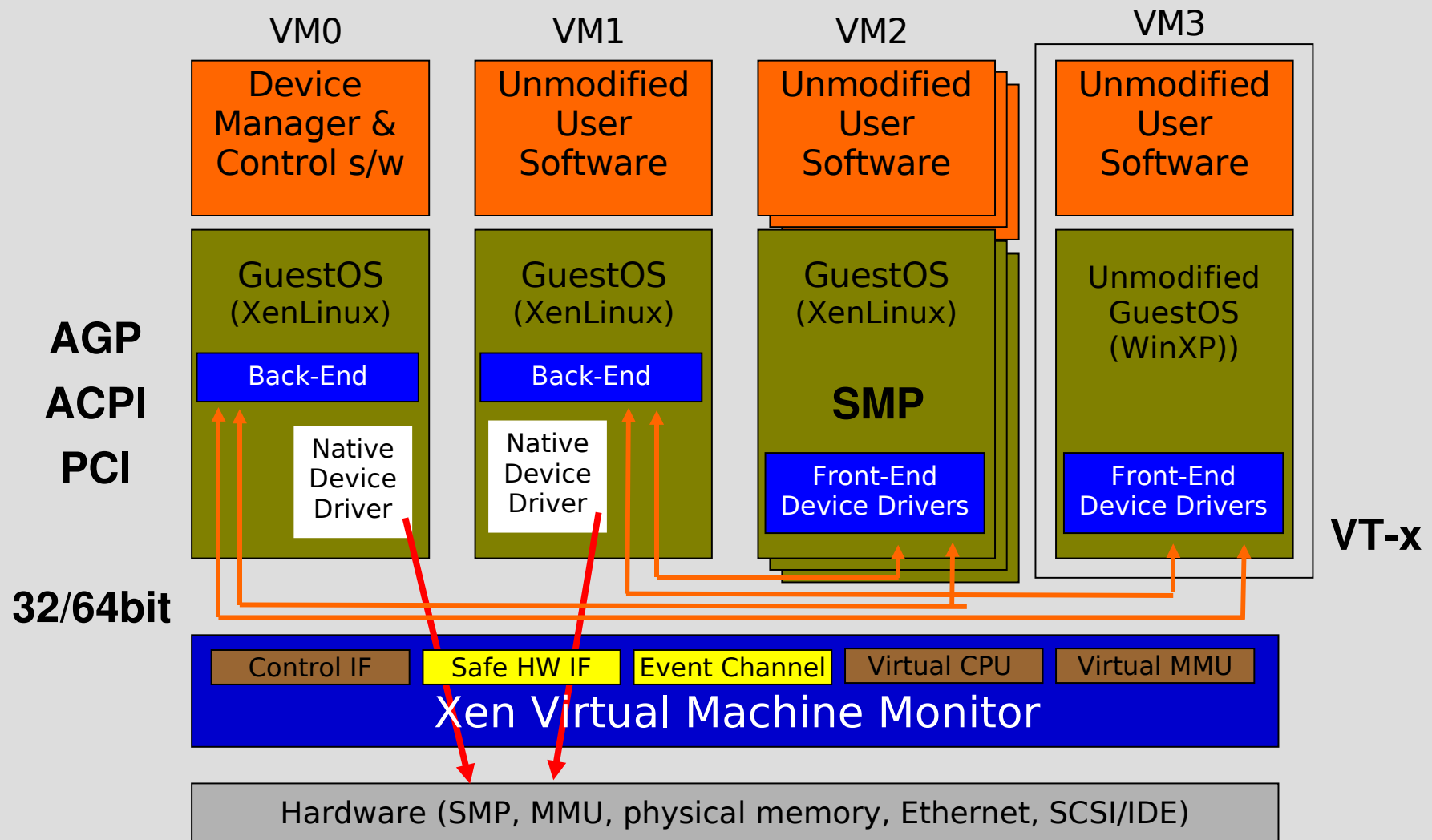
- Xen refers to virtual machines as *domains*
- The host OS runs in a special domain called domain-0 (Dom0)
- The guest OSes run in *User Domains (DomUs)*.
- Domains have a domain number and a domain name.
 - Name is fixed; number varies (except for Dom0)
- But all Domains are guests of the *Hypervisor*

Xen Overview - Hypervisor

- Under Xen, the bootloader does not load the OS!
- It loads the Xen Hypervisor (Virtual Machine Monitor)
- The Hypervisor then loads the host OS
- From `/boot/grub/menu.lst`

```
title Kernel-2.6.18.8-0.1-xen
  root (hd0,5)
  kernel /boot/xen.gz <- Hypervisor
  module /boot/vmlinuz-2.6.18.8-0.1-xen root=/dev/sda6 vga=0x31a ...
  module /boot/initrd-2.6.18.8-0.1-xen
```

Xen Overview - Architecture V3



Xen Overview – Dom0 Services

- `xend`
 - manipulates the Dom0 environment to accommodate Xen
- `xenddomains`
 - Auto start/stop DomU domains defined in `/etc/xen/auto/`
 - I personally create symbolic links from `/etc/xen/auto` to `/etc/xen/vm` (Where `yast[2]` creates domains)

Xen Overview - xm

- xm is a text mode tool for managing domains
- Runs in Dom0
- The 'list' command lists the domains

```
gateway:~ # xm list
```

Name	ID	Mem(MiB)	VCPUs	State	Time(s)
Domain-0	0	1242	2	r-----	83315.2
server	2	512	1	-b-----	3062.8

```
gateway:~ #
```

Xen Overview - xm

- Also supports a 'console' command

```
gateway:~ # xm console server
```

```
lists login: root
```

```
Password:
```

```
Last login: Sat Apr 7 07:57:58 PDT 2007 from 192.168.2.10 on pts/5
```

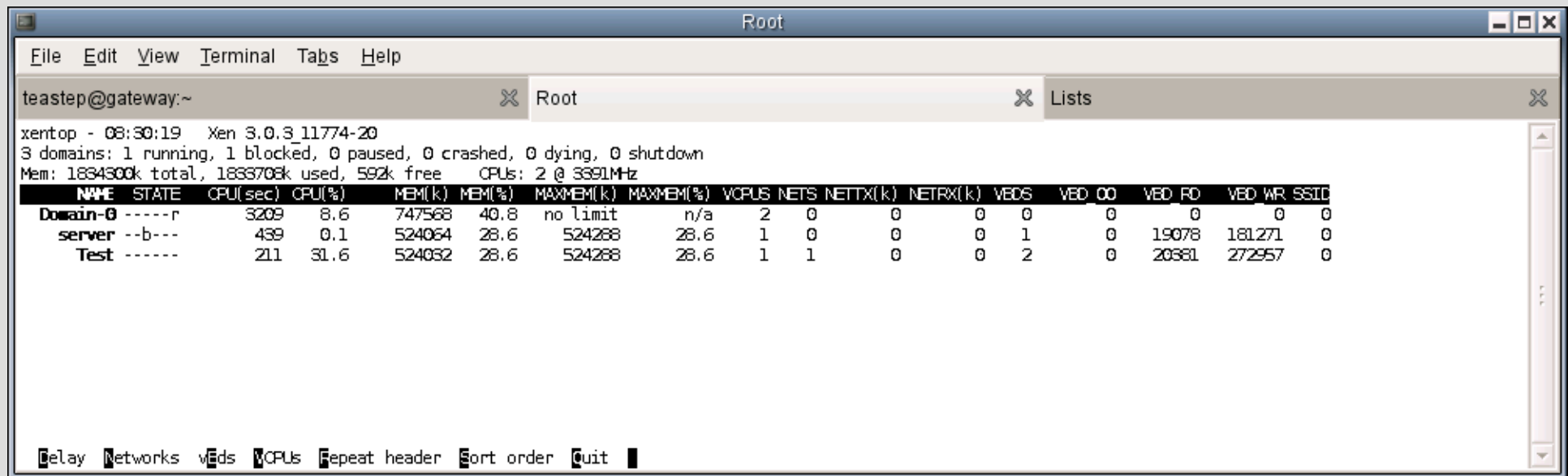
```
Have a lot of fun...
```

```
lists:~ # ps aux
```

USER	PID	%CPU	%MEM	VSZ	RSS	TTY	STAT	START	TIME	COMMAND
root	1	0.0	0.0	808	76	?	Ss	Apr05	0:00	init [5]
root	2	0.0	0.0	0	0	?	S	Apr05	0:00	[migration/0]
root	3	0.0	0.0	0	0	?	SN	Apr05	0:00	[ksoftirqd/0]

Xen Overview - xm

- xm top



The screenshot shows a terminal window titled "Root" with a menu bar (File, Edit, View, Terminal, Tabs, Help) and a tab bar (Root, Lists). The terminal output displays the results of the "xm top" command. It shows system statistics and a table of domain information.

```
xentop - 08:30:19 Xen 3.0.3_11774-20
3 domains: 1 running, 1 blocked, 0 paused, 0 crashed, 0 dying, 0 shutdown
Mem: 1834300k total, 1833708k used, 592k free CPUs: 2 @ 3391MHz
```

NWE	STATE	CPU(sec)	CPU(%)	MEM(k)	MEM(%)	MAXMEM(k)	MAXMEM(%)	VCPUS	NETS	NETTX(k)	NETRX(k)	VEDS	VED OO	VED FO	VED WR	SSID
Domain-0	-----r	3209	8.6	747568	40.8	no limit	n/a	2	0	0	0	0	0	0	0	0
server	--b---	439	0.1	524064	28.6	524288	28.6	1	0	0	0	1	0	19078	181271	0
Test	-----	211	31.6	524032	28.6	524288	28.6	1	1	0	0	2	0	20381	272957	0

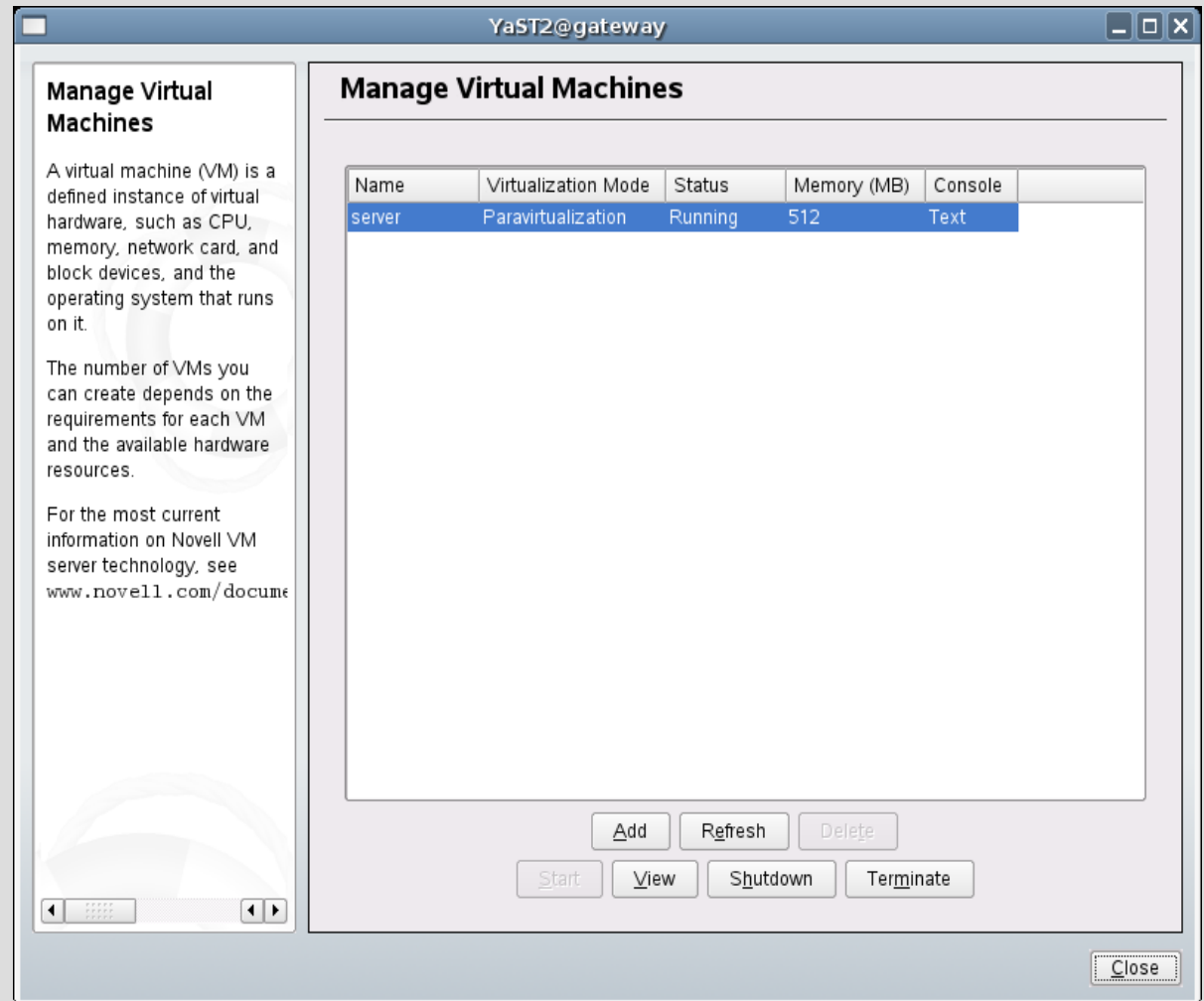
Delay Networks vEds CPUs Repeat header Sort order Quit

DomU Installation

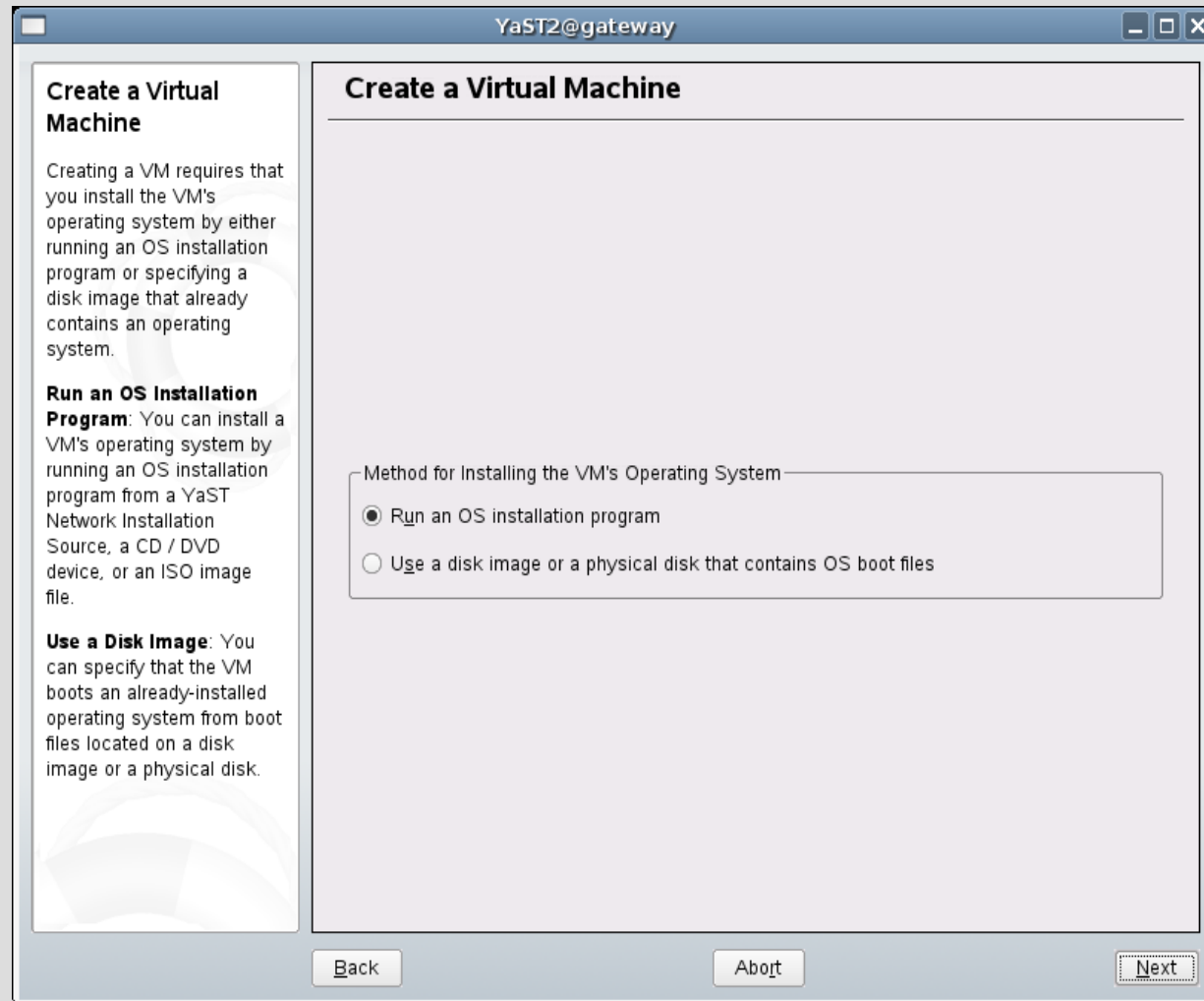
- Use an Installer
 - OpenSuSE (my personal choice)
 - Fedora
 - Other Linux distributions
- Use an Image
 - <http://jailtime.org/>
- Roll your own

DomU Installation - Yast2

- Yast2 xen
- Select 'Add'



DomU Installation - Yast2



DomU Installation - Yast2

- Click “Run an OS Installation Program” and you get a choice of doing a SuSE Install or another distro's install (I've only done SuSE Installs).
 - OpenSuSE install is text mode in an XTERM after the initial configuration screens.
- Click “Use a disk image or a physical disk that contains OS boot files” to install an image.

DomU Installation - Yast2

The screenshot displays the YaST2@gateway installation environment. The terminal window on the left shows the execution of the `fdisk /dev/sda` command, which partitions the disk. The output shows the disk size (160.0 GB) and the resulting partition table entries. The graphical window on the right, titled "Preparing Installation of the Virtual Machine", shows a progress bar at 75% and a "Language" selection dialog. The dialog lists various languages, with "English (US)" selected. The terminal output is as follows:

```
teastep@gateway:~
gateway:~ # fdisk /dev/sda

The number of cylinders for this disk is 160041688.
There is nothing wrong with that, but you should know that
cylinders are not always the same size. In particular, on
and could in certain setups cause problems with:
1) software that runs at boot time (e.g., BIOS)
2) booting and partitioning software like cfdisk,
   (e.g., DOS FDISK, OS/2 FDISK)

Command (m for help): m
Command action
 a toggle a bootable flag
 b edit bsd disklabel
 c toggle the dos compatibility flag
 d delete a partition
 l list known partition types
 m print this menu
 n add a new partition
 o create a new empty DOS partition table and use it as the default
 p print the partition table
 q quit without saving changes
 s create a new empty Sun disk label
 t change a partition's system id
 u change display/entry units
 v verify the partition table
 w write table to disk and exit
 x extra functionality (expert mode)

Command (m for help): p

Disk /dev/sda: 160.0 GB, 160041688 sectors, 255 heads, 63 sectors/track, 19457 cylinders
Units = cylinders of 16065 * 512 = 8225760 bytes

   Device Boot      Start         End      Size   File System  Mount Options
   -----
 /dev/sda1  *         1         1     512B
 /dev/sda2             18618      19457  8225760B ext3
 /dev/sda3             5484      18617  8225760B ext3
 /dev/sda5             5484      18617  8225760B ext3
 /dev/sda6             5746      18617  8225760B ext3
 /dev/sda7             8296      19457  8225760B ext3
 /dev/sda8            10396      19457  8225760B ext3
 /dev/sda9            10598      19457  8225760B ext3

Partition table entries are not in disk order

Command (m for help): q
gateway:~ #
```

The graphical window shows the "Preparing Installation of the Virtual Machine" step. The progress bar is at 75%. The "Language" dialog is open, showing a list of languages: Dansk, Deutsch, Eesti, English (UK), English (US) (selected), Espanol, Francais, Greek, Hindi, and Hrvatski. The dialog also includes instructions: "Choose the Language to use during installation and for the installed system. Click Next to proceed to the next dialog. Nothing will happen to your computer until you confirm all your settings in the last installation dialog. You can select Abort at any time to abort." Buttons for [Back], [Abort], and [Next] are visible at the bottom of the dialog.

DomU Installation - Yast2

The screenshot shows a terminal window titled "Xen - Test" with a cyan header bar containing "YaST @ linux" and "Press F1 for Help". The main content area has a blue background and is titled "Package Installation".

On the left, a white box contains the text: "Please wait while packages are installed."

The main area displays a table of installation statistics:

Media	Size	Packages	Time
Total	2.36 GB	667	45:53
CD 1	2.36 GB	667	45:53

Below the table, the following packages and their installation progress are shown:

- openobex-1.3-22.x86_64.rpm (installed size 55.91 KB) -- Open Source Implementation of the Object Exchange (OBEX) Protocol
- Downloading pam (download size 725.54 KB)
- pam-0.99.6.3-24.x86_64.rpm (installed size 2.29 MB)
- pam-0.99.6.3-24.x86_64.rpm (installed size 2.29 MB)

Progress bars are shown for the last two items:

- A yellow bar for the first pam package is at 100%.
- A yellow bar for the second pam package is at 12%.

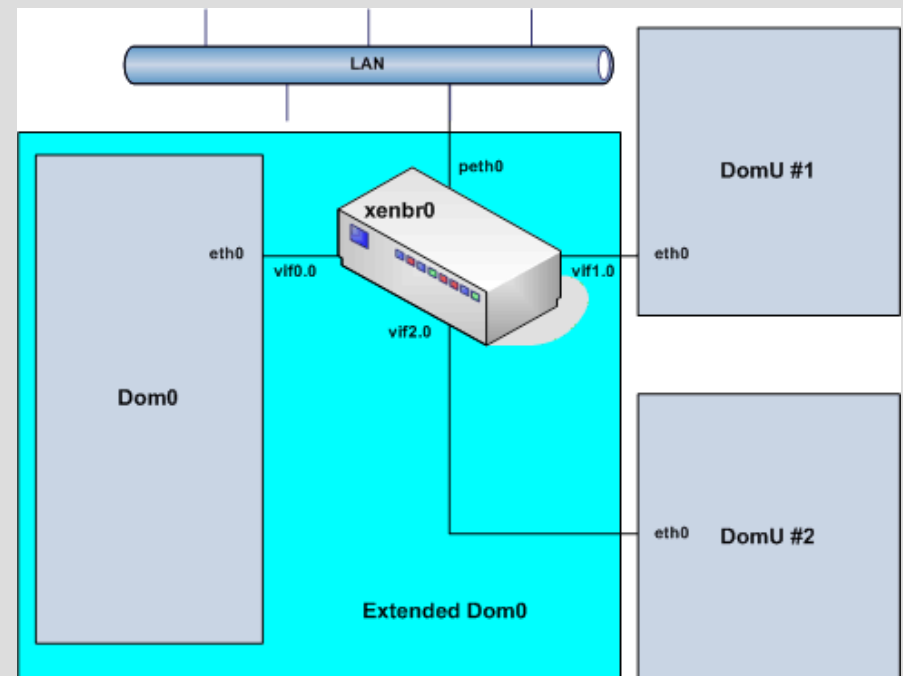
At the bottom, there are three navigation options: [Back], [Abort], and [Next].

Xen Network Options

- Bridged – Default
 - See <http://www.shorewall.net/Xen.html>
 - Also <http://www.shorewall.net/XenMyWay.html>
- Routed
 - See <http://www.shorewall.net/XenMyWay-Routed.html>
- Nat

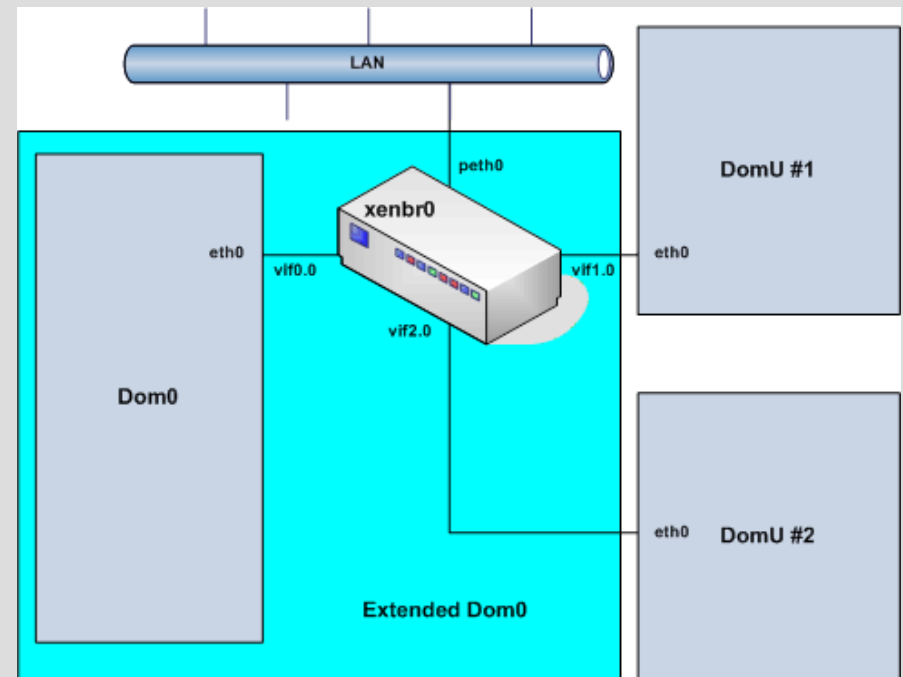
Xen Network Options - Bridged

- Default
- Pros
 - Simple
 - DomU's can use local DHCP Server
 - Works well for creating multiple servers in one physical system
 - Distro Installers assume it



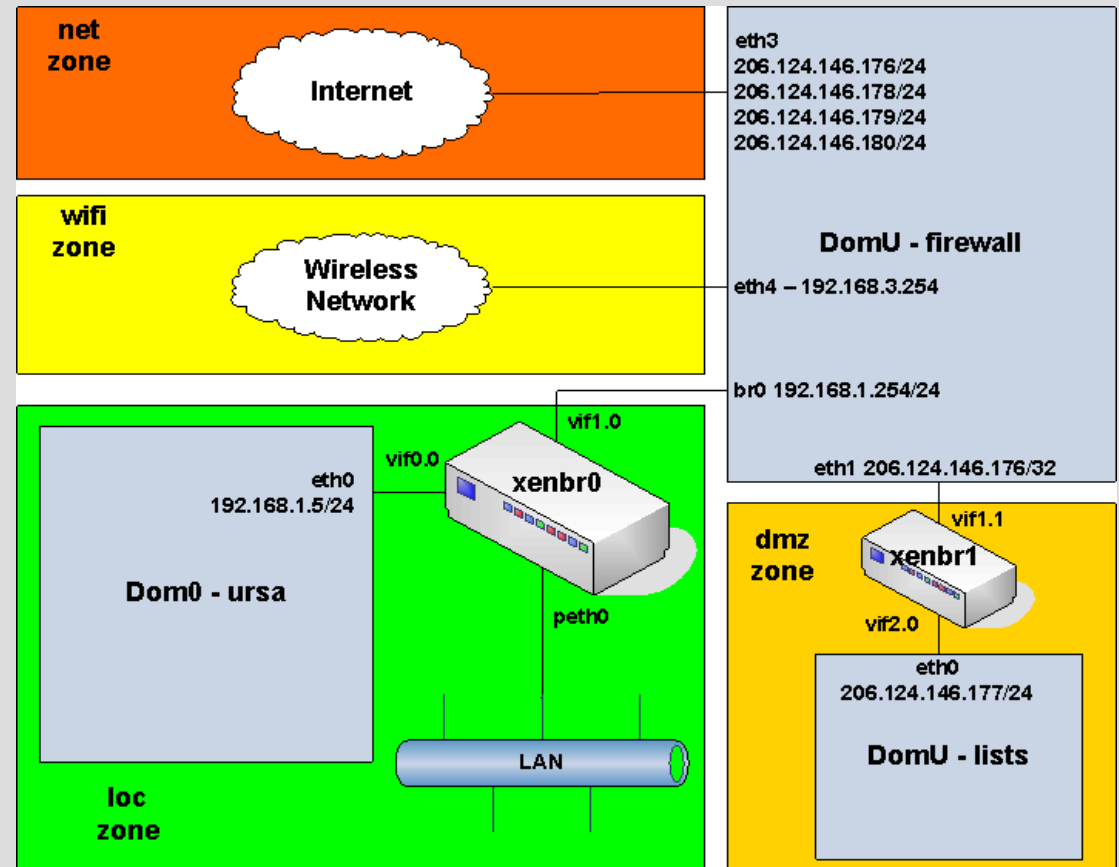
Xen Network Options - Bridged

- Cons – Awkward to firewall
 - Bridge is visible in Dom0
 - Beginning with Kernel 2.6.20, dealing with bridges gets a lot harder



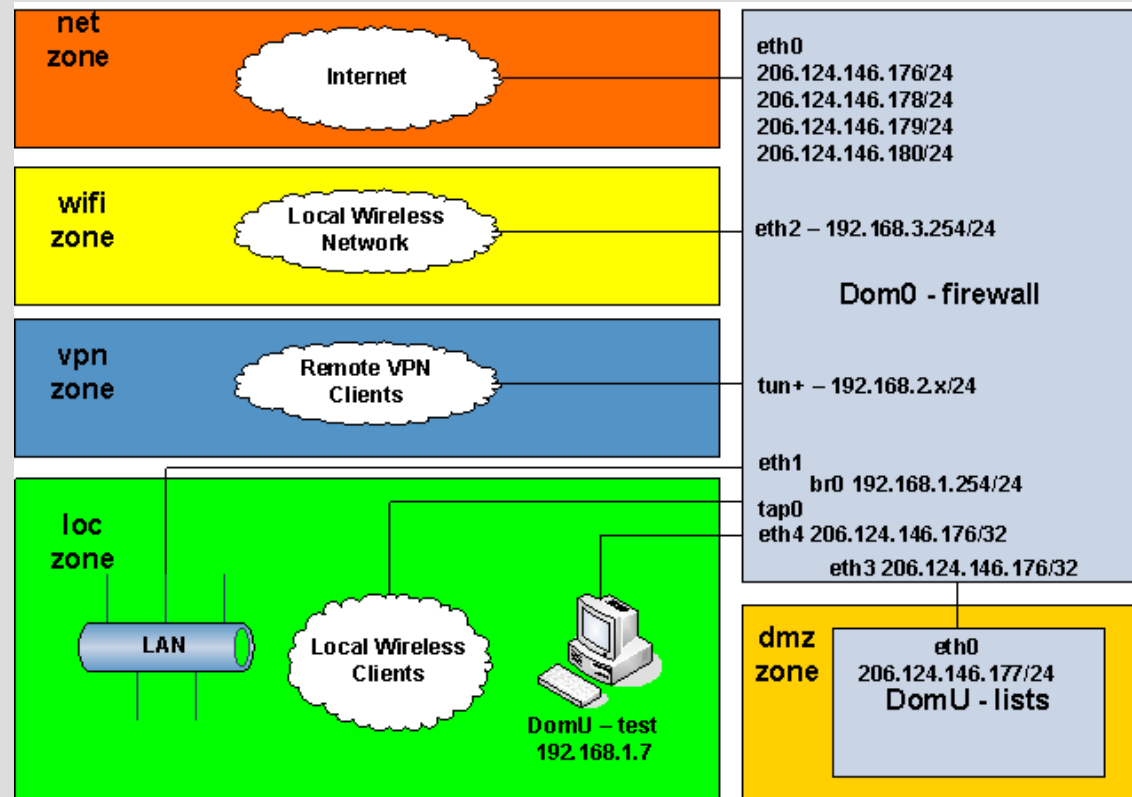
Xen Network Options - Bridged

- Firewall environment can be made more conventional if you run the firewall in a DomU
- Uses *device delegation*



Xen Network Options - Routed

- Dom0 acts as a router for DomUs
- Pros
 - No Bridges
- Cons
 - Little Documentation
 - DomU Installation awkward



Xen Network Options - Nat

- Like the routed configuration except that NAT is used rather than routing.
- I haven't used this option since Shorewall configures NAT

Q & A