Subject: Re: [RFC][PATCH 1/2] add user namespace [try #2]
Posted by ebiederm on Tue, 12 Sep 2006 15:06:28 GMT
View Forum Message <> Reply to Message

Kirill Korotaev <dev@sw.ru> writes:

> 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 ipcs */
>>>>+#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.
> :))) I suggest to write HOWTO-select-unused-bits in CodingStyle :))
>
>> Plus what other namespaces are on the todo list?
>> We have network, and pid, and time.
> I think more.
>
> proc-ns,
> sysfs-ns,
> printk-ns or syslog-ns?: syslog should be virtualized
> and more...

I don't think those meet the criteria for namespaces.
But certainly there is work we need to do there.

> semi-namespaces:
> fs-ns (should regulate which filesystems are accessiable from container, but
> probably this is not exact name space... need to think over...),

I think the problem there is the same as allowing untrusted users the ability to mount filesystems, in which case we just tag filesystems that are safe for untrusted users to use.

> dev-ns (should regulate which devices are accessiable from container)

Yes. Devices certainly have global names that we need to bring under control. The easy solution is just to limit CAP_SYS_MKNOD but we may need something more.

One of the pieces that needs consideration when it comes to permissions is the plan9 style of permission control.   Where file have an initial owner, and if someone else needs access to them you chmod, chown them so that everyone who needs to has access. I think that is an simpler model to get right than to have a bunch of special cases.

Eric