New Requirements and Thoughts for AI Data Center Networks

Presenter: Weiqiang Cheng (China Mobile)

IETF-117

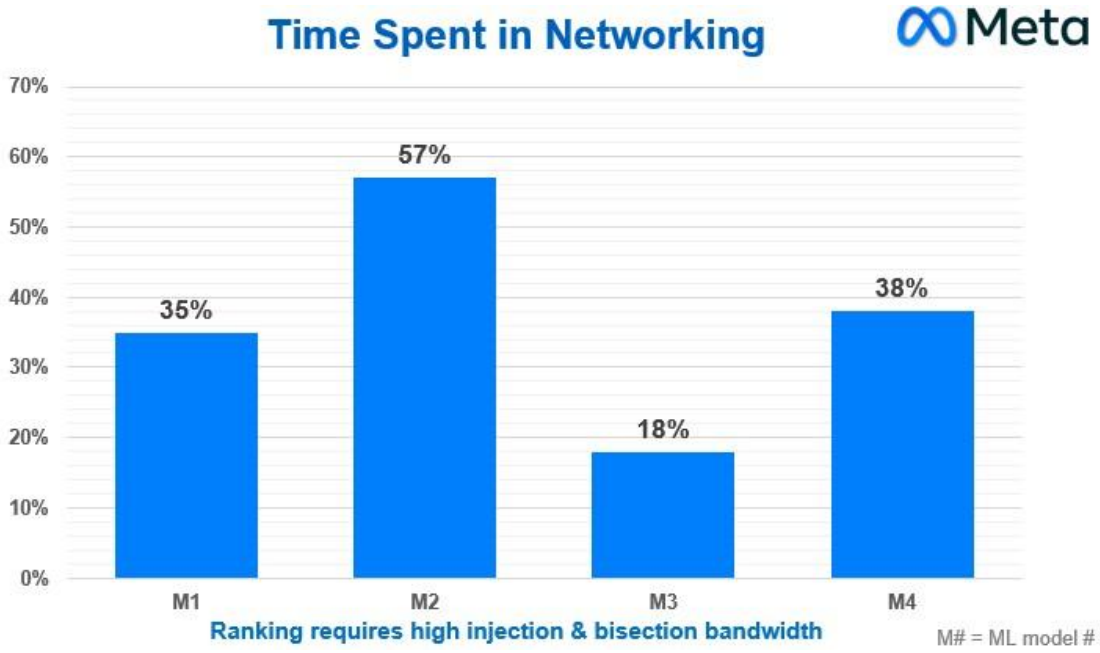
Networks Problem in AI Data Center

Challenge 1: Large-scale training network

- 10 Thousands of GPUs with high-speed interconnectivity
- Thousands of RDMA QP connections
- Access bandwidth ≥ 100G

Challenge 2: High-performance requirements

- Sensitivity to packet loss
- Significant impact from latency and jitter
- Sub-millisecond fault recovery

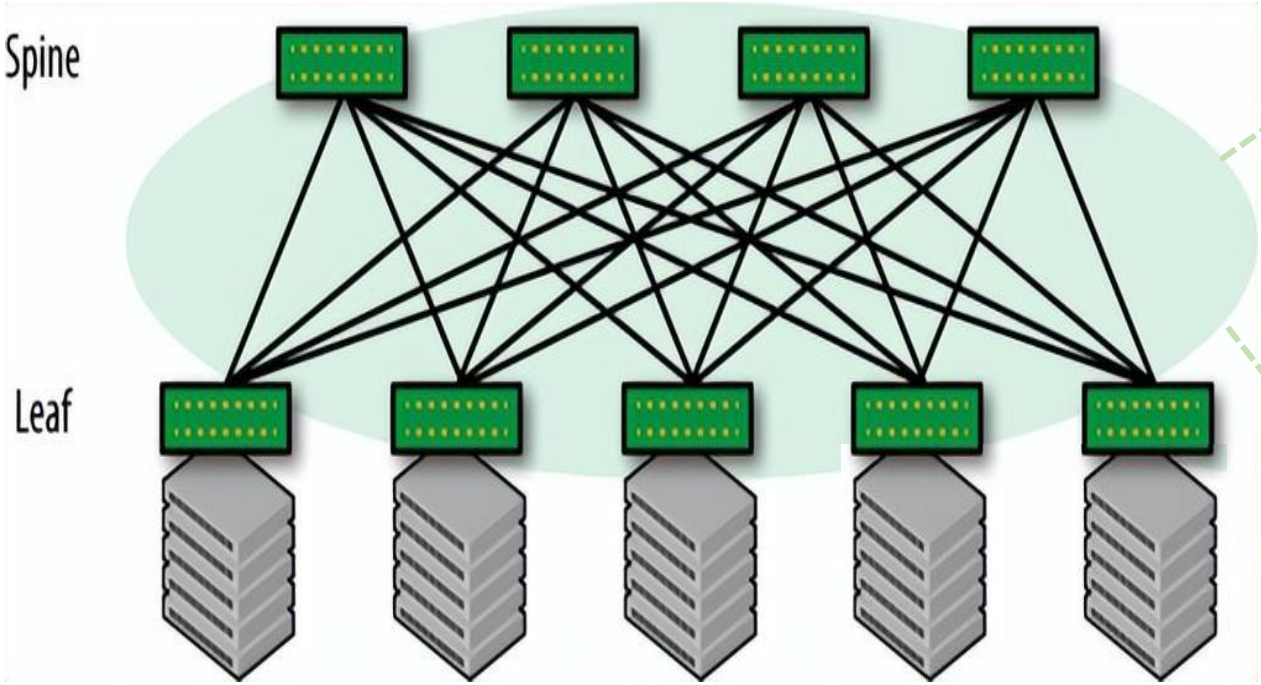


2022 OCP keynote by Alexis Bjorlin, VP, Infrastructure at Meta

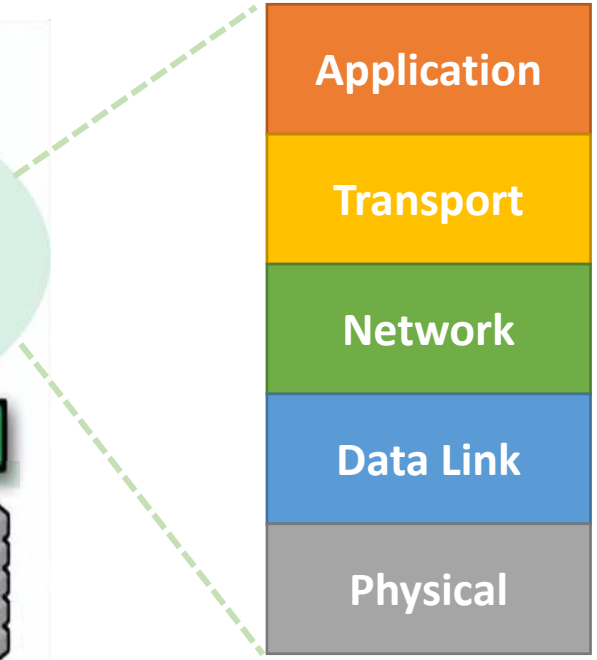
Traditional networks results in low bandwidth utilization, significant tail latency effects, and complex system designs. Network performance becomes a bottleneck in improving the efficiency of AI training.

Possible solutions for AI data center networks

Topology Changes



Stack Changes



...

Better RDMA

Simplified Protocols

Efficient Load balance

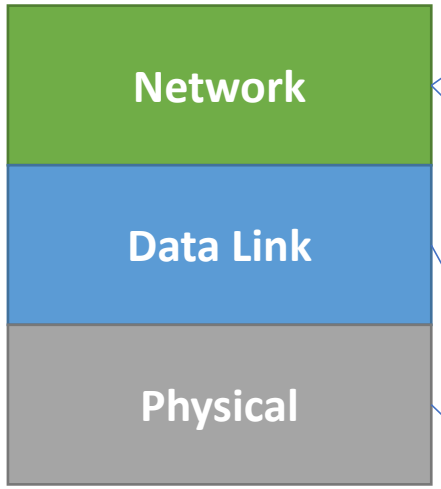
Lower Latency

Overview of some attempts in Networks

New Topology

New Topology: Dragonfly, Dragonfly+ ...
https://datatracker.ietf.org/doc/draft-wang-rtgwg-dragonfly-routing-problem/
https://datatracker.ietf.org/doc/draft-agt-rtgwg-dragonfly-routing/

New Protocols



SRv6 for DCN: source-driven path programming multicast rather than hop by hop establishing multicast tree.
https://www.ietf.org/archive/id/draft-cheng-rift-srv6-extensions-01.txt
https://datatracker.ietf.org/doc/draft-ietf-spring-srv6-srh-compression/

MSR6 for AI Multicast offloading: source-driven path programming multicast rather than hop by hop establishing multicast tree.
https://datatracker.ietf.org/meeting/116/materials/slides-116-bier-07-bier-multicast-use-case-in-dc-00.pdf
https://www.ietf.org/archive/id/draft-liu-multicast-for-computing-storage-00.txt

Globle Scheduling Ethernet(GSE) : Packet Spraying based on packet container and Lower Latency FEC in Ethernet layer are described in GSE white paper post by CMCC. You can scan the QR to download the white paper.

Brief introduction to Globle Scheduling Ethernet(GSE)

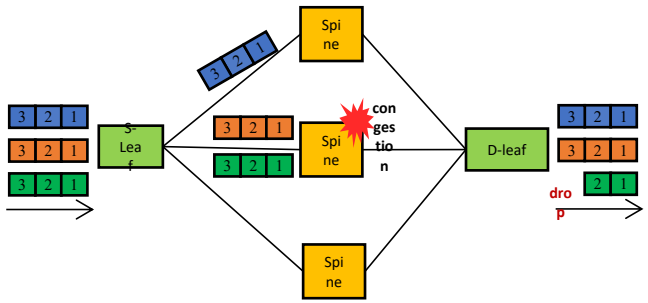
In May 2023, 15 partners worldwide jointly released the "Globle Scheduling Ethernet (GSE) White Paper", include China Mobile, Tencent, Intel, Marvell, Huawei, ZTE, Ruijie, H3C, FiberHome, Spirent, Tsinghua University, Centec and others

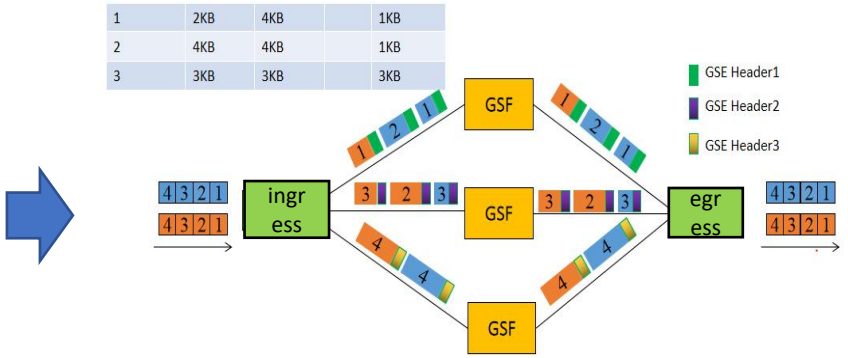
Container Based Packet Spraying

At ingress, no hashing is used to select a link for a stream, and packets are evenly distributed across all available links. At egress, packets within a flow need to be re-ordered. Packet containers are introduced to ensure bytes balance.

Enhanced VoQ

Ingress Virtual Output Queues (VOQs) store packets for specific destinations and traffic classes. when granted, Ingress node can transmit packet, ensuring efficient packet transmission and preventing oversubscription.



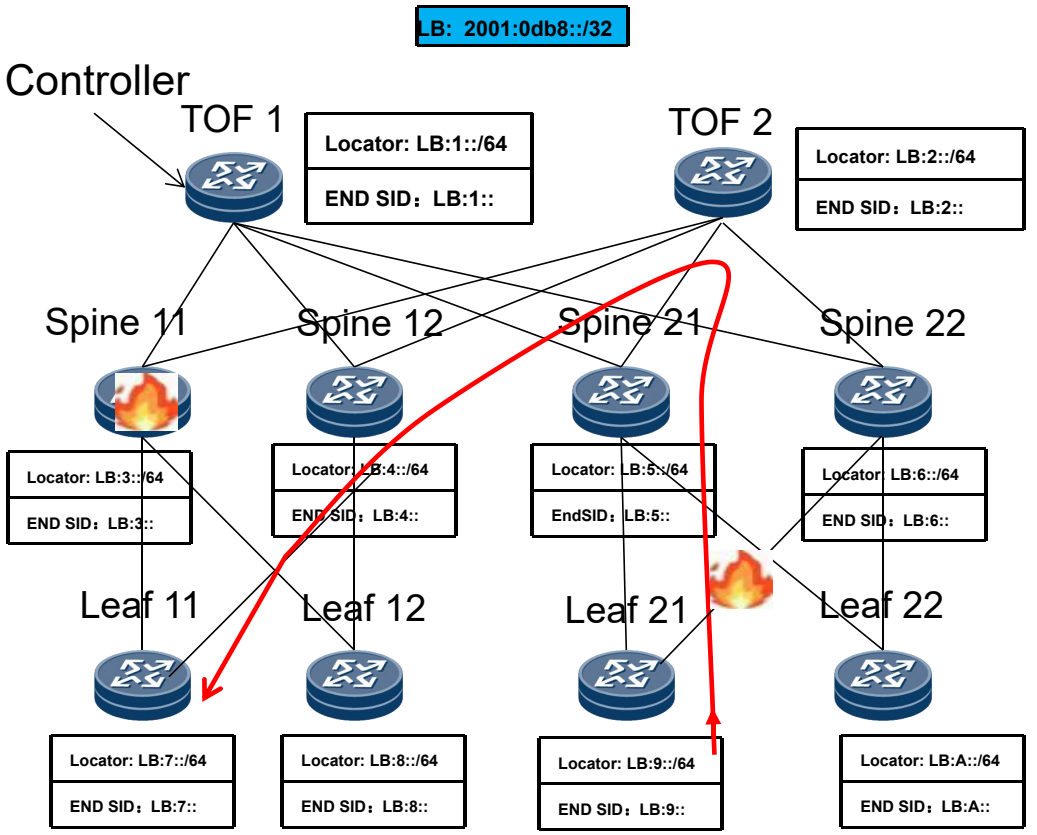


Compressed SRv6 in DC network

Using compressed SRv6 in data center networks brings several advantages

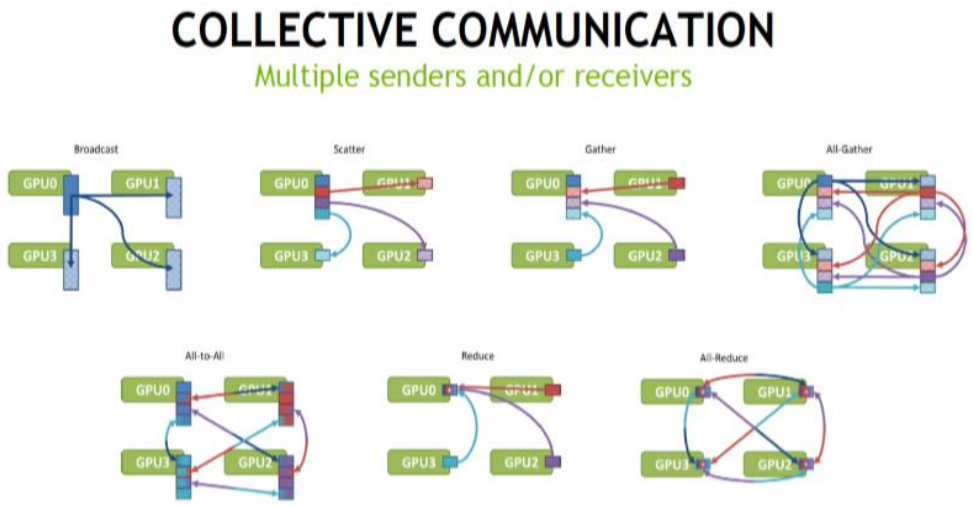
- Simplified end-to-end protocols: Currently, multiple tunneling technologies such as MPLS and VxLAN are used separately in the data center and inter-data center. SRv6 can simplify end-to-end protocols.
- Enhanced TE capabilities: SRv6 enables easier load balancing and facilitates adaptive routing.
- Better Service Function Chaining (SFC): SRv6 inherently supports SFC and can be considered for simplified service provisioning through end-to-end orchestration with WAN networks.
- Enable Network Programmability: Customers have the complete flexibility to program the SID in SRH to enable simplified network programming.

SRv6 in data center networks with RIFT

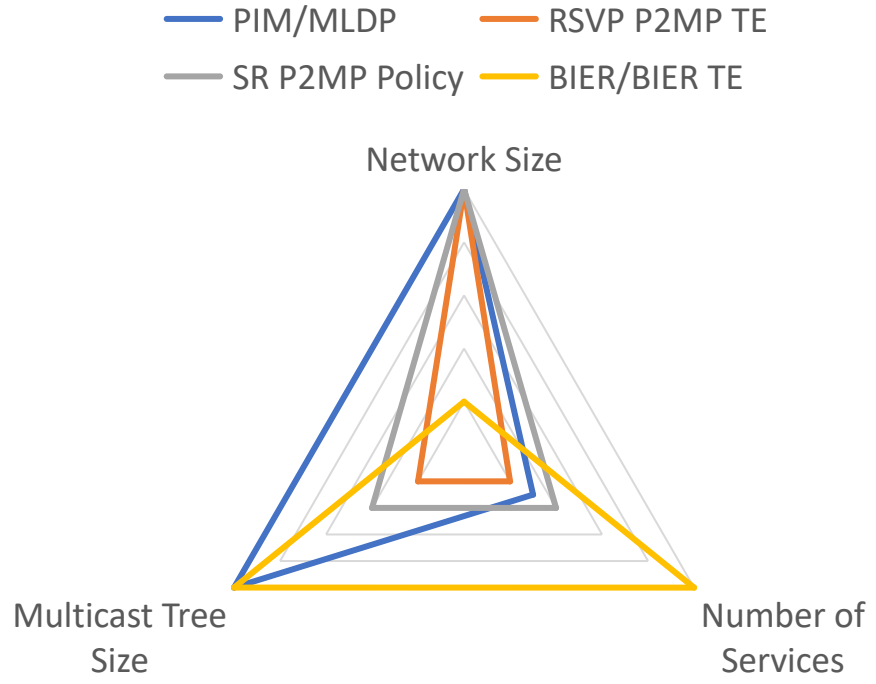


New Multicast is required for AI DC

- ✓ Current data centers are using unicast to emulate multicast, where replication is done by the application layer rather than the network device;



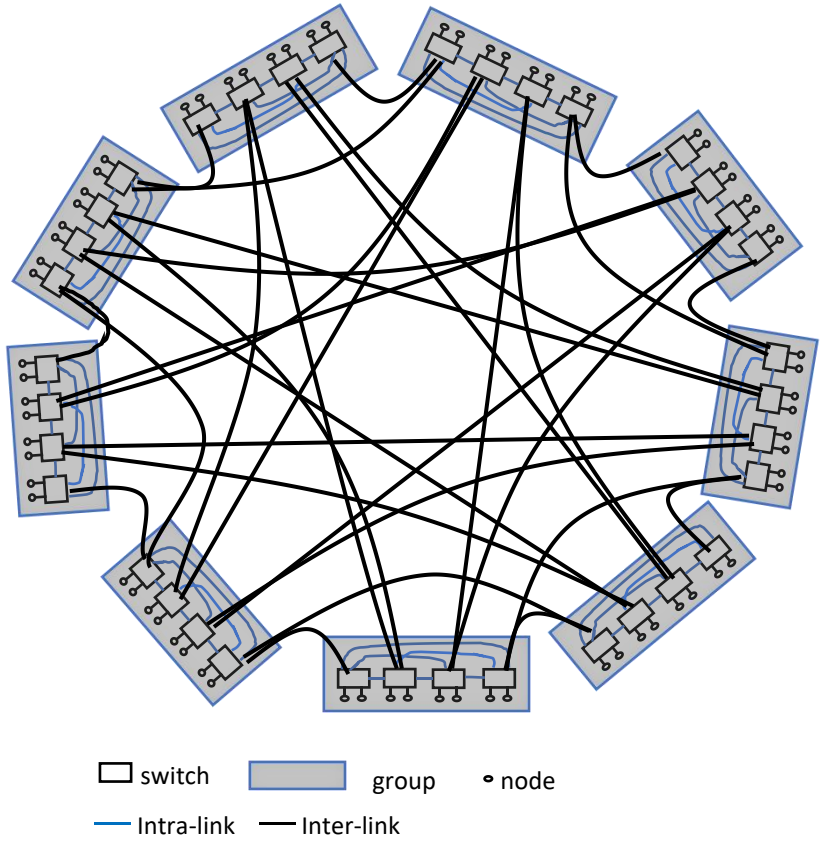
✓existing multicast technologies is not satisified with multicast requirements in AI training DC



Multicast Source routing over IPv6 (MSR6): source-driven path programming multicast rather than hop by hop establishing multicast tree.

Possible new Topology and Routing

The Dragonfly topology is evaluated by industry



Problem analysis of routing for Dragonfly topology

- Problem 1:

The existing routing protocols can not support dynamic load balancing mechanisms.

- Problem 2:

Lacks standards for network congestion and a notification mechanism for remote congestion, which makes it difficult to timely exchange congestion information.

Next Steps

- Questions or Comments ?