
Subject: Re: [patch -mm 09/17] nsproxy: add namespace flags
Posted by [ebiederm](#) on Mon, 11 Dec 2006 20:02:15 GMT
[View Forum Message](#) <> [Reply to Message](#)

Cedric Le Goater <clg@fr.ibm.com> writes:

```
> Eric W. Biederman wrote:
>> Cedric Le Goater <clg@fr.ibm.com> writes:
>>
>>>> /*
>>>> + * namespaces flags
>>>> + */
>>>> +#define NS_MNT 0x00000001
>>>> +#define NS_UTS 0x00000002
>>>> +#define NS_IPC 0x00000004
>>>> +#define NS_PID 0x00000008
>>>> +#define NS_NET 0x00000010
>>>> +#define NS_USER 0x00000020
>>>> +#define NS_ALL (NS_MNT|NS_UTS|NS_IPC|NS_PID|NS_NET|NS_USER)
>>> hmm, why _another_ set of flags to refer to the
>>> namespaces?
>>> well, because namespaces are a new kind in the kernel
>>
>> Gratuitous incompatibility.
>
> ?
```

Changing the numbers for no good reason. We can easily keep the existing numbers.

```
>>>> is the clone()/unshare() set of flags not sufficient
>>>> for that?
>>> because we are reaching the limits of the CLONE_ flags.
>>
>> Not really. There are at least 8 bits that clone cannot use
>> but that unshare can.
>
> please, could you list them ?
```

CSIGNAL. There are a several others as well that will never mean anything in an unshare context:

I believe the 10 flags below are also nonsense from an unshare perspective.

```
CLONE_PTRACE 0x00002000
CLONE_VFORK 0x00004000
CLONE_PARENT 0x00008000
CLONE_SETTLS 0x00080000
CLONE_PARENT_SETTID 0x00100000
CLONE_CHILD_CLEARPID 0x00200000
```

CLONE_DETACHED 0x00400000
CLONE_UNTRACED 0x00800000
CLONE_CHILD_SETTID 0x01000000
CLONE_STOPPED 0x02000000

>>>> if so, shouldn't we switch (or even better change?
>>>> the unshare() too) to a new set of syscalls?
>>> unshare_ns() is a new syscall and we don't really need a
>>> clone anyway. nop ?
>>
>> Huh? Clone should be the primary. There are certain namespaces
>> that it are very hard to unshare, without creating a new process.
>
> You just said above that clone had less available flags than
> unshare ...

I did. It is just easier to support clone than unshare.

> anyway, could you elaborate a bit more ? I have the opposite
> feeling and you gave me that impression also a few month ago.
>
> No problem for me, i just want a way to use this stuff without

My feeling is basically that there are some things that we
can do much more cleanly at process creation time.

>From an implementation standpoint unshare is fairly nasty because
it keeps things from being invariants across the lifetime of a
process. Which means it contains more races and is harder to support.
That's why we can't unshare we can share with clone right now.

Eric

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