Subject: Re: [PATCH] pidns: Place under CONFIG_EXPERIMENTAL (take 2)
Posted by ebiederm on Sat, 27 Oct 2007 03:46:59 GMT
View Forum Message <> Reply to Message

Andrew Morton <akpm@linux-foundation.org> writes:

>> On Sat, 27 Oct 2007 04:04:08 +0200 Adrian Bunk <bunk@kernel.org> wrote:
>> > be happy to hear if someone has a better idea.
>>
>> There is a difference between "complete the feature" and "early adopters
>> to start playing with the feature" on the one side, and making something
>> available in a released kernel on the other side.
>>
>> For development and playing with it it can depend on BROKEN (perhaps
>> with the dependency removed through the first -rc kernels), but as soon
>> as it's available in a -final kernel the ABI is fixed.
>>
>
> Yes, if we're not 100% certain that the interfaces are correnct and unchanging
> and that the implementation is solid, we should disable the feature at Kconfig
> time.

Reasonable. So far things look good for a single pid namespace. Multiple
pid namespaces look iffy.

> The best option would be to fix things asap. But assuming that option isn't
> reasonable and/or safe, we can slip a `depends on BROKEN' into -rc6 then
> resume development for 2.6.25.

I think we can make a lot of progress but there is enough development
yet to do to reach the target of correct and unchanging interfaces,
with a solid interface. That unless we achieve a breakthrough I
don't see us achieving that target for 2.6.24.

The outstanding issues I can think of off the top of my head:
- signal handling for init on secondary pid namespaces.
- Properly setting si_pid on signals that cross namespaces.
- The kthread API conversion so we don't get kernel threads
  trapped in pid namespaces and make them unfreeable.
- At fork time I think we are doing a little bit too much work
  in setting the session and the pgrp, and removing the controlling
  tty.
- AF_unix domain credential passing.
- misc pid vs vpid sorting out (autofs autofs4, coda, arch specific
  syscalls, others?)
- Removal of task->pid, task->tgid, task->signal->__pgrp,
  tsk->signal->__session or some other way to ensure that we have
  touched and converted all of the kernel pid handling.

- flock pid handling.

It hurts me to even ponder what thinking makes it that
CONFIG_EXPERIMENTAL isn't enough to keep a stable distro
from shipping the code in their stable kernel, and locking us into
trouble.

With that said. I think I should just respin the patchset now and add
the "depends on BROKEN".

The user namespace appears to need that treatment as well.

The network namespace has so little there and it already depends
on !SYSFS so I don't think we are going to run into any trouble
with it. Happily I managed to parse that problem differently,
so I could slice of the parts of the networking stack that
had not been converted.

Eric

_____