Network Notification Problem Statement & Discussion

Weiqiang Cheng - China Mobile

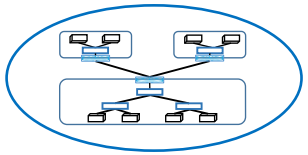
Network for AI: Overview

Network for AI consists of three parts: Scale-Up network, Scale-Out network, and DCI network, which together form a network with over 100,000 cards.

1 DCI Network

N*400Gbps Between AI data center,
Near-zero loss

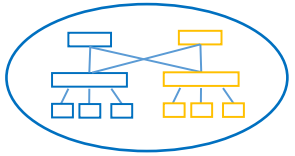
Mesh



2 Scale-Out Network

Over 10K GPU cards, ultra-high bandwidth,
ultra-low latency and jitter,
lossless network

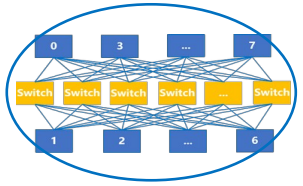
Fat Tree

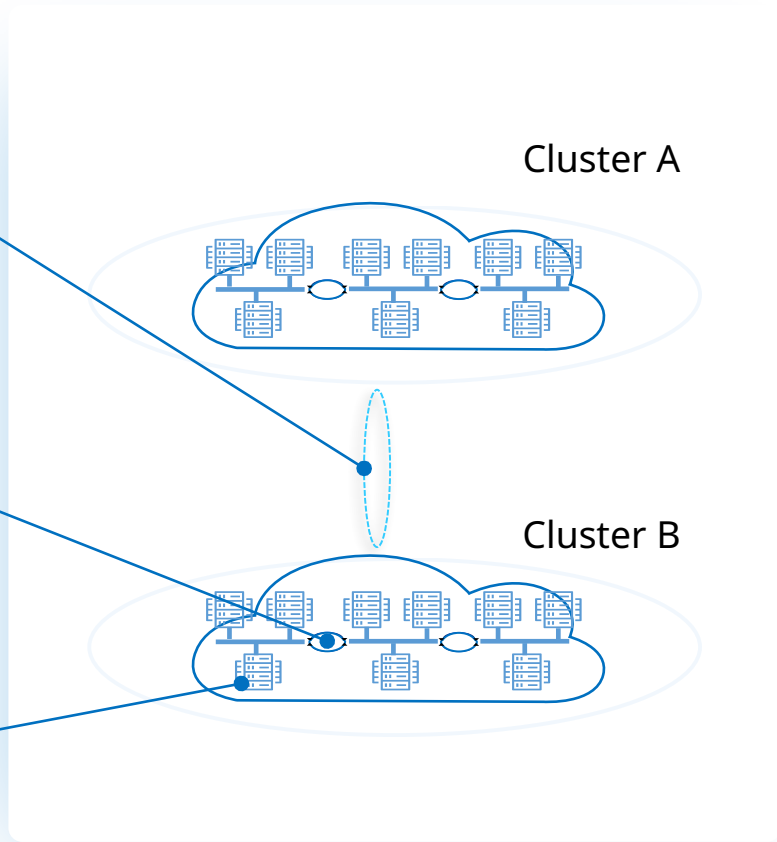


3 Scale-Up Network

Expands from one node with 8 cards to a super node with 128 cards,
bandwidth from 100GBps to 800GBps

Full mesh

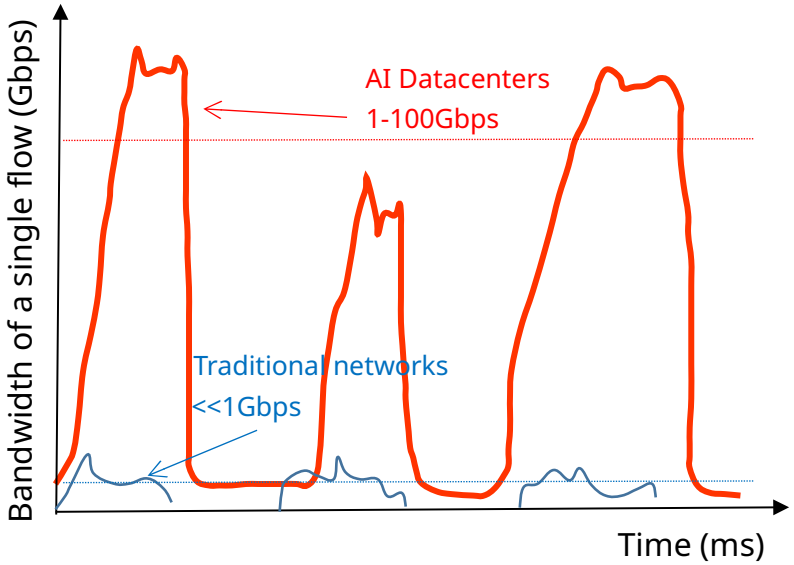


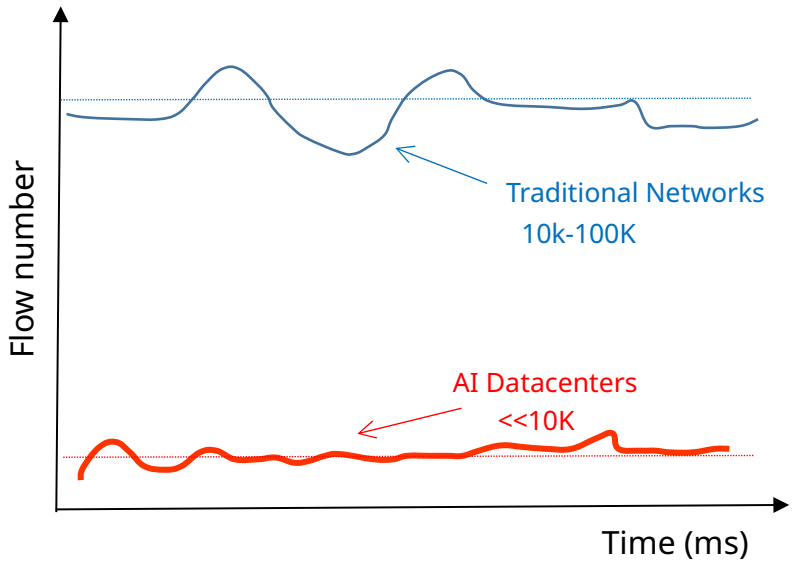


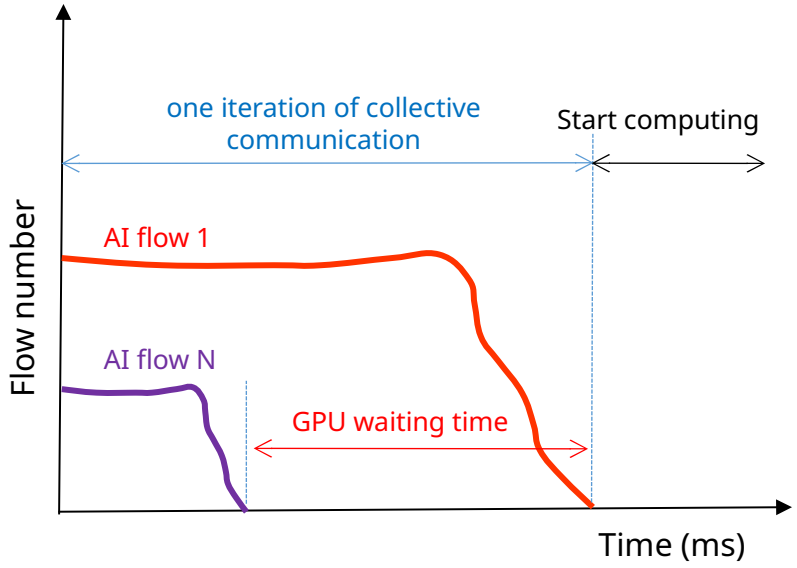
Traffic Model of AI Networks

The traffic pattern of AI datacenter networks is significantly different from that of traditional networks.

- High Performance: High-BW, Low Latency, Busty, Lossless







Per-Flow Characteristics

- *AI Computing:*
  High bandwidth (1-100Gbps), periodic bursts, long-lived flows
- *Traditional Networks:*
  Low bandwidth (<<1Gbps), random micro-bursts, short-flow dominated

Flow Number

- *AI Computing:*
  Relatively stable (typically <10K)
- *Traditional Networks:*
  Larger variable range (typically 10K-100K)

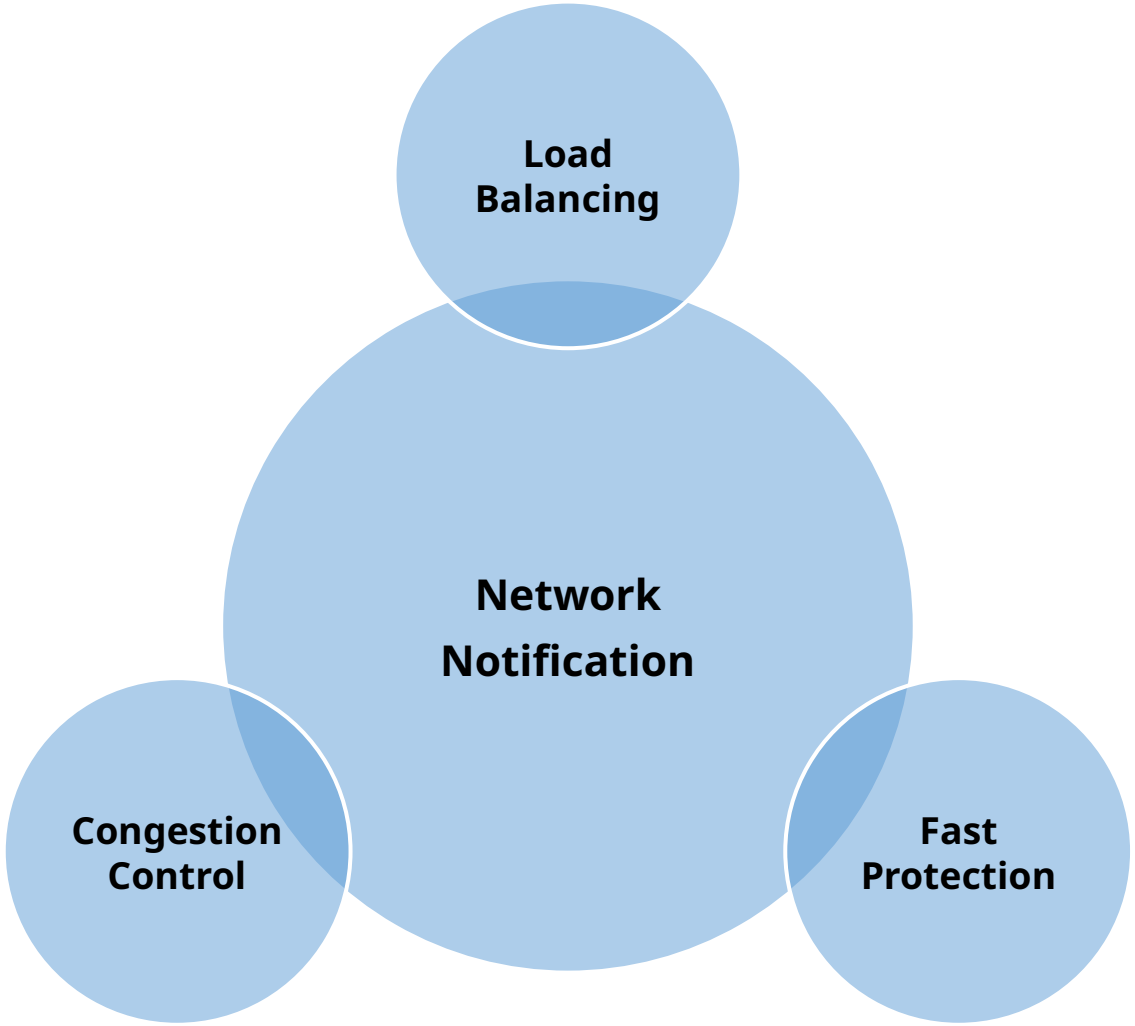
Communication Behavior

- *AI Computing:*
  Collective communication with inter-flow synchronization
- *Traditional Networks:*
  Point-to-point communication without synchronization

Application Areas for Network Notification

Network notification is supposed to be used in the following areas:

- **Load Balancing:** Enables rapid dissemination of network states and link information to assist in balancing traffic across multiple paths
- **Fast Protection:** Facilitates fast awareness of link or node failures, supporting quicker protection switching and reduced performance loss.
- **Congestion Control:** Helps inform upstream nodes of downstream congestion or performance degradation, enabling timely traffic shaping or rate adjustment.



Motivation for Network Notification

- **Network Demands**
  - Load balancing/Congest control/Fast protection/Capacity awareness call for **more frequent, flexible, and detailed notification mechanisms** to convey network information—supporting faster reroute, congestion response, and load balancing.

- **Need for Unified Notification**
  - Depending on different requirements, notification may differ in format, scope, and function. This leads to **redundancy, integration complexity, and limited interoperability.**

Problem Statement #1 – Existing Notification Mechanisms Inadequate for New Requirements

Gaps in existing notification mechanisms:

- Traditional IETF protocols (IGP/BGP) only focus on routing-related information. Even with some network performance-related extensions, it is difficult to capture and notify real-time network status.
- Using Telemetry (SNMP/IOAM) as a measurement method is practicable, but it still requires a notification mechanism to deliver the measurement results to the action-executing node.
- No event-driven notification, millisecond-level reroute is not enough for latency/loss-sensitive Traffic.

Problem Statement #2 – Proprietary Notifications Block Cross-Vendor Interop

- Notification mechanisms for Load Balancing, Fast Protection and Congestion Control today are mostly vendor-specific, lacking open standards, ***Take notification for load balance as example:***

NLB/GLB, ARN, CONGA, Centralized TE + enhanced ECMP , from different vendors,

use different notification methods

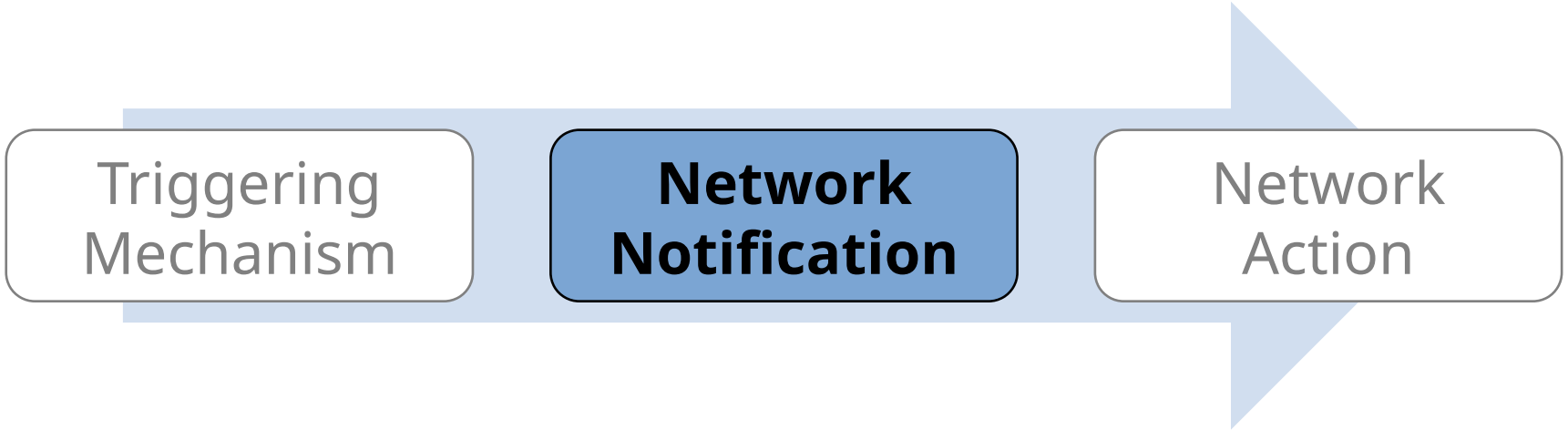
Vendor lock-in complicates multi-vendor AI Networks.

Problem Statement #3 – No Unified Procedure for Notifications

- **No Unified Information Model:** Lack of a standard information model defining what data should be carried for each network event.
- **Inconsistent Encoding and Encapsulation Methods:** Some reuse existing protocols (IGP, BGP); others use proprietary formats
- **Undefined Propagation Scope and Policies:** Some notifications should stay local, while others may require wider advertisement.
- **Absent Control & Management Framework:** The deployment of notification mechanisms necessitates control signaling and configuration support.

Summary

The complete notification process is shown below:



- Our scope focuses on **Network Notification.**
- We suggest **standardizing Network Notification** while reusing established mechanisms for other parts, to keep networks open.

Thanks!