Subject: Re: [PATCH 0/2] resource control file system - aka containers on top of nsproxy!
Posted by Sam Vilain on Thu, 08 Mar 2007 00:50:01 GMT
View Forum Message <> Reply to Message

Srivatsa Vaddagiri wrote:
> container structure in your patches provides for these things:
>
> a. A way to group tasks
> b. A way to maintain several hierarchies of such groups
>
> If you consider just a. then I agree that container abstraction is
> redundant, esp for vserver resource control (nsproxy can already be used
> to group tasks).
>
> What nsproxy doesn't provide is b - a way to represent hierarchies of
> groups.
>

Well, that's like saying you can't put hierarchical data in a relational database.

The hierarchy question is an interesting one, though. However I believe it first needs to be broken down into subsystems and considered on a subsystem-by-subsystem basis again, and if general patterns are observed, then a common solution should stand out.

Let's go back to the namespaces we know about and discuss how hierarchies apply to them. Please those able to brainstorm, do so - I call green hat time.

1. UTS namespaces

Can a UTS namespace set any value it likes?

Can you inspect or set the UTS namespace values of a subservient UTS namespace?

2. IPC namespaces

Can a process in an IPC namespace send a signal to those in a subservient namespace?

3. PID namespaces

Can a process in a PID namespace see the processes in a subservient namespace?

Do the processes in a subservient namespace appear in a higher level namespace mapped to a different set of PIDs?

4. Filesystem namespaces

Can we see all of the mounts in a subservient namespace?

Does our namespace receive updates when their namespace mounts change? (perhaps under a sub-directory)

5. L2 network namespaces

Can we see or alter the subservient network namespace's interfaces/iptables/routing?

Are any of the subservient network namespace's interfaces visible in our namespace, and by which mapping?

6. L3 network namespaces

Can we bind to a subservient network namespace's addresses?

Can we give or remove addresses to and from the subservient network namespace's namespace?

Can we allow the namespace access to modify particular IP tables?

7. resource namespaces

Is the subservient namespace's resource usage counting against ours too?

Can we dynamically alter the subservient namespace's resource allocations?

8. anyone else?

So, we can see some general trends here - but it's never quite the same question, and I think the best answers will come from a tailored approach for each subsystem.

Each one *does* have some common questions - for instance, "is the namespace allowed to create more namespaces of this type". That's probably a capability bit for each, though.

So let's bring this back to your patches. If they are providing visibility of ns_proxy, then it should be called namesfs or some such. It doesn't really matter if processes disappear from namespace aggregates, because that's what's really happening anyway. The only problem is that if you try to freeze a namespace that has visibility of

things at this level, you might not be able to reconstruct the
filesystem in the same way. This may or may not be considered a problem,
but open filehandles and directory handles etc surviving a freeze/thaw
is part of what we're trying to achieve. Then again, perhaps some
visibility is better than none for the time being.

If they are restricted entirely to resource control, then don't use the
nsproxy directly - use the structure or structures which hang off the
nsproxy (or even task_struct) related to resource control.

Sam.

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