Subject: Re: [PATCH 14/15] Destroy pid namespace on init's death
Posted by serue on Thu, 02 Aug 2007 19:13:35 GMT
View Forum Message <> Reply to Message

Quoting Oleg Nesterov (oleg@tv-sign.ru):
> On 08/02, sukadev@us.ibm.com wrote:
> >
> > Oleg Nesterov [oleg@tv-sign.ru] wrote:
> > |
> > | This means that we should take care about multi-thread init exit,
> > | otherwise the non-root user can crash the kernel.
> > |
> > | >From reply to Kirill's message:
> > |
> > | > Still. A non-root user does clone(CLONE_PIDNS), then clone(CLONE_THREAD),
> >
> > Agree we should fix the crash. But we need CAP_SYS_ADMIN to clone
> > pid or other namespaces - this is enforced in copy_namespaces() and
> > unshare_nsproxy_namespaces()
>
> Hmm. sys_unshare(CLONE_PIDNS) doesn't (and shouldn't) work anyway, but
> I don't see the CAP_SYS_ADMIN check in copy_process()->copy_namespaces()
> path.
>
> Perhaps I just missed it (sorry, I already cleared my mbox, so I can't
> look at theses patches), but is it a good idea to require CAP_SYS_ADMIN?
> I think it would be nice if a normal user can create containers, no?

For pid namespaces I can't think of any reason why CAP_SYS_ADMIN should
be needed, since you can't hide processes that way. Same for uts
namespaces.

However for ipc and mount namespaces I'd want to think about it some
more. Any case I can think of right now where they'd be unsafe, is
unsafe anyway, so maybe it's fine...

-serge