Add servers into CRON

LANET@LSU April 2012 (lxue2@tigers.lsu.edu)

(In this document, servers are HP DL160G6)

Step 1. Moving servers to Frey Computing Center

Move, move, move!

Step 2. Setup the light out manager port for serial monitoring

Turn on the server, press F10 into bios, change following options:

1. Advanced
   a. Power efficiency mode -> Performance (performance first)
   b. CPU configuration
      i. Intel HT Tech -> Disable (Intel HT Tech does not support Linux well)
   c. IPMI configuration
      i. Set LAN
         1. BMC LAN configuration
            a. IP -> 192.168.0.X (depend on current available monitoring IP)
            b. Subnet -> 255.255.255.0
   d. Bios serial console -> enable
      i. Serial port mode -> 115200
      ii. Terminal type -> VT100

2. Boot
   a. Boot setting
      i. Quiet boot -> disable
   b. Boot
      i. Boot order
         1. 1st boot -> IPA GE slot (set as 1G management NIC)
         2. Disable all others

After setting bios like this, the light out manager port is enabled

The commands examples of light out manager port are:

Log on the manager port:

telnet 192.168.0.25
user:admin
password:admin
Restart server:

```bash
./-> cd system1
./system1/-> reset
```

Start console:

```
ESC + Q
```

End console:

```
ESC + ( 
```

To set the remote console type:

```bash
boss# cd /tftpboot/
boss# cd freebsd.newnode/
boss# cd boot/
boss# vi loader.conf.orig
    console="comconsole"
#console="vidconsole"
#console="nullconsole"
```

**Step 3. Set the hard disks as RAID 0**

Restart the server, press F8 into HP smart array management

1. Delete logical drive
2. Create logical driver
   a. Choose 4 hard disks
   b. Choose RAID 0
3. Select boot volume
   a. Direct Attached Storage
      i. Logical drive 01

After set RAID 0, you can see a total volume of the RAID. In our case, RAID 0 is 587GB.

**Step 4. Delete nodes from CRON**

Because we have 10 machines to delete this time in CRON, if you do not have nodes to delete, skip this step 3 and continue step 4.

```bash
%/usr/testbed/sbin/withadminprivs /usr/testbed/sbin/sched_reserve emulab-ops hwdown pc17
Checking if pc17 is reserved...Reserved - Scheduling next reservation...
%/usr/testbed/sbin/withadminprivs nfree emulab-ops reloading pc17
Moving [Node: pc17] to [Experiment: emulab-ops/hwdown]
%/usr/testbed/sbin/withadminprivs /usr/testbed/sbin/deletenode pc17
*** WARNING:
This will erase all evidence that pc17 ever existed in the testbed!
Are you SURE you want to continue? y
```
Removing pc17...

Step 5. Create a PC type

Now you need to create a PC type for HP DL160G6 server. Following picture is an example for PC type attributes.

Note that for disktype, it is showed on the console when you restart server

Delay_capacity means how many delay nodes this node can accommodate

Set frequency as 2700

![Node Attributes Table]

Step 6. Add new nodes into CRON

Click “Adding Testbed Nodes” under “Administration”, turn on one server, and you will see CRON detects one new server as following picture:
Click on “ID (61)”, fill out “Identifier” and “Type”, and modify mac addresses and switch port connections as following. Note that data ports (mxge) should always be numbered after the control ports (igb), so the sequence of the ports should be changed like in the following picture.

Click “Update node” and then “Back to the new node list”, choose this changed pc, check “Ignore unconnected interfaces”, and click “Create selected nodes”.

Done!

Step 7. Customize the image for new server

This is the most challengeable part.
First of all, you need to go to Frey building with an image CD/USB, and install that image on the server, and then to customize it. I suppose you know how to install a raw clean image.

Only two notices:

1. For partition, make two partitions, one for your use, and one for swap which is the same size as memory size.
2. should install grub on /dev/sdal1, not /dev/sda; for HP DL160G6 server its /dev/cciss/c0d0p1

Following are the things after installing the clean image.

This step is an example for Ubuntu 10.04.4

First reserve a node:
withadminprivs sched_reserve emulab-ops hwdown pc15

Open console of that pc
telnet 192.168.0.25
admin
admin
./~-> cd system1
./system1/~-> reset
open console, when following shows quick type a key
Type a key for interactive mode (quick, quick!)
Enter boot spec:

You need to do this several times during this step, there are two options,
Option1. For booting the OS on the hard disk, which should be the image you want to customize
Enter boot spec:part:1

Option2. For booting the management MFS, which is freebsd, and you need freebsd when you zip the image for CRON
Enter boot spec:loader:/tftpboot/freebsd

use boss root to login to pc
boss# ssh pc15

use root on pc15 (passwd root)

root@pc15:~# aptitude update
root@pc15:~# df -h

<table>
<thead>
<tr>
<th>Filesystem</th>
<th>Size</th>
<th>Used</th>
<th>Avail</th>
<th>Use%</th>
<th>Mounted on</th>
</tr>
</thead>
<tbody>
<tr>
<td>/dev/sdal1</td>
<td>92G</td>
<td>772M</td>
<td>87G</td>
<td>1%</td>
<td>/</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>192K</td>
<td>4.0G</td>
<td>1%</td>
<td>/dev</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>0</td>
<td>4.0G</td>
<td>0%</td>
<td>/dev/shm</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>36K</td>
<td>4.0G</td>
<td>1%</td>
<td>/var/run</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>0</td>
<td>4.0G</td>
<td>0%</td>
<td>/var/lock</td>
</tr>
</tbody>
</table>
LANET GROUP LSU

none                  4.0G     0  4.0G   0% /lib/init/rw
none                  92G    772M  87G   1%

uncomment "#" from /etc/inputrc as below to enable backward history search:
"\e[5~": history-search-backward
"\e[6~": history-search-forward

echange /etc/apt/source.list to US mirror site:
Argonne National Laboratory http ftp rsync 10 Gbps Up to date
http://mirror.anl.gov/pub/ubuntu/
deb http://mirror.anl.gov/pub/ubuntu/ lucid main
deb-src http://mirror.anl.gov/pub/ubuntu/ lucid main

fix MBR: In FreeBSD MFS
sysctl kern.geom.debugflags=16 # needed if you are running FBSD8
fdisk -u aacd0  (for HP DL160G6, its da0)
Then go through the interactive prompts:
don't change what BIOS thinks
don't change partition 1&2
do change partition 3
    set sysid/start/size all to 0
don't explicitly set beg/end
    yes, you are happy with this partition
do the same for partitions 3 and 4 (i.e., zero them)
don't change the active partition
do write the new partition table

"geom not found: XXX" is negligible.

Now imagezip for the whole disk should be happy.

with serial console enable

pc15 need to goto FreeBSD first, and go to boss
boss# ssh pc15 imagezip /dev/da0 ->
/usr/testbed/images/template/pcHPDL160G6-UBUNTU10-64-v1.ndz
Slice 3 is unused, NOT SAVING.
Slice 4 is unused, NOT SAVING.
99999580160 input (2384191488 compressed) bytes in 101.764 seconds
Image size: 204472320 bytes
22.343MB/second compressed
Finished in 104.671 seconds

test ==> succeed
boss# ssh pc16 imageunzip - /dev/da0 < /usr/testbed/images/template/
pcHPDL160G6-UBUNTU10-64-v1.ndz
Wrote 57999884288 bytes (1709064192 actual) in 18 seconds
983 max bufs, 1048576 max memory
981 buffers split
root@pc15:~# df -h
Filesystem Size Used Avail Use% Mounted on
/dev/sda1 459G 877M 435G 1% /
none 4.0G 188K 4.0G 1% /dev
none 4.0G 0 4.0G 0% /dev/shm
none 4.0G 36K 4.0G 1% /var/run
none 4.0G 0 4.0G 0% /var/lock
none 4.0G 0 4.0G 0% /lib/init/rw

OK, now pc should go back to Ubuntu (not freebsd!!)

aptitude safe-upgrade

install necessary kernel tools:
aptitude install linux-headers-$(uname -r) fakeroot kernel-wedge
build-essential makedumpfile kernel-package libncurses5 libncurses5-dev dkms
aptitude build-dep linux

root@pc15:~# aptitude clean

root@pc15:~# mkdir -p /etc/kernel/postinst.d/
root@pc15:~# cp /usr/share/doc/kernel-package/examples/etc/kernel/postinst.d/initramfs
/etc/kernel/postinst.d/initramfs
root@pc15:~# mkdir -p /etc/kernel/postrm.d/
root@pc15:~# cp /usr/share/doc/kernel-package/examples/etc/kernel/postrm.d/initramfs
/etc/kernel/postrm.d/initramfs

Install necessary software:
GNU make. make -version. On Linux it is the standard make. It is included in Ubuntu, version: GNU Make 3.81 (lucid)
Python. python -V. It is included in Ubuntu, Python 2.6.5 (lucid)
Boost. It is a little harder to install boost in Linux, use aptitude install libboost-dev, version: libboost-dev 1.40.0.1 (lucid)
Dhcclient. It is included in Ubuntu, Internet Systems Consortium DHCP Client V3.1.3. (lucid)
Perl. "perl -v" It is included in Ubuntu, This is perl, v5.10.1 (*)
built for x86_64-linux-gnu-thread-multi
Ethtool. Installed by "aptitude install ethtool", version 6
Yacc and byacc. Yacc is included in Ubuntu, and byacc is installed by "aptitude install byacc", 'byacc -V' byacc - 1.9 20080827
OpenSSL is included in ubuntu: openssl version OpenSSL OpenSSL 0.9.8k
25 Mar 2009 (lucid)
Rsync. Rsync is included in Ubuntu by default. rsync version 3.0.7 protocol version 30 (lucid)
emulab installation list:

```sh
aptitude install ethtool autoconf2.13 libboost-dev libboost-graph-dev
libpcap-dev perl-suid byacc libssl-dev flex ntp tcsh rpm nfs-common
htop
```

```
root@pc15:~# aptitude clean

root@pc15:~# df -h
```

<table>
<thead>
<tr>
<th>Filesystem</th>
<th>Size</th>
<th>Used</th>
<th>Avail</th>
<th>Use%</th>
<th>Mounted on</th>
</tr>
</thead>
<tbody>
<tr>
<td>/dev/sda1</td>
<td>459G</td>
<td>1.1G</td>
<td>435G</td>
<td>1%</td>
<td>/</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>192K</td>
<td>4.0G</td>
<td>1%</td>
<td>/dev/shm</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>0</td>
<td>4.0G</td>
<td>0%</td>
<td>/dev/shm</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>68K</td>
<td>4.0G</td>
<td>1%</td>
<td>/var/run</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>0</td>
<td>4.0G</td>
<td>0%</td>
<td>/var/lock</td>
</tr>
<tr>
<td>none</td>
<td>4.0G</td>
<td>0</td>
<td>4.0G</td>
<td>0%</td>
<td>/lib/init/rw</td>
</tr>
</tbody>
</table>

Install Utah's pubsub headers and libraries. Using latest pubsub-0.9.
```
p15# wget http://www.emulab.net/downloads/pubsub-0.9.tar.gz
pc15# tar zxvf pubsub-0.9.tar.gz
pc15# cd pubsub-0.9
pc15# make client
pc15# make install-client
```

aptitude update
aptitude clean

remove network interface orders in /etc/udev/rules.d/
```
root@pc1:/etc/udev/rules.d# rm 70-persistent-net.rules
```
To view a list of loadable kernel modules on a system, as well as their status, run:
```
# lsmod
```

Mixed Up Devices, Sound/Network Cards Changing Order Each Boot
Because udev loads all modules asynchronously, they are initialized in a different order. This can result in devices randomly switching names. For example, with two network cards, you may notice a switching of designations between eth0 and eth1.
following this link:
http://www.paulgraydon.co.uk/geeky/nixtricks/changing-the-order-modules-load-in-ubuntu/ to change the driver module order, the blacklist should be /etc/modprobe.d/blacklist.conf
in /etc/modprobe.d/blacklist.conf:
```
# blacklisted the following modules to ensure load order follows
/etc/modules
# Don't remove without consulting ccui
blacklist myri10ge
```
in /etc/modules:
```
loop
install myri10ge NIC driver by using dkms
1) sudo aptitude install dkms

# tar -C /usr/src -zxvf mxge/myri10ge-linux.1.5.3.tgz

linux.1.5.3:
(in Makefile)
MYRI10GE_LRO=1
MYRI10GE_BUILTIN_FW=1
MYRI10GE_ALLOC_ORDER=2
MYRI10GE_RX_SKBS=0
MYRI10GE_VPUMP=0
MYRI10GE_THROTTLE=0
MYRI10GE_JUMBO=1
MYRI10GE_KERNEL_LRO=1
MYRI10GE_PRIV_LRO=0
MYRI10GE_PKTCTNT=0
MYRI10GE_KABI=0

3) Add the module to DKMS
   # dkms add -m myri10ge -v linux.1.5.3
   # dkms build -m myri10ge -v linux.1.5.3
   # dkms install -m myri10ge -v linux.1.5.3
4) Confirm driver has been added:
   % dkms status
       myri10ge, linux.1.5.2, 2.6.32-22-generic, x86_64: installed
5) Load the driver
   # modprobe myri10ge
6) To uninstall:
   # dkms remove -m myri10ge -v linux.1.5.3 --all
   # rm -rf /usr/src/myri10ge-linux.1.5.3
   # rmmod myri10ge

# network stack tuning in /etc/sysctl.conf
net.ipv4.ip_forward = 1
net.ipv4.tcp_timestamps = 0
net.ipv4.tcp_sack = 1
LANET GROUP LSU

net.ipv4.tcp_dsack = 1
net.ipv4.tcp_fack = 1
net.ipv4.tcp_window_scaling = 1
net.ipv4.tcp_app_win = 31
net.ipv4.tcp_mem = 65536 131072 262144
net.ipv4.udp_mem = 65536 131072 262144
net.ipv4.tcp_rmem = 8192 134217728 268435456
net.ipv4.tcp_wmem = 8192 134217728 268435456
net.ipv4.tcp_moderate_rcvbuf = 1
net.ipv4.tcp_no_metrics_save = 1
net.core.rmem_max = 134217728
net.core.wmem_max = 134217728
net.core.optmem_max = 1048576
net.core.netdev_max_backlog = 15000

/sbin/ifconfig eth4 txqueuelen 15000 # take effect at run time

finished install myri10ge NIC driver

root@pc15:~# df -h
Filesystem Size Used Avail Use% Mounted on
/dev/sda1 459G 1.1G 435G 1% /
none 3.9G 176K 3.9G 1% /dev
none 3.9G 0 3.9G 0% /dev/shm
none 3.9G 64K 3.9G 1% /var/run
none 3.9G 0 3.9G 0% /var/lock
none 3.9G 0 3.9G 0% /lib/init/rw

before install new kernel and emulab client side, back up the image again:
PC need to goto freebsd, not Ubuntu!!!

boss# ssh pc1 volatile imagezip /dev/da0 >
/usr/testbed/images/template/pchpdl160G6-UBUNTU10-64-V2.ndz
Slice 3 is unused, NOT SAVING.
Slice 4 is unused, NOT SAVING.
499999866880 input (8997855232 compressed) bytes in 377.724 seconds
Image size: 314572800 bytes
22.718MB/second compressed
Finished in 392.031 seconds

PC go back to Ubuntu , not FreeBSD!!
In folder "src/client", you should have the newest CRON source.
But it should
root@pc15:~# mkdir tmp
root@pc15:~# scp -rp ccui@ops:/src/client ~/tmp/

(DEC 2011 version:
node1:/src/client> pwd
/users/linxue/src/client
node1:/src/client> ls
CRON-src-basedonEmulab-stable-December2011.tgz)
root@pc15:/~tmp/client# tar zxvf Jun08-2011CRON-src-basedonEmulab-stable-April2011.tgz

pc15# cp -p /etc/passwd /etc/group /etc/shadow /etc/gshadow
/etc/hosts ~/tmp/client/CRON-src-basedonEmulab-stable-April2011/tmcd/ubuntu10

In the Dec 2011 version,

root@pc15:/~tmp/client/CRON-src-basedonEmulab-stable-April2011/tmcd/ubuntu10/

pc15# mkdir ~/tmp/obj
pc15# cd ~/tmp/obj/
root@pc45:/~tmp/obj# /root/tmp/client/CRON-src-basedonEmulab-stable-April2011/configure --with-
TBDEFS=/root/tmp/client/CRON_definition_file/defs/CRON-publicNet

Please cd into your source directory and run:
git submodule init
git submodule update
Then rerun configure in your object tree
apt-get install git-core

In the Dec 2011 version,
/root/tmp/client/emulab-stable/configure --with-
TBDEFS=/root/tmp/client/emulab-stable/defs/CRON-publicNet

root@pc15:/~tmp/obj# make client
root@pc15:/~tmp/obj# make client-install

in ccui/linxue/cchiu’ folders, there is a folder "image_customize" copy "image_customize" into "/root/tmp/client/".

pc15# cp -rp /root/tmp/client/image_customize/ssh_host_* /etc/ssh/
pc15# chmod 0600 /etc/ssh/ssh_host_dsa_key
pc15# chmod 0600 /etc/ssh/ssh_host_rsa_key
root@pc15:/~tmp/client/image_customize# chmod 0640 client.pem
pc15# cp -p /root/tmp/client/image_customize/{client,emulab}.pem /etc/emulab/
pc15# cp -p /etc/passwd /etc/group /etc/shadow /etc/gshadow
/etc/hosts /etc/emulab/

Finalize image:
sudo apt-get clean

root@pc15:~# userdel -r yourusername
Cleanup the filesystem prior to making the image.
clean up /usr/obj/, remove created user account, remove all source
code
cd /usr/local/etc/emulab
# ./prepare
root@pc15:~# rm -rf ~/*
root@pc15:~# aptitude clean
root@pc15:~/.ssh# vim authorized_keys          ===add root's public key
#  DO NOT EDIT! This file auto generated by Emulab.Net account software.
#  Please use the web interface to edit your public key list.
#  ssh
  rsa
AAAAB3NzaC1yc2EAAAABIAwAAAAQEA63VhWAXu72TlI7HO3aststGd9QdFU9r1LANhoLxkD43U117yg2SLLo+cD03F9f3XEvBWRQwrBstzEGDJbiNFFMLrDh+r/MyO+SlXt78JGe yfWQQWXgzTzVHYfEOW9fDoq7SEG1MwF7ELSIIsXnAwa6XMmlt6sIARulotHhLGrhKvd65ZEdpojf
VmNHfLweHXWG5hV6tz7X99sDqJykD1zch6tXWF8IKUkUH95Dtxb3BzSt/Q0ggMhqr107pyIEaxXwGbXFyazgsBtmeWDSixY12YLosOTkY762w59FrpW5KzKHgKqcXAQxJkXTsdBuTc0
7zvIoHemBmMeC6Dvh5Q== root@boss.cron.cct.lsu.edu

root@pc15:~# df -h
Filesystem           Size  Used Avail Use% Mounted on
/dev/sda1             46G  1.1G  43G   3% /
none                  3.9G  208K  3.9G   1% /dev
none                  3.9G   3.9G  0% /dev/shm
none                  3.9G  40K  3.9G   1% /var/run
none                  3.9G   0  3.9G   0% /var/lock
none                  3.9G   0  3.9G   0% /lib/init/rw

back up a final clean image, the final image!!:
 pc15 need go to Freebsd!!
 boss# ssh pc15 imagezip  /dev/da0  - >
 /usr/testbed/images/template/pcHPDL160G6-UBUNTU10-64-final.ndz
  Slice 3 is unused, NOT SAVING.
  Slice 4 is unused, NOT SAVING.
  50000330752 input (1928785920 compressed) bytes in 76.167 seconds
  Image size: 305135616 bytes
  24.150MB/second compressed
  Finished in 77.645 seconds

back up a customized image:
boss# ssh pc15 imagezip  /dev/da0  - > /usr/testbed/images/
pcHPDL160G6-UBUNTU10-64.ndz
  Slice 3 is unused, NOT SAVING.
  Slice 4 is unused, NOT SAVING.
  499999866880 input (9188380672 compressed) bytes in 386.290 seconds
  Image size: 363855872 bytes
  22.684MB/second compressed
  Finished in 400.592 seconds
Step 8. Create image type

Create a new Image Descriptor as in following figure:

![Image Descriptor Form]

Step 9. FINAL TEST!!!

test by experiment swapin==> succeed!!!!!

Step 10. Enjoy the new servers!!!