Subject: Re: [PATCH 0/2] Fix /proc/net in presence of net namespaces
Posted by ebiederm on Sun, 02 Mar 2008 02:03:28 GMT
View Forum Message <> Reply to Message

- The experience from vserver, planetlab and OpenVZ is that it is good to be able to monitor processes in other namespaces.

- The linux experience says filesystems are a good way to do that.

- So we really want to filesystem monitoring interfaces to depend on the filesystem mount options instead of current.

- Starting with making /proc and sysctls depend on current is a cheap way to get things up and going.

- When I consider breaking things up into multiple filesystems I run across the occasional file that depends on multiple namespaces. uids in /proc/sysvipc/* for example. Luckily I have yet to find any directory structures that depend on more then one namespace.

  Maybe that can be handled properly by capturing multiple namespaces at mount time but I am a bit leery of that.

- The visibility of namespaces should be match the visibility of the processes that use them.   Access control of course can be more restricted.

- We want to see how namespaces connect to tasks.

Therefore.

/proc/net, /proc/sys, /proc/sysvipc, and probably a few others should migrate under /proc/<pid>/task/<tid> (not under /proc/<pid> so we can finally straighten out the task group vs task issue).

Todays problem of course is /proc/net/

What I had intended to implement was:
/proc/current -> /proc/<pid>/task/<tid>
(A new symlink to the task directory)

/proc/net -> /proc/current/net
(like /proc/mounts)

The only downside of placing files under the task directory is that we use a lot more dentries for /proc.

....

Optimizations.

If the dentry pressure is significant and we don't have data from
other namespaces in the files causing us to want to present the
information differently for different processes I support using
an id and a per namespace upper level directory. With a symlink
into there from the task directories.

/proc/<pid>/task/<tid>/net -> ../../.../netns/<netns id>

The id I would use is a struct pid because that makes the id useful
for userspace monitoring and control applications and because we
can migrate it.

In my view /proc/netns/<pids> would be implemented like
/proc/<pids> with readdir and lookup returning different contents
based upon the pid namespace captured when we mounted proc.

Further struct pid would be enhanced so that as long as we have
a namespace using a struct pid as an id we would not free that pid_nr
in any of the pid namespaces. Just like we do with process groups
and sessions today.

I think for the network namespace and network /proc files that
optimization is safe. I seem to recall checking and not finding any
ids from other namespaces in the files under /proc/net.

I will try for some more detailed replies.

Eric

____