A OSF Framework for Artificial Intelligence (AI) Network

draft-hcl-rtgwg-osf-framework-00

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

IETF-120

Background

AI training network is a critical component in the field of artificial intelligence. It is a computer network system specifically designed for training and optimizing AI models. With large-scale datasets and optimization algorithms, AI training networks continuously drive the learning and evolution of AI models to adapt to changing environments and demands. In the field of AI, training networks play a crucial role, providing strong support and foundations for technologies such as deep learning, machine learning, and neural networks. The development of AI training networks lays a solid foundation for the progress and application of AI technology, while also promoting its widespread use and development in various industries.

Features of AI network:

- Fewer flows
- Bursty traffic

Requirements of AI network:

- Ultra-high bandwidth demand
- Stability demand
- Low latency demand

https://datatracker.ietf.org/doc/html/draft-hcl-rtgwg-ai-network-problem-00

Problem Analysis

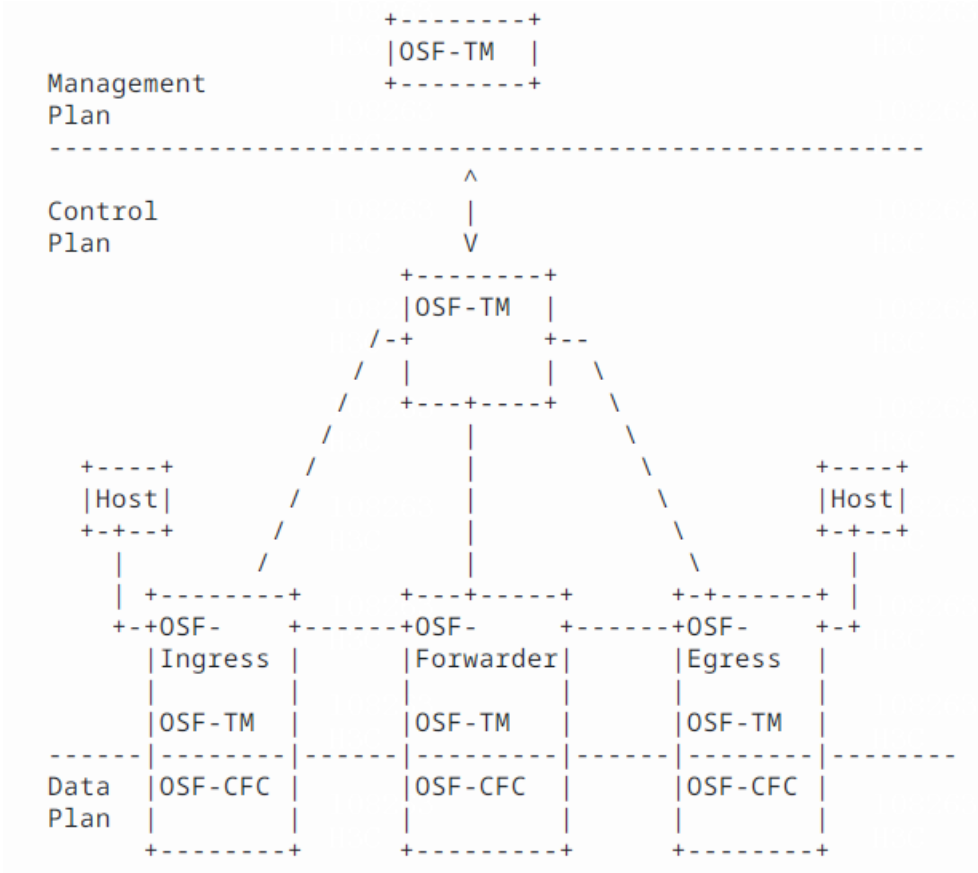
Existing Solution		Open Scheduled Fabric
Load Balance - Hash collision problem • N-tuple hash algorithm		**Data spraying**
Congestion control - Low efficiency • ECN (Explicit Congestion Notification) • PFC (Priority-based Flow Control) technologies		**Credit Request Control**
Network reliability - Remote link failure • Equal-Cost Multipath (ECMP) • Fast Reroute (FRR) • BGP PIC (Prefix Independent Convergence)		**topology management with link quality information**

Open Scheduled Fabric

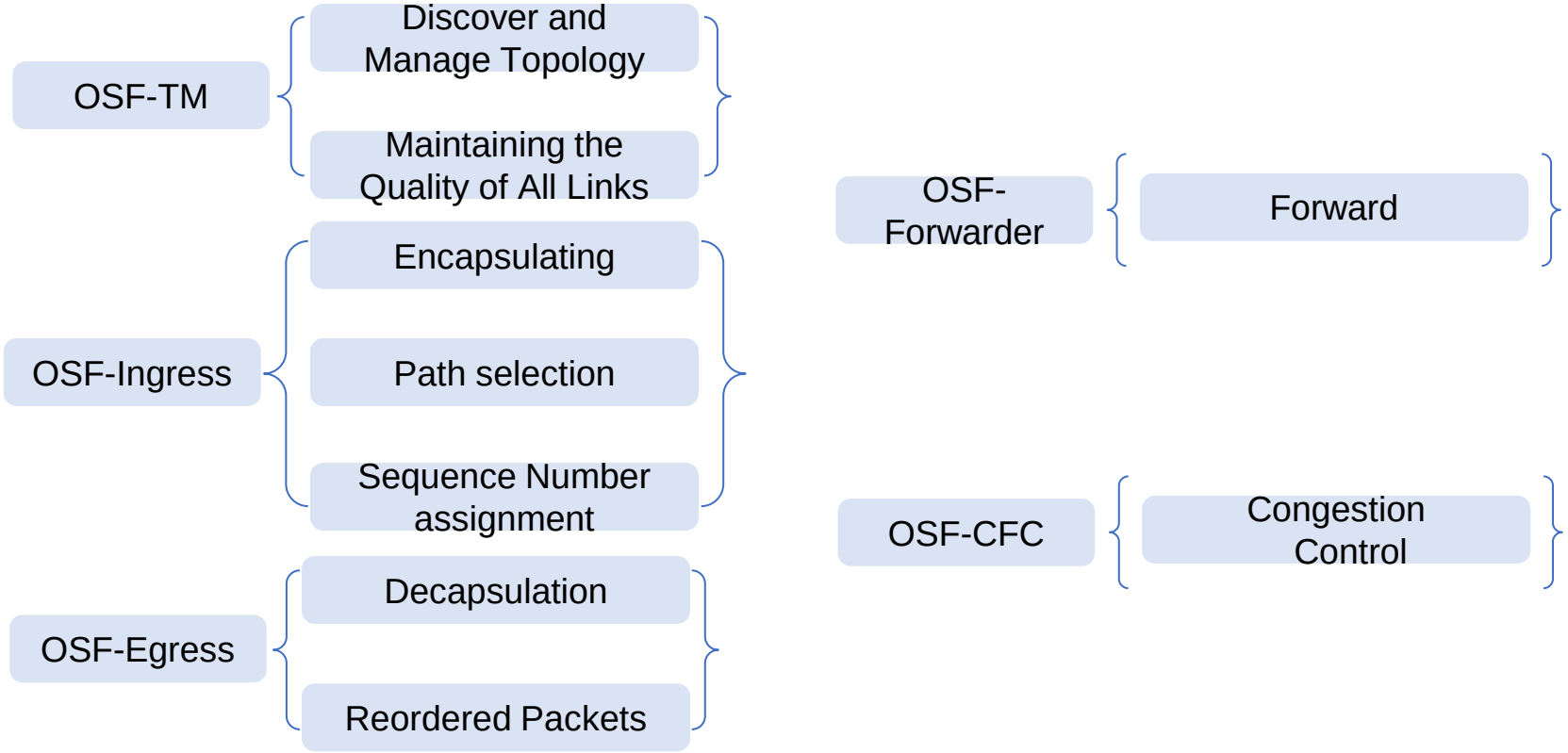
The OSF network architecture has redesigned network functional components to meet the new requirements of AI networks. Additionally, it has re-engineered the implementation mechanisms in three areas: topology management, load balancing, and network congestion control, providing solutions to various issues encountered in AI model training from a network perspective.

OSF Functional Components

- OSF-TM: OSF Topo Manager
- OSF-Ingress: OSF Ingress router
- OSF-Egress: OSF Egress router
- OSF-Forwarder: OSF Forwarder Router
- OSF-CFC: OSF Credit-based Flow Control



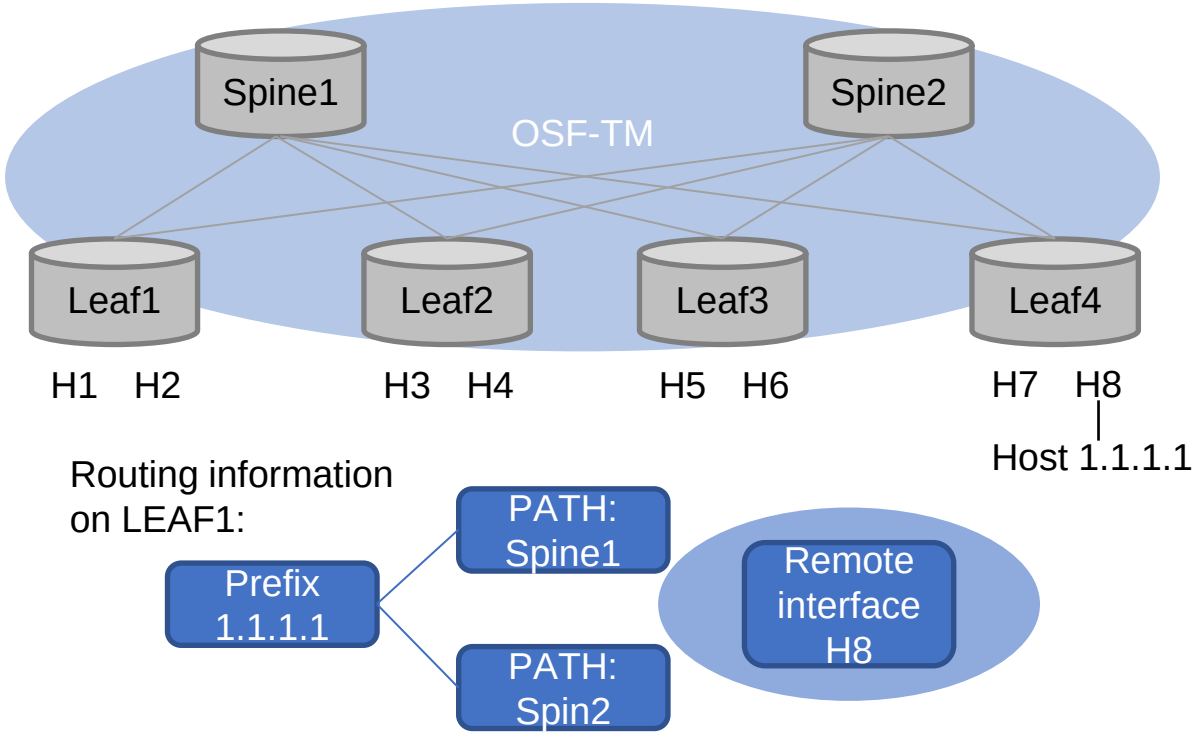
Open Scheduled Fabric



OSF Control Plane Workflow

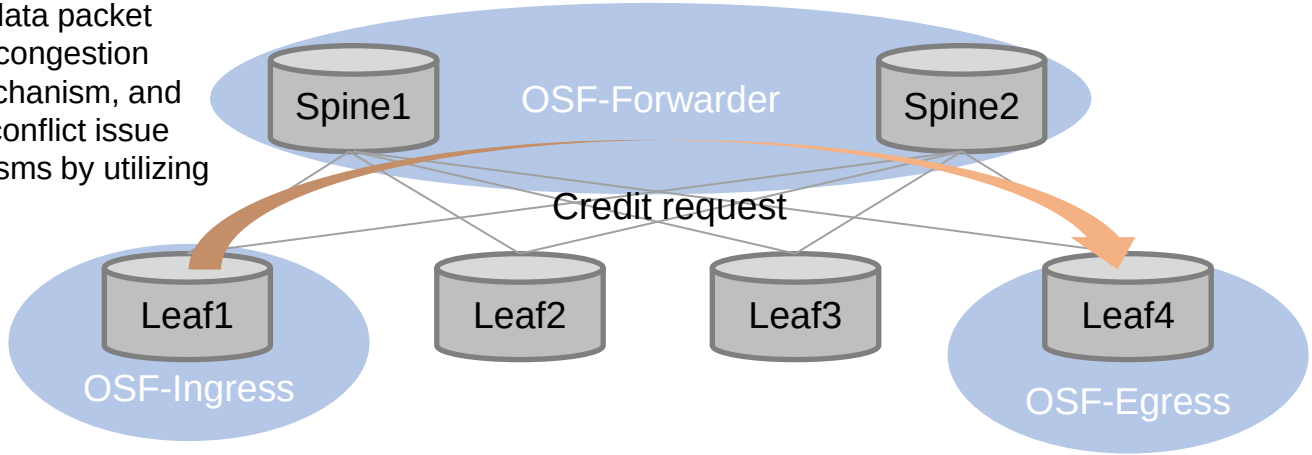
OSF-TM is responsible for collecting the network topology with link quality information and synchronizing it across the entire network. To provide more information for the data plane to perform data spraying, OSF-TM is also responsible for associating remote outbound interface information with route information.

- Step1:Collect link quality information and topology information, and synchronize them to all devices.
- Step2:Collect the outbound interface information of local routes and synchronize it with other devices.
- Step3:Associate remote interface information with routes.



OSF Data Plane Workflow

In the OSF network architecture, data packet forwarding first ensures end-to-end congestion control through an authorization mechanism, and secondly, resolves the polarization conflict issue inherent in traditional hash mechanisms by utilizing data traffic spraying.



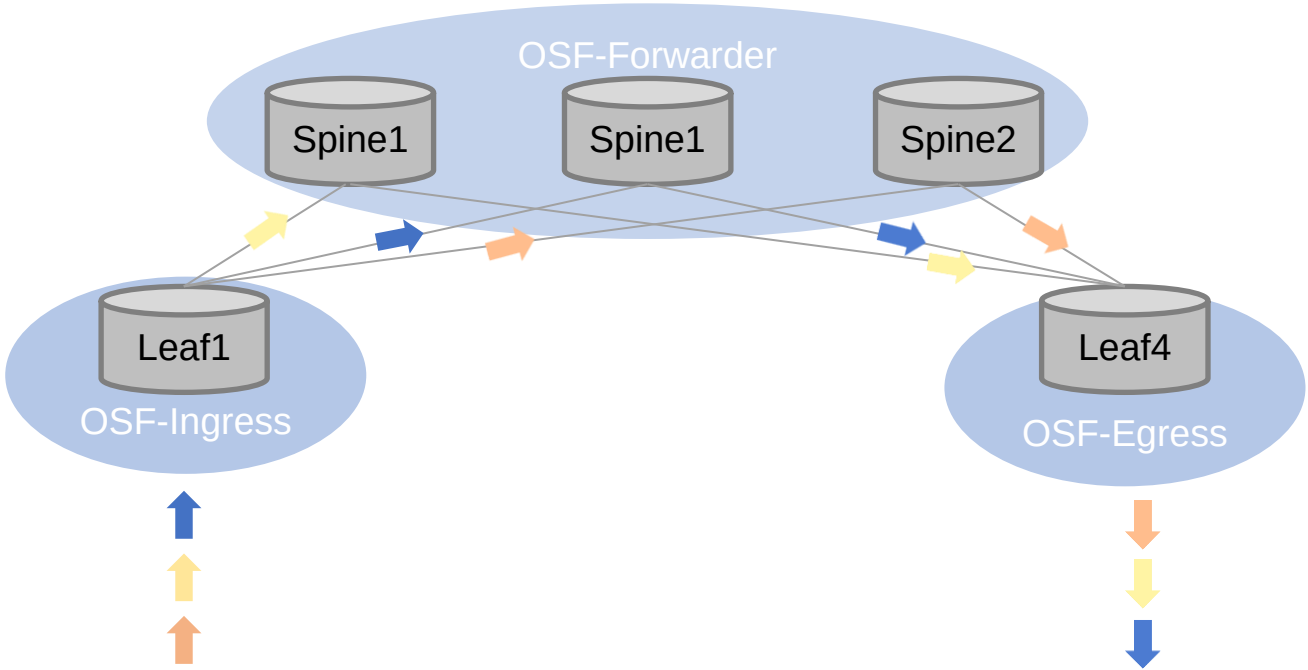
When the OSF-Ingress is preparing to send a data packet...

- Step1:The OSF-Ingress makes an Credit request to the OSF-Egress.
- Step2:The OSF-Egress reply Credit ACK.
- Step3: OSF-Ingress assigns a number to each data packet based on the remote outbound interface information of the route.

OSF Data Plane Workflow

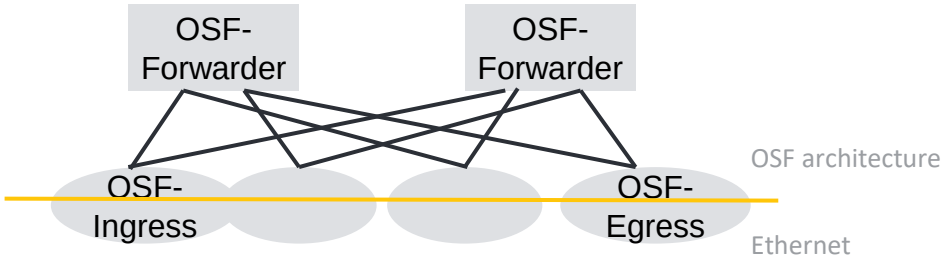
After OSF-Ingress obtains the send credit..

- Step1: OSF-Ingress assigns a number to each data packet based on the remote outbound interface information of the route.
- Step2: OSF-Ingress evenly sprays the data packets to OSF-Forwarder(ECMP path), which then forwards them to OSF-Egress.
- Step3: OSP-Egress reorder the packets based on the received sequence numbers.
- Step4: OSF-E sends the reordered data packets to the HOST.



USE CASE

A complete OSF network architecture can be constructed using the DDC system. In a SPINE-LEAF network topology, SPINE acts as the OSF-FORWARDER, while LEAF serves as the OSF-INGRESS/OSF-EGRESS.



Next Step

- Any questions or comments are Welcomed
- Continue improving the OSF network architecture