SRv6 uSID for Hyperscalers SDN use cases

Guohan Lu

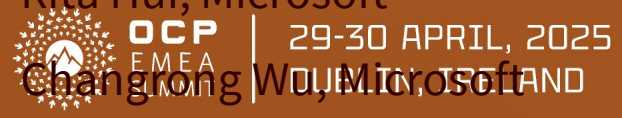
Partner Software Engineering Manager, Microsoft

With contributions from

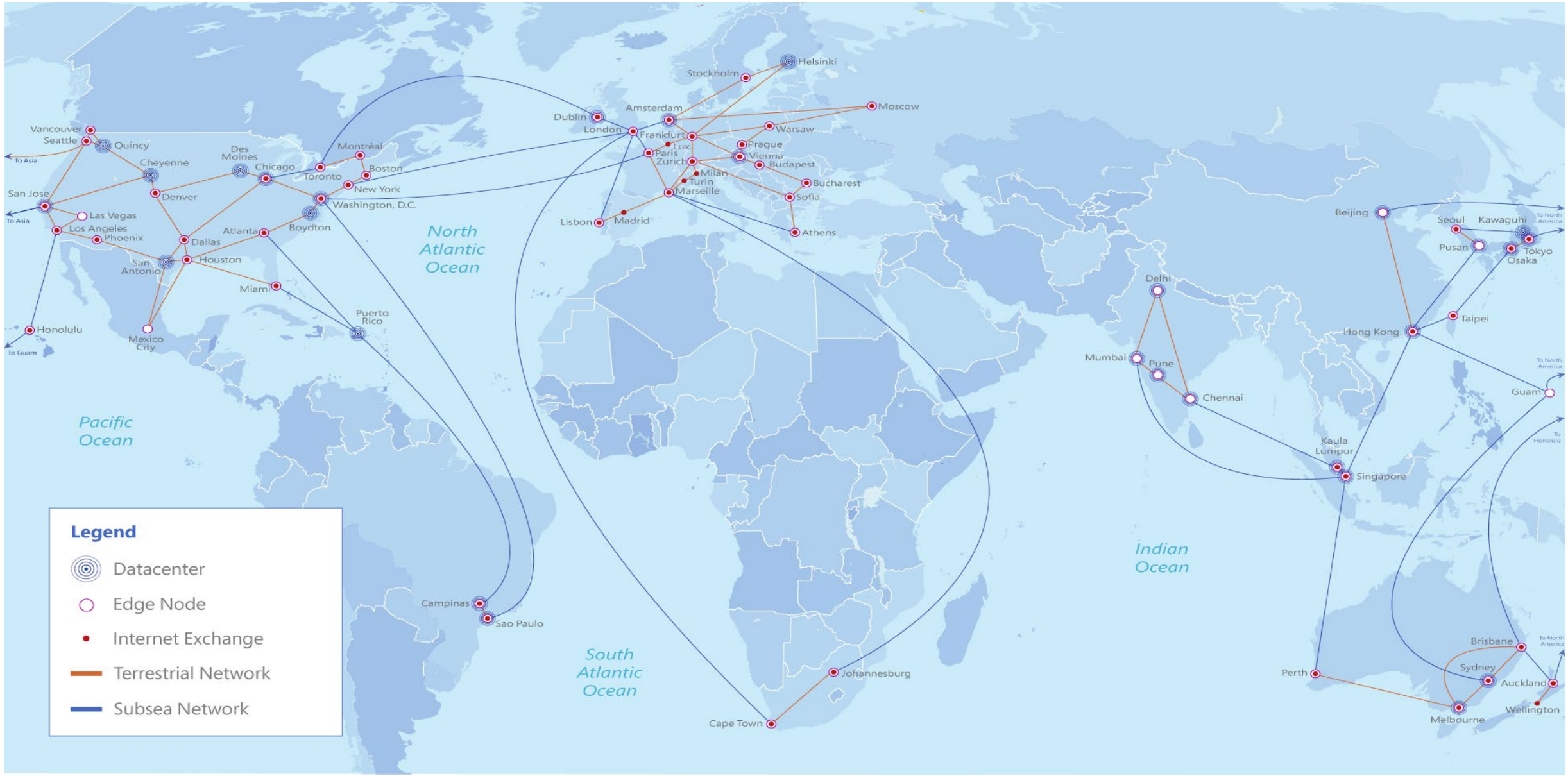
Ahmed Abdelsalam, Cisco

Rita Hui, Microsoft

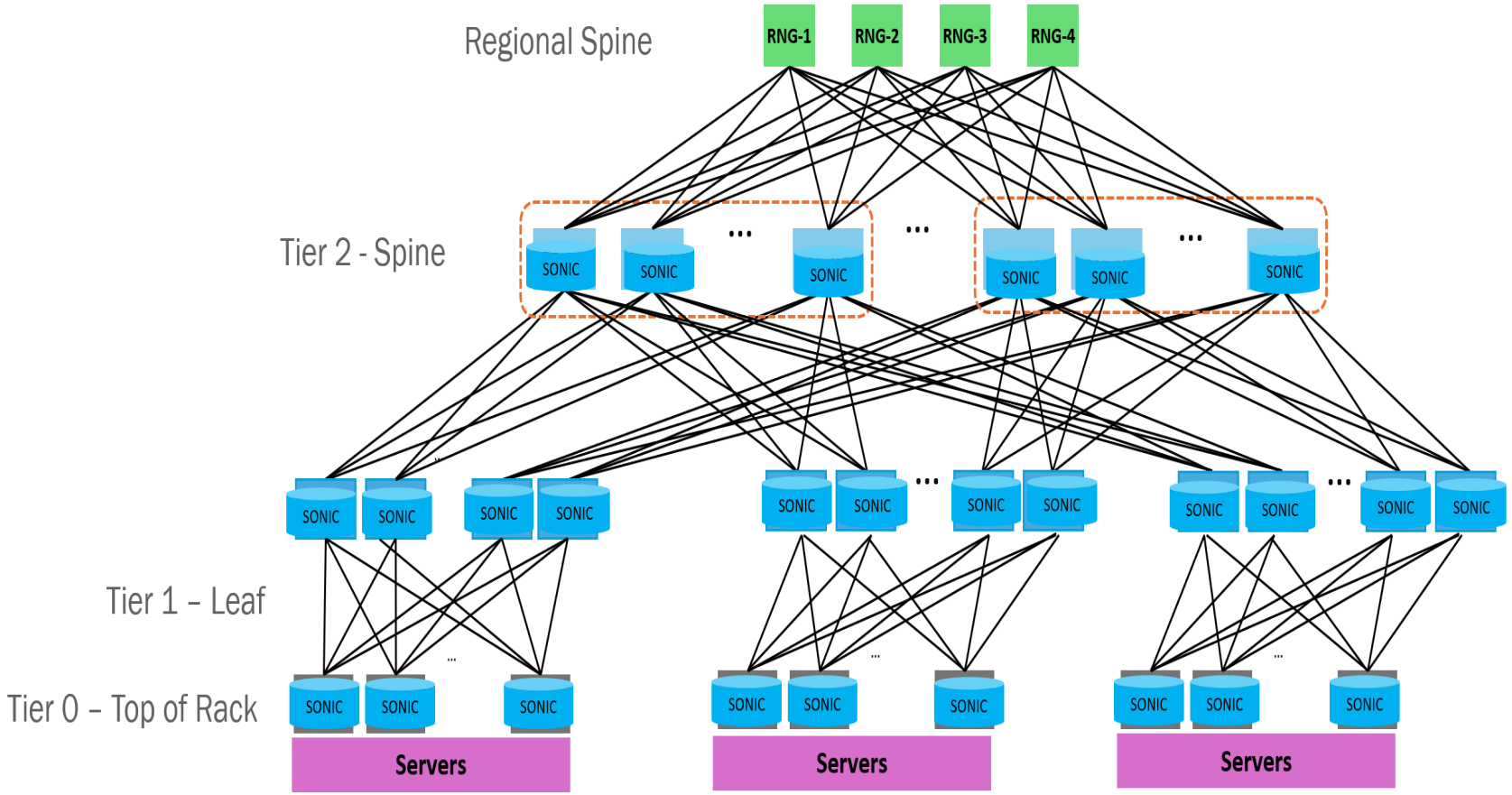
Changrong Wu, Microsoft



Microsoft Global Cloud Network



SONiC Is Powering Cloud At Scale



In Microsoft

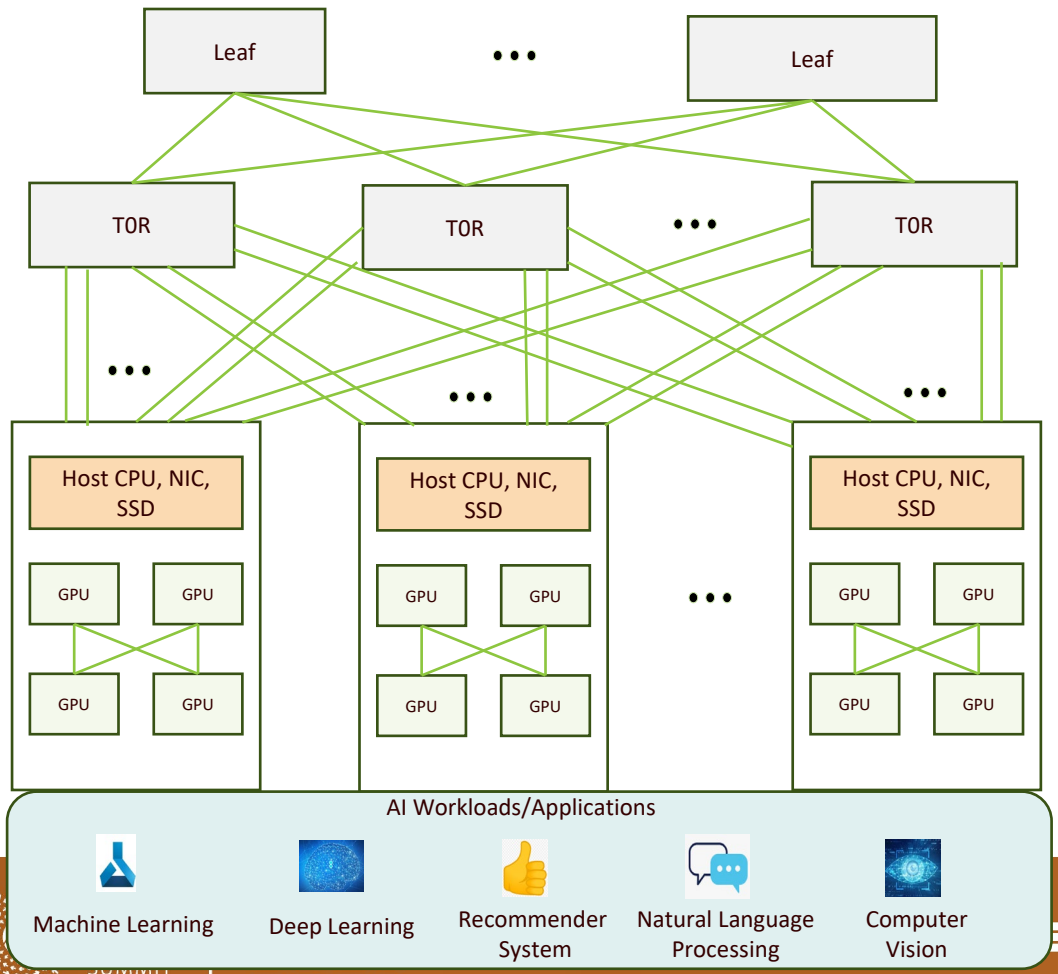
> 90% T0 switches running SON

> 80% T1 switches running SON

Started rolling out T2 with SONi in 2024

Artificial Intelligence in the Cloud

Raising the Bar for Hyperscale Datacenter Networks



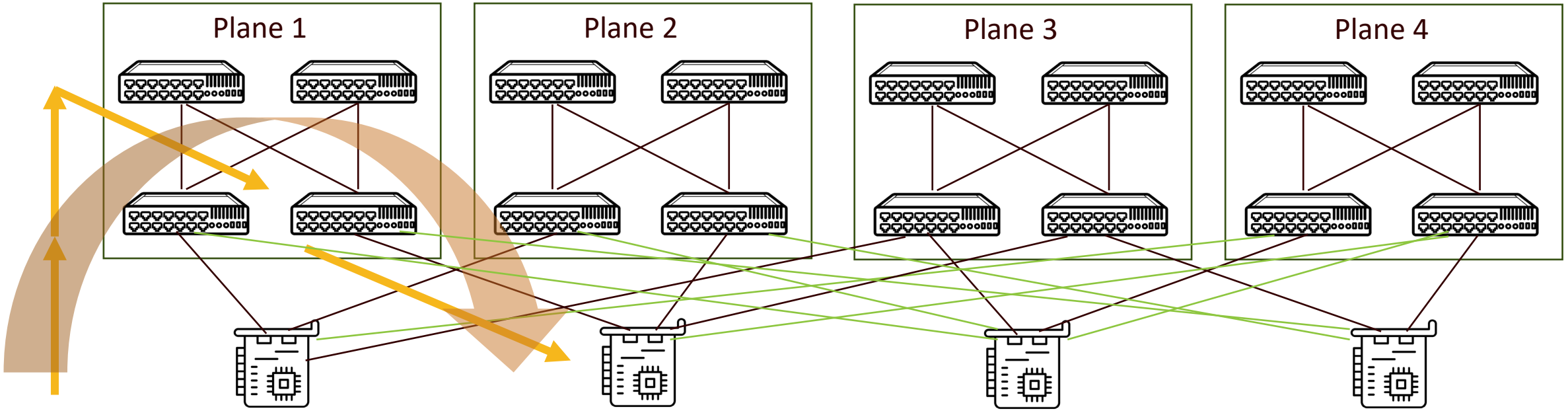
New Traffic Pattern:

- Small number of large flows
- Periodic bursts of data sent synchronously

Challenges:

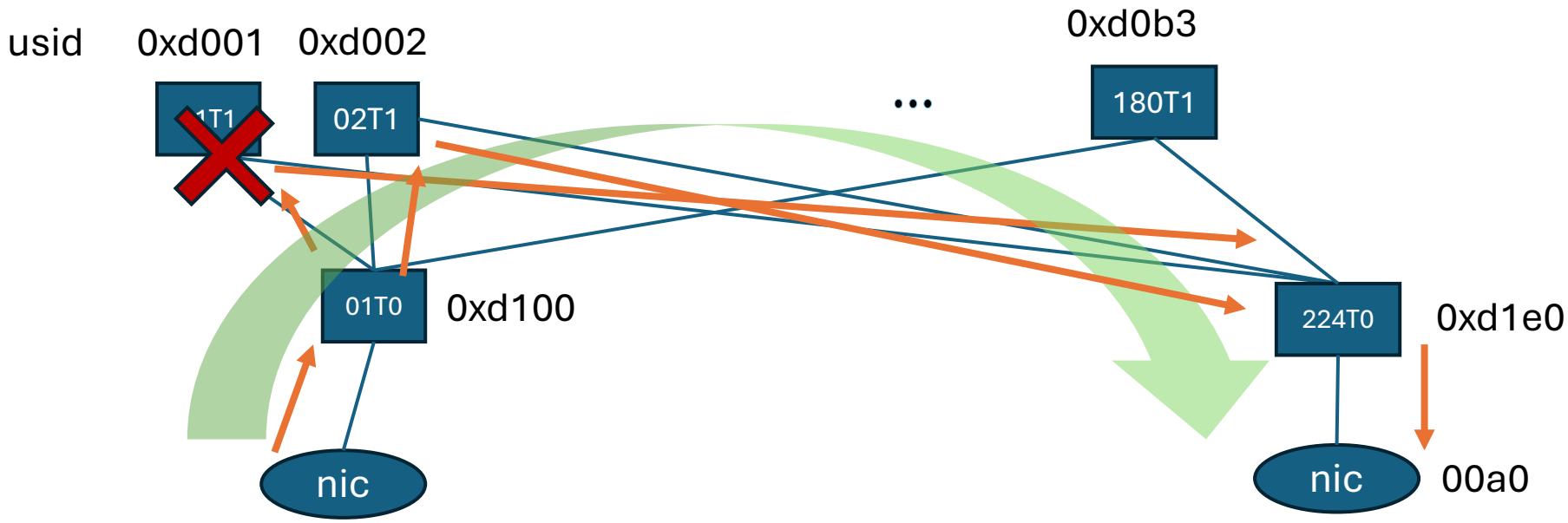
- Traditional passive ECMP-based load balancing mechanisms suffered from low entropy problem.
- Failures of communications in LLM training are costly
  - An epoch of training is blocked until the synchronized collective communications of last epoch finishes
  - GPU hours are expensive resources because of tight supply
  - If an ongoing training job crashes, all progress since the last checkpoint may be lost.

SRv6 for Backend Network



- Provides fined-grained network control based on source routing
- Enables path enumeration for traffic management
- Integration with AI workloads flow scheduling provides optimal network performance
- Allow source to quickly reroute upon path failures or congestion

SRv6 for 2-layer topology using uSID



SRv6 packet from NIC

DstIPv6 *fcbb:bbbb:d100:d001:d1e0:00a0::*

32-block

1st hop 2nd hop 3rd hop 4th Hop

When congestion or failure is detected on 01T1

NIC switches to new path by change the uSIDs encoded in the packet header

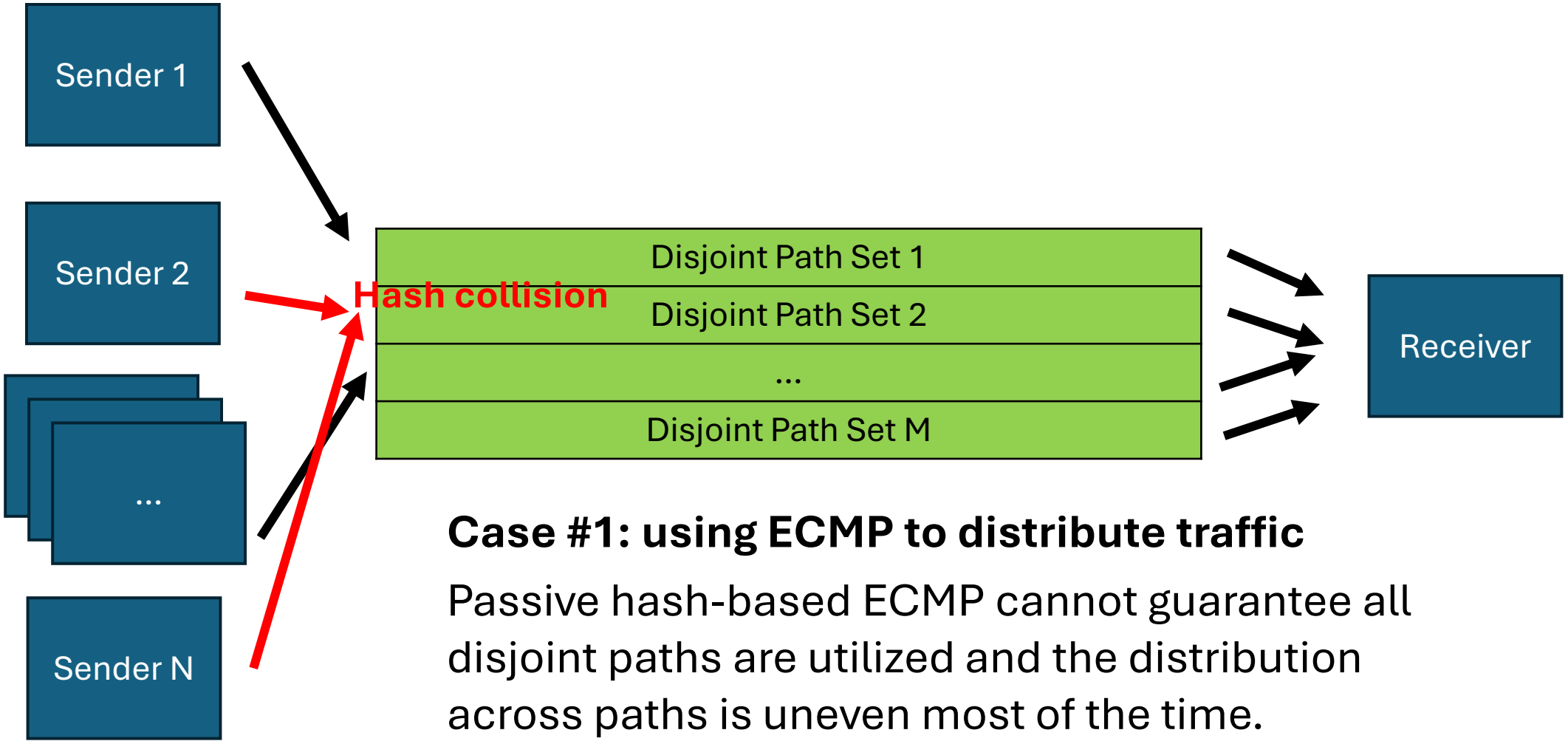
fcbb:bbbb:d100:d002:d1e0:00a0::

32-block

1st hop 2nd hop 3rd hop 4th Hop

Path control via SRv6 maximize utilization

(Using incast as an example)

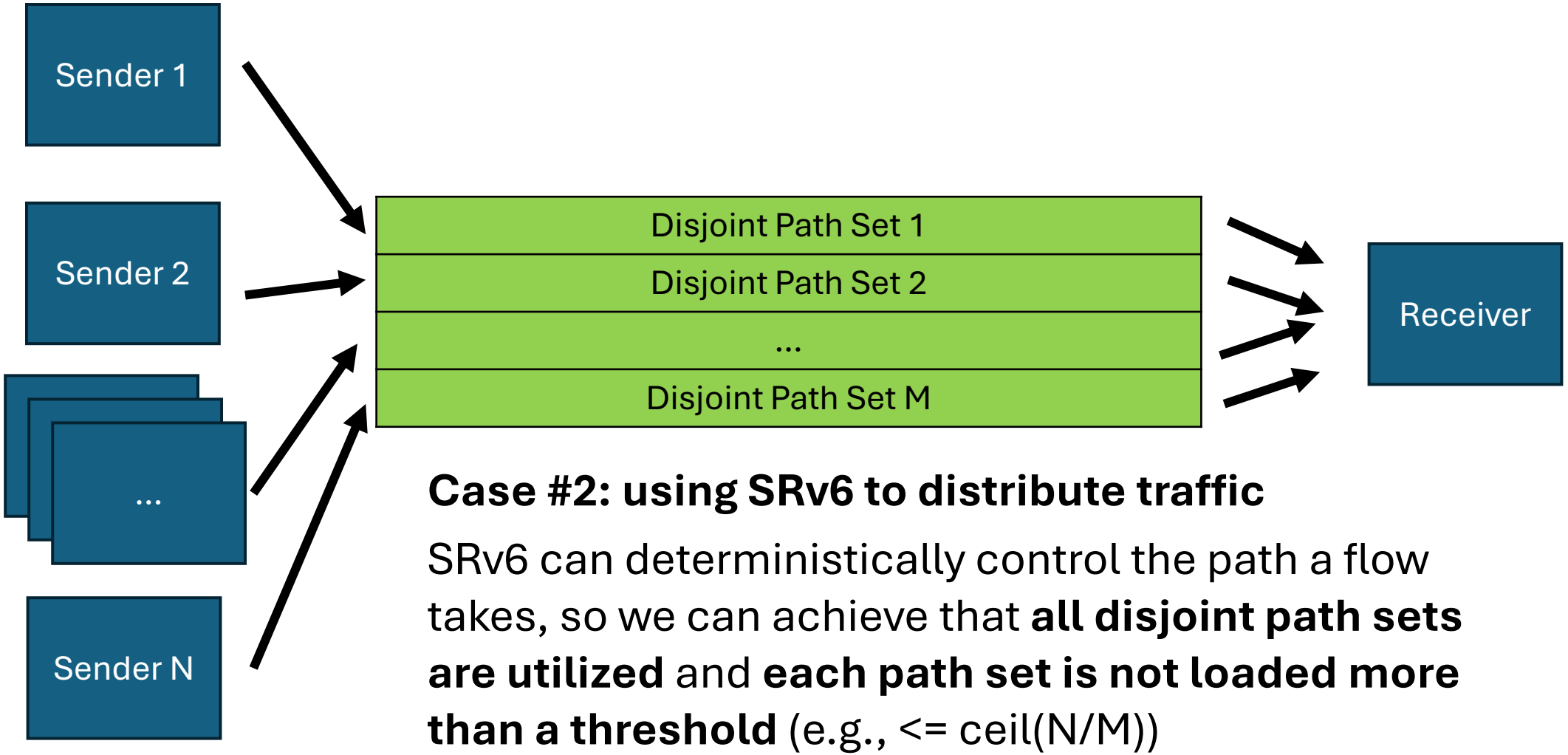


Case #1: using ECMP to distribute traffic

Passive hash-based ECMP cannot guarantee all disjoint paths are utilized and the distribution across paths is uneven most of the time.

Path control via SRv6 maximize utilization

(Using incast as an example)



Case #2: using SRv6 to distribute traffic

SRv6 can deterministically control the path a flow takes, so we can achieve that **all disjoint path sets are utilized** and **each path set is not loaded more than a threshold** (e.g., <= ceil(N/M))

SRv6 with static uSID in SONiC

SDN Controller

SRv6 Configs

IPv6 Static Routes

```
segment-routing
  srv6
    static-sids
      sid fcbb:bbbb:20::/48 locator loc1 behavior uN
```

SONiC

BGP docker

FRR

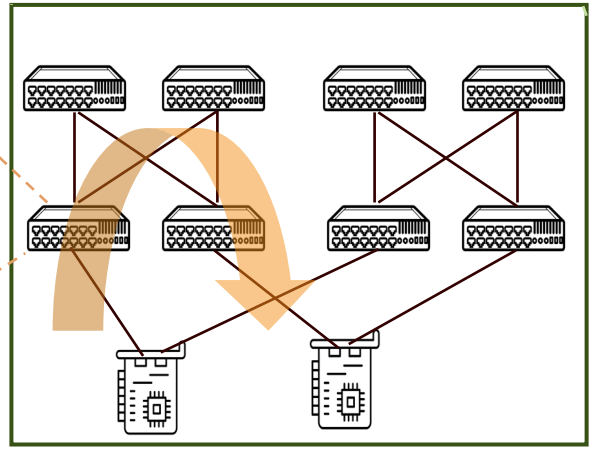
FRR config

bgpcfgd

config DB

SRV6_MY_SIDS
SRV6_MY_LOCATORS

HW Switch

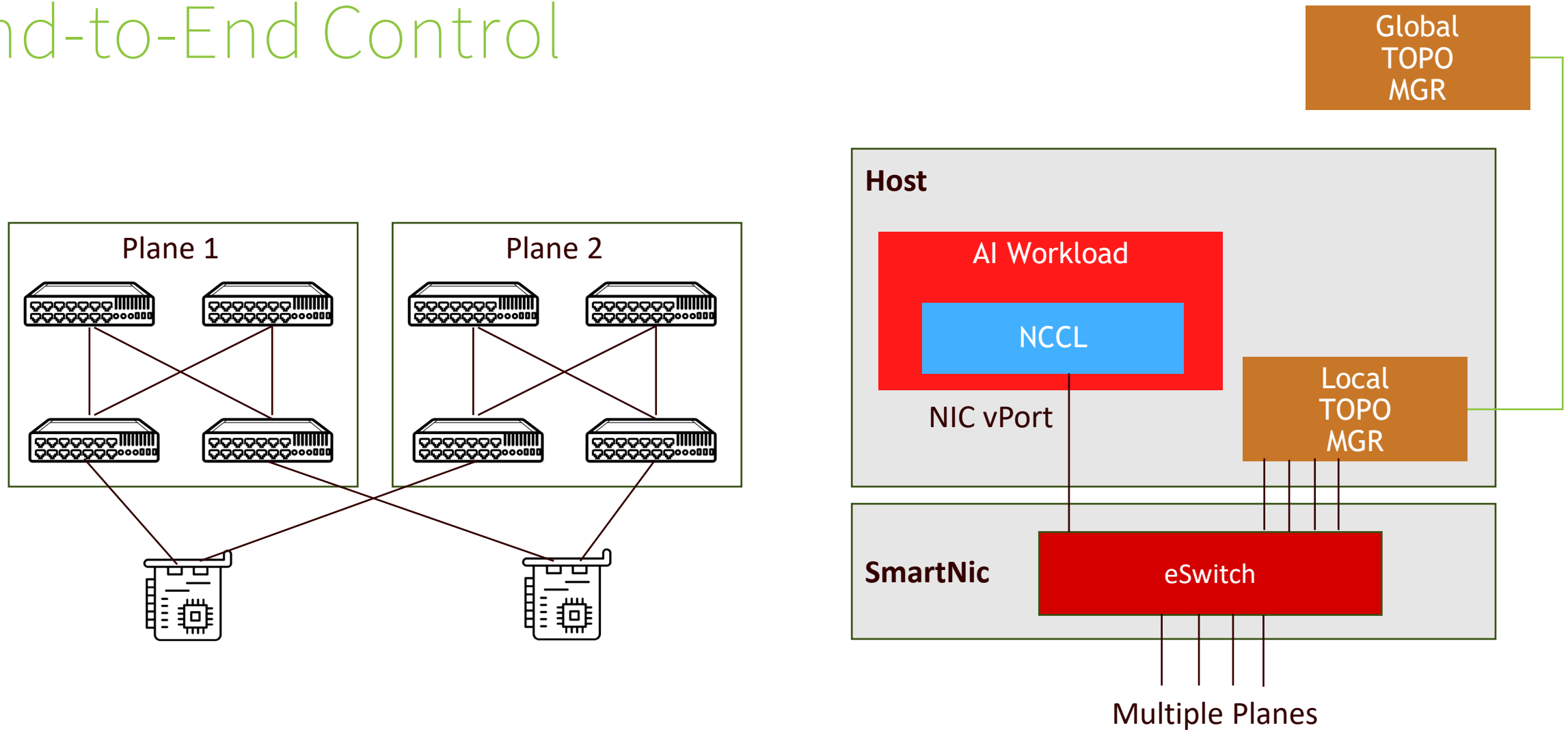


Config DB

SRV6_MY_LOCATORS Table
serves as a configuration source for SRv6 locators

SRV6_MY_SIDS Table
configures segment identifiers

End-to-End Control



The NIC/DPU has many switch functionalities in multi-plane networks

Inviting contributions to SONiC community in all areas

- SONiC/SAI
- Hardware platforms
- Testing and tooling
- Download, test, deploy!

- Project Wiki with latest specification : https://sonicfoundation.dev/
- Source Code: https://github.com/sonic-net/SONiC/blob/master/sourcecode.md
- Becoming a contributor: https://github.com/sonic-net/SONiC/wiki/Becoming-a-contributor
- Mailing list: https://lists.sonicfoundation.dev/g/sonic-dev
- SONiC Community meeting: https://sonic-net.github.io/SONiC/Calendar.html
- SONiC Routing Workgroup: sonic-wg-routing@lists.sonicfoundation.dev | Home

Thank You

OCP EMEA SUMMIT | 2025