Subject: Feisty VE breaks Edgy HN
Posted by smckown on Wed, 27 Jun 2007 15:51:14 GMT
View Forum Message <> Reply to Message

I have created an OpenVZ setup for testing. This is the HN configuration:

Dell PowerEdge 1800, dual Xeon
OpenVZ kernel 2.6.18-028stab035.1 from debian.systs.org (openvz stable)
vzctl version 3.0.16-5dso1
Ubuntu 6.10 "Edgy" server

On this system, and using the assistance of http://wiki.openvz.org/Physical_to_VE, I was able to migrate a quite old Mandrake 9.1 physical server to a VE, which runs well with no problems. So, I believe the HN/OpenVZ configuration to be sound.

However, runing a Feisty VE does create a problem. I have read through the related topic at http://forum.openvz.org/index.php?t=tree&th=2297&mid =11810&&rev=&reveal= and Upstart bug #87173 (https://bugs.launchpad.net/upstart/+bug/87173).

The Feisty VE will start with the most notable changes:

Apply the file descriptor patch to the VE's upstart:
http://codebrowse.launchpad.net/~keybuk/upstart/main/revisio
n/scott%40netsplit.com-20070313191319-gztu8c0r0sjla0hp?start
_revid=scott%40netsplit.com-20070316171800-scmrd6w9r22uf4me
Change the kill -USR1 1 in the VE's /etc/init.d/mountall.sh script to use TERM instead.

However, the VE's init process and upstart's domain socket seem to be leaking between the VE and the HN. On the HN before starting the VE:

```
sysadmin@pe18001:~$ ps -ef | grep init | grep -v grep
root         1     0  0 08:45 ?        00:00:00 /sbin/init splash
sysadmin@pe18001:~$ sudo netstat -anp | grep init | grep -v grep
unix  2      [ ]         DGRAM                    4771     1/init              @/com/ubuntu/upstart
sysadmin@pe18001:~$ sudo initctl list
tty1 (start) running, process 6552 active
tty2 (start) running, process 6553 active
tty3 (start) running, process 6554 active
tty4 (start) running, process 6555 active
tty5 (start) running, process 6556 active
tty6 (start) running, process 6557 active
rc-default (stop) waiting
rc0 (stop) waiting
rc0-halt (stop) waiting
rc0-poweroff (stop) waiting
rc1 (stop) waiting
```

```
rc2 (stop) waiting
rc3 (stop) waiting
rc4 (stop) waiting
rc5 (stop) waiting
rc6 (stop) waiting
rcS (stop) waiting
rcS-sulogin (stop) waiting
logd (start) running, process 4115 active
control-alt-delete (stop) waiting
sulogin (stop) waiting
ttyS0 (start) running, process 6562 active
sysadmin@pe18001:~$
```

Start the VE:

```
sysadmin@pe18001:~$ sudo vzctl start 132
Starting VE ...
Mount partition ... done
VE is mounted
Adding IP address(es): 172.16.0.132
Setting CPU units: 1000
Configure meminfo: 49152
File resolv.conf was modified
VE start in progress...
sysadmin@pe18001:~$ sudo vzlist
      VEID      NPROC STATUS  IP_ADDR         HOSTNAME
       132          5 running 172.16.0.132    -
sysadmin@pe18001:~$
```

PS - the extra output "Mount partition ... done" is from a custom vps.mount that mounts the VE's private LVM LV.

Now that the VE is running, things look a bit weird for HN's init and its domain socket. Note the extra init process and the extra domain socket:

```
sysadmin@pe18001:~$ sudo ps -ef | grep init
root         1     0  0 08:45 ?        00:00:00 /sbin/init splash
root      8629     1  0 09:07 ?        00:00:00 init
sysadmin  9049  6755  0 09:07 ttyS0    00:00:00 grep init
sysadmin@pe18001:~$ sudo netstat -anp | grep init
unix  2      [ ]         DGRAM                    27118    8629/init           @/com/ubuntu/upstart
unix  2      [ ]         DGRAM                    4771     1/init              @/com/ubuntu/upstart
sysadmin@pe18001:~$ sudo initctl list
    (hangs for minutes, must hit ^C to interrupt)
sysadmin@pe18001:~$
```

Inside the VE, things look pretty good, but I would expect only one domain socket for init. Note the second one has no domain name.

```
sysadmin@pe18001:~$ sudo vzctl enter 132
entered into VE 132
root@ubuntuvm:/# ps -ef
UID        PID  PPID  C STIME TTY          TIME CMD
root         1     0  0 15:07 ?        00:00:00 init
root     10048     1  0 15:07 ?        00:00:00 /sbin/syslogd
root     10078     1  0 15:09 ?        00:00:00 vzctl: pts/0
root     10079 10078  0 15:09 pts/0    00:00:00 -bash
root     10092 10079  0 15:09 pts/0    00:00:00 ps -ef
root@ubuntuvm:/# netstat -anp | grep init
unix  2      [ ]         DGRAM                    27118    1/init              @/com/ubuntu/upstart
unix  2      [ ]         DGRAM                    27637    1/init
root@ubuntuvm:/# ll /proc/1/fd
total 8
lrwx------ 1 root root 64 Jun 27 15:10 0 -> /dev/null
lrwx------ 1 root root 64 Jun 27 15:10 1 -> /dev/null
lrwx------ 1 root root 64 Jun 27 15:10 2 -> /dev/null
lr-x------ 1 root root 64 Jun 27 15:10 3 -> pipe:[27117]
l-wx------ 1 root root 64 Jun 27 15:10 4 -> pipe:[27117]
lrwx------ 1 root root 64 Jun 27 15:10 5 -> socket:[27118]
lr-x------ 1 root root 64 Jun 27 15:10 6 -> inotify
lrwx------ 1 root root 64 Jun 27 15:10 7 -> socket:[27637]
root@ubuntuvm:/# initctl list
control-alt-delete (stop) waiting
logd (stop) waiting
rc-default (stop) waiting
rc0 (stop) waiting
rc1 (stop) waiting
rc2 (stop) waiting
rc3 (stop) waiting
rc4 (stop) waiting
rc5 (stop) waiting
rc6 (stop) waiting
rcS (stop) waiting
rcS-sulogin (stop) waiting
sulogin (stop) waiting
root@ubuntuvm:/# runlevel
N 2
root@ubuntuvm:/# logout
exited from VE 132
sysadmin@pe18001:~$
```

Once the VE is stopped, the HN's init processes and domain socket listing appear to return to

normal:

```
sysadmin@pe18001:~$ sudo vzctl stop 132
Stopping VE ...
VE was stopped
VE is unmounted
sysadmin@pe18001:~$ ps -ef | grep init
root         1     0  0 08:45 ?        00:00:00 /sbin/init splash
sysadmin 12308  6755  0 09:29 ttyS0    00:00:00 grep init
sysadmin@pe18001:~$ sudo netstat -anp | grep init
unix  2      [ ]         DGRAM                    4771     1/init              @/com/ubuntu/upstart
sysadmin@pe18001:~$ sudo initctl list
tty1 (start) running, process 6552 active
tty2 (start) running, process 6553 active
tty3 (start) running, process 6554 active
tty4 (start) running, process 6555 active
tty5 (start) running, process 6556 active
tty6 (start) running, process 6557 active
rc-default (stop) waiting
rc0 (stop) waiting
rc0-halt (stop) waiting
rc0-poweroff (stop) waiting
rc1 (stop) waiting
rc2 (stop) waiting
rc3 (stop) waiting
rc4 (stop) waiting
rc5 (stop) waiting
rc6 (stop) waiting
rcS (stop) waiting
rcS-sulogin (stop) waiting
logd (start) running, process 4115 active
control-alt-delete (stop) waiting
sulogin (stop) waiting
ttyS0 (start) running, process 6562 active
sysadmin@pe18001:~$
```

BTW, networking seems to work fine and since the command captures above I've installed openssh on the VE and it also works correctly.

I suspect an OpenVZ virtualization error here, but don't know enough to confirm this. I also suspect this only happens because the HN OS and the VE OS both use a domain socket of the same 'name' or something. Can I get some help? I'm happy to do some more testing here or send my Feisty VE.

PS - I've documented the process of bringing up OpenVZ on Edgy and building the Feisty template. After this problem is solved and I've done some more testing, I plan to contribute a couple of wiki pages and a template cache for others.

Thanks,
Steve

Subject: Re: Feisty VE breaks Edgy HN
Posted by Vasily Tarasov on Thu, 28 Jun 2007 11:48:32 GMT
View Forum Message <> Reply to Message

Thank you for your future contributions

First of all, where can I download your Feisty VE? I will start it on my CentOS host-node and check, how many sockets this VE will create on my mashine.

Thanks,
Vasily.

Subject: Re: Feisty VE breaks Edgy HN
Posted by smckown on Fri, 29 Jun 2007 19:26:34 GMT
View Forum Message <> Reply to Message

Hi Vasily,

I sent you the link to the VE via PM. If you didn't get it, please let me know. Thanks again!

Subject: Re: Feisty VE breaks Edgy HN
Posted by hoppaz on Tue, 01 Jul 2008 17:13:04 GMT
View Forum Message <> Reply to Message

Is there a fix available for this problem?

Using 2.6.18-53.1.19.el5.028stab053.14PAE/fedora core 9/upstart.

Subject: Re: Feisty VE breaks Edgy HN
Posted by smckown on Tue, 01 Jul 2008 17:27:17 GMT
View Forum Message <> Reply to Message

I replaced upstart with sysvinit in my Feisty template, from which I create all my VEs. This strategy also works for Gutsy VE's, but I never looked closer to see if Gutsy VE's would work without changing out upstart for sysvinit.

Here's one link. Several more via google.
http://ubuntuforums.org/showthread.php?t=456414.

Best,
Steve

Subject: Re: Feisty VE breaks Edgy HN
Posted by hoppaz on Wed, 02 Jul 2008 14:38:02 GMT
View Forum Message <> Reply to Message

I patched 3 files in upstart sources and changed the path /com/ubuntu/upstart to /ve/com/ubuntu/upstart for ve. While upstart is running with the old path in ve0...

This fixes the initctl-problem.

I think this is a security issue between upstart and openvz.

Question to developer: Is this a problem which has to get fixed in openvz?

File Attachments
1) upstart-ve.patch, downloaded 338 times

Subject: Re: Feisty VE breaks Edgy HN
Posted by hoppaz on Thu, 03 Jul 2008 09:05:58 GMT
View Forum Message <> Reply to Message

I mailed to upstart developer Scott James Remnant and described the problem. He thinks that this problem should get solved in openvz.
