Subject: Re: [PATCH 14/15] Destroy pid namespace on init's death
Posted by Sukadev Bhattiprolu on Thu, 02 Aug 2007 18:36:08 GMT
View Forum Message <> Reply to Message

Oleg Nesterov [oleg@tv-sign.ru] wrote:
| On 08/02, sukadev@us.ibm.com wrote:
| >
| > Oleg Nesterov [oleg@tv-sign.ru] wrote:
| > | > | > | > + if (pid_ns != &init_pid_ns) {
| > | > | > | > +   zap_pid_ns_processes(pid_ns);
| > | > | > | > +   pid_ns->child_reaper = init_pid_ns.child_reaper;
| > |
| > | OOPS. I didn't notice this before, but this is not right too (regardless
| > | of multi-threaded init problems).
| > |
| > | We should not "reset" ->child_reaper here, we may have exiting tasks
| > | which will re-parent their ->children to global init.
| > |
| > | No, we are still /sbin/init of this namespace even if we are exiting,
| > | ->child_reaper should point to us, at least until zap_pid_ns_processes()
| > | completes.
| >
| > Yes, we are resetting the reaper _after_ zap_pid_ns_processes() completes
| > right ? (all other processes in the namespace must have exited).
|
| OOPS again :) Can't understand how I managed to misread this code.
|
| This means that we should take care about multi-thread init exit,
| otherwise the non-root user can crash the kernel.
|
| >From reply to Kirill's message:
|
| > Still. A non-root user does clone(CLONE_PIDNS), then clone(CLONE_THREAD),

Agree we should fix the crash. But we need CAP_SYS_ADMIN to clone
pid or other namespaces - this is enforced in copy_namespaces() and
unshare_nsproxy_namespaces()

Suka