Subject: Re: Network namespaces without isolation
Posted by Andreas B Aaen on Fri, 04 Jul 2008 15:07:28 GMT
View Forum Message <> Reply to Message

On Friday 04 July 2008 11:52, Eric W. Biederman wrote:
> Andreas B Aaen <andreas.aaen@tietoenator.com> writes:

> Answering part of your question. As currently designed you can use
> multiple network namespaces in a single task, and you can place each vlan
> interface in different network namespace. However the current model is
> most cumbersome for doing so.
>
> You can use unshare instead of clone which is a little easier.

How do you actually use multiple name spaces in the current implementation in the same task if you refer to them indirectly through pids?
So if I need 500 network namespaces then I need to fork 500 processes.

> A socket option sounds like a nice idea.

And quite easy to implement except for the handling of which network namespaces you should be allowed to talk to.

> The two challenges are what names to use to refer to network namespaces
> and how to get network namespaces to persist.

Exactly. In my current proof of concept implementation indexed/named network namespaces are created through an extended netlink interface instead of the clone/unshare calls. Delete of the namespaces are also through a netlink interface. E.g.:

ip netns add 1
(adds a network namespace with the "name" index 1)
ip netns del 1
(deletes it again)

> There have been a number of discussions about identifiers none of which
> have led to any sort of agreement. One of the goals in the design is
> that we don't introduce new global identifiers allowing us to ultimately
> have nested containers.

In this case this means that the index' should be a namespace of itself just like pids. It seems to be overkill. At lest for my purpose.

> So far we have been referring to namespaces indirectly by the pids of the
> processes which are using them.

Right. And with namespaces into namespaces and usage of pid namespaces you

could have two different namespaces named with the same numerical value of
pid.

> > It would also be nice to be able to see the network statistics from all
> > the namespaces through the proc filesystem at least in an uncloned
> > (isolated) namespace.
>
> Currently this is possible by looking at /proc/<pid>/net.

Which was what lead me to the question of how you can have more name spaces in
a single task with the current implementation.

> > So you would be able to see the network statistics in
> > /proc/net/ns/<index>/

Or maybe this should have been /proc/<pid>/net/<index>/ ?

> One of the things we have tried to do is to keep the number of new
> interfaces to a minimum.

Sure.

> If we can work out the details on how to do that cleanly it seems totally
> reasonable to enhance network namespaces in that direction. You are not
> the first to express those kind of requirements, and probably won't be the
> last.

So it seems that we need to restart the naming discussion.

Regards,
--
Andreas Bach Aaen System Developer, M. Sc.
Tieto Enator A/S tel: +45 89 38 51 00
Skanderborgvej 232 fax: +45 89 38 51 01
8260 Viby J Denmark andreas.aaen@tietoenator.com
_____