Subject: Network Namespace status
Posted by ebiederm on Thu, 13 Sep 2007 19:12:08 GMT
View Forum Message <> Reply to Message

Now that the network namespace work is partly merged I figure
a short status summary of where everything is at is in order.

David Miller has merged the core of the network namespace work
and that probably needs to sit just a little while to make certain
we don't have unexpected breakage.

Before enabling multiple instances of the network namespace
it is necessary to sort through a few last user interface issues.

In Greg KH's tree there is work from Tejun and myself that decouples
the sysfs dentry tree from the kobject tree, and Tejun is actively
working on completing that decoupling. From the current sysfs state
it takes just a handful of patches to support multiple super_blocks
each displaying the network devices for a different network namespace.
And the last round of patches that did that Tejun and I almost agree
upon. That support is needed before we can allow network devices
to exist in anything except the initial network namespace.

In Andrew's tree there is the start of my sysctl cleanup. Basically
just an additional sanity check in register_sysctl_table and a bunch
of fixes to avoid the errors that sanity check has found. Pending
I have a few more general cleanups and code to support multiple
network namespaces. Last we talked Andrew said I have sent
him enough sysctl changes for now, and to wait until after the
merge window before sending more.

The proc support in the net-2.6.24 tree is reasonable from the
direction of the networking code. Currently I am looking at
"current->net_ns" and resolving /proc/net based upon that. Long term
we want to refactor that code so that "current->net_ns" is captured
when we mount /proc. So the network namespace state can be monitored
from outside applications, and so that we aren't playing dangerous
games with the vfs dentry trees.

The final blocker to having multiple useful instances of network
namespaces is the loopback device. We recognize the network namespace
of incoming packets by looking at dev->nd_net. Which means for
packets to properly loopback within a network namespace we need a
loopback device per network namespace. There were some concerns
expressed when we posted the cleanup part of the patches that allowed
for multiple loopback devices a few weeks ago so resolving this one
may be tricky.

Looking into my patch queue I have:
5 patches for cleaning up and making a per network namespace loopback device.
4 patches for making rtnetlink message processing per network namespace
1 patch for making AF_UNIX per network namespace
1 patch for making AF_PACKET per network namespace

The ipv4 part of my patchset is currently working but it needs some
more cleanup and reordering of patches before it is ready to go anywhere.
Nothing has been done for ipv6, but the changes should very much parallel
ipv4.

The other protocols I haven't even looked at yet.

Eric

_____

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