Subject: Re: L2 network namespace benchmarking
Posted by dev on Wed, 28 Mar 2007 07:49:26 GMT

View Forum Message <> Reply to Message

>>The loss of performances is very noticeable inside the container and
>>seems to be directly related to the usage of the pair device and the
>>specific network configuration needed for the container. When the
>>packets are sent by the container, the mac address is for the pair
>>device but the IP address is not owned by the host. That directly
>>implies to have the host to act as a router and the packets to be
>>forwarded. That adds a lot of overhead.
>
>
> Well it adds measurable overhead.
>
>
>>A hack has been made in the ip_forward function to avoid useless
>>skb_cow when using the pair device/tunnel device and the overhead
>>is reduced by the half.
>
>
> To be fully satisfactory how we get the packets to the namespace
> still appears to need work.
>
> We have overhead in routing. That may simply be the cost of
> performing routing or there may be some optimizations opportunities
> there.
>
> We have about the same overhead when performing bridging which I
> actually find more surprising, as the bridging code should involve
> less packet handling.
>
> Ideally we can optimize the bridge code or something equivalent to
> it so that we can take one look at the destination mac address and
> know which network namespace we should be in. Potentially moving this
> work to hardware when the hardware supports multiple queues.
yes, we can hack the bridge, so that packets coming out of eth devices
can go directly to the container and get out of veth devices from
inside the container.

> If we can get the overhead out of the routing code that would be
> tremendous. However I think it may be more realistic to get the
> overhead out of the ethernet bridging code where we know we don't need
> to modify the packet.
Why not optimize both? :)

Thanks,
Kirill