Subject: Re: L2 network namespaces + macvlan performances Posted by Rick Jones on Mon, 09 Jul 2007 16:59:06 GMT

View Forum Message <> Reply to Message

- > Between the "normal" case and the "net namespace + macvlan" case,
- > results are about the same for both the throughput and the local CPU
- > load for the following test types: TCP_MAERTS, TCP_RR, UDP_STREAM, UDP_RR.

>

- > macvlan looks like a very good candidate for network namespace in these
- > cases.

>

- > But, with the TCP_STREAM test, I observed the CPU load is about the
- > same (that's what we wanted) but the throughput decreases by about 5%:
- > from 850MB/s down to 810MB/s.
- > I haven't investigated yet why the throughput decrease in the case.
- > Does it come from my setup, from macvlan additional treatments, other? I
- > don't know yet

Given that your "normal" case doesn't hit link-rate on the TCP_STREAM, but it does with UDP_STREAM, it could be that there isn't quite enough TCP window available, particularly given it seems the default settings for sockets/windows are in use. You might try your normal case with the test-specific -S and -s options to increase the socket buffer size:

netperf -H 192.168.76.1 -i 30,3 -l 20 -t TCP_STREAM -- -m 1400 -S 128K -S 128K

and see if that gets you link-rate. One other possibility there is the use of the 1400 byte send - that probably doesn't interact terribly well with TSO. Also, it isn't (?) likely the MSS for the connection, which you can have reported by adding a "-v 2" to the global options. You could/should then use the MSS in a subsequent test, or perhaps better still use a rather larger send size for TCP_STREAM|TCP_MAERTS - I myself for no particular reason tend to use either 32KB or 64KB as the send size in the netperf TCP_STREAM tests I run.

A final WAG - that the 1400 byte send size interacted poorly with the Nagle algorithm since it was a sub-MSS send. When Nagle is involved, things can be very timing-sensitive, change the timing ever so slightly and you can have a rather larger change in throughput. That could be dealt-with either with the larger send sizes mentioned above, or by adding a test-specific -D option to set TCP_NODELAY.

happy benchmarking,		
rick jones		
Containers mailing list		

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

Page 2 of 2 ---- Generated from OpenVZ Forum