Subject: Re: [RFC][PATCH 5/7] VPIDs: vpid/pid conversion in VPID enabled case
Posted by Cedric Le Goater on Fri, 03 Feb 2006 17:05:04 GMT
View Forum Message <> Reply to Message

Alexey Kuznetsov wrote:

> Frankly speaking, using pair (container, pid) was the first thing, which
> we did (year ago), so that from viewpoint of core the switch
> is not a big deal. :-) However, it was rejected by several reasons:
>
> 1. Replacing all the references to pid with pair (container, pid) is quite
> expensive. F.e. it is possible that a task has a pid from one container,
> but it is in process group and/or session of another container,
> and its controlling terminal owner by another container. Grr..

If that happens, it also means your container is not fully isolated which
is also a challenge for the vpid approach when you try to migrate. nop ?

If i take your example with the external process group, what would happen
if the process group leader dies and then you try to migrate that container
? How would you restore the processes in your container that are refering a
dead external process group leader ?

Everything is possible but "loose" isolation on pid raises a lot of issues
on vpids at restart. I would stick to a real strict isolation and forbid
such cases. And, in that case, it's much easier to use the pair approach
(container, pid).

We've been living with the vpid approach also for years and we found issues
that we haven't solve at restart. So we think we might do a better job with
another. But, this still needs to be confirmed :)

> So, the structures are bloated, the functions get additional arguments.
> And all this is for no real purpose, the functionality comparing with
> VPID is even reduced.

i don't see much changes, when you query a task by pid, you only look in
your *current* container pidspace.

some areas in the kernel use directily pids, true. Eric Biederman really
knows well his job on this topic. Many thanks. But, that could be fixed.

> 2. It is very inconvenient not to see processes inside VPS from host system.
> To do ps, strace, gdb etc. we have to move inside VPS. With VPID approach I can
> gdb even "init" process of VPS in a way invisible to VPS, see?

that's another container model issue again. your init process of a VPS
could be the real init. why do you need a fake one ? just trying to

understand all the issues you had to solve and I'm sure they are valid.

> Well, and main problem is that gui administration and monotoring tools,
> which were existing for ages stop to work and require a major rewrite.
> Does it answer to question about plans for moving away?
>
> To summarize: (container, pid) approach looks clean and consistent.
> At first sight I loved it, even thought it will solve some of problems
> with inter-container access control. But the devil is in details,
> I have to learn this again and again: access control must be separate
> of real engine, otherwise you get something which does not satisfy anyone.

hmm, I'm not completely satisfied :) but we'll work this out, we'll find a
way to agree on something.

C.
