Subject: Re: [PATCH 1/2] pidns: Don't allow new pids after the namespace is dead.
Posted by Oleg Nesterov on Tue, 15 Feb 2011 18:30:28 GMT
View Forum Message <> Reply to Message

On 02/15, Daniel Lezcano wrote:
>
> In the case of unsharing or joining a pid namespace, it becomes
> possible to attempt to allocate a pid after zap_pid_namespace has
> killed everything in the namespace. Close the hole for now by simply
> not allowing any of those pid allocations to succeed.

Daniel, please explain more. It seems, a long ago I knew the reason
for this patch, but now I can't recall and can't understand this change.

> --- a/include/linux/pid_namespace.h
> +++ b/include/linux/pid_namespace.h
> @@ -20,6 +20,7 @@ struct pid_namespace {
>   struct kref kref;
>   struct pidmap pidmap[PIDMAP_ENTRIES];
>   int last_pid;
> + atomic_t dead;

Why atomic_t? It is used as a plain boolean.

And I can't unde

> --- a/kernel/pid.c
> +++ b/kernel/pid.c
> @@ -282,6 +282,10 @@ struct pid *alloc_pid(struct pid_namespace *ns)
>   struct pid_namespace *tmp;
>   struct upid *upid;
>
> + pid = NULL;
> + if (atomic_read(&ns->dead))
> + goto out;
> +

So why this is needed?

If we see ns->dead != 0 we are already killed by zap_pid_ns_processes()
which sets ns->dead = 1.

Oleg.