Subject: Re: [patch -mm 10/17] nsproxy: add unshare_ns and bind_ns syscalls
Posted by Herbert Poetzl on Sat, 09 Dec 2006 04:18:14 GMT
View Forum Message <> Reply to Message

On Fri, Dec 08, 2006 at 12:26:49PM -0700, Eric W. Biederman wrote:
> clg@fr.ibm.com writes:
>
> > From: Cedric Le Goater <clg@fr.ibm.com>
> >
> > The following patch defines 2 new syscalls specific to nsproxy and
> > namespaces :
> >
> > * unshare_ns :
> >
> > enables a process to unshare one or more namespaces. this
> >   duplicates the unshare syscall for the moment but we
> > expect to diverge when the number of namespaces increases
>
> Are we out of clone flags yet? If not this is premature.

no, but a different nevertheless related question:
does anybody, except for 'us' use the unshare() syscall?

because if not, then why not simply extend that one
to 64bit and be done, we probably won't need a clone64()
but if we find we do (at some point) adding that with
the new flags would be trivial ...

OTOH, we could also just add an unshare64() too

anyway, we _will_ run out of flags in the near future

> > * bind_ns :
> >
> > allows a process to bind
> > 1 - its nsproxy to some identifier
> > 2 - to another nsproxy using an identifier or -pid
>
> NAK
>
> Don't use global identifiers. Use pids. i.e. struct pid * for your
> identifiers. Is there is a reason pids are unsuitable?

yes, see my reply in the other mail

> I'm also worried about the security implications of switching
> namespaces on a process.
> That is something that needs to be looked at very closely.

Linux-VServer currently uses a capability to prevent
changing between namespaces (a very generic one) but
it probably makes sense to add something like that
in general ... btw, did I mention that the capability
flags are running out too?

> These two changes certainly don't belong in a single patch,
> and they certainly use a bit more explanation.
> syscalls are not something to add lightly.

well, and they will take ages to get into mainline
for all archs, or has that changed sine we reserved
sys_vserver()?

> Because they must be supported forever.

I'm not sure about that, most archs 'reuse' syscalls
when there is no user left ...

best,
Herbert

> Eric
> ____
> Containers mailing list
> Containers@lists.osdl.org
> https://lists.osdl.org/mailman/listinfo/containers

____
Containers mailing list
Containers@lists.osdl.org
https://lists.osdl.org/mailman/listinfo/containers