

Open Source Virtualization

About Me

Dan Deighton
CISSP, CISA, RHCE,...
Co-founder of Aplura

ddeighton@aplura.com



Agenda

- Overview of Open Source Virtualization
- Real World Example
- Tips, Tricks and Gotchas
- Demo

Reasons to Virtualize

- Cost Savings
 - Server Consolidation
 - Fully Utilize Hardware Investment
 - Lower Admin Cost
- Test Environment
 - Cost and Time Savings
- Training Environment
 - Cost and Time Savings

Reasons to Virtualize (cont)

- Green Computing Movement
- Security
 - Increased Availability
 - Isolate Applications/Services
- It is COOL!

Why Not?

- Need Maximum Performance
 - Standalone OS will outperform a Virtual OS
- Security
 - Smart Malware can detect VME and react
 - "Break-out" of the Guest OS is possible

Types of Virtualization

- Application Level Virtualization
- Emulation
- Full Virtualization
- Hardware Enabled Virtualization
- Paravirtualization
- OS-Level Virtualization

Terms

- Hypervisor (Virtual Machine Monitor, VMM):
 - Manages Virtual Environments
 - Type 1 – Runs directly on Hardware
 - Type 2 – Runs within an OS environment
- VME – Virtual Machine Environment, Guest
- Dom0 – Xen term for Privileged Domain
 - Controls other domains
 - By default, only domain with hardware access
- DomU – Xen term for VME

Application Level Virtualization

- Isolated environment for each Virtual instance
- Single Host OS (only 1 OS license required)
- Examples:
 - Sun Java VM
 - MS SoftGrid
 - Trustware BufferZone*

Emulation

- Simulates All Hardware
- Run Unmodified Guests
- Can Emulate a Different Architecture
- Examples:
 - PearPC
 - Bochs
 - Qemu without Acceleration

Full Virtualization

- Simulates Hardware to Run Unmodified Guests
- VME uses the same Architecture as the Host
- Examples:
 - VMWare WS
 - QEMU w/ KQEMU
 - Virtual PC
 - Virtual Iron
 - KVM
 - VirtualBox*

Hardware-enabled Full Virtualization

- Full Virtualization + ability to offload some work
- Allows "near native" performance
- Intel-VT or AMD-V
 - `egrep -e "vmx|svm" /proc/cpuinfo`
- Examples:
 - VMWare Fusion (and other versions?)
 - Parallels
 - Xen using HVM

OS Level Virtualization

- Host and all VMEs run the same OS
- Same kernel is reused for each VME
- Examples:
 - Virtuozzo/OpenVZ*
 - Solaris Containers (or Zones)
 - FreeBSD jails

Paravirtualization

- Virtual OS aware that it is virtual
- VME collaborates with Hypervisor
- Uses an API to interact w/ host
- Guests must be modified
- Runs on "regular" hardware
- Examples:
 - Xen
 - Sun Logical Domains

Aplura Case Study

- Hazardous Mail Mitigation Service
 - Linux Hosted Mail Servers
 - Physical Systems Hosted at Data Center
 - Need Room to Grow
- Other Managed Services in the Future

Aplura Case Study

- The Problems:
 - Limited Rack Space
 - Each New RU Costs More
 - Need for Multiple Systems w/ option to expand
 - Need to Maximize Server Utilization
 - Need to Isolate Services
 - Wanted Flexibility

Aplura Case Study

- Virtualization Options
 - OpenVZ
 - Xen

Aplura Case Study

- OpenVZ
 - Open source
 - Basis for Virtuozzo (Commercial Version)
 - Fast
 - Live Migration
 - Need custom kernel (provided by project)
 - Major distros do not include OpenVZ

Aplura Case Study

- Xen
 - Open source
 - XenSource (Commercial Version)
 - Major distros starting to support it
 - Red Hat, Debian, Sun, etc
 - Flexibility to install different Operating Systems
 - Paravirt and Full Virtualization
 - Live Migration
 - Not as scalable as OpenVZ

Aplura Case Study

- Xen is our winner
 - Distro support is a big advantage
 - Debian provides xen tools and kernels
 - Big company support also a huge plus
 - Red Hat committed to Xen. Contributing with libvirt, virtsh and VirtManager
 - Sun working with Xen. Solaris Dom0 (host) and paravirt DomU (guest) possible.

Aplura Case Study

- Good Decision?
 - May 2007 – Xen 3.1 released with new features
 - 32bit-on-64bit guest support
 - COW disk support (borrowed from qemu)
 - July 2007 – XenSource kernel patches in mainstream starting with 2.6.23
 - July 2007 - Security Issues with Virtualization (including Xen)

Aplura Case Study

- Our Solution
 - Dell PowerEdge 1850, 2GB RAM, Hardware RAID
 - Hardened Debian 4.0 as Dom0
 - Debian 4.0 as DomU, each in a LV
 - Created standard image for additional DomUs
 - Option to use other OS for DomU

Aplura Case Study

- Issues:
 - Overall Smooth Install
 - PAE mismatch
 - Kernel and Xen Hypervisor must match
 - That bit me once
 - Limited IP addresses required NAT
 - Configured NAT in Xen Config
 - Trick was to modify DomU scripts to open/close ports in firewall

Aplura Case Study

- Lessons Learned
 - More RAM is good
 - Be careful with Distro upgrades
 - Caused PAE mismatch
 - Use LVMs
 - Snapshots
 - Less overhead than a loopback file image
 - Xen Networking is not straight-forward
 - Xen has worked very well for our purposes

Roadmap to Success



Instead, Avoid Problems

The screenshot shows a Minesweeper game window titled "Mines". The window has a menu bar with "Game", "Settings", and "Help". A small "X" icon is visible in the top center. The game board is a 16x16 grid. The top row is blue. The rest of the grid contains numbers (1, 2, 3) and mines (black bomb icons). A yellow starburst icon is on the cell at row 4, column 11, indicating a mine was just revealed. The numbers are color-coded: green for 2, blue for 1, and red for 3. The status bar at the bottom shows "Flags: 0/40" and "Time: 00:00:02".

			2	1	1	1									
			1			1	1	2	1	2					
			2	1					1						
			1						1	3					
			2	1	1				1						
			1						1	1	1				
			1	1					1						
			1						2						
			1						1						
			1						1						
			1	1					1	1					
			1	1					1						
			1						2						
			2	1					1	2					
			2	1					1	2					
			1						1						

Recommendations

- Hardware
 - Lots of RAM (the more, the better)
 - VMEs on non-system disk
 - RAID stripe is even better
 - For Full Virtualization:
 - Intel-VT (Vanderpoole)
 - AMD-V (Pacifica)

More Recommendations

- **Disable Unneeded Services**
 - Should do that anyway
- **Use LVM**
 - Easy Backups with Snapshots
 - Easy to Expand with ext3
 - Less overhead than a looped back filesystem
- **Build and Reuse Stock Images**
 - Faster Deployment

Tips

- For Debian Installations
 - Use debootstrap
 - Fast install
 - Works well
 - Requires post-configuration
- For RPM-based distros:
 - Use virt-install or virt-manager
 - Performs complete install
 - rpmstrap not well maintained

More Tips

- losetup is useful when dealing with file images
- kpartx is even better
 - Part of multipath tools
 - Normally used by hotplug on block devices
 - Works with Virtual Block Devices (VBDs)

Simple Tricks

- Unique MAC address based on date
 - `echo 0A:$(printf "%02X:%02X:%02X:%02X:%02X" $(date +"%-y %-m %-d %-H %-M"))`
- Create a large disk image quickly
 - `dd if=/dev/zero of=NAME.img bs=1M seek=4096k count=1`
- Convert file image to LVM image
 - `bzcat <image>.bz2 | dd of=/dev/VG/LV bs=5M`
 - Then, run `fdisk` on the partition

More Tricks

- Convert VMWare Image to raw disk image
 - Use qemu-img from qemu project
 - `qemu-img convert -f vmdk <image>.vmdk -O raw <image>.raw`
 - May need to "Clean" the image after it is converted
 - Add modules
 - Install xen libraries

More Tricks

- Convert Xen image to Other Platform
 - qemu-img
 - vditool (convert to VirtualBox format)
 - VMWare Converter

One More Trick

- Use PCI Hardware from inside DomU
 - Use lspci to determine pci id
 - Disable in Dom0
 - Disable at boot with pciback.hide option
 - Disable in /etc/modprobe.conf
 - Enable in DomU
 - Use pci option in config file

Gotchas!

- Video Drivers
 - Both ATI and Nvidia will not compile with Xen
- Mixing Virtualization Products
 - Can't run VirtualBox or Vmware on XEN
 - Probably a good thing

Things That Got Me

- NAT issue
 - Needed to disable the transmit checksum in DomU
 - `ethtool -K eth0 tx off`
- Run disk-based VMs on ext3 filesystem
 - Corruption on XFS partition
- PAE mismatch
 - Debian kernel changed to PAE
 - Xen w/ PAE not installed automatically

Other Issues

- Xen Documentation is Terrible
 - Unorganized Wiki
 - Can't find Xen 3.1 docs
- Network Setup can be a Pain
 - libvirt is helping
- Inconsistencies In Full Virtualization

Demos

- kpartx
- Generate MAC address
- Windows on XEN

Parting Thoughts

- Xen + Laptop = Headache
- Be Patient
- Huge Improvements in the near future
- For Now:
 - Use VirtualBox or VMWare on Desktops and Laptops
 - Xen, OpenVZ or VMWare Server on Servers

Resources

General

- [Virtualization at Wikipedia](#)
- [Red Hat Virtualization HQ](#)
- [KVM vs. Xen and VMWare](#)

Resources (cont.)

Sources of virtual appliances

- rpath.org
- <http://virtualappliances.net>
- VMTN
- <http://jailtime.org/>

Resources (cont.)

Conversion

- VMWare to VirtualBox

Resources (cont.)

Cool Virtualization Software

- Trustware BufferZone
- OpenVZ
- Xen Source
- VirtualBox
- VMWare
- Qemu

More Xen Resources

- [HVM compatible Processors](#)
- [Another Xen Networking Guide](#)
- [Virtualization Dashboard](#)