Subject: Re: linux-2.6.20-openvz tree
Posted by dev on Fri, 23 Mar 2007 08:51:47 GMT
View Forum Message <> Reply to Message

Carl-Daniel Hailfinger wrote:
> On 22.03.2007 16:59, Kirill Korotaev wrote:
>
>>Speaking about upstream merges:
>>there are 2 network virtualization implementations currently exist.
>>I'm not sure how much time it will take to merge this work,
>>it is very much depends on netdev@ maintainers. Maybe 2-3 month.
>
>
> OK, so the target is 2.6.22 or 2.6.23, if the usual time between
> releases is used as a basis for the estimation.
>
>
>>Why are you interested in that? Do you want to use some particular
>>feature?
>
>
> Yes. I currently use Linux policy routing for ONE machine performing
> double/triple/... NAT. Many people state that this is impossible,
> but it works fine unless two connections from the different subnets
> have identical 5-tuples. In that case, the connection tracking code
> gets confused. Unfortunately, the 5-tuple used by connection tracking
> and NAT has no means to incorporate the NF mark, so I hope I can
> use different containers for that.
>
> However, last time I checked, all network virtualization attempts
> did NOT consider one aspect I consider important for double NAT and
> virtual routers: Efficiency. Once I use virtualization, I am
> constrained to virtual network interfaces and suffer the overhead
> of multiple routing/bridging decisions for one packet.
> It would be great if I could make physical interfaces accessible
> in a VE without resorting to bridging or routing. For example,
> move eth0 and eth1 to one VE, eth2 and eth3 to another VE and
> keep eth4 under control of the HN.
This was possible for years in OpenVZ:

man vzctl
http://openvz.org/documentation/mans/vzctl.8

Network devices control parameters
--netdev_add name
  move network device from VE0 to a specified VE
--netdev_del name
  delete network device from a specified VE

this is exactly the thing you are talking about:
you can move eth0 and eth1 to one VE, eth2 and eth3 to another VE
and keep eth4 to HN.

And sure, this removes overhead of virtual network devices, additional
routing/bridging etc.
At the same time you can use separate NAT/firewall,routing,arp tables
inside each VE.

Isn't it the thing you want?

> I admit that most of this can be done with policy routing and NF
> marks, but connection tracking cares about neither of them.

Regards,
Kirill
