Subject: Re: [PATCH 11/15] Signal semantics
Posted by Pavel Emelianov on Mon, 30 Jul 2007 09:34:57 GMT

View Forum Message <> Reply to Message

sukadev@us.ibm.com wrote:
> Serge E. Hallyn [serue@us.ibm.com] wrote:
> | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > Pavel Emelianov [xemul@openvz.org] wrote:
> | > | Oleg Nesterov wrote:
> | > | >(What about ptrace_attach() btw? If it is possible to send a signal to init
> | > | > from the "parent" namespace, perhaps it makes sense to allow ptracing as
> | > | > well).
> | > |
> | > | ptracing of tasks fro different namespaces is not possible at all, since
> | > | strace utility determines the fork()-ed child pid from the parent's eax
> | > | register, which would contain the pid value as this parent sees his child.
> | > | But if the strace is in different namespace - it won't be able to find
> | > | this child with the pid value from parent's eax.
> | > |
> | > | Maybe it's worth disabling cross-namespaces ptracing...
> | >
> | > I think so too. Its probably not a serious limitation ?
> |
> | Several people think we will implement 'namespace entering' through a
> | ptrace hack, where maybe the admin ptraces the init in a child pidns,
> | makes it fork, and makes the child execute what it wants (i.e. ps -ef).
> |
> | You're talking about killing that functionality?
>
> No. I was only thinking in terms of debugging container init and missed
> the namespace entering part.
>
> Pavel, I am not sure I understand your comment about being unable to
> ptrace() a child ns.

Ok. We have a task with pid 100, which tries to clone the new namespace.
This task fork-s and we have a new task with a couple of pids (101, 1).
Then this "init" forks again and we have the third task with pids (102, 2).
The problem is that when the 3rd task appears the return value from fork(),
that is - the new task's pid as it is seen by it's parent (2nd task), will
go to eax register (for i386) and it will be 2! But the prtaces from the
initial namespace cannot address this task with pid 2.

> BTW, I am able to gdb a process (incl container-init) from parent ns now.

Debugging separate processes is possible, but when this task starts forking
with namespaces creation - this becomes impossible.

> |
> | -serge
>