

---

Subject: Re: [PATCH] Allow signalling container-init  
Posted by [serue](#) on Thu, 09 Aug 2007 14:42:14 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Quoting Daniel Pittman (daniel@rimspace.net):

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

> > Quoting Daniel Pittman (daniel@rimspace.net):

> > > sukadev@us.ibm.com writes:

>

> [...]

>

> > > TODO: Ideally we should allow killing the container-init only from

> > > ancestor containers and prevent it being killed from that or

> > > descendant containers. But that is a more complex change and

> > > will be addressed by a follow-on patch. For now allow the

> > > container-init to be terminated by any process with sufficient

> > > privileges.

> >

> > This will break, as far as I can see, by allowing the container root to

> > send signals to init that it doesn't expect.

> >

> > Yes, in the end what we want is for a container init to receive

> >

> > 1. all signals from a (authorized) process in a parent

> > pid namespace.

> > 2. for signals sent from inside it's pid namespace, only

> > exactly those signals for which it has installed a

> > custom signal handler, no others.

> >

> > In other words to a process in an ancestor pid namespace, the init of a

> > container is like any other process. To a process inside the namespace

> > for which it is init, it is as /sbin/init is to the system now.

>

> That makes sense.

>

> > Actually achieving that without affecting performance for all

> > signalers is nontrivial. The current patchset is complex enough that

> > I'd like to see us settle on non-optimal semantics for now, and once

> > these patches have settled implement the ideal signaling.

>

> I appreciate that. I figured to make you aware that this will make it

> impossible to run upstart and, probably, other versions of init in your

> container as expected.

>

> Since this was a somewhat subtle bug to track down it is, I think, work

> documenting so that people trying to use this code are aware of the

> limitation.

Agreed. I do think it is documented in the code and in changelogs.  
Maybe it's worth adding a Documentation/ file describing how to use the  
pid namespaces, ideal semantics, and current shortcomings, for people  
who want to use+test the feature rather than work with the kernel code.

-serge

---

Containers mailing list

Containers@lists.linux-foundation.org

<https://lists.linux-foundation.org/mailman/listinfo/containers>

---