Subject: Re: [RFC][PATCH] Devices visibility container
Posted by Cedric Le Goater on Tue, 25 Sep 2007 12:25:24 GMT
View Forum Message <> Reply to Message

Hello Eric !

Eric W. Biederman wrote:
> Pavel Emelyanov <xemul@openvz.org> writes:
>
>> At KS we have pointed out the need in some container, that allows
>> to limit the visibility of some devices to task within it. I.e.
>> allow for /dev/null, /dev/zero etc, but disable (by default) some
>> IDE devices or SCSI discs and so on.
>
> NAK
>
> We do not want a control group subsystem for this.

we will need one way to configure the list of available devices from
user space. Any proposal ?

> For the short term we can just drop CAP_SYS_MKNOD.

Sure. Pavel is working on something mid-term ;)

> For the long term we need a device namespace for application
> migration, so they can continue to use devices with the same
> major+minor number pair after the migration event.

Hmm, yes. I can imagine that for some big database application using
raw devices but it only means that the same device must be present
upon restart. I don't see any identifier virtualization issues.

> Things like
> ensuring a call to stat on a given file before and after the migration
> return the exact same information sounds compelling. So I don't think
> this is even strictly limited to virtual devices anymore. How many
> applications are there out there that memorize the stat data on a file
> and so they can detect if it has changed?

that we need to support of course, otherwise we would break things
like tail.

> If we need something between those two it may make sense to enhance
> the LSM or perhaps introduce an alternate set security hooks. Still
> if we are going to need a full device namespace that may be a little
> much.

serge's implementation using security hooks should help us choose
the right approach.

Thanks !

C.

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