Subject: Re: NFS causing oops when freeing namespace
Posted by ebiederm on Thu, 18 Jan 2007 04:46:15 GMT
View Forum Message <> Reply to Message

Oleg Nesterov <oleg@tv-sign.ru> writes:

> On 01/17, Eric W. Biederman wrote:
>>
>> Cedric Le Goater <clg@fr.ibm.com> writes:
>> >
>> > your first analysis was correct : exit_task_namespaces() should be moved
>> > above exit_notify(tsk). It will require some extra fixes for nsproxy
>> > though.
>>
>> I think the only issue is the child_reaper and currently we only have one of
>> those. When we really do the pid namespace we are going to have to revisit
>> this. My gut feel says that we won't be able to exit our pid namespace until
>> the process is waited on. So we may need to break up exit_task_namespace into
>> individual components.
>
> I agree, but please note that the child_reaper is not the only issue.

To be clear I believe the only issue keeping us from moving exit_namespaces
back where it used to be is the child reaper as that is the only part of
the pid namespace that has been implemented.

There is more when we revisit this.
> Think
> about sub-thread which auto-reaps itself. I'd suggest to add the comment in
> do_exit() after exit_notify() to remind that the task is really dead now, it
> has no ->signal, it can't be seen in /proc/, we can't send a signal to it, etc.

A very interesting case is what happens when we reparent a zombie. I think
that needs the child reaper and it happens well after exit_namespaces is currently
being called.

In the very stupid test we need our struct pid that identifies the process until
we are reaped. Therefore our pid namespace must continue to exist, even if we
don't keep a pointer to it in struct nsproxy.

A comment after exit_notify would certainly be useful.

Eric

_____
Containers mailing list
Containers@lists.osdl.org
https://lists.osdl.org/mailman/listinfo/containers