Subject: Re: dev_change_netns on a tunnel device
Posted by ebiederm on Tue, 19 Jun 2007 17:07:46 GMT
View Forum Message <> Reply to Message

"Sapan Bhatia" <sapan.bhatia@gmail.com> writes:

> Hi Eric,
>
> Are the current semantics of tunnel devices (ipip, ip_gre etc.) with
> respect to changing the current netns correct? These devices have an
> ip_tunnel structure associated with each net_device, which is
> currently a global (not per_net). The result is that you can set up in
> container X a tunnel with endpoints associated with a device in
> container Y in a different network namespace. This feature seems
> useful, because you can hook up containers on 2 machines with one of
> these tunnels, without using etun (or like) devices, but intuitively,
> it seems that the tunnel module should be listening for a
> DEV_UNREGISTER and reset it when the device is migrated

Good question. Examining the ipip case here is my conclusion.

- Because the directing of packets is based on ip address and multiple
  network namespaces are allowed to use the same ip addresses then
  the decode needs to take the network namespace into account.

- The only thing that would prevent the migration semantics from being
  correct is if you could manipulate a migrated tunnel in such a way that you
  could do something nasty to the source namespace.

  Since a tunnel change command is also a tunnel rename command that should
  force any ipip tunnel into using ip addresses from the current
  namespace, which makes it safe, and thus unable to affect the source
  namespace.

So semantically I believe the tunnel semantics are essentially
correct. However I believe there are several places where the code
needs to be updated to correctly implement those semantics.

Eric

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