Subject: Re: NFS causing oops when freeing namespace
Posted by ebiederm on Wed, 17 Jan 2007 22:30:16 GMT

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.

Eric

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

Subject: Re: NFS causing oops when freeing namespace
Posted by Trond Myklebust on Wed, 17 Jan 2007 22:46:42 GMT

On Wed, 2007-01-17 at 15:30 -0700, 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.

It makes little sense, afaics, to have an interruptible sleep in
something like lockd_down() if you have no pid space or signal handling.

That isn't the only place where the process has to wait in an NFS
unmount, BTW. Things like rpc client cleanup (waiting for all the RPC
tasks to complete) also tend to lead to interruptible sleeps.

Trond

Subject: Re: NFS causing oops when freeing namespace
Posted by Oleg Nesterov on Wed, 17 Jan 2007 22:55:24 GMT
View Forum Message <> Reply to Message

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. 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.

Oleg.