
Subject: Re: Re: [RFC][PATCH 1/2] add user namespace [try #2]
Posted by [Herbert Poetzl](#) on Tue, 12 Sep 2006 10:54:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Mon, Sep 11, 2006 at 10:59:04AM +0200, Cedric Le Goater wrote:

> Herbert Poetzl wrote:

> > On Thu, Sep 07, 2006 at 02:01:00PM -0600, Eric W. Biederman wrote:

> > > Kirill Korotaev <dev@sw.ru> writes:

> > >

> > > > BTW...

> > > >

> > > > > --- 2.6.18-rc4-mm3.orig/include/linux/sched.h

> > > > > +++ 2.6.18-rc4-mm3/include/linux/sched.h

> > > > > @@ -26,6 +26,7 @@

> > > > > #define CLONE_STOPPED 0x02000000 /* Start in stopped state */

> > > > > #define CLONE_NEWUTS 0x04000000 /* New utsname group? */

> > > > > #define CLONE_NEWIPC 0x08000000 /* New ipc */

> > > > > + #define CLONE_NEWUSER 0x10000000 /* New user */

> > > > we have place for 3 namespaces more only. Does anyone have a plan

> > > > what to do then? I warned about this at the beginning when we

> > > > were discussing the interfaces and this flags soon going to be

> > > > exhausted, so probably it is time to do something in advance...

> > > Actually there is another unused bit in the middle :)

> > > Plus there are a bunch of bits that unshare can use but clone can't.

> > > Plus what other namespaces are on the todo list?

> > > We have network, and pid, and time.

> > > What else?

> > >

> > > resource (could be limits and/or accounting),

> > > lightweight-net, (maybe fs in contrast to vfs)

> > >

> > > I guess we're reaching the limits anyway and it would not leave much

> > > room in the clone flags for other features not related to containers.

> > >

> > > It's not like we're adding one or two, we would take at least 6 : uts,

> > > ipc, user, pid, net, time, etc. I'm sure ideas to extend the list will

> > > come when this is in use ...

as I said, I'd opt for having a new clone() syscall in addition to the existing one, with a separate 64bit set of flags to decide what namespaces should be created or cloned. there is no problem with putting 'important' or generally 'useful' flags (like for example for pid, uts or lightweight network isolation) into the existing clone call (will require a simple mapping if done properly) so that they can be used with 'older' libc interfaces too

I know, it would be 'nice' to keep the existing clone() interface, but I think it already has become a complication we should avoid (and we have not even used up all the available flags :)

are there any strong arguments against having a new clone() syscall, which I was missing so far?

TIA,
Herbert

> C.

> _____

> Containers mailing list

> Containers@lists.osdl.org

> <https://lists.osdl.org/mailman/listinfo/containers>
