Subject: Re: [RFC][PATCH 1/2] add user namespace [try #2]
Posted by dev on Sat, 16 Sep 2006 12:05:09 GMT
View Forum Message <> Reply to Message

>>>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.
Well, it is hard to say what is the criteria...

>>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.
You need some groupping mechanisms, don't you?
Say, I need to allow isofs for containers 1,2,5,6
and ext3 for containers 2,3,4,5

>>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.

CAP_SYS_MKNOD is not an option.

Can you please propose how to organize it?

You can check how it is implemented in OpenVZ in kernel/vecalls.c
devperms_struct
real_get_device_perms_ve()
real_setdevperms()

BTW, taking a look near this code, I found another bunch of interesting
functionality - statistics (e.g. real_update_load_avg_ve).

Though load avg statistics logically belong to pspace namespace there is a lot of other stats which can not be associated so easily with the namespaces.

> 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.

it is Linux :)

Thanks,
Kirill