High-Performance Networking in Distributed Quasi Real-time Systems

Summer Student Lightning Talks

Daniel Lupu

Supervisor: Fabrice Le Goff

15/09/2022
Introduction

• ATLAS TDAQ will undergo an upgrade to take advantage of HL – LHC

• Dataflow sub-system is evaluating TCP/IP for network communication

• Goal: characterize performance of a 100GbE test bench
Network Topology

- 8 servers running CentOS
- Switch: QFX5120-32C
- NIC: Mellanox ConnectX-5
- Jumbo frames: 9000 bytes

Receiving machine:
- CentOS Stream 8
- 1× AMD EPYC 7302P
- NIC firmware and driver updated to the latest version

Tools: `iperf` command and `Python` in a GNU/Linux environment
Single Connection – MSS

- The Maximum Segment Size (MSS) is the largest amount of data in a TCP segment, bounded by the frame size.
- Almost linear increase in throughput.
- Receiver CPU core saturated for all values.
- In power-saving mode throughput decreased by 35%.
Multiple Connections

*Many-to-one traffic pattern*

At least three connections are needed to achieve a throughput more than 90 Gbit/s.

For a more realistic scenario to a DAQ system, 15 connections per machine were started, for a total of 105 senders to a single receiver.
Multiple connections - issues

- Connections dropping to 0 Gbit/s

- `rx_discards_phy` counter increased after each test

<table>
<thead>
<tr>
<th></th>
<th>rx_discards_phy (#packets discarded)</th>
<th>Standard deviation</th>
<th>Discarded data mss x #packets</th>
<th>Received data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>2968204</td>
<td>17368</td>
<td>24.7 GB</td>
<td>306.2 GB</td>
</tr>
<tr>
<td>Custom coalescing</td>
<td>162289</td>
<td>6790</td>
<td>1.35 GB</td>
<td>336.3 GB</td>
</tr>
</tbody>
</table>
Multiple connections – TCP parameters

Congestion control algorithm: dctcp

Reducing TCP retransmission timeout

Congestion control algorithm: dctcp

Aggregated throughput: 75.40 Gbit/s

Reducing TCP retransmission timeout

Aggregated throughput: 73.71 Gbit/s
Multiple connections – NIC settings

Interrupt coalescing

Aggregated throughput: 89.67Gbit/s
Conclusions

CPU plays a major role in a high-performance networking scenario, both for single and multiple connections.

Applying data center aware settings can be a sensible choice in a high-performance networking distributed system.
QUESTIONS?

daniel.lupu@studenti.unipd.it