Subject: Re: [PATCH 01/11] SYSCTL: export root and set handling routines
Posted by ebiederm on Mon, 19 Dec 2011 16:37:19 GMT
View Forum Message <> Reply to Message

Stanislav Kinsbursky <skinsbursky@parallels.com> writes:

>> In practice what this means is that register_net_sysctl_table should
>> work for any sysctl file anywhere under /proc/sys. I think
>> register_net_sysctl_table is the right solution for your problem. The
>> only possible caveat I can think of is you might hit Al's performance
>> optimizations and need to create a common empty directory first with
>> register_sysctl_paths.
>
> Sorry, but I forgot to mention one more important goal I would like to achieve:
> I want to manage sysctl's variables in context of mount owner, but not viewer one.
> IOW imagine, that we have one two network namespaces: "A" and "B". Both of them
> have it's own net sysctl's root. And we have per-net sysctl "/proc/sys/var".
> And for ns "A" variable was set to 0, and for "B" - to 1.
> And B's "/proc/sys/var" is accessible from "A" namespace
> ("/chroot_path/proc/sys/var" for example).
> With this configuration I want to read "1" from both namespaces:
> owner "B" (/proc/sys/var) and "A" ("/chroot_path/proc/sys/var").
> Looks like simple using of register_net_sysctl_table doesn't allow me this,
> because current net ns is used. And to achieve this goal I need my own sysctl
> set for SUNRPC like it was done for network namespaces.

Doing that independently of the rest of the sysctls is pretty horrible
and confusing to users.   What I am planning might suit your needs and
if not we need to talk some more about how to get the vfs to do
something reasonable.

>> That said since I am in the process of rewriting things some of this
>> may change a little bit, but hopefully not in ways that immediately
>> effect the users of register_sysctl_table.
>>
>> Don't use register_net_sysctl_ro_table.   I think what the implementors
>> actually wanted was register_net_sysctl_table(&init_net, ...) and didn't
>> know it.
>>
>> Don't put subdirectories in your sysctl tables. Use a ctl_path to
>> specify the entire directory where the files should show up. Generally
>> the code is easier to read in that form, and the code is simpler to deal
>> with if we don't have to worry about directories.
>>
>> Don't play with the sysctl roots. It is my intention to completely kill
>> them off and replace them by moving the per net sysctl tree under
>> /proc/<pid>/sys/.   Leaving behind symlinks in /proc/sys/net and I guess
>> ultimately in /proc/sys/sunrpc/ and /proc/sys/fs/nfs... Which actually

>> seems to better describe your mental model.
>>
>
>
> I'm afraid, that this approach this not allow me to achieve the goal, mentioned
> above, because current->nsproxy->net_ns will be used during lookup.
> Or maybe I misunderstanding here?

What I hope to do is to stop using current, and to behave like /proc/net. Aka a per process view under /proc/<pid>/sys that matches the namespaces of the specified process.

The VFS really hates my use of current in the sysctl case, and I intend to stop.

I need to run and catch my plane. It doesn't look like I will have access to this email address for the next two weeks :(

Eric