Purpose

This document supplements load balancing part of the chapter ‘Requirements and Challenges of AI computing Networks’ in AICN report draft v0.1 (1-24-0022-00-ICne-aicn-report-draft-v0-1).

Author: Jieyu Li

Requirements and Challenges of AI computing Networks

Scale

Efficiency

Load balancing

✓ ***Brief introduction of LB***

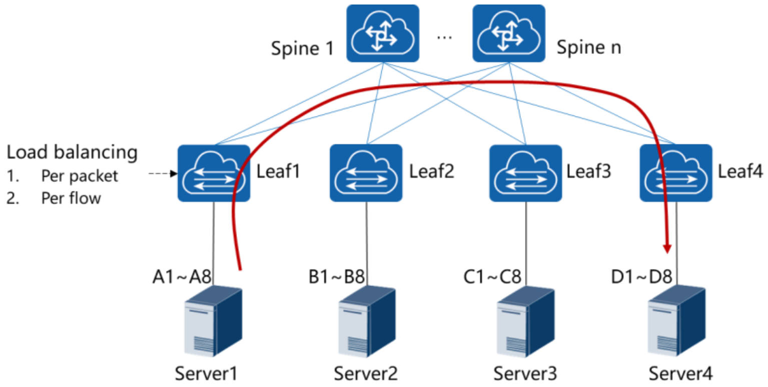
Modern datacenter networks generally provide multiple forwarding paths for each end pairs. Load balancing (LB) is a kind of technologies aiming at fully utilizing these redundant paths. LB can effectively relief the congestion hotpot intra network and raise the overall throughput by distributing flows or packets among multiple paths.

✓ ***The requirements decided by AI workload feature: good at balancing a small number of Elephant flows. → ECMP works badly → per-packet LB is the trend.***

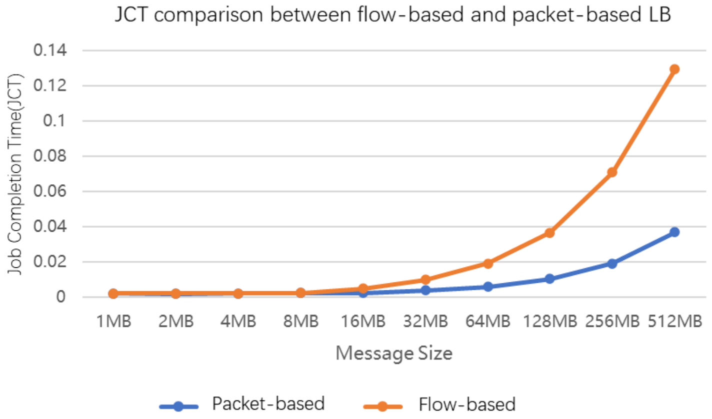
The effectiveness of a LB scheme is tightly related to the network traffic pattern. As analyzed in the former chapter, the AI traffic is mainly composed of a small number of large bandwidth flows. It’s hard for the most conventional LB algorithm ECMP to evenly distribute few elephant flows restricting by its flow-based granularity. It is almost coming into a consensus that AI network need a more fine-grained load balancing scheme to service these elephant flows. Per-packet LB solution is widely considered as the technology trend to avoid per-flow LB’s drawbacks for AI network.

The work of [1] conduct a simple experiment to verify that per-packet LB performs better on Job completion time (JCT) than per-flow.

- **Experiment settings:** The topology is the classic two-layer clos network. There are 4 servers. Each server has 8 GPUs and 8 NICs. Running 8 jobs that is between A1 and D1, A2 and D2, … A8 and D8 respectively.



- **Results:** Figure shows the JCT of per-flow and per-packet LB under different message size. The per-packet LB achieve shorter JCT obviously with the message size increasing. When the message size is 512 MB, JCT of per-packet LB is about one-third of flow-based LB.



[1] 1-24-0004-05-ICne-load-balancing-challenges-in-ai-fabric.

✓ ***The challenges toward trade-off between finer granularity LB and out-of-order problems***

The direct side-effect of per-packet LB is causing packets of a flow arriving at receiver out of order. These out-of-order packets lead to troubles mainly in two aspects:

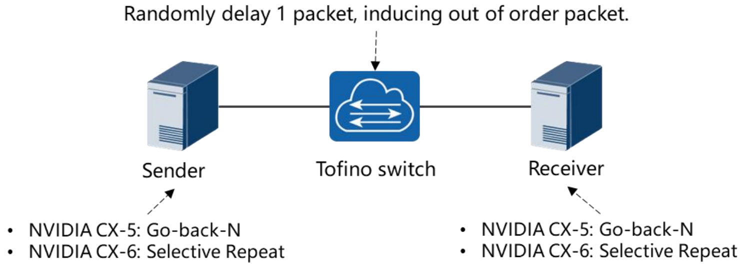
1) **Re-ordering:** End-side device or NIC may need to re-order out-of-order packets, which causes severe scalability problem especially for hardware-based protocol like RDMA.
2) **Packet loss recovery:** Lowering the efficiency of packet loss recovery.
    - **Loss recovery protocol Go-back-N (GBN) and Selective Repeat (SR):** In initial design, RDMA NIC (RNIC) adopt Go-back-N protocol to ensure reliable data transfer. The receiver only keeps track of the expected sequence number to receive next, and simply discard all the out of order packets. To provide the ability to deal with out of order packets, selective repeat protocol has been supported in NVIDIA ConectX-6 (CX6) NIC. SR allows RNIC buffer the out of order packets and selectively retransmit the unreceived packet.
    - **Packet loss can’t be detected fast:** Although SR can deal with out of order packets, the receiver cannot directly determine whether the packet is lost or just delayed. SR relies on the timeout mechanism to detect packet loss, which is inefficient and

causes the sender to significantly reduce its sending rate once loss happen.

- **Packet loss can't be located:** If the packets loss is caused by silent network device failure, it' hard for servers to locate the error as having no knowledge of the packets forwarding path.

The work of [2] quantify how out-of-order packets affect RDMA performance, here list some key information.

- **Experiment settings:** The sender and receiver are connected by Tofino programmable switch, and both equipped with an NVIDIA Mellanox ConnectX-5 (CX5) or ConnectX-6 (CX6) RNICs that support Go-back-N and SR, respectively.
  To induce out of-order packet arrivals, selecting randomly a packet from the RDMA flow and recirculating it in the switch before forwarding it.


Randomly delay 1 packet, inducing out of order packet.
Sender
Tofino switch
Receiver
• NVIDIA CX-5: Go-back-N
• NVIDIA CX-6: Selective Repeat
• NVIDIA CX-5: Go-back-N
• NVIDIA CX-6: Selective Repeat

- **Results:** Compares the Flow Completion Time (FCT) for short (10 KB) and long (1MB) flows.
  RDMA is highly sensitive to even a single out-of-order packet arrival. CX6 enabled SR perform better than CX5 enabled GBN due to fewer retransmissions, but is still greatly impacted by the out-of-order packet.
  Compared FCT under in-order delivering, SR's performance [3]:
  - In 10KB message, P50(50th percentile) of FCT is 1.65 times longer than in-order delivering, and P99(99th percentile) of FCT is 1.25 times longer.
  - In 1MB message, P50 of FCT is 2.65 times longer than in-order delivering, and P99 of FCT is 3.27 times longer.

[2] Song C H, Khooi X Z, Joshi R, et al. Network Load Balancing with In-network Reordering Support for RDMA[C]//Proceedings of the ACM SIGCOMM 2023 Conference. 2023: 816-831.
[3] https://www.youtube.com/watch?v=SlCJBGpn_4I