Subject: Re: [RFC][PATCH 1/3] Pid ns helpers for signals
Posted by Oleg Nesterov on Thu, 30 Aug 2007 08:21:48 GMT
View Forum Message <> Reply to Message

On 08/29, sukadev@us.ibm.com wrote:
>
> +static int ancestor_pid_ns(struct pid_namespace *ns1, struct pid_namespace *ns2)
> +{
> + int i;
> + struct pid_namespace *tmp;
> +
> + if (ns1 == NULL || ns2 == NULL)
> + return 0;
> +
> + if (ns1->level >= ns2->level)
> + return 0;

Looks like this check is not needed, because

> + tmp = ns2->parent;
> + for (i = tmp->level; i >= ns1->level; i--) {
> + if (tmp == ns1)
> + return 1;
> + tmp = tmp->parent;
> + }
> +
> + return 0;
> +}

"for ()" does the necessary comparison. The same for descendant_pid_ns().

But do we really need two different functions with 2 arguments? Afaics, we
only need is_current_ancestor_pid_ns(tsk).

kill_something_info() needs is_current_ancestor_or_the_same_pid_ns(tsk) which
could be trivially implemented using the previous one.

> --- 2.6.23-rc3-mm1.orig/include/linux/sched.h 2007-08-27 20:04:23.000000000 -0700
> +++ 2.6.23-rc3-mm1/include/linux/sched.h 2007-08-28 23:12:54.000000000 -0700
> @@ -1290,6 +1290,23 @@ static inline pid_t task_ppid_nr_ns(stru
> return pid_nr_ns(task_pid(rcu_dereference(tsk->real_parent)), ns);
> }
>
> +static inline struct pid_namespace *pid_active_ns(struct pid *pid)
> +{
> + if (pid == NULL)
> + return NULL;
> +

> +	return pid->numbers[pid->level].ns;
> +}

The function itself is racy, this pid can be already freed. Perhaps not
a problem currently, afaics it is only used on signal sending path, the
receiver is locked (or the task was found under tasklist), and another
task is current.

> static inline struct pid_namespace *task_active_pid_ns(struct task_struct *tsk)
> {
> - return tsk->nsproxy->pid_ns;
> + /*
> + * If task still has its namespaces (i.e it is not exiting)
> + * get the pid ns from nsproxy.
> + */
> + if (tsk->nsproxy)
> + return tsk->nsproxy->pid_ns;

Racy. Needs task_lock() or rcu_lock(). Please see below,

> + /*
> + * If it is exiting but has not been reaped, get pid ns from
> + * pid->numbers[]. Otherwise return NULL.
> + */
> + return pid_active_ns(task_pid(tsk));

Why can't we just use this chunk and avoid using ->nsproxy?

Oleg.

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