Subject: Re: + remove-the-likelypid-check-in-copy_process.patch added to -mm tree
Posted by Oleg Nesterov on Sat, 17 Mar 2007 15:09:49 GMT
View Forum Message <> Reply to Message

On 03/17, Eric W. Biederman wrote:
>
> Oleg Nesterov <oleg@tv-sign.ru> writes:
>
> > --- a/init/main.c~explicitly-set-pgid-and-sid-of-init-process
> > +++ a/init/main.c
> > @@ -783,6 +783,7 @@ static int __init init(void * unused)
> > */
> > init_pid_ns.child_reaper = current;
> >
> > + __set_special_pids(1, 1);
> > cad_pid = task_pid(current);
> >
> > smp_prepare_cpus(max_cpus);
> >
> > Nice changelog :)
> >
> > The patch looks good, except __set_special_pids(1, 1) should be no-op.
> > This is a child forked by swapper. copy_process() was changed by
> > use-task_pgrp-task_session-in-copy_process.patch
> > , but signal->{pgrp,_session} get its value from INIT_SIGNALS ?
> >
> > Could you explain this as well? Some other changes I missed?
>
> As I recall the patch series started with modifying attach_pid
> to take a struct pid pointer instead of a pid_t value. It means
> fewer hash table looks ups and it should help in implementing the pid
> namespace.
>
> 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?

> 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.

> Anyway the point of removing the likely(pid) check was that it didn't
> look necessary any longer. But as you have correctly pointed putting
> it on the task list and incrementing the process count for the idle
> threads is probably still a problem.

Yes. Note also that the parent doing fork_idle() is not always swapper,
it is just wrong to do attach_pid(PIDTYPE_PGID/PIDTYPE_SID) in this case.
example: arch/x86_64/kernel/smpboot.c:do_boot_cpu()

> So while we are much better we
> still have some use for the if (likely(p->pid)) special case.

Yes, I think this change should be dropped for now.

> Is that enough to bring you up to speed?

Thanks for your explanations!

Oleg.

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