Subject: Re: [PATCH 1/3] Signal semantics for /sbin/init
Posted by Sukadev Bhattiprolu on Fri, 14 Sep 2007 03:00:53 GMT
View Forum Message <> Reply to Message

Oleg Nesterov [oleg@tv-sign.ru] wrote:
| On 09/13, Cedric Le Goater wrote:
| >
| > Oleg Nesterov wrote:
| > > On 09/10, sukadev@us.ibm.com wrote:
| > >> (This is Oleg's patch with my pid ns additions. Compiled and unit tested
| > >> on 2.6.23-rc4-mm1 with other patches in this set. Oleg pls update this
| > >> patch if necessary and sign-off)
| > >
| > > Sukadev, my apologies. This patch does need some changes,
| > >
| > >> Notes:
| > >>
| > >> - Blocked signals are never ignored, so init still can receive
| > >>   a pending blocked signal after sigprocmask(SIG_UNBLOCK).
| > >>   Easy to fix, but probably we can ignore this issue.
| > >
| > > I was wrong. This should be fixed right now. I _think_ this is easy,
| > > and I was going to finish this patch yesterday, but - sorry! - I just
| > > can't switch to "kernel mode" these days, I am fighting with some urgent
| > > tasks on my paid job.
| > >
| > To respect the current init semantic,
|
| The current init semantic is broken in many ways ;)
|
| > shouldn't we discard any unblockable
| > signal (STOP and KILL) sent by a process to its pid namespace init process ?

Yes. And Patch 1/3 (Oleg's patch) in the set I sent, handles this already
(since STOP and KILL are never in the task->blocked list)

| > Then, all other signals should be handled appropriately by the pid namespace
| > init.

|
| Yes, I think you are probably right, this should be enough in practice. After all,
| only root can send the signal to /sbin/init.

I agree - the assumption that the container-init will handle these
other signals, simplifies the kernel implementation for now.

| On my machine, /proc/1/status shows that init doesn't have a handler for
| non-ignored SIGUNUSED == 31, though.
|
| But who knows? The kernel promises some guarantees, it is not good to break them.
| Perhaps some strange non-standard environment may suffer.
|
| > We are assuming that the pid namespace init is not doing anything silly and
| > I guess it's OK if the consequences are only on the its pid namespace and
| > not the whole system.
|
| The sub-namespace case is very easy afaics, we only need the "signal comes from
| the parent namespace" check, not a problem if we make the decision on the sender's
| path, like this patch does.

Yes, patches 2 and 3 of the set already do the ancestor-ns check. no ?

|
| Oleg.

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