Subject: Re: [RFC][PATCH 06/16] Define is_global_init()
Posted by Dave Hansen on Tue, 29 May 2007 16:54:32 GMT
View Forum Message <> Reply to Message

On Fri, 2007-05-25 at 13:44 -0700, sukadev@us.ibm.com wrote:
> Dave Hansen [hansendc@us.ibm.com] wrote:
> | On Thu, 2007-05-24 at 13:24 +0400, Pavel Emelianov wrote:
> | > > | > +int is_global_init(struct task_struct *tsk)
> | > > | > +{
> | > > | > + return (task_active_pid_ns(tsk) == &init_pid_ns && tsk->pid == 1);
> | > > |
> | > > | This can OOPS if you pass arbitrary task to this call...
> | > > | tsk->nsproxy can already be NULL.
> | > >
> | > > Hmm. You are right. btw, this could be a bisect issue. Patch 9 of uses
> | > > pid_ns from pid->upid_list and removes nsproxy->pid_ns.
> | >
> | > Yes, but that patch is not good either.
> | > task_pid(tsk) may become NULL as well and this will oops.
> |
> | Have you reviewed the call paths to make sure this can actually happen
> | in practice?
>
> task_pid() can be NULL when we are tearing down the task structure in
> release_task() and in the tiny window between detach_pid() and attach_pid()
> in de_thread().
>
> I think task_pid() is safe as long as it is called for 'current'. (we should
> probably add some comments)

If we only call it for "current", then perhaps we should just change the
function so that it doesn't take any arguments. That way nobody can
screw it up.

> I will double check my code, but I think all my calls to task_pid() and hence,
> to task_active_pid_ns() are safe, except for two cases:
>
>   a) is_global_init(). There are a few calls to process other than
>   current, but not sure if they are a problem.
>
>   For instance in current code, unhandled_signal() checks
>   tsk->pid == 1 and proceeds to derefernce tsk->sighand.
>
>   If task_pid() is NULL because the task was in release_task(),
>   then so is tsk->sighand.

Really? Are there barriers or locks to make this happen? Can you be
sure that compiler or cpu re-ordered code will keep this true?

>   b) the temporary check I added in check_kill_permissions().
>   (I need to address Serge's comment here anyway).
>
> To make is_global_init() more efficient and independent of task_pid(),
> can we steal a bit from task_struct->flags ? Like PF_KSWAPD, and there
> are unused bits :-)

It feels to me like we're adding too many hacks on hacks here. Let's
define the problem, because it sounds to me like we don't even know what
it really is. What is the problem here, again?

-- Dave

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