Gap Analysis, Problem Statement, and Requirements in AI Networks

draft-hcl-rtgwg-ai-network-problem-00

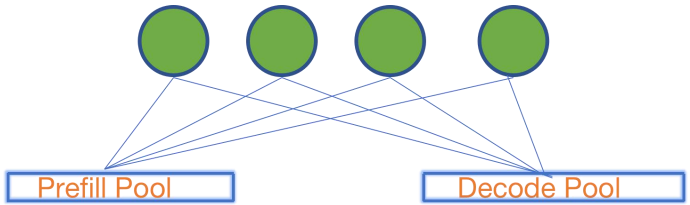
PengFei Huo (ByteDance) (Presenter)
Gang Chen (ByteDance)
Changwang Lin (New H3C Technologies)
Syed Hasan Raza Naqvi (Broadcom)

New context and challenges for AI

- Why make a big model?
- Decoder model represented by GPT has become the mainstream
- Different walls
- How does the network handle 3D parallel traffic

TP	AllReduce	can't be overlaped
DP	AllReduce	some can be overlaped
PP	Send&Recv	some can be overlaped

- MOE architecture model becomes mainstream
- Large EP domain strategies are gradually becoming mainstream - eg. Deepseek
- Inference separation deployment - PD separation



Requirements

Requirements of AI network:

- **Ultra-high bandwidth demand**

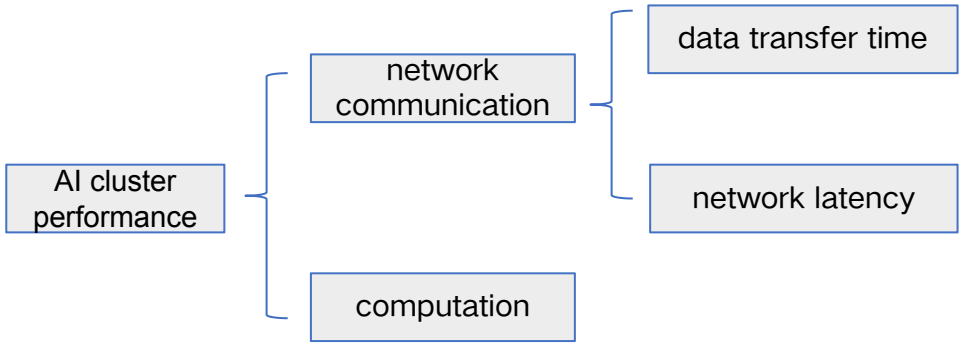
In AI training scenarios with large models, there will be a massive amount of communication data, which imposes higher bandwidth requirements on the network.

- **Stability demand**

Due to the long training time of large models, any failure during the training process can result in prolonged downtime, significantly affecting the efficiency of AI training.

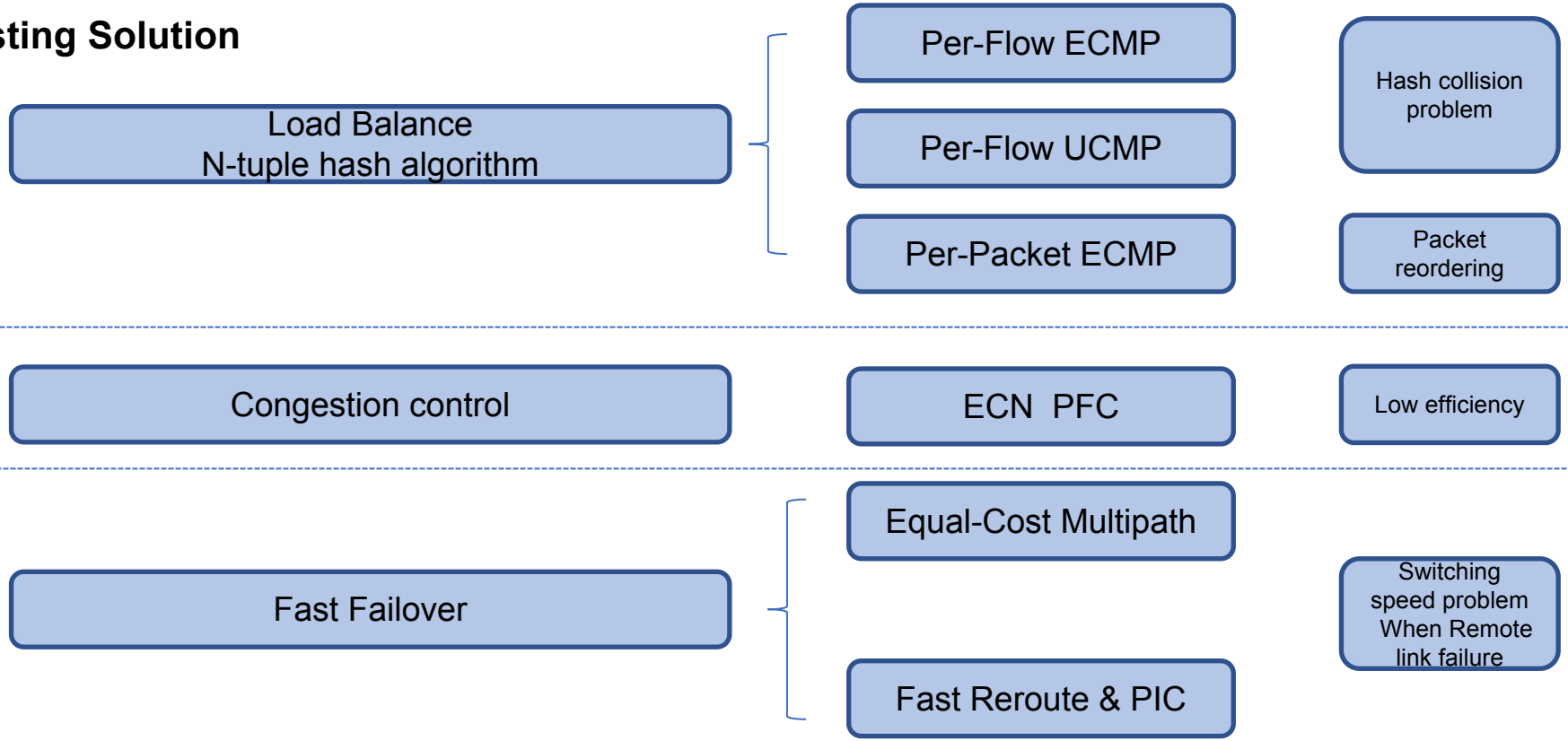
- **Low latency demand**

Higher network latency indicates a lower proportion of time spent on GPU computing. Thus, minimizing latency is crucial in AI training networks, often caused by network congestion.



Problem Analysis

Existing Solution



Requirements for AI network Mechanisms

- **New Load Balancing Mechanisms**

Capable of performing load balancing based on data packets to avoid the imbalance caused by the relatively small number of flows and bursty traffic in AI training networks

- **New Congestion Control Mechanisms**

To avoid the inflexibility of current congestion control mechanisms and achieve global, end-to-end congestion control.

- **Fast Failover Mechanisms**

The need for new fast failover mechanisms that can quickly detect faults, rapidly notify remote endpoints, and enable rapid global fault handling mechanisms.