Load balancing challenges in AI fabric

Weiqiang Cheng (China Mobile)

Ruixue Wang (China Mobile)

Jieyu Li (China Mobile)

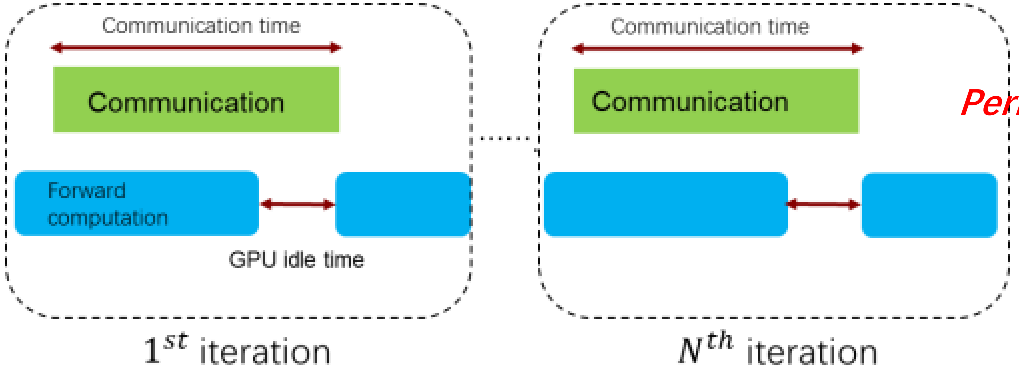
Yadong Liu (Tencent)

AI Traffic pattern challenge

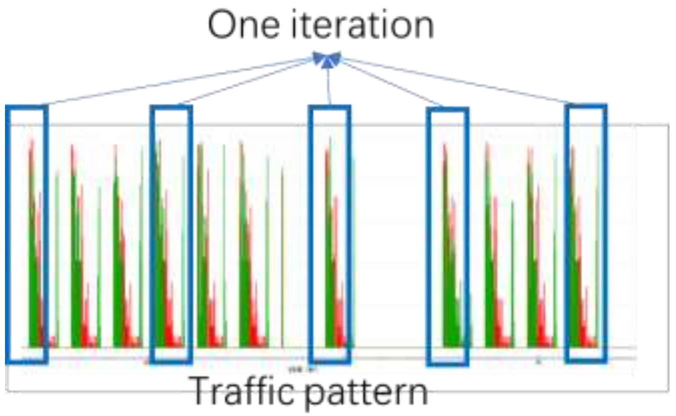
Traditional DC Traffic pattern

- Many asynchronous small BW flows.
- Chaotic pattern averages out to consistent load.

AI (All-to-all Collective) Traffic Pattern



Periodical



Burst

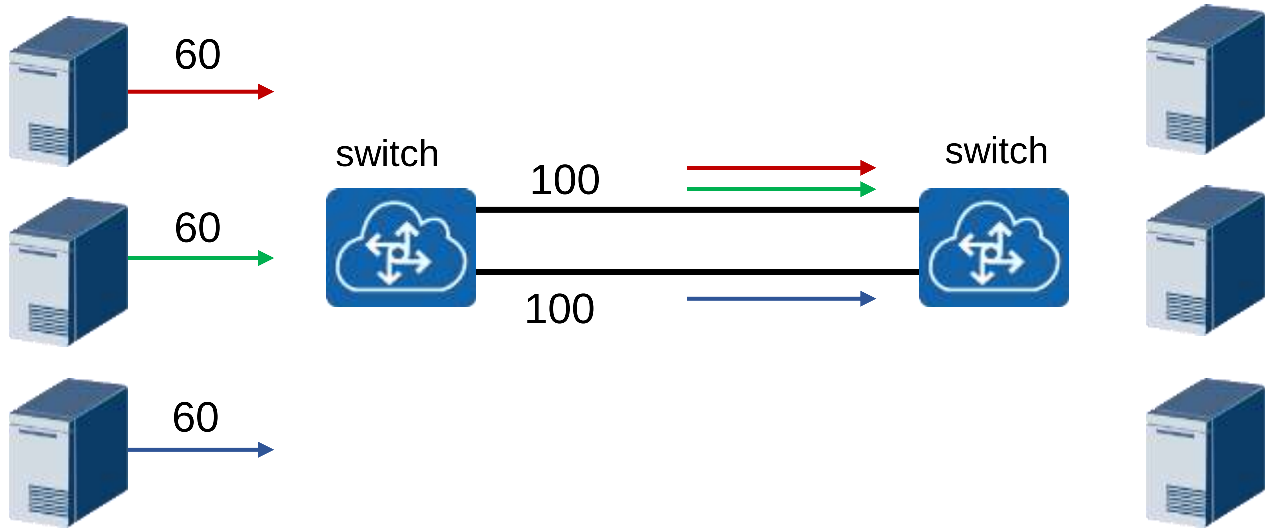
- Few synchronous high BW flows.
- Synchronization magnifies long tail latency and bad load balancing decisions.

Data from Cisco's public whitepaper.

Traditional flow-based LB perform poorly

- **Flow-based load balancing means switches distribute packets to multiple paths in the flow granularity, and Packets within a flow take the same forwarding path.**

- **The inherent drawback of flow-based LB is its coarse granularity:**
  - It does not take into account the size of different flows;
  - Especially in AI fabric, it's hard to balance the few and high bandwidth flows well.



Example

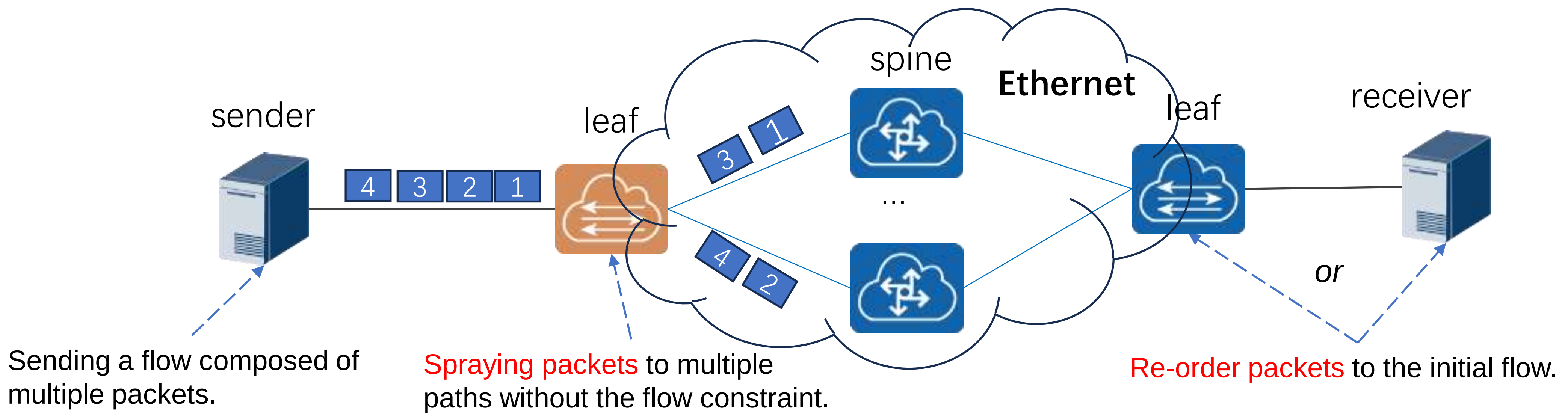
- When the number of flows is not divisible by the number of available paths, it' s impossible to get a optimal balance using flow-based LB.

Packet-spray-based LB become the trend for AI fabric（1）

- **Packet spray means switches distribute each packet to multiple paths independently, making the load on the network more balanced than flow-based.**

- **There are several routes supporting packet spray:**
  - Cell-based in dedicated network or ethernet-based： Standardization → ✓ **Ethernet-based.**
  - NIC-driven or Network-driven: Applicable to different scenarios. → Focus on **network-driven** solution in this document.

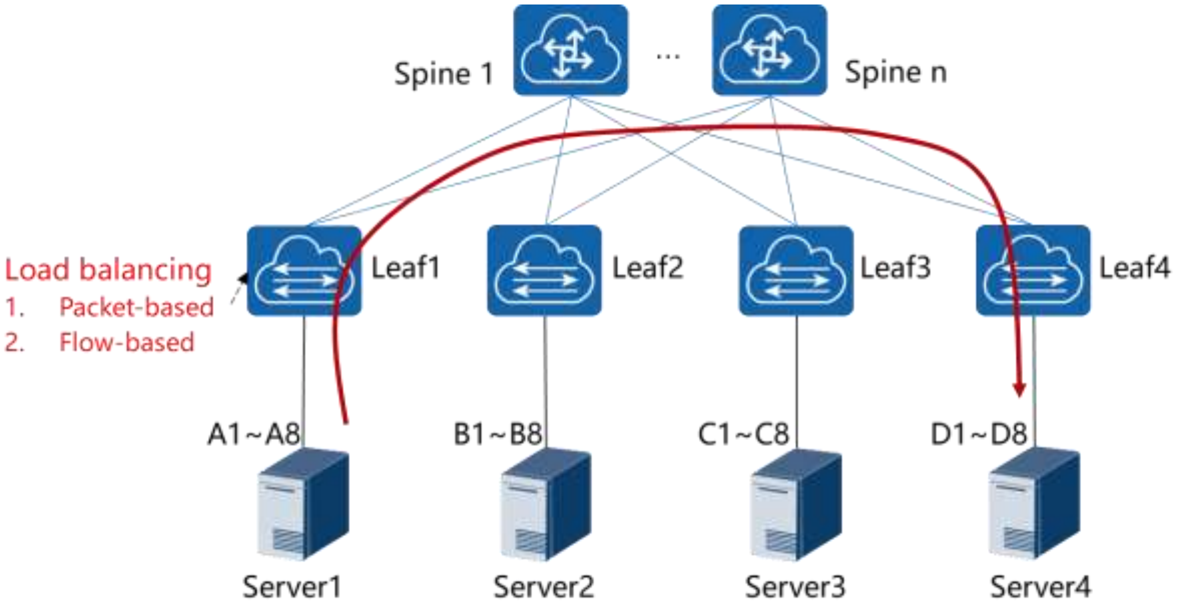
- **Basic Architecture of network-driven packet spraying in ethernet:**



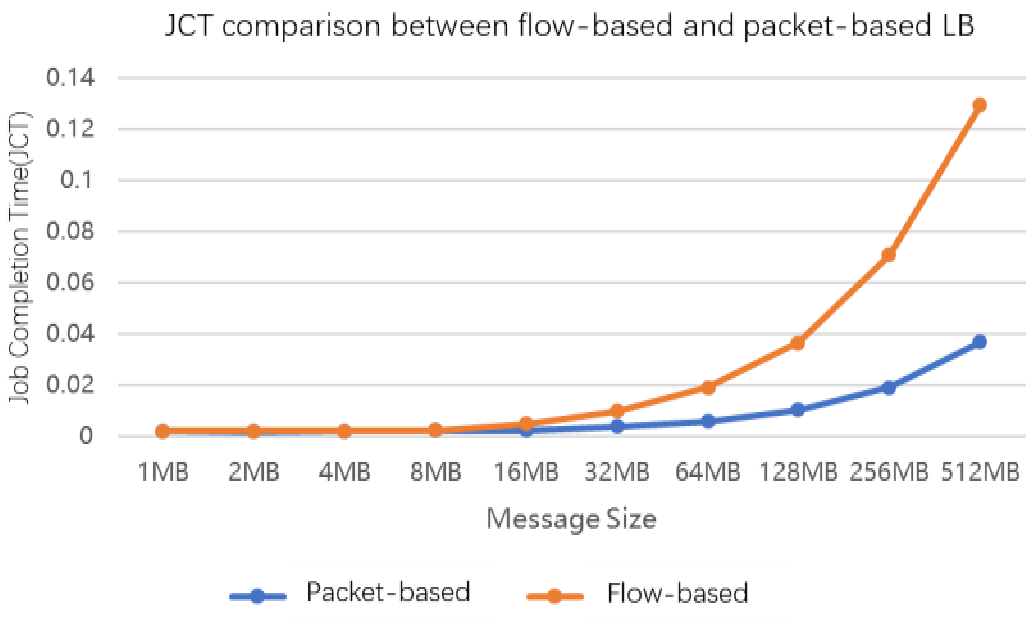
Packet-spray-based LB become the trend for AI fabric（2)

- **We conduct an experiment to evaluate the performance of flow-based and packet-spray-based LB.**

Experiment settings



Results

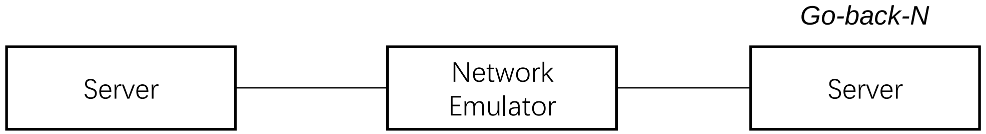


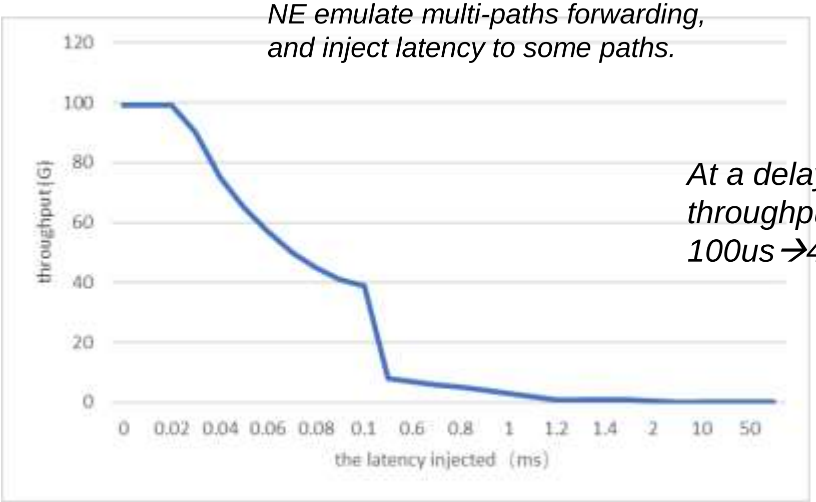
- The topology is the classic two-layer clos network, 4 servers, 8GPU with 8 NICs in a server.
- There are 8 jobs running: A1~D1、A2~D2….A8~D8.

- Testing the task completion time (JCT) of flow-based and packet-spray-based load balance under different message size.
- In a 512MB scenario, JCT of packet-spray-based LB is reduced to about one-third compared to flow-based.

Challenges in Packet-based LB

- **The main side-effect of packet-spray-based LB is causing packets of a flow arriving at receiver out of order:**
  - Re-order problem.
  - **Reliability problem: Loss-detection and retransmission;**
- **Out-of-order cause performance degradation significantly under Go-back-N mechanism.**
  - The mainstream RNIC adopt Go-back-N mechanism to provide reliability.
  - A lot of out-of-order packets may trigger frequently Go-back-N, resulting in a precipitous decline in throughput, as shown in the right emulation.





At a delay of about 20us, throughput begins to decline. 100us→40%, 500us→10%

- **RNIC can adopt Selective ACK to improve GO-back-N, but still existing problems hindering performance.**
  - The receiver can not directly determine whether the packet is lost or just out of order through the PSN,
  - relying on the timeout mechanism to detect packet loss reduces the sending rate.
  - Accurate fast-retransmit is necessary, but only by receiver is often not possible.

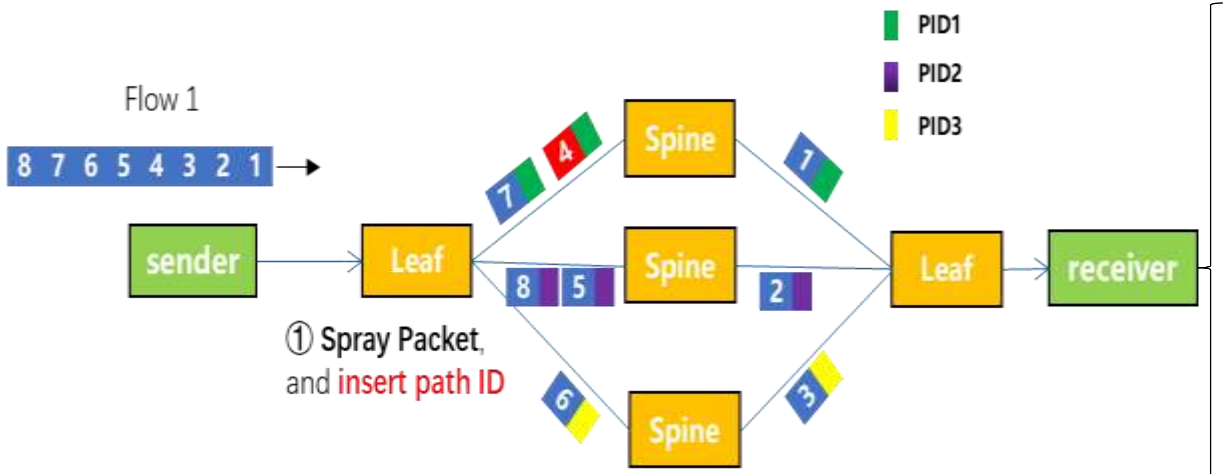
- **A preliminary conclusion is that processing out-of-order packets exclusively on the receiver NIC can hardly achieve optimal performance.**

Network can do more…

- **In packet-spray-based LB, the root difficulty of receiver dealing with out of order packets is that it does not know the forwarding path and state of each packet.**
- **An intuitive solution is that network provide receiver the path information of packet forwarding to help loss detection and fast retransmission.**

Key idea: network device insert the path information(e.g. Path ID) into packet header, so that the receiver can detect the loss more quickly and execute fast retransmission.

Example



② Update the receiving window of flow 1, *assume the 'hole' is packet 4*:

PSN	1	2	3	4	5	6	7	8
state	1	1	1	0	1	1	1	0

③ Update the max receiving PSN of each path of flow 1:

- Path 1: maxRcvPSN[1]:7
- Path 2: maxRcvPSN[2]:5
- Path 3: maxRcvPSN[3]:6

④ Compare the hole number with maxRcvPSN of each path:

- If hole number < maxRcvPSN of all paths → Packet 4 loss

Current industrial support for packet-based LB

① Cisco：Silicon one

- *"An alternate solution is to use a fully scheduled fabric to connect the GPUs. This approach sprays packets across all links and reoders the packets at the exit, so no network congestion builds up. Simply put, it allows for an ideal interconnect under all traffic conditions."*
- Solutions - Cisco Silicon One Product Family White Paper - Cisco

② Broadcom：Tomahawk 5

- *"It supports both per-packet spray and flowlet modes of operation and can be enabled selectively for different traffic types with ineligible flows falling back to hash-based ECMP. DLB is successfully deployed in multiple networks today."*
- Cognitive routing in the Tomahawk 5 data center switch (broadcom.com)

③ Juniper：Junos OS

- *"Randomly sprays the packets to the aggregate next hops to ensure that the next hops are equally loaded resulting in packet reordering."*
- Per-packet random spray load balancing | Juniper Networks Pathfinder Feature Explorer

……

- **The mainstream chip venders（Cisco、Broadcom、Nvidia..） have supported the packet spraying to balance load, but their solutions are different. → standardization of packet-spray-based load balancing on ethernet is needed.**

Summary

- **Introduce the drawbacks of traditional flow-based ECMP for AI fabric, and packet-spray-based load balancing become the trend.**
- **Analyze the challenges bring to receiver in packet-spray-based load balancing.**
- **Network can assist receiver to solve the challenges.**
- **Potential Standard Requirements：** Need to standardize path information in L2 for network-assisted fast retransmission, such as path ID.

Next Action

- **Propose a Study/Work Item：** Packet-Spray-Based AI Fabric
- **Scope：** Packet-Spray-Based Load Balance 、Packet-Spray-Based Congestion Control、Reorder、Ethernet QoS、Telemetry、high-precision OAM、Protection etc.

Thank You !