Subject: Re: [PATCH 1/2] Extend sys_clone and sys_unshare system calls API
Posted by Pavel Machek on Thu, 24 Jan 2008 17:50:33 GMT
View Forum Message <> Reply to Message

On Thu 2008-01-24 09:46:57, Dave Hansen wrote:
> On Thu, 2008-01-24 at 18:37 +0100, Pavel Machek wrote:
> > > Did we decide not to do something with a variable number of
> > arguments?
> > >
> > >   sys_clonefoo(unsigned long *flags, int len);
> >
> > That is evil, because that means strace can no longer reliably print
> > flags being used (for example).
>
> Ahhh. Just like it can't print strings for "buf"?
>
> write(int fd, char *buf, size_t len)
>
> Man, strace is stupid! ;)

I said _reliably_.

Guess what happens on smp when one process will periodically overwrite
buf, and second will do the syscall?

It sucks for write, but it would suck much more for clone, where flags
drastically change the behaviour.

(Yep, it is possible to do reliably, see subterfugue, but it is a lot
of work and lot of overhead).

> If that's *really* a concern, why don't we just pass, say 4 or 5 longs
> in:
>
> sys_clonebig(unsigned long flags0, unsigned long flags1,
>   unsigned long flags2, unsigned long flags3);

Works for me.
        Pavel
--
(english) http://www.livejournal.com/~pavelmachek
(cesky, pictures) http://atrey.karlin.mff.cuni.cz/~pavel/picture/horses/blog.html

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