Subject: Re: [RFC][PATCH 5/6] Define helper functions to unshare pid namespace
Posted by serue on Sun, 11 Mar 2007 13:21:23 GMT
View Forum Message <> Reply to Message

Quoting Eric W. Biederman (ebiederm@xmission.com):
> sukadev@us.ibm.com writes:
>
> > From: Sukadev Bhattiprolu <sukadev@us.ibm.com>
> > Subject: [RFC][PATCH 5/6] Define helper functions to unshare pid namespace
> >
> > Define clone_pid_ns() and unshare_pid_ns() functions that will be
> > used in the next patch to unshare pid namespace.
> >
> > Changelog:
> > - Rewrite of orignal code in -lxc from Cedric Le Goater to enforce
> > setsid() requirement on unshare().
>
> Why do we need a setsid() before we unshare?

If we don't do that, then the session and pgrp leaders need to get
pulled into the new namespace.

Previous versions did that, and eventually we want to support that
again, but for now to keep the rfc patches simpler this seemed the
better way to go.

We will want that for checkpoint-restart ("application") containers, to
preserve normal shell control.

> I know it is almost always the correct thing to do but what requires
> the setsid?
>
> Doing the setsid before we switch pid namespaces appears the wrong
> order to me.
>
> I am not convinced that unshare can be done safely for a pid
> namespace. Changing the meaning or definition of pid on a running
> process is questionable.

Hmm, interesting notion. On the one hand, the process explicitly asked
for the change, so it's not like it's going to get confused. So on that
basis alone I would think we should support it. On the other hand, I
can't think of anything that would ever require it - vservers will want
to clone off a fresh init. Well, maybe it keeps things shorter for
application containers. User asks shell to do
 run_container do_my_calculation
where run_container unshares and execs do_my_calculation. Adding a
clone in there seems unnecessary.

-serge

_____

Containers mailing list
Containers@lists.osdl.org
https://lists.osdl.org/mailman/listinfo/containers