Subject: Re: [RFC][PATCH 5/6] Define helper functions to unshare pid namespace
Posted by ebiederm on Sun, 11 Mar 2007 18:07:23 GMT
View Forum Message <> Reply to Message

"Serge E. Hallyn" <serue@us.ibm.com> writes:

> 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.

Yes. We need a case for dealing with pids that are not in the current
pid namespace. And the 0 return from pid_nr should work in that case.
Otherwise I don't think we really care.

So I don't think there is a special case in the code to worry about.

>> 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.
glibc cache pids. If you change them without forking things that you think
are syscalls are going to start returning the wrong values. So saying you asked
for it is no guarantee that it won't 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.

I think there are cases where unshare could make sense. I really think unshare of a pid namespace needs separate consideration from the clone case. Where things get really are multiple unshares of the pid namespace from the same process and things like that.

unshare of a pid namespace if it makes sense probably leads to Herbert's really light-weight guests with a single process.

So I think a separate conversation of what we want unshare of the pid namespace to mean needs to happen before we merge it because we have some choices and it isn't obvious.

Eric

_____

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