
Subject: Re: [PATCH 1/3] change clone_flags type to u64
Posted by [Cedric Le Goater](#) on Thu, 10 Apr 2008 12:25:31 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello Andi,

Andi Kleen wrote:

> sukadev@us.ibm.com writes:

>

>> From: Sukadev Bhattiprolu <sukadev@us.ibm.com>

>> Subject: [lxc-dev] [patch -lxc 1/3] change clone_flags type to u64

>>

>> This is a preliminary patch changing the clone_flags type to 64bits

>> for all the routines called by do_fork().

>

> I must admit I was always a little sceptical of giving every tiny

> namespaceable kernel feature its own CLONE flag (and it's own

> CONFIG option). What was the rationale for that again?

I guess that was a development rationale. Most of the namespaces are in use in the container projects like openvz, vserver and probably others and we needed a way to activate the code.

Not perfect I agree.

> With your current strategy are you sure that even 64bit will

> be enough in the end? For me it rather looks like you'll

> go through those quickly too as more and more of the kernel

> is namespaced.

well, we're reaching the end. I hope ! devpts is in progress and mq is just waiting for a clone flag.

> Also I think the user interface is very unfriendly. How

> is a non kernel hacker supposed to make sense of these

> myriads of flags? You'll be creating another

> CreateProcess123_extra_args_extended()

> in the end I fear.

well, the clone interface is a not friendly interface anyway. glibc wraps it and most users just use fork().

We will need a user library, like we have a libpthread or a libaio, to effectively use the namespaces features. This is being worked on but it's another topic.

> Wouldn't it be better to just partition all this into

> fewer more understandable larger feature groups? I think

- > that would be much nicer from pretty much all perspectives
- > (kernel maintenance, user interface sanity, not needing
- > clone128/256 in the end etc.)

Yes. this make sense. Most of the namespaces have dependencies between each other.

- > Some consolidation on the CONFIGs would be good too. I just
- > cannot imagine it really makes sense to configure everything
- > so fine grained and this is just asking for random compile
- > breakage on randconfig.

yes. definitely agree.

but we still need a way to extend the clone flags because none are left.
would you say that the clone64 is the right way to go or should we rather go in the direction hpa proposed :

<http://lkml.org/lkml/2008/4/9/318> :

- > If you're going to make it a 64-bit pass it in as a 64-bit number,
- > instead of breaking it into two numbers. Better yet, IMO, would
- > be to pass a pointer to a structure like:
- >
- > struct shared {
- > unsigned long nwords;
- > unsigned long flags[];
- > };
- >
- > ... which can be expanded indefinitely.

if we could agree on some new interface, we could then make sure we are not abusing it.

Thanks,

C.

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