Subject: Re: strict isolation of net interfaces
Posted by Daniel Lezcano on Fri, 30 Jun 2006 15:22:51 GMT
View Forum Message <> Reply to Message

Eric W. Biederman wrote:
> Daniel Lezcano <dlezcano@fr.ibm.com> writes:
>
>
>>Serge E. Hallyn wrote:
>>
>>>>Quoting Cedric Le Goater (clg@fr.ibm.com):
>>>>
>>>>
>>>>>we could work on virtualizing the net interfaces in the host, map them to
>>>>>eth0 or something in the guest and let the guest handle upper network layers ?
>>>>>
>>>>>lo0 would just be exposed relying on skbuff tagging to discriminate traffic
>>>>>between guests.
>>>>
>>>>This seems to me the preferable way. We create a full virtual net
>>>>device for each new container, and fully virtualize the device
>>>>namespace.
>>
>>I have a few questions about all the network isolation stuff:
>

It seems these questions are not important.

>
> So far I have seen two viable possibilities on the table,
> neither of them involve multiple names for a network device.
>
> layer 3 (filtering the allowed ip addresses at bind time roughly the current vserver).
>   - implementable as a security hook.
>   - Benefit no measurable performance impact.
>   - Downside not many things we can do.

What things ? Can you develop please ? Can you give some examples ?

>
> layer 2 (What appears to applications a separate instance of the network stack).
>   - Implementable as a namespace.

what about accessing a NFS mounted outside the container ?

>   - Each network namespace would have dedicated network devices.
>   - Benefit extremely flexible.

For what ? For who ? Do you have examples ?

> - Downside since at least the slow path must examine the packet
> it has the possibility of slowing down the networking stack.

What is/are the slow path(s) you identified ?

> For me the important characteristics.
> - Allows for application migration, when we take our ip address with us.
> In particular it allows for importation of addresses assignments
> mad on other machines.

Ok for the two methods no ?

> - No measurable impact on the existing networking when the code
> is compiled in.

You contradict ...

> - Clean predictable semantics.

What that means ? Can you explain, please ?

> This whole debate on network devices show up in multiple network namespaces
> is just silly.

The debate is not on the network device show up. The debate is can we
have a network isolation ___usable for everybody___ not only for the
beauty of having namespaces and for a system container like.

I am not against the network device virtualization or against the
namespaces. I am just asking if the namespace is the solution for all
the network isolation. Should we nest layer 2 and layer 3 vitualization
into namespaces or separate them in order to have the flexibility to
choose isolation/performance.

> The only reason for wanting that appears to be better management.
> We have deeper issues like can we do a reasonable implementation without a
> network device showing up in multiple namespaces.

Again, I am not against having the network device virtualization. It is
a good idea.

> I think the reason the debate exists at all is that it is a very approachable
> topic, as opposed to the fundamentals here.
>
> If we can get layer 2 level isolation working without measurable overhead
> with one namespace per device it may be worth revisiting things. Until

> then it is a side issue at best.

I agree, so where are the answers of the questions I asked in my
previous email ? You said you did some implementation of network
isolation with and without namespaces, so you should be able to answer...

-- Daniel