

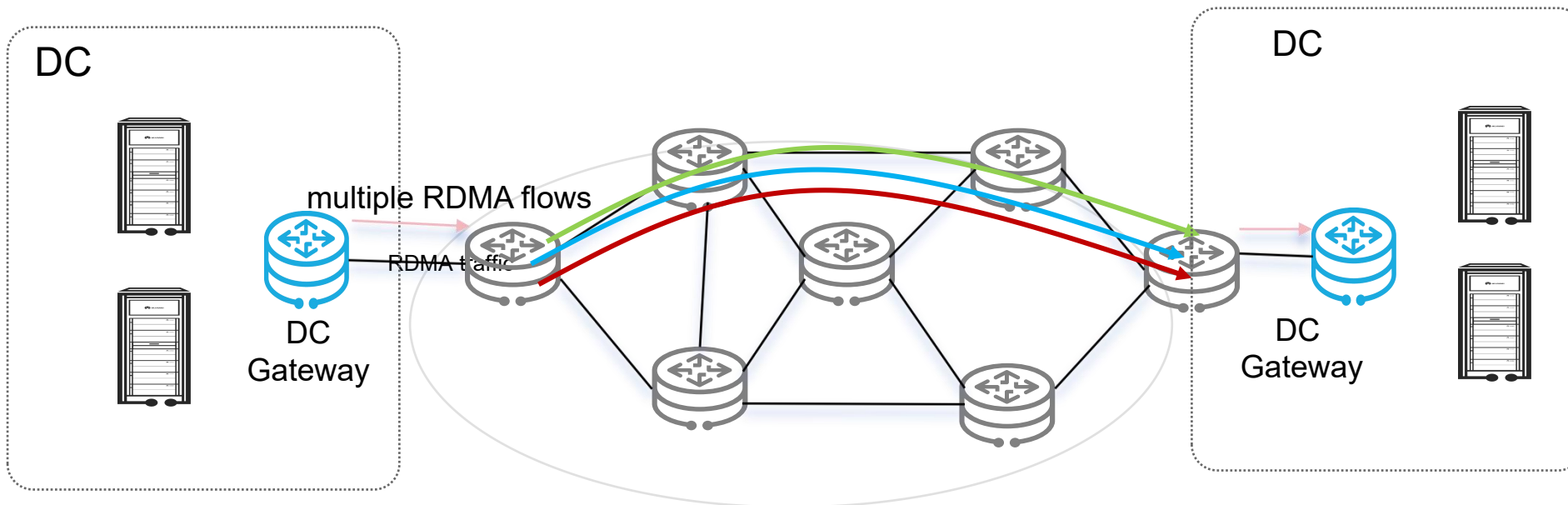
A RoCEv2 Flow-Level Load Balancing Method Based on the IPv6 Flow Label

Jiayuan Hu(Presenter), Jie Dong, Xia Gong

2026.7

Problem Statement: The "Elephant Flow" Dilemma in RoCEv2

- **The Challenge:** AI/HPC workloads heavily rely on RoCEv2 for high-throughput, low-latency data transfer. However:
 - Multiple parallel RDMA sessions from the same host **share an identical 5-tuple** (Src/Dst IP, Src/Dst Port, Protocol).
 - Traditional ECMP hash based on the **5-tuple treats these distinct flows as one single flow.**
- **The Result:** "Elephant flows" are all hashed to the **same network path**. This leads to severe congestion, packet loss, and significant performance degradation.



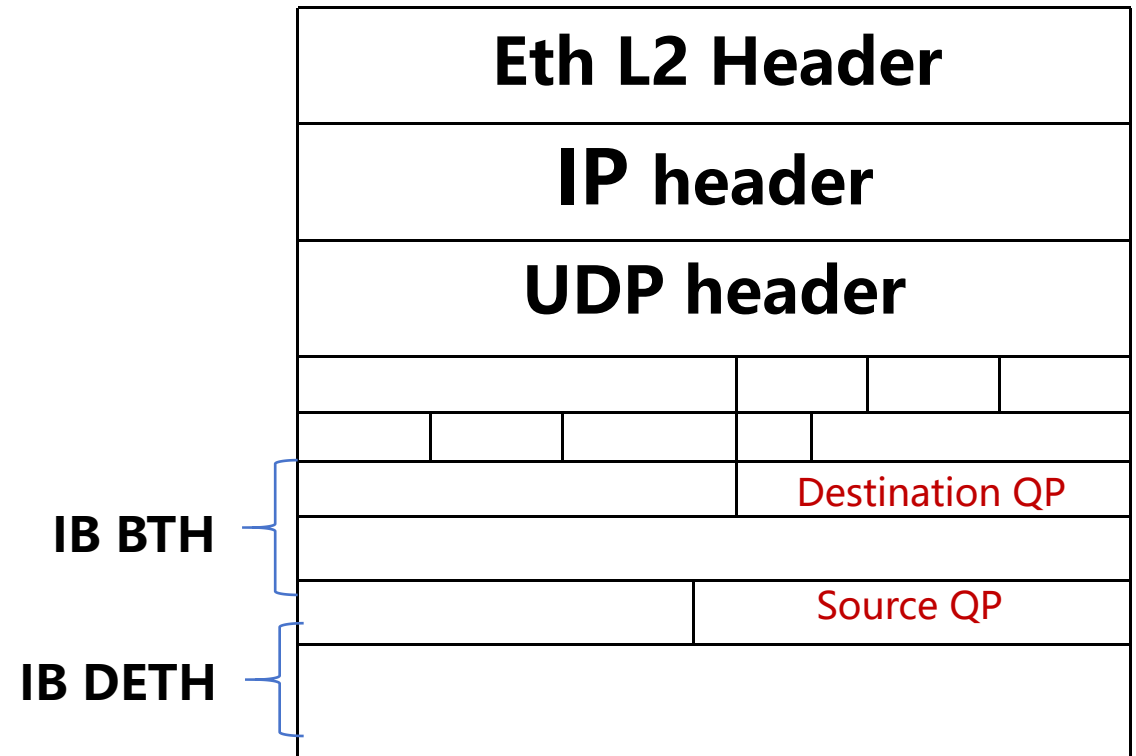
The Core Idea: Extracting Entropy from the QP

The **Queue Pair (QP)** number is the unique identifier for an RDMA connection.

■ Our Proposed Method:

- Step 1: The ingress network device parses the RoCEv2 packet.
- Step 2: It extracts the Source QP and Destination QP (24-bit each).
- Step 3: These QP values, combined with other entropy, are hashed to generate the Flow Label.

This creates the necessary differentiation that the 5-tuple cannot provide!



Solution Overview: Flow Label Generation Process

■ Hash Input Construction:

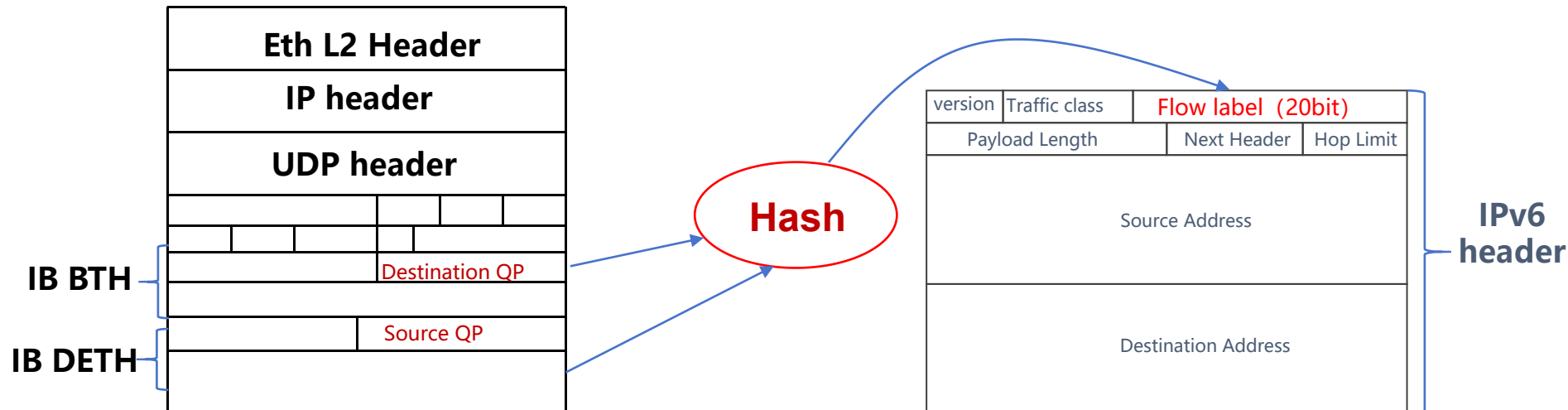
- QP Pair: **Src_QP (24-bit) + Dst_QP (24-bit)**
- Entropy from IPv6 Addresses:
Lower 16 bits of IPv6 Src Addr + Lower 16 bits of IPv6 Dst Addr (for collision avoidance).

■ Hash Algorithm: CRC32

- Fast, hardware-friendly (widely supported in ASICs), low collision rate.
- Initial register: **0xFFFFFFFF**

■ Flow Label Population:

- Take the lower 20 bits of the 32-bit CRC hash result.
- Write this into the IPv6 Flow Label field of the packet.



Implementation Benefits & Security Considerations

■ Implementation Point:

- Performed by the **ingress network device**.
- Transparent to end-host operating systems and applications.

■ Benefits:

- **Fine-Grained Flow Load Balancing:** Achieves true "5-tuple + Flow Label" level ECMP for RoCEv2 traffic.
- **Reduced Congestion:** Breaks up "elephant flows," distributing them across multiple paths.
- **Improved Throughput:** Maximizes utilization of multi-path network topologies.

■ Security:

- Modifies **only the IPv6 Flow Label field**.
- Does not alter IP addresses or port numbers, avoiding issues with firewalls or NAT.
- Standard security considerations from RFC 6437 apply.

Next steps

- Ask for more reviews and comments

Thank you!