
Subject: Re: [RFC][PATCH 0/7] Clone PTS namespace

Posted by [serue](#) on Thu, 10 Apr 2008 01:59:55 GMT

[View Forum Message](#) <> [Reply to Message](#)

Quoting Eric W. Biederman (ebiederm@xmission.com):

> On Tue, 2008-04-08 at 17:53 -0700, H. Peter Anvin wrote:

> > sukadev@us.ibm.com wrote:

> > > Devpts namespace patchset

> > >

> > > In continuation of the implementation of containers in mainline, we need to

> > > support multiple PTY namespaces so that the PTY index (ie the tty names) in

> > > one container is independent of the PTY indices of other containers. For

> > > instance this would allow each container to have a '/dev/pts/0' PTY and

> > > refer to different terminals.

> > >

> >

> > Why do we "need" this? There isn't a fundamental need for this to be a

> > dense numberspace (in fact, there are substantial reasons why it's a bad

> > idea; the only reason the namespace is dense at the moment is because of

> > the hideously bad handling of utmp in glibc.) Other than indicies, this

> > seems to be a more special case of device isolation across namespaces,

> > would that be a more useful problem to solve across the board?

>

> In short application migration. When you move a running applicaiton

> from one machine to another you want to be able to keep the same pseudo
> devices.

>

> The isolation that you have noticed is also an important application and

> like the rest of the namespaces if we can solve the duplicate identifier

> problem needed to restore checkpoints we also largely solve the

> isolation problem.

>

> This problem is much larger then ptys. ptys are the really in your face

> aspect of it. There are a more pseudo devices in the kernel and it is

> the device number to device mapping that we are abstracting. So this

> really should be done as a device namespace not a pty namespace.

>

> I would be happy if the first version of the device namespace could not

> map anything but pty's (assuming an incremental implementation path). I

> really don't think we should do a special case for each kind of device.

Sounds like we're all agreed on this and just doing

s/CLONE_NEWPTS/CLONE_NEWDEV/ on the current patchset suffices for now.

But,

> Oh and just skimming the patch summary I'm pretty certain this

> implementation breaks /sys/class/tty/ptyXX/uevent. Which is another

> reason why it would be good to have a single device namespace. So we

> only to capture one more namespace and figure out how to deal with it
> when mounting sysfs.

Feh, so of course sysfs would have the most interactions for a device namespace, but now we have pty, network, and user namespace all needing some sort of sysfs solution. For a quickfix for CONFIG_USER_SCHED+CONFIG_USER_NS, I just moved /sys/kernel/uids/<uid> to /sys/kernel/uids/<usersns_address>/<uid>. But what would be a *good* general solution?

In -s /sys /proc/self/sys?

-serge

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>
