Subject: Re: [PATCH] pidns: Place under CONFIG_EXPERIMENTAL (take 2)
Posted by ebiederm on Fri, 26 Oct 2007 21:59:29 GMT
View Forum Message <> Reply to Message

"Kir Kolyshkin" <kir@swsoft.com> writes:

> Eric,
>
> Could you please hold off the horses a bit and wait till Pavel Emelyanov
> returns? It means next Monday; he's currently at a conference whose organisers
> don't provide internet access.

When we decided to go top down (i.e. user interface first) instead of
bottom up with the pid namespace implementation it was my
understanding that we had agreed we would make the pid namespaces
depend on CONFIG_EXPERIMENTAL so that we wouldn't be stuck forever
supporting early ABI mistakes.

So to my knowledge the conversation has already happened. I believe
something in the confusion of trying to use these options to shrink
the kernel and the futility of that, caused whatever config options
we had before to be dropped.

Further I was happy to let Pavel and Suka work on this code because
the appeared to know what they were doing and it freed me to do other
things. I don't think there are any mysteries in what we are trying
to do that I need them to explain.

> I feel it makes great sense to review/discuss patches first on containers@
> first before submitting directly to lkml/Linus.

My feel before starting to review the pid namespace patches was that
the work was essentially done except a handful of minor details. Upon
closer examination, I found that not to be the case. My rough fix
queue had 25 or so patches as of last night to fix pid namespace
issues.

I have no confidence that we will fix all of the pid namespaces issues
before 2.6.24-final. I do think we can get most of them fixed.

Given that most of the remaining issues are integration issues
with the rest of the kernel having the code merged should make
it much easier to see what is going on and actually fix things.
So I am not in favor of reverting this code despite seeing numerous
problems.

> Speaking of this particular patch -- I don't understand how you fix
> "innumerable little bugs" by providing stubs instead of real functions.

> Sent from my BlackBerry; please reply to kir@openvz.org

It doesn't fix the bugs it avoids them because there is no way to
get to the them and trigger them. So far I have yet to find a bug
that is a problem with only a single pid namespace in the kernel.

Since everyone agrees that there are at least some deficiencies in
the current pid namespace I think this makes sense, to mark
the code as EXPERIMENTAL and have a way for people who care to
shut it off just so they don't have to worry about new issues.

As far as how the config option is implemented I don't much care
so long as I get the -EINVAL when I pass CLONE_NEWPID as root.

Essentially this patch is part of a defense in depth against pid
namespace problems hitting people. This patch is my first line
of defense. Actually fixing all of the rest of the known bugs
is the other line.

Eric

_____

Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers