Subject: Re: + remove-the-likelypid-check-in-copy_process.patch added to -mm tree
Posted by Oleg Nesterov on Sat, 17 Mar 2007 17:17:47 GMT
View Forum Message <> Reply to Message

On 03/17, Eric W. Biederman wrote:
>
> Oleg Nesterov <oleg@tv-sign.ru> writes:
>
> > On 03/17, Oleg Nesterov wrote:
> >>
> >> > Well the initial kernel process does not have a struct pid so when
> >> > it's children start doing:
> >> > attach_pid(p, PIDTYPE_PGID, task_group(p));
> >> > attach_pid(p, PIDTYPE_SID, task_session(p));
> >> > We will get an oops.
> >>
> >> So far this is the only reason to have init_struct_pid. Because the
> >> boot CPU (swapper) forks, right?
> >
> > Damn. I am afraid I was not clear again :) Not init_struct_pid, but
> >
> > + .pids = { \
> > + [PIDTYPE_PID] = INIT_PID_LINK(PIDTYPE_PID), \
> > + [PIDTYPE_PGID] = INIT_PID_LINK(PIDTYPE_PGID), \
> > + [PIDTYPE_SID] = INIT_PID_LINK(PIDTYPE_SID), \
> > + }, \
> >
> > for INIT_TASK().
> >
> >> > So a dummy unhashed struct pid was added for the idle threads.
> >> > Allowing several special cases in the code to be removed.
> >> >
> >> > With that chance the previous special case to force the idle thread
> >> > init session 1 pgrp 1 no longer works because attach_pid no longer
> >> > looks at the pid value but instead at the struct pid pointers.
> >> >
> >> > So we had to add the __set_special_pids() to continue to keep init
> >> > in session 1 pgrp 1. Since /sbin/init calls setsid() that our setting
> >> > the sid and the pgrp may not be strictly necessary. Still is better
> >> > to not take any chances.
> >>
> >> Yes, yes, I see. But my (very unclear, sorry) question was: shouldn't we
> >> change INIT_SIGNALS then? /sbin/init inherits ->pgrp == ->_session == 1,
> >> in that case __set_special_pids(1,1) does nothing.
> >
> > ... and thus /sbin/init remains attached to the .pids above, no?
>

> The problem is that we dynamically allocate the struct pid for
> pid_t == 1 when we fork init.
>
> Which means we don't have access to it at compile time so we can
> no longer make INIT_SIGNALS set ->gprp == ->session == 1.

Yes! I meant we should change INIT_SIGNALS(), currently it does

```
#define INIT_SIGNALS(sig) {
  ...
  .pgrp          = 1,
  { .__session      = 1},
```

and this confuses (I think) set_special_pids(1,1) above. Because
__set_special_pids() still deals with pid_t, not "struct pid".

Unless I missed something, we should kill these 2 initializations
above.

Oleg.

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