Subject: Re: [patch -mm 0/4] mqueue namespace
Posted by serue on Fri, 20 Jun 2008 14:50:31 GMT
View Forum Message <> Reply to Message

Quoting Eric W. Biederman (ebiederm@xmission.com):
> Cedric Le Goater <clg@fr.ibm.com> writes:
>
> > Hello !
> >
> > Here's a small patchset introducing a new namespace for POSIX
> > message queues.
> >
> > Nothing really complex a part from the mqueue filesystem which
> > needed some special care
>
> This looks stalled.

It actually isn't really - Cedric had resent it a few weeks ago but had troubles with the mail server so it never hit the lists. I think Dave made a few more changes from there and was getting ready to resend again. Dave?

> I have a brainstorm that might takes a totally
> different perspective on things.
>
> The only reason we don't just allow multiple mounts of mqueuefs to
> solve this problem is because there is a kernel syscall on the path.
>
> If we just hard coded a mount point into the kernel and required user
> space to always mount mqueuefs there the problem would be solved.
>
> hard coding a mount point is unfortunately violates the unix rule
> of separating mechanism and policy.
>
> One way to fix that is to add a hidden directory to the mnt namespace.
> Where magic in kernel filesystems can be mounted. Only visible
> with a magic openat flag. Then:
>
> fd = openat(AT_FDKERN, ".", O_DIRECTORY)
> fchdir(fd);
> umount("./mqueue", MNT_DETACH);
> mount(("none", "./mqueue", "mqueue", 0, NULL);
>
> Would unshare the mqueue namespace.
>
> Implemented for plan9 this would solve a problem of how do you get
> access to all of it's special filesystems. As only bind mounts
> and remote filesystem mounts are available. For linux thinking about

> it might shake the conversation up a bit.

It is unfortunate that two actions are needed to properly complete the unshare, and we had definately talked about just using the mount before. I forget why we decided it wasn't practical, so maybe what you describe solves it...

But at least the current patch reuses CLONE_NEWIPC for posix ipc, which also seems to make sense.

-serge

_____

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