



Contribution ID: 160

Type: **not specified**

Can the Linux networking stack be used with very high speed applications?

Monday, September 12, 2022 12:30 PM (30 minutes)

Ethernet networking speeds continue to increase – 100G is common today for both NICs and switches, 200G has been available for years, 400G is the current cutting edge with 800G on the horizon. As the speed of the physical layer increases how does S/W scale - specifically, the Linux IP/TCP stack? Is it possible to leverage the increasing line-rate speeds for a single flow? Consider a few example data points about what it means to run at 400 Gbps speeds:

1. the TCP sequence number wraps 12.5 times a second - i.e., wrapping every 80 msec, and
2. at an MTU of 1500B, to achieve 400G speeds the system needs to handle 33M pps - i.e., a packet arrives every 30 nsec (for reference, an IPv4 FIB lookup on a modern Xeon processor takes ~25nsec).

We used an FPGA based setup with an off-the-shelf server and CPU to investigate the Linux networking stack to determine how fast it can be pushed and how well it performs at high rates for a single flow. With this setup tailored specifically to push the kernel's TCP/IP stack, we were able to achieve a rate of more than 670 Gbps (application data rate) and more than 31 Mpps (different tests) for a single flow. This talk discusses how we achieved those rates, the lessons learned along the way and what it suggests are core requirements for deploying very high speed applications that want to use the Linux networking stack. Some of the topics are well established such as the need for GRO, TSO, zerocopy and a reduction of system calls; others are not so prominent. This talk presents a systematic and comprehensive review of the effect of variables involved and serves as a foundation for future work.

I agree to abide by the anti-harassment policy

Yes

Primary author: AHERN, David

Presenter: AHERN, David

Session Classification: eBPF & Networking

Track Classification: eBPF & Networking Track