



Xen-Users Common Questions Guide

January 2010

Table of Contents

Introduction.....	3
Support Tools.....	3
How To Guide Links.....	3
Guest Related Questions.....	4
Guest Conversion.....	4
Console.....	4
Drivers.....	5
Domain0	5
DomainU.....	6
Devices.....	10
Installation Questions.....	11
File Systems.....	11
32bit vs 64 bit.....	12
Networking Questions.....	13
Bridging.....	13
IP Determination.....	13
NAT.....	14
SSL/VPN.....	14
General.....	14
High Availability Questions.....	16
Services.....	16
Performance Questions.....	17
Security Questions.....	17
Design/Misc Questions.....	17
Nested Xen.....	17
How does Xen Work.....	17

Introduction

This document is a Xen.org community effort to gather the most commonly asked questions from the xen-users emailing list and other support tools to assist new and experienced Xen hypervisor users with problems that frequently arise. If you would like to add content to this document, please send an email to stephen.spector@xen.org for updates.

For those users interested in trying Xen without installing the application, a Live CD version is available at <http://wiki.xensource.com/xenwiki/LiveCD>.

Support Tools

The following sites are available for Xen hypervisor support:

- xen-users mailing list - <http://lists.xensource.com/mailman/listinfo/xen-users>
- IRC Freenode on ##xen & 2 years searchable data - <http://www.zentific.com/irclogs/>
- Xen 3.0 Documentation - <http://bits.xensource.com/Xen/docs/user.pdf>
- Stack Oveflow - <http://stackoverflow.com/>
- Complete email history of all xen mailing lists – <http://xen.markmail.org>
- Language Specific Support
 - Japanese - <http://lists.xensource.com/mailman/listinfo/xen-japanese>
 - Italian - <http://forum.xen-it.org/>
 - Portuguese - <http://groups.google.com/group/xen-br>

How To Guide Links

The Xen.org community Wiki has a “HowTo Page” with various information sources at <http://wiki.xensource.com/xenwiki/HowTos>. Updates to this Wiki page are continuous so check back often for new How Tos.

Sample topics internally within the Wiki are:

- Xen Networking - <http://wiki.xensource.com/xenwiki/XenNetworking>
- Xen Ubuntu Breezy - <http://wiki.xensource.com/xenwiki/UbuntuBreezyHowTo>
- Xen FreeBSD - <http://wiki.xensource.com/xenwiki/FreeBSDdomU>

Sample topics with external links are:

- Adding USB Devices to Xen HVM - <http://www.virtuatopia.com/index.php>
- PDF Running Xen Tutorial - <http://www.packtpub.com/files/Xen-Virtualisation-Sample-Chapter-Chapter-2-Running-Xen.pdf>
- Xen Dom0 Root FileSystem over NFS - <http://www.mirkomariotti.it/skels/xen0nfs.html>

- Xen on Fedora - <http://fedoraproject.org/wiki/Tools/Xen>
- Create and Install CentOS on Xen - <http://wiki.centos.org/HowTos/Xen/InstallingCentOSDomU>
- http://www.virtuatoxia.com/index.php/Configuring_a_VNC_based_Graphical_Console_for_a_Xen_Paravirtualized_domainU_Guest

Guest Related Questions

Guest Conversion

Q (G1.0): How do I convert a Centos HVM Guest to a PV Guest?

A (G1.0): Creating a Centos HVM domU with working PV drivers : <http://pastebin.com/fb6fe631>
 Converting HVM guest to PV guest : <http://pastebin.com/f6a5022bf>

If you follow both parts correctly you should have a working PV domU. If anything goes wrong during conversion process, you should still be able to boot the previous HVM domU config if you select the non-xen kernel (second entry) from grub menu.list.

Console

Q (G2.0): I have an Xen image that was built for a graphical console (VNC). Is there any way to change it to the non-graphical console (xen console)?

A (G2.0): For HVM guest, you need to enable serial port on domU config file (example here: <http://pastebin.com/fb6fe631>), and setup domU to use serial port (ttyS0 on Linux) by modifying (for Linux domU) /boot/grub/menu.lst, /etc/inittab, and /etc/securetty.

If it's PV guest, you need to set up domU to use xen console (which is xvc0 on current xen version, hvc0 on pv_ops kernel). It's similar to setting up domU for serial console, you just need to change ttyS0 to hvc0. An example of domU setup that can use both xvc0 and vnc console is here : <http://pastebin.com/f6a5022bf>

Q (G2.1): How do I remove an active virtual machine?

A (G2.1): *xm shutdown* or *xm delete*

Q (G2.2): How do I run xm console to a WindowsXP DomU?

A (G2.2): You can't xm console to that (I'm not sure you can xm console to any hvm, but I know you can't to one that doesn't have a console).

Q (G2.3): I start a new DomainU (Guest) and some text scrolls by for launching the guest but then it just sits there with **Continue** and no actions takes place?

A (G2.3): The console for this new DomainU is not properly available for you; fix this by adding `xtra="xencons=tty"` in the configuration file. This will bring up a login screen directly for your new DomainU.

Q (G2.4): One of our CentOS 5.3 randomly reboots, at different times of the day, and I can't see why it's doing it. I have looked through the logs, but don't see anything in there that shows me why it has rebooted. How can I debug this?

A (G2.4): The problem is that when the box panics, it stops syslogd, so you don't get the panic output in /var/log. The best way to fix this is to setup a logging serial console.

Drivers

Q (G3.0): What are the GPLPV Drivers and where can I get them?

A (G3.0): A collection of open source Windows PV drivers that allow Windows to be para-virtualized. They are currently being implemented under the leadership of James Harper. More information on these drivers at:

- <http://wiki.xensource.com/xenwiki/XenWindowsGplPv/Installing>
- <http://lists.xensource.com/archives/html/xen-users/2009-04/msg00058.html>
- <http://meadowcourt.org/downloads/>

Q (3.1): How can I tell if the GPLPV Drivers are loaded correctly?

A(G3.1): If the drivers are installed correctly there should be a Xen device under 'System Devices' in device manager.

Domain0

Q (G4.0): Why cannot I see all my RAM on my Dom0?

A (G4.0): Domain 0 is a paravirt VM in reality, so the amount of ram you allocate to it is what you will see when using local tools like free, /proc/meminfo, top, etc.

To see the full system ram, you need to use the xm tools... and in this case, 'xm info' which will show you all the system resources, as opposed to the resources available to dom0.

Also, you have 16GB ram on the system... you probably already know this, but be aware that without a PAE enabled kernel (if you're using 32bit Xen) you'll only see 4GB of this. PAE will allow you to use up to 16, or maybe 32 (I don't remember what the upper limit for PAE enabled Xen is off the top of my head).

Q (G4.1): Is there any way of checking DomU's I/O from Dom0?

A (G4.1): iostat (Debian: sysstat-package)

Q (G4.2): Can I allocate one CPU to the Dom0 exclusively?

A (G4.2): Add this to the kernel boot line - *dom0_max_vcpus=1 & dom0_vcpus_pin*

Edit /etc/xen/xend-config.sxp - set "(dom0-cpus 1)"

Reboot Dom0

To try on an active system without a Reboot -

xm vcpu-set 0 1 xm vcpu-pin 0 0 0

Q (G4.3): Running *xm info* I see the following memory available; what does the free memory mean?

```
total_memory : 2046 free_memory : 5
```

A (G4.3): Free_memory from "xm info" shows memory not allocated to any domain (including dom0). "free", "top" (or whatever) shows free memory on that particular domain (in your case, dom0). You can adjust memory allocation per domain using "xm mem-set".

DomainU

Q (G5.0): My DomU does not fully start; it shows the following output stopping at Continue...

```
$ sudo xm console test
io scheduler cfq registered
RAMDISK driver initialized: 16 RAM disks of 4096K size 1024 blocksize
```

```
Xen virtual console successfully installed as xvc0 Event-channel device installed. netfront: Initialising
virtual ethernet driver. i8042.c: No controller found. mice: PS/2 mouse device common for all mice
TCP bic registered NET: Registered protocol family 1
```

```
NET: Registered protocol family 17
```

```
Using IPI No-Shortcut mode xen-vbd: registered block device major 8 blkfront: sda2: barriers enabled
XENBUS: Device with no driver: device/console/0 Freeing unused kernel memory: 140k freed
kjournald starting. Commit interval 5 seconds
```

```
EXT3-fs: mounted filesystem with ordered data mode.
```

```
*****
*****
** WARNING: Currently emulating unsupported memory accesses **
**   in /lib/tls glibc libraries. The emulation is   **
**   slow. To ensure full performance you should   **
**   install a 'xen-friendly' (nosegneg) version of **
**   the library, or disable tls support by executing **
**   the following as root:                         **
**   mv /lib/tls /lib/tls.disabled                 **
** Offending process: modprobe (pid=663)          **
*****
*****
```

Continuing...

A (G5.0): It might only be that you don't have a VPS physical console, and that your VPS is fully booted, but you can't see it. There are few things to check.

First, check that your VPS has a "console" device in /dev. Mount your domU filesystem in the dom0, go in /dev and do:

/dev/MAKEDEV console

If you are using a modern Xen kernel and hypervisor, you should check the parameters of the startup file. Check that it has the following option:

```
extra = "4 TERM=xterm xencons=tty console=tty1"
```

Then start your VPS and watch it booting. Note that once it's booted up, you should check that it has a xen friendly libc6 installed (in Debian, you would do "apt-get install libc6-xen").

Q (G5.1): is there a way to set the credit-scheduler's limits and weights per domU in the domU configuration file?

A (G5.1): weight= in the xm config file works, unless you are using RHEL or CentOS.

Q (G5.2): How do I start an application from within a DomU?

A (G5.2): Well, you could always just log in to that VM, open a terminal and run the program. Or you could SSH or telnet in to the VM, start a screen session and run the program.

A VM acts just like any other server, so the procedure for starting programs and executing commands locally and remotely are exactly the same as doing so on any computer.

Q (G5.3): My domUs are in a permanent 'b' (blocked) status as shown by 'xm list', even though they are functioning just fine. That's not normal, is it?

A (G5.3): It's normal for them to show as blocked when they aren't actively running something - in the same way that any process on a 'normal' machine will show as blocked when it's waiting for input, each guest will show as blocked when it's got nothing to do. Give something a processor intensive task to do and you'll find it changes state to running (at least some of the time).

Q (G5.4): Is there any way, to get the name of a domU from the network-common script?

A (G5.4): hostname=\$(xenstore_read "\$XENBUS_PATH/domain" | tr -- '_.:/+' '-----')

Q (G5.5): Is it possible to increase the screen resolution of my xen guest Windows Vista?

A (G5.5): On current Xen, with stdvga=1 & videoram=16, resolutions up to 2048x1536x32 are possible. All that said, the RDP suggestion is probably a better way to access the guest in any case.

Q (G5.6): How to install Solaris via HTTP as a para guest?

A (G5.6): Solaris 10 can only be used as HVM guest. OpenSolaris can be used as PV guest, installed from iso. You can't install it from http. Once you have it installed, you also need zfs support for pygrub (either that, or manually copying kernel and boot archive to dom0)

Boris provides some nice examples on his site : <http://bderzhavets.wordpress.com/>

Q (G5.7): Is it possible to find out the specific vnc Display Number of a domU?

A (G5.7): virsh vncdisplay domU_name_or_id

```
xenstore-ls /local/domain/domU_id/console
```

Q (G5.8): I am trying to create a guest domain. I specified the configurations in `/etc/xen-tools/xen-tools.conf` and I ran `$sudo xen-create-image --hostname=virtualrouter1 --role=udev` the output is: `sudo: xen-create-image: command not found`

A (G5.8): Make sure you installed the Xen tools, for example: `apt-get install xen-tools`

Q(G5.9): I'm trying to assign a dynamic hostname to a xen instance as follows:

Cfg file

==== ==

```
kernel = "/root/vmlinuz-2.6.18-128.1.14.el5xen"  
ramdisk = "/root/initrd-2.6.18-128.1.14.el5xen.img"  
memory = 512  
hostname = "uniquehostname"  
#hostname = " uniquehostname.xen.org"  
name = "my-vm-name"
```

.....

In both the cases, the instance is unable to get the correct hostname..

A (G5.9): From "`xm create --help_config`" : `hostname=NAME` Set the kernel IP hostname. `interface=INTF` Set the kernel IP interface name. `dhcp=off|dhcp` Set the kernel dhcp option.

On most Linux distros, kernel hostname and IP address is ignored, making it somewhat useless. You need to use your normal distro method to set hostname on domU (`/etc/sysconfig/network` on RHEL)

Q(G5.10): As far as I can see, there is something different between using '`xm create`' and '`xm new`' followed by '`xm start`'. It's something to do with data being stored in XenStore. I couldn't suspend the one started with '`xm create`'. Could someone please explain the effective difference between the two and when '`create`' should be used instead of '`new`' and vice-versa.

A(G5.10): `xm create` -> domU configuration is NOT managed by xend. Usually using config files on `/etc/xen`. This is the easiest method to use for beginners, as you have a config file that you can edit manually. The default on RHEL5 (which uses Xen 3.1+).

"`xm new`" and "`xm start`" -> domU configuration is managed by xend. You change values using commands like "`xm block-attach`", which can modify settings online. No config file to edit manually. The default on current versions of Xen.

G (G5.11): I have problem with domU clock. It lose 30 minutes each day. How can i synchronize it with dom0 clock?

A (G5.11): Is this PV domU? If yes, setting `/proc/sys/xen/independent_wallclock` to 0 (the default) should make it sync with dom0. You only need ntp on dom0, and domUs will follow.

The alternative, set `/proc/sys/xen/independent_wallclock` to 1 and run ntp on domU. If this is a HVM

dom0, running ntp on domU is your friend.

Also, check

<http://tuttodebian.blogspot.com/2008/05/xen-clocksource0-time-went-backwards.html> to see if your system experience similar symptoms.

G (G5.12): I would like to set sched-cred parameters on my domU configuration file. How can i do that?

A (G5.12): `cpu_cap & cpu_weight`

Run "`xm create --help_config`" for details, and read <http://wiki.xensource.com/xenwiki/CreditScheduler>

G (G5.13): Is it possible to increase guest memory without reboot?

A (G5.13): You can do a "`xm mem-set <Domain> <memory>`" for a PV domU, but you had to set `maxmem` higher than current assignment beforehand.

G (G5.14): Is it possible to take an already created domU sparse file and make it a non sparse file?

A (G5.14): `cp --sparse=never orig.img new.img`

G (G5.15): I have tried to change CD ISO images during a HVM install using the following commands but it doesn't work. After changing the CD ISO image, it doesn't detect the new ISO image.

```
(qemu) eject -f hdc (qemu) change hdc /media/hitachi/cd-rom-image.iso
```

A (G5.15): Use `xm block-list <domid>` to find the cdrom be-path for the domain, for example:

```
xm block-list 5 Vdev BE handle state evt-ch ring-ref BE-path 768 0 0 4 9 16383
/local/domain/0/backend/vbd/5/768 5632 0 0 1 -1 -1 /local/domain/0/backend/vbd/5/5632
```

Having identified the cdrom device (5632) you can check what iso image it is connected to:

```
xenstore-read /local/domain/0/backend/vbd/5/5632/params
```

(nothing returned)

To connect a new iso image:

```
xenstore-write /local/domain/0/backend/vbd/5/5632/params /mnt/gl3-
tb1_store/MWWin2003R2SvrStdX86_BX2SVOL_EN.iso
```

And you can now see that it is connected:

```
xenstore-read /local/domain/0/backend/vbd/5/5632/params /mnt/gl3-
tb1_store/MWWin2003R2SvrStdX86_BX2SVOL_EN.iso
```

This method works with both emulated devices and with `gplpv` drivers.

Q (G5.16): Is it possible to set the xen to boot the domU one by one when server starts, as currently we have 20 domU, and if boot them together, the the hard disk will be very very slow.

A (G5.16): `cd /etc/xen/config/..... && for i in * do (start VM,). sleep 60 (or whatever time you think is right to start a VM) done`

Q (G5.17): I use FluidVM on some of our VPS host nodes, and the management server

has crashed, so now I need to recover the running VM's, somehow. FluidVM deploys the domU's on the hostnode dynamically from a database, i.e. there's no `/etc/xen/vps1` (for example) config files. The

domU's are still running on the servers, and I now want to create config files for them, while they're running.

How would I be able to do this?

For example, here's a list of running VM's from one of the servers:

```
root@usaxen02:[~]$ xm list
Name                               ID Mem(MiB) VCPUs State  Time(s)
AndriesBurger_39_cronos           90  255    1 -b---- 42.4
Bruce_18_carmen                    60  255    1 r----- 3528327.5
Domain-0                           0  3433   4 r----- 1116681.7
Rudi_14_mars                       40  3007   2 -b---- 953036.3
Rudi_44_vps2                       93  255    1 -b---- 22.9
```

Is there any way to create a config file, /etc/xen/AndriesBurger_39_cronos, from the running domU AndriesBurger_39_cronos ?

A (G5.17): You can use "xm list -l" to dump the configuration in SXP format; then you should be able to use "xm new" or "xm create" with the "-F" option to load an SXP-based config file. See the "xm" man page for more info - that's where I dug up this.

Q (G5.18): How to set up Xen DomU as Windows 2008 Server on a CentOS Dom0 machine?

A (G5.18): Start using the normal way that you usually do when you install a HVM domU, whether it's virt-manager/virt-install or using manually-created config file. One additional thing to note is that for 64bit HVM domUs you need to make sure that acpi, apic, and pae is set to 1 on domU config file.

Once you get that Win2008 fully installed, you can install GPLPV driver later to improve performance.

Devices

Q (G6.0): For my xen domUs I'm using a mixture of local physical partitions (with LVM) and iSCSI disks. For local partitions, I don't have any problem, because LVM volumes are always the same. But for iSCSI disks, devices are assigned in the order they are connected, so I can't be sure that device that now is /dev/sdb (for example) will always be /dev/sdb.

So, is there any way to identify the physical device in the domU configuration not as phy:/dev/sdb, but something like phy:label=fslabel? Or is there any other solution to this problem?

A (G6.0): I go with phy:/dev/disk/by-path/ip-*-iscsi-iqn.* If you assign iscsi luns directly as domU's fs without additional partitioning, you could probably also use /dev/disk/by-label/* or /dev/disk/by-uuid/*

Installation Questions

File Systems

Q (I1.0): Is there a way to have a shared root file system amongst a set of Xen servers?

A (I1.0): You can install the OS in an LVM partition and use it shared across all the xen domUs when you use the partition as 'r' instead of 'w' when defining the disk.

But you have to do all tasks needed for read-only root filesystem.

Like 1. mount ramfs in /tmp and in /var ...

http://en.opensuse.org/How-To_Make_the_root_filesystem_read-only

I use 2 disks in Xen with one as read-only mounted as / and the other is the data partition. I have a need to have scratch partition with pre-populated data and for this I create a LV and put data into it (eg:- software etc..) and then create a snapshot of this volume and send it as rw to the xen machine. This way my original software partitions are intact and also the changes (may be damaging) done in the xen volumes are lost once the snapshot grows to 100%.

Q (I1.1): I installed two Debian web server which run a phpbb3 forum. One stays on a Xen paravirtualized domU (512 MB of ram, 1 vcpu, disc on a raw file file:/home/vale/debian.img,hda,w) on OpenSuse 11.0 and one run on a hyper-v virtual machine (512 MB 1 cpu) build on Windows Server 2008 R2. The performances on PV are very poor than hyper-v. `ab -n 3000 -k -c50`

<http://site.lan/phpBB3/> returns 13,22 req/sec on PV domU and 38,37 req/sec on hyper-v.

Why?

I installed O.S. guest as a HVM domain, then I installed linux-xen-image files and I use them for `vmlinuz` and `initrd`. I also installed `libc6-xen`.

Xen PV config file:

```
name='pv'
ramdisk='/home/vale/initrd.img-2.6.18-6-xen-686'
kernel='/home/vale/vmlinuz-2.6.18-6-xen-686'
bootloader=""
vif=['mac=00:16:3e:33:37:4f, bridge=xenbr0']
vcpus=1
memory=512
disk=['file:/home/vale/pv.img,hda,w']
on_reboot='restart'
on_crash='restart'
extra=""
root='/dev/hda1'
platform='xen'
```

A (I1.1): I'm assuming that phpBB3 is relatively I/O intensive (since it uses db, which I assume you also installed on the same host). In that case, your bad numbers are probably because of this

```
disk=['file:/home/vale/pv.img,hda,w']
```

On Xen, file:/ is not recommended, and you should use tap:aio:/ instead for file-backed storage. Then again, another user reported that even tap:aio isn't good enough

<http://lists.xensource.com/archives/html/xen-users/2009-01/msg00820.html>

So in short, if you use Xen PV, you might want to consider using LVM/partition-backed storage.

Q (I1.2): Is it possible to start a VM that contains just gpxe (which when started, will get an image from a provisioning server and will load that image)

A (I1.2): In this article, we'll show you the processes to setup PXE boot environment for Xen host (hypervisor + dom0) and Xen guest, both PV (Para-Virtualized) guest and HVM (Hardware-assisted Virtual Machine). Details at <http://os-drive.com/files/docbook/xen-pxeboot.html>.

Q (I1.3): I tried to resize a disk of my data guest from 100 to 400 GB. I did an lvresize /dev/xendata/data-disk -L 400G and it works. I started the Guest and did an df -h to check the size but there are still 100 G

A (I1.3): The container is bigger but the filesystem isn't. Resizing an LV doesn't make the FS any bigger.

Log into the DomU and do a resize2fs <device>. You can do this while it's mounted as long as the filesystem is getting bigger.

Oh, and if you've partitioned the LV inside the guest, you'll also need to resize the partition (BEFORE you do a resize2fs, etc.). There are two ways to do this - the safest is to use parted, which works if you're using ext2/ext3 (and a couple other of the most popular filesystems - reiser, I think). The other method is to delete the partition and recreate it with the extended end points. This isn't quite as safe and requires that 1) you're start point for the partition is exactly the same as it was before, and 2) the partition is the last (or only) one on the LV.

Q (I1.4): I want use xen with dynamic slices. For example, I have 20 domU based on FreeBSD, xen hypervisor 3.3.1, Debian Lenny dom0 system. All domUs have 80Gb LVM partitions, but really they use 20 of this 80Gb and I want to create more domU's. How can I do it? I know that some virtualisation have possibility to do dynamic slices(4 example Virtul box)

A (I1.4): Do you mean storage overcommit? That is, assign more storage to domU than what you actually have?

If yes, it's not a matter of Xen vs VirtualBox. It's a matter of what storage backend you use. If you use one of these: - sparse raw file (with file: or tap:aio:) - qcow - vmdk/vdisk (I think full support is only in newer Xen or Opensolaris) - zvol (on Opensolaris) then you can overcommit storage. But if you use disk/partition/LVM for domU storage, you won't be able to.

32bit vs 64 bit

Q (I2.1): Is there anyway to install 64Bit Linux DomU on 32Bit Linux Dom0?

A (I2.1): Types of domU that can be run depends mostly on hypervisor, and not dom0. So if you have 64bit hypervisor, you should be able to run 32 and 64bit PV and HVM domUs, regardless whether dom0 is 32 or 64bit.

If you have 32bit dom0 and 32bit hypervisor, you should be able to run 64bit HVM domU, but not 64bit PV domU.

Networking Questions

Bridging

Q (N1.0): Which Mechanism is used by Xen bridging to handle packets coming from various VMs to forward them to their destination

A (N1.0): Nothing. Xen by itself does not handle bridge. dom0 OS does that.

On Linux dom0 : <http://www.linuxfoundation.org/en/Net:Bridge>

On opensolaris dom0: <http://opensolaris.org/os/project/crossbow/>

IP Determination

Q (N2.0): I want to know the IP of a running VM in XEN.. Is there any way to have this without login to that VM..

A (N2.0): Find domU's mac. This can be easy (if your domU config specify a static MAC).

The easy way to get domU's IP address, you can look at domUs config file (if you specify it), or you can try running this:

```
xm network-list domU_name
```

if you get this line

```
Idx BE MAC Addr. handle state evt-ch tx-/rx-ring-ref BE-path 0 0 00:16:3E:F7:D6:E7 0 4  
6 16238/16237 /local/domain/0/backend/vif/163/0
```

Then domU's MAC is 00:16:3E:F7:D6:E7

The hard way to find out your MAC from a bridge, since your bridge is called eth0 you can try:

```
xm list, note the domain ID (the number)
```

- *brctl showstp eth0* that should show which interface is identified as which "port". For example if your domU has an ID 163, look for the lines that has "vif163.0" or "tap163.0". If the line looks like this

```
vif163.0 (11)
```

then that vif is identified as port 11 on the bridge.

- *brctl showmacs eth0* Look for the port corresponding to the port above. If you get this line

```
11 00:16:3e:f7:d6:e7 no 0.96
```

then on port11 (where your domU interface is) there's a MAC address 00:16:3e:f7:d6:e7.

Now that you have domU's mac, you try snooping the bridge for that MAC. For example :

```
# tcpdump -n -i eth0 ether src 00:16:3e:f7:d6:e7 tcpdump: verbose output suppressed, use
-v or -vv for full protocol decode listening on eth0, link-type EN10MB (Ethernet), capture size 96 bytes
15:54:56.419482 IP 10.0.0.10 > 10.0.0.1: ICMP echo reply, id 5443, seq 1, length 64
15:54:57.422349 IP 10.0.0.10 > 10.0.0.1: ICMP echo reply, id 5443, seq 2, length 64
```

Then you know that domU has IP address 10.0.0.10.

NAT

Q (N3.0): I managed to configure NAT on dom0 but this does not work properly. Outgoing traffic from domU is seen with the original domU ip address instead of the dom0 ip address and the requests can't get back to the domU.

A (N3.0): I figured out MASQUERADING was not set.

The following rule needs to be set:

```
iptables -t nat -A POSTROUTING -o eth0 -j MASQUERADE
```

SSL/VPN

Q (N4.0): Way can't i use openvpn with a xen guest I can't load the tun module

A (N4.0): From openvpn perspective, the requirements for openvpn on xen domU is the same as openvpn on native Linux. If you can't load tun module, then you need to get a kernel that supports it.

The easiest way is to use a distro that supports it. For example, I'm using RHEL/Centos 5.3 domU, loaded with pygrub, and they can run openvpn just fine.

Another alternative is compile your own kernel with tun/tap support.

General

Q (N5.0): I want to ask how to mount one storage device to 2 guests? When I try to create vm handle everything is fine I can create one vdi and vbds for every guest. When I start first machine everything is ok, but when I try to start second one it says that

```
ERROR: 2 INTERNAL_ERROR Device 2051 (vbd) could not be connected.
Device /dev/mapper/test_vg-test64_454 is mounted in a guest domain,
and so cannot be mounted now.
```

Idea how I can share one device to two or more virtual machines? I don't want network solution like nfs,iscsi ... etc but instead use ocfs2. How I can set mode to w! ?

A (N5.0): First of all, you DO realize that sharing a block device without some kind of cluster file system could lead to data corruption? If you want to share it anyway, you can try changing the mode to "r"(for read only) or "w!" (to force read-write multiple mount).

in your domU.cfg:

'phy:/dev/data/bla-disk,sda1,w' => 'phy:/dev/data/bla-disk,sda1,r'

Q (N5.1): I'm looking for a way to monitor network activities of processes in Guest OS. I want to get a list of Guest OS processes that open TCP connections to other machines (like "lsof" command).

A (N5.1): If you're thinking about doing on from dom0, that's not possible. You need something that runs on domU for that, possibly by using snmpd and extending it to run "netstat -anp --tcp". Other host (including dom0) can then collect the information using snmp.

Also, have a look at Versiera, it provides what you are looking for including, user IDs, inbound/outbound communications, IPv4, IPV6, etc. There are many more capabilities. Versiera is not open-source, but the Internet self-manage service is free.

Q (N5.2): I'm attempting to gather stats on usage of the "metal", by which I mean the physical host's hardware. I would like to know the CPU, IO, and network stats for the hardware.

A (N5.2): All DomU's IO passes through Dom0. there you can measure all you want.

- for disk IO, if you use phy: devices, you can use iostat to see the usage of each device. if you use file-based backends, it would be easier to check the userspace daemons. maybe iotop would help. in any case, if aggregate usage is all you need, just measure the disk usage seen at Dom0
- for network, if you can measure at peth0, that would give you the aggregate usage. if you need stats for each DomU, check the respective tun devices.

Q (N5.3): I have some trouble finding the best solution to my networking requirements.

I want to have the following things:

-dom0 : have 2 physical networks devices * 1 eth with public IP (static) * 1 eth with private IP (static)

-domU * 1 eth with private IP

-I also want openVPN solution to let people outside the private network have access to it.

-A DHCP server is required so that domU get their IP from it.

I have debian lenny and xen 3.2 installed and working. Actually openvpn and dhcp are on the dom0. All is fine *except* that domU don't have access to internet (this is my main problem). My current config use network-bridge netdev=eth1 (eth1 have a static private ip).

It is perhaps better to have dhcp and openvpn server in a domU, feel free to suggest what you think is the best choice (and the config that go with it) :)

A (N5.3): When designing such setups, with bridge networking, I often find it easier to think of dom0 as a switch or router, and domU like any other physical server on your network.

In your setup you're making dom0 act as router/firewall. Your problem is that probably you haven't setup ip forwarding and NAT on dom0 to allow domU internet access.

Note that (if you want) you could also have dom0 act like a switch. In that scenario you'd need another domU, with two network interfaces connected to both dom0's eth0 and eth1, acting as router/firewall.

Q (N5.4): I have been trying to get a HVM DomU running and being able to connect to a vlan. I am starting to get the feeling that at least the hw emulations I have tried do not support vlans. Also all the

things I have found online would have me creating the vlans inside the Dom0 and pushing them to the DomU's as regular interface

A (N5.4): You could always have a trunk port in your dom0 and create bridges for each VLAN for xen. You can even script it so you can add it to boot time.

If you have the VLAN trunk set up, you can create bridges as follows.

For this example, my trunk interface is on eth0 and the vlan I am adding is 2.

```
# vconfig add eth0 2 # brctl addbr xenbr2 # brctl addif xenbr2 eth0.2 # ifconfig eth0.2 up # ifconfig xenbr2 up
```

Now all you would add in the domU configuration file is:

```
vif=['bridge=xenbr2']
```

And you would be on VLAN 2.

Otherwise I'm pretty sure you would have to pass through the network card to get VLAN access.

You can also script this and give it a space separated list of VLANs and loop it through. I will leave this up to you though.

High Availability Questions

Services

Q (HA1.0): What software exists for Xen to handle high availability?

A (HA1.0): Here are several tools that currently exist:

- CentOS Cluster
- Project Kemari - <http://www.osrg.net/kemari/>
- Project Remus - <http://blog.xen.org/index.php/2009/04/02/project-remus-released/>
- Linux Cluster Resource Manager - http://clusterlabs.org/mediawiki/images/f/fb/Configuration_Explained.pdf

Q (HA1.1): I'm not sure about how snapshots works on Xen. For example, if I snapshot a DomU with 10GB HD will take, for example, 30 seconds. But, if I snapshot a DomU with a 100GB HD will take longer (I guess).

So, I wanna know how the snapshot works on xen. What if I want to move a snapshotted with a 100GB HD from a Dom0 to another Dom0? I've to move a 100GB file?

A (HA1.1): You could try this approach:

<http://www.infohit.net/blog/post/using-lvm2-snapshots-to-provide-rollback-functionality-for-xen.html>

Makes snapshots quite quick.

Performance Questions

Security Questions

Q(S1.0): If I install minimal linux for XEN in dom0 and a periphery firewall in domU and other applications in other instances of domU, is it possible to restrict/bind the network card to domU having periphery firewall and from there forward packets for dom0 or for other domUs?

Is this possible? If so, is it secure? Or does dom0 always have direct access to Network Card and needs a separate firewall? And packets will always route from dom0 to all domUs ?

What are the issues involved?

Q(S1.0): The approach I've used at home is to hide a network card from Dom0 (see pic-back.hide) and pass it through to a DomU which then sees it as a native interface. I then run a firewall in the DomU and the outside traffic does NOT go through Dom0. The route for packets is then :

real i/f -> DomU (firewall) -> VIF -> int bridge [Dom0 | VIF -> DomU]

From security perspective, this is the same as having an L2 switch (when dom0's bridges have no IP address) or L3 switch (when dom0's bridges have an IP address)

Q(S1.1): I want to use a Disk Encryption and the complete physikal Disk in a DomU. I prefer Truecrypt or Loop-aes. i will going to test loop-aes cause it should have the better performance.

But, did anybody here using truecrypt or loop-aes ? What is the better one, in the fact of speed ?

A(S1.1): dm-crypt/luks is one option, and performs about the same or better than loop-aes. Also it's less problematic because it doesn't use loop devic

Design/Misc Questions

Nested Xen

Q (D1.0): Can I run Xen within Xen?

A (D1.0): Yes, You can install Xen on the base system, create a HVM domain, and install Xen in that guest domain. Note that the inner Xen system will not (yet) support HVM again.

How does Xen Work

Q (D3.0): How does Xen process System Calls on para-virtualized guest?

A (D3.0): When ever a system call is invoked via interrupt or sys center control gets transferred to the kernel (ring 0), which is then handled via system call handler. System call never goes to libc but Libc is a library that provides POSIX interface to the user space applications and in a way wrapper for invocation of a system call.

System call interrupt based [i386]: During booting process, linux kernel of a domU register's its IDT with Xen Hypervisor via `HYPervisor_set_trap_table(trap_table);` [arch/i386/kernel/traps-xen.c]. Xen maintains two IDT's, one global IDT (its own) and other per domain IDT. Xen uses global IDT to register the entire trap handler except for system call handler (int

0x80). When a VM gets scheduled, its system call handler (from per domain IDT table) is registered with the processor. Hence when a domain/VM executes a system call, its own handler is executed.

Implementation differs for x86_64: Xen registers its own system call handler with the processor and from that handler routes the request to VM/Domain specific handler.

Q (D3.1): How does Xen process System Calls on fully virtualized guest (HVM)?

A (D3.1): For HVM domU there is no change in the behavior of the system call. HVM is only supported for Intel-VT and AMD-SVM processors. These processors are virtualization aware. Virtualization aware processors provide a new ring (Root-Ring 0) with higher privilege for VMM and Guest OS continues to run with the same privilege (as without Xen) in Non-Root Ring 0. Guest OS can issue the system calls the way it used to without Xen.

Q (D3.2): Can I run various DomU operating systems on a different Dom0 operating system?

A (D3.2): Yes.

Q (D3.3): Just curious to know, if there is any way that given a terminal to a box, we can determine is it a physical machine or a virtual machine ?

A (D3.3): You should be able to get some useful information from the DMI, e.g:

```
% for i in system-manufacturer system-product-name system-version\ system-serial-number; do echo -n "$i: "; sudo dmidecode -s $i; done
```

```
system-manufacturer: Xen system-product-name: HVM domU system-version: 3.3.1 system-serial-number: 89e5915f-dead-beef-cefd-46904ea94c4a
```

OR

Probably checking kernel process, check your process table for:

```
[xenwatch] [xenbus]
```

Another clue is checking the kernel suffix, for example:

```
# uname -r 2.6.24-23-xen
```

and the proc files:

```
/proc/xen/capabilities
```

Q (D3.4): What is STUBDOM ?

A (D3.4): Stubdoms are lightweight 'service' or 'driver' domains. The initial purpose was to offload qemu (for hvm guests) out of dom0.

So with stubdoms you can run hvm guest qemu in a separate stubdom, which boosts performance and makes it more secure.

stubdoms can also run for example pv-grub for pv guests, making it more secure compared to pygrub, which always runs in dom0.

<http://wiki.xensource.com/xenwiki/StubDom>

Presentation about stubdoms at Xen Summit:

http://www.xen.org/files/xensummitboston08/SamThibault_XenSummit.pdf

<http://blog.xen.org/index.php/2008/08/28/xen-33-feature-stub-domains/>

<http://blog.xen.org/index.php/2008/08/28/xen-33-feature-hvm-device-model-domain/>

<http://lxr.xensource.com/lxr/source/stubdom/README>

Q (D3.5): When is hardware virtualization used in Xen? Is it required?

A (D3.5): Xen uses hardware virtualization for HVM guests. Xen will not launch a HVM guest unless hardware virtualization is turned on.