Subject: Re: [RFC] network namespaces
Posted by ebiederm on Tue, 05 Sep 2006 17:09:21 GMT
View Forum Message <> Reply to Message

> This family of containers are used too for HPC (high performance computing) and
> for distributed checkpoint/restart. The cluster runs hundred of jobs, spawning
> them on different hosts inside an application container. Usually the jobs
> communicates with broadcast and multicast.
> Application containers does not care of having different MAC address and rely on
> a layer 3 approach.

Ok I think to understand this we need some precise definitions.
In the normal case it is an error for a job to communication with a different
job.

The basic advantage with a different MAC is that you can found out who the
intended recipient is sooner in the networking stack and you have truly
separate network devices. Allowing for a cleaner implementation.

Changing the MAC after migration is likely to be fine.

> Are application containers comfortable with a layer 2 virtualization ? I don't
> think so, because several jobs running inside the same host communicate via
> broadcast/multicast between them and between other jobs running on different
> hosts. The IP consumption is a problem too: 1 container == 2 IP (one for the
> root namespace/ one for the container), multiplicated with the number of
> jobs. Furthermore, lot of jobs == lot of virtual devices.

First if you hook you network namespaces with ethernet bridging
you don't need any extra IPs.

Second don't see the conflict you perceive between application containers
and layer 2 containment.

The bottom line is that you need at least one loopback interface per non-trivial
network namespace. One you get that having a virtual is no big deal. In
addition network devices don't consume less memory than a process. So lots
of network devices should not be a problem.

Eric