
Subject: Re: [PATCH 1/15] Move exit_task_namespaces()
Posted by [Oleg Nesterov](#) on Thu, 26 Jul 2007 16:47:24 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 07/26, Pavel Emelyanov wrote:

>
> Make task release its namespaces after it has reparented all his
> children to child_reaper, but before it notifies its parent about
> its death.
>
> The reason to release namespaces after reparenting is that when task
> exits it may send a signal to its parent (SIGCHLD), but if the parent
> has already exited its namespaces there will be no way to decide what
> pid to dever to him - parent can be from different namespace.
>
> The reason to release namespace before notifying the parent it that
> when task sends a SIGCHLD to parent it can call wait() on this taks
> and release it. But releasing the mnt namespace implies dropping
> of all the mounts in the mnt namespace and NFS expects the task to
> have valid sighand pointer.
>
> Signed-off-by: Pavel Emelyanov <xemul@openvz.org>
>
> ---
>
> exit.c | 5 ++++-
> 1 files changed, 4 insertions(+), 1 deletion(-)
>
> diff -upr linux-2.6.23-rc1-mm1.orig/kernel/exit.c
> linux-2.6.23-rc1-mm1-7/kernel/exit.c
> --- linux-2.6.23-rc1-mm1.orig/kernel/exit.c 2007-07-26
> 16:34:45.000000000 +0400
> +++ linux-2.6.23-rc1-mm1-7/kernel/exit.c 2007-07-26
> 16:36:37.000000000 +0400
> @@ -788,6 +804,10 @@ static void exit_notify(struct task_stru
> BUG_ON(!list_empty(&tsk->children));
> BUG_ON(!list_empty(&tsk->ptrace_children));
>
> + write_unlock_irq(&tasklist_lock);
> + exit_task_namespaces(tsk);
> + write_lock_irq(&tasklist_lock);

No.

We "cleared" our ->children/->ptrace_children lists. Now suppose that another thread dies, and its forget_original_parent() choose us as a new reaper before we re-take tasklist.

I'll try to read other patches tomorrow, but I can't avoid a stupid question: can we have a CONFIG_ for that? This series adds a lot of complications.

OK, I guess the answer is "it is very difficult to achieve", but can't help myself :)

Oleg.
