Optimal Path Utilization in Multi Plane Fabric

Motivation – *Jai Kumar (Broadcom)*

Deployment Usescase – *ChenChen Oi (Bytedance)*

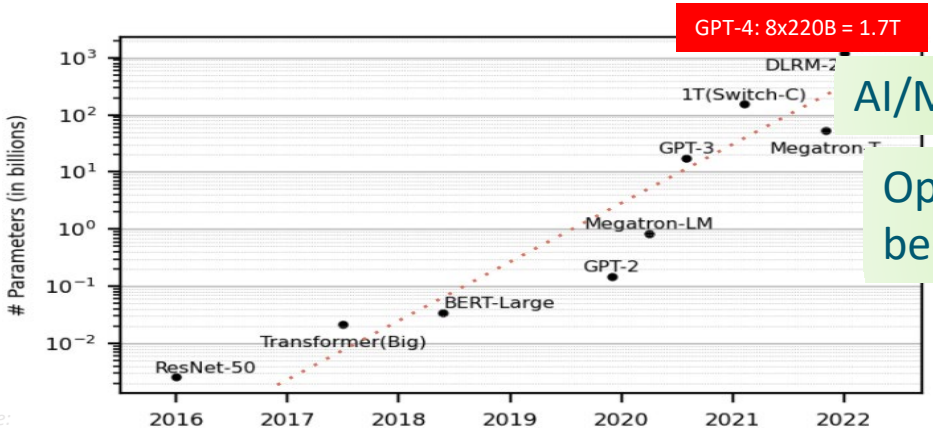
Topology – *Wenda Ni (Bytedance)*

SAI Enhancements – Jai Kumar (Broadcom)

OCP GLOBAL SUMMIT | OCT 15-17, 2024
SAN JOSE, CA

Emerging Trends

- Changing Deep Learning Workloads
- Newer Collective Frameworks like MCR-DL, MoE, ZERO and more, are emerging with emphasis on parallelism
- Instead of just AllReduce/Broadcast, more collectives are used
- Newer Demands on Network like data reduction, AR, low latency, resiliency



AI/ML Fabric need to Scale Out Significantly

Optimal path selection in AI/ML Fabric becomes critical

Source:
https://www.researchgate.net/figure/Deep-learning-model-size-trend-in-terms-of-model-parameters-billions-from-2016-to-2022_fig1_370825733

Optimal Path Selection

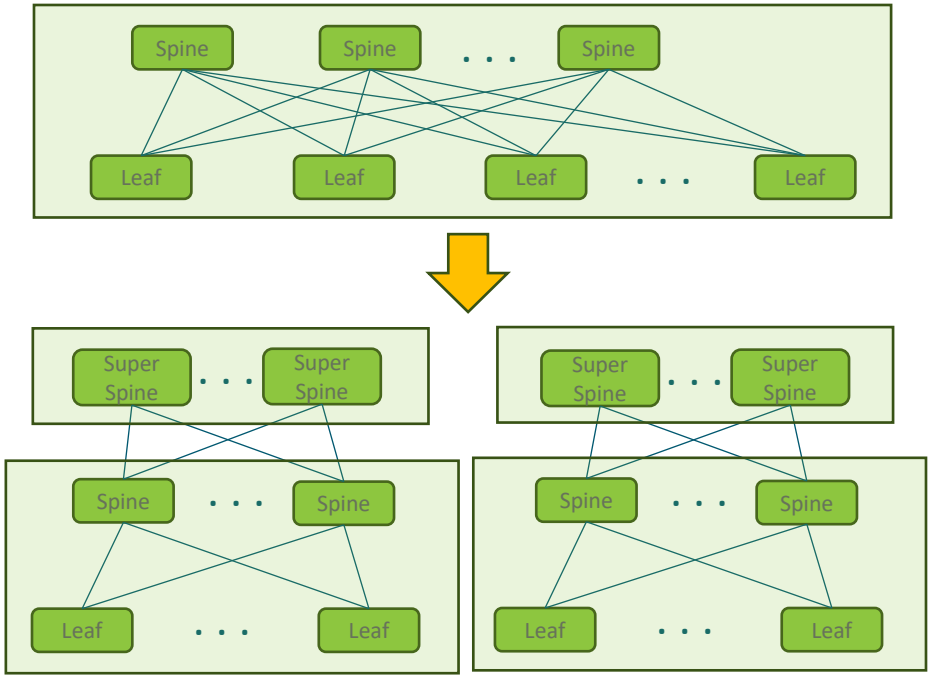
- Typical path selection is local to the fabric element's link state for e.g. ECMP, WECMP AR etc
- This doesn't take account for link quality downstream the packet path
- BGP does provide ability to learn and provide optimal and sub optimal path but is too slow
- Is there a better way to do path selection ???

Today we will talk about methods to pick an optimal local link in the AI/ML fabric that accounts for remote link quality as well.
We will qualify this with the measurements in actual deployment showing increase in app performance , resiliency and availability of the fabric.

Deployment Usecase And Topology

Large Scale AI Fabric

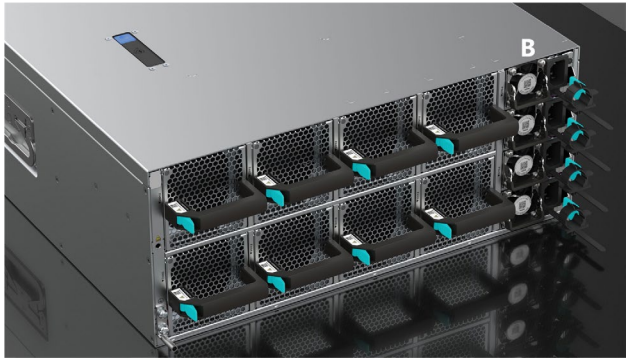
- Support max 128K GPUs
- Three tiers for large scale cluster
- RDMA over Converged Ethernet
- Loss-less transmission
- Better load balancing
- Rail-optimized network



Switch for High Performance Network

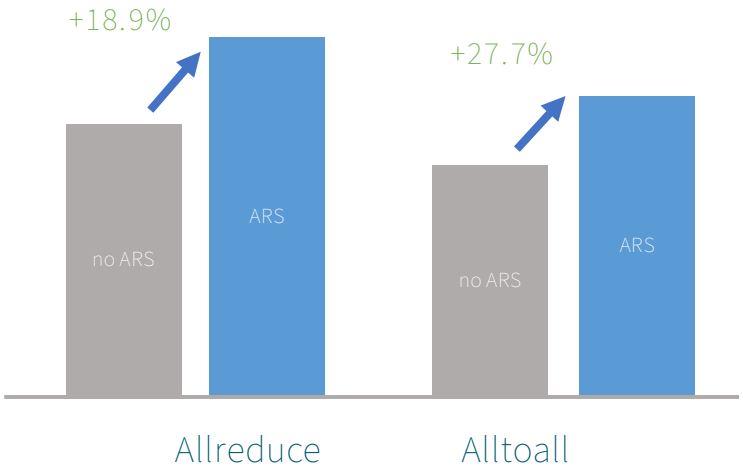
- Hardware platform
  - 51.2T switch asic
  - 64x800G OSFP
  - All ports support LPO (IL < 7db)
  - 1 MAC PCB, PHY-Less design

- Advanced features based on SONiC
  - Optimized adaptive routing with ARS
  - Inband telemetry under end-network fusion
  - High-precision traffic and congestion monitor
  - Warm upgrade tool covered all online bugs



Adaptive Routing and Switching (ARS) Deployment

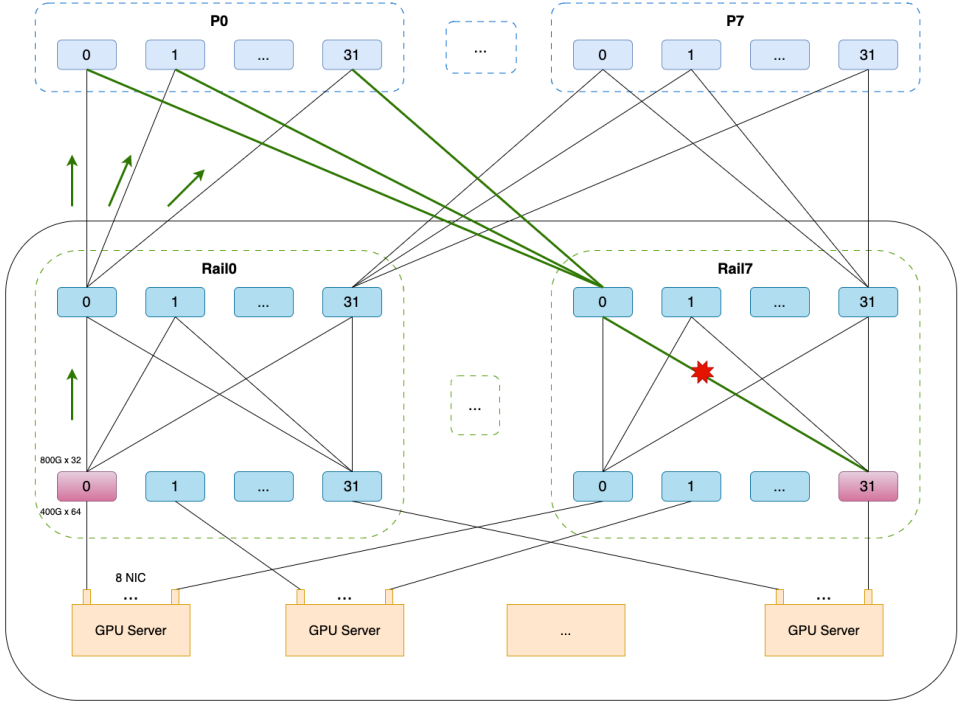
- Higher throughput
  - 3%~12% higher avg BW utilization
  - Lower queue congestion
- Faster link failover
  - Packet loss time < 0.5ms
- Multi-compatibility
  - Flowlet mode with non-AR NIC
  - Packet spray mode with AR NIC



Challenges

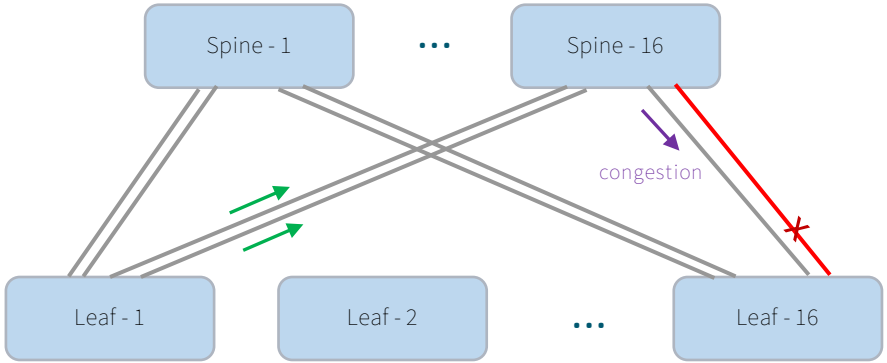
- Remote downlink congestion
  - Cannot select the optimal path under three tiers network
  - Greater impact under ARS flowlet mode

- Remote downlink failure
  - Long packet loss time caused by multi-hop routing convergence



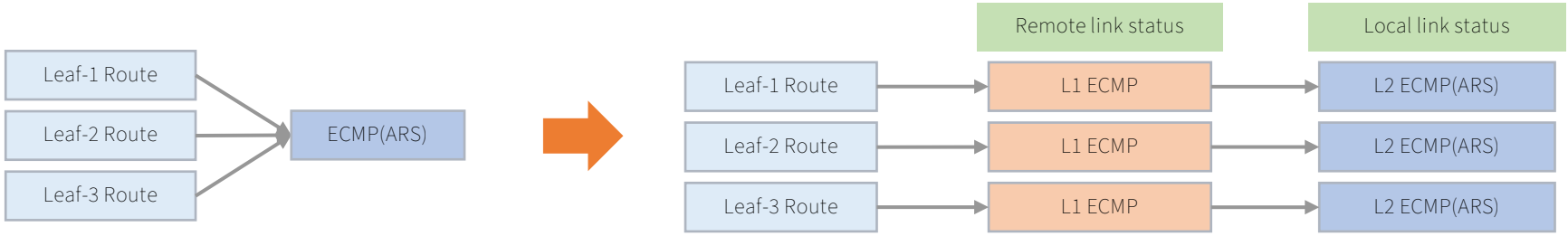
Challenges

- Load balancing under asymmetric topology
  - In a small radix pod with a multi-link topology, link failures may result in different numbers of uplink and downlink.
  - Such asymmetric topology may cause congestion due to backpressure and HOLB.

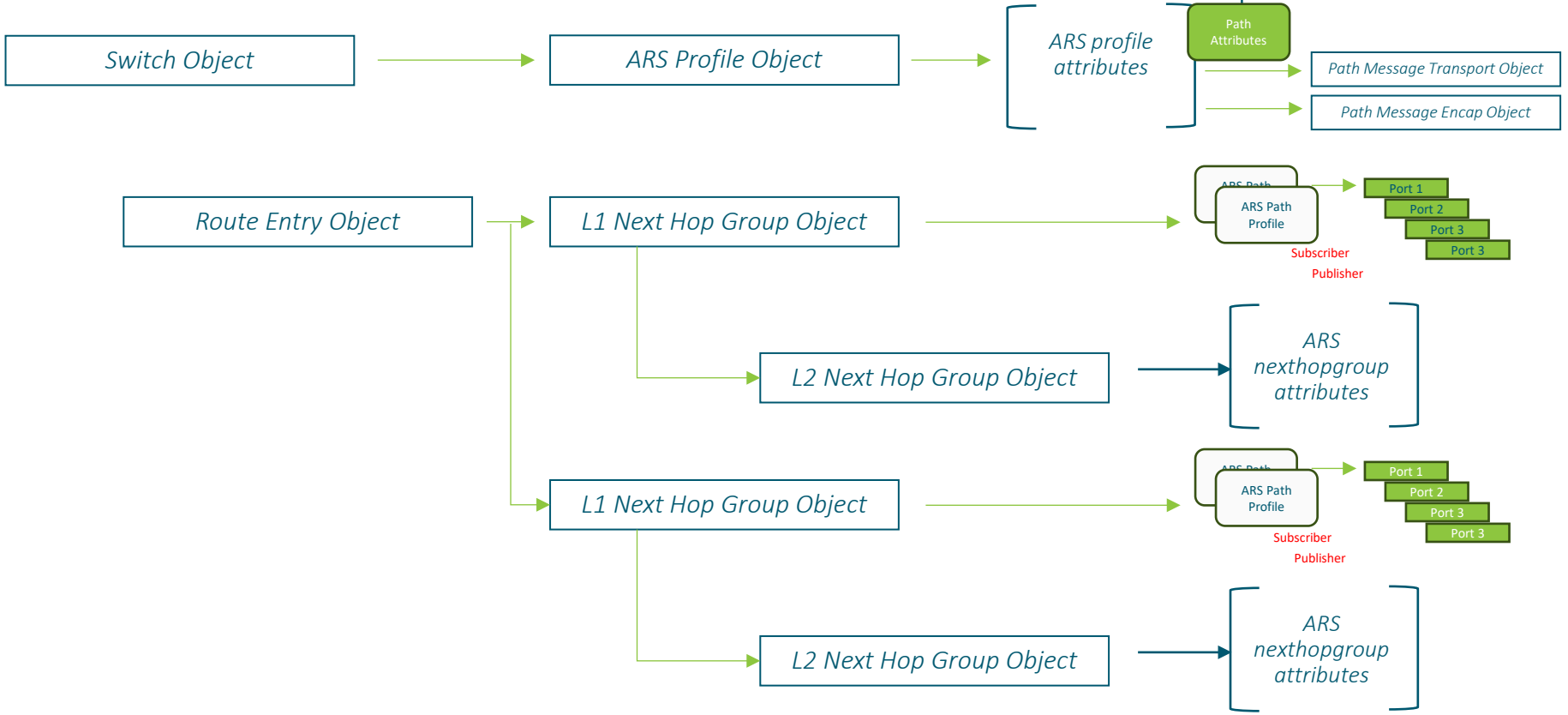


Load Balancing Based on Global Topology

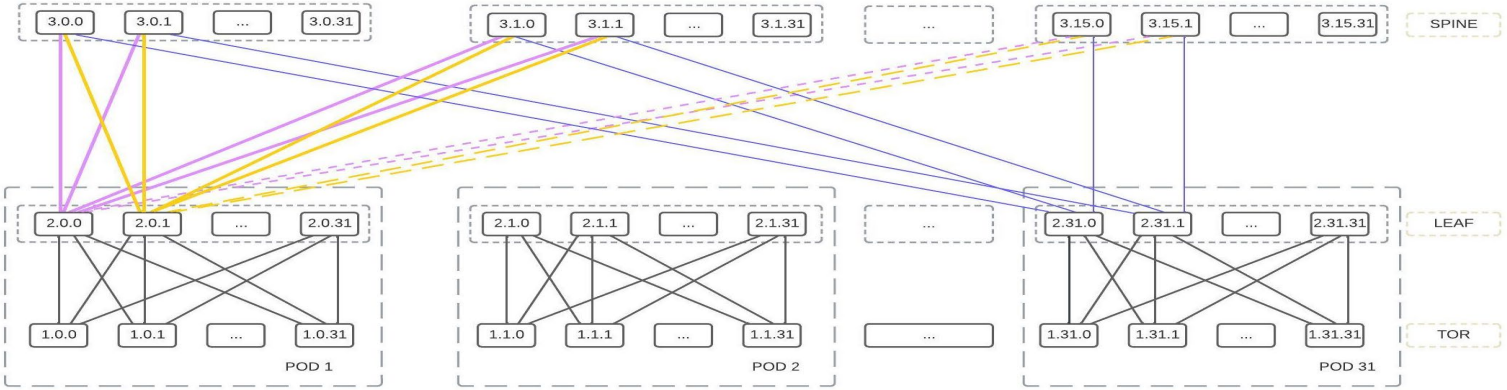
- Link state protocol based on global topology
  - Associate global link topology with BGP routing to form point-to-point ECMP Group.
- Full path status update
  - The remote link status is updated through the micorsecond level notification message.
  - The optimal path selection of AR ECMP is based on the quality of the local path and the remote path.



SAI Enhancements: ARS Path Profile Object



Path Characterization – 3 Tier Network



Node 1.0.0: Remote Quality of 1.x and 3.x nodes (One step lookahead)

L1 ECMP groups are create with remote nexthop type for 1.x.x

L1G1 till L1G31 ->[1.0.1] till [1.0.31] - - - - - Total 31 groups

L2 ARS ECMP groups

L1G0 -> L2G0.0 -> [2.0.0, .. . , 2.0.31]

L1G2 -> L2G0.1 -> [2.0.0, .. . , 2.0.31]

. .

L1G31-> L2G0.31 -> [2.0.0, .. . , 2.0.31]

Monitoring Ports:

No Monitoring ports

Publishing Ports:

No Publishing ports

L1 ECMP groups are create with remote nexthop type for 3.x.x

L1G0 -> [3.0.0]

L1G1 -> [3.0.1] . . .

L1G30 -> [3.15.0]

L1G31 -> [3.15.1]

L2 ARS ECMP group

L2G0 -> [2.0.0, 2.0.1] . . .

L1G480 -> [3.0.30]

L1G481 -> [3.0.31] . . .

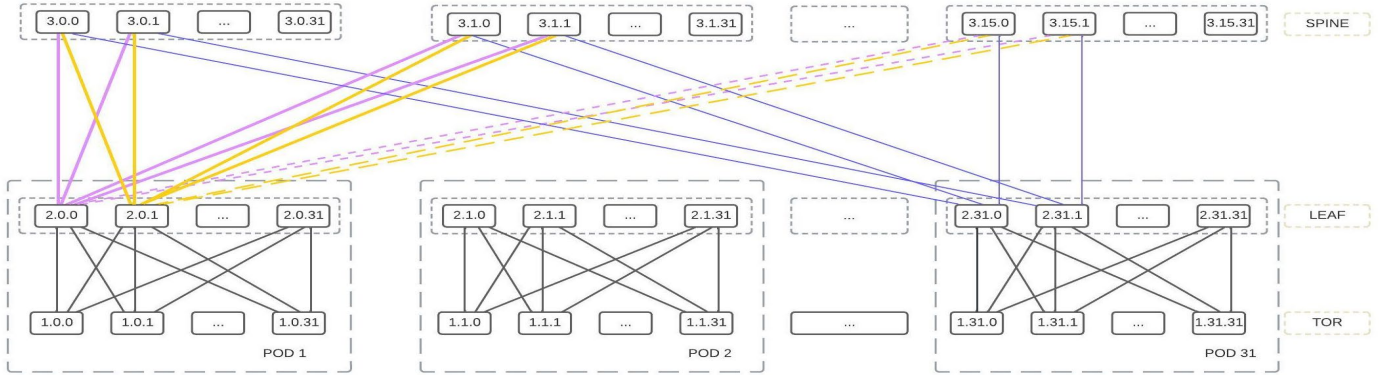
L1G510 -> [3.15.30]

L1G511 -> [3.15.31]

L2 ARS ECMP group

L2G15 -> [2.0.30, 2.0.31]

Path Characterization – 3 Tier Network ..contd



Node 2.0.0: Remote Quality of 2.x.x nodes

L1 ECMP groups are create with remote nexthop type

L1G0 -> [2.0.1]

L1G1 -> [2.1.0]

L1G2 -> [2.1.1] . . .

L1G62 -> [2.31.1] -> Total 63 groups

L2 ECMP group is created with ARS enabled.

L2G0 -> [NH3.0.0, NH3.0.1, . . . NH3.15.0, NH3.15.1]

Standalone [NH1.0.0, NH1.0.1, . . . NH 1.0.31] nexthops for downstream traffic

Monitoring and Publishing Ports:

[1.0.0, 1.0.1, … 1.0.31] monitored ports are published to members of L2G0 as well as to [1.0.0, 1.0.1, ... 1.0.31]

L2G0 member ports are monitored and published to [1.0.0, 1.0.1 … [1.0.31]

path_profile_obj1 =

SAI_ARS_PATH_PROFILE_ATTR_MON_PORT_LIST =
[1.0.0, 1.0.1, ..., 1.0.30, 1.0.31]
SAI_ARS_PATH_PROFILE_ATTR_PUB_PORT_LIST =
[1.0.0, 1.0.1, ..., 1.0.30, 1.0.31] +
[3.0.0, 3.0.1, 3.1.0, 3.1.1, ...,3.15.0, 3.15.1]
SAI_ARS_PATH_PROFILE_ATTR_REMOTE_PATH_ID_LIST =
[1.0.0.x, 1.0.1.x, ..., 1.0.30.x, 1.0.31.x]
SAI_ARS_PATH_PROFILE_ATTR_TYPE =
SAI_ARS_PATH_PROFILE_TYPE_BOTH

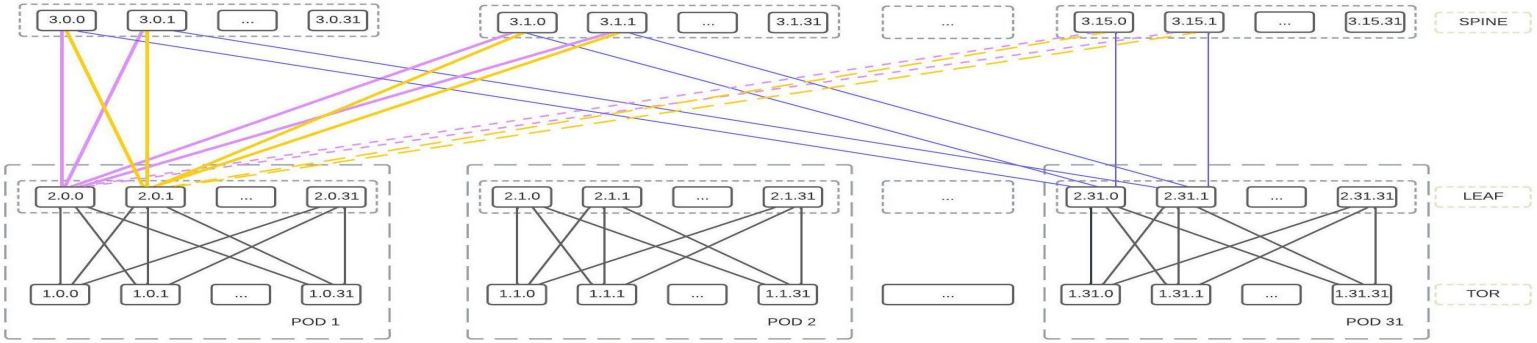
path_profile_obj2 =

SAI_ARS_PATH_PROFILE_ATTR_MON_PORT_LIST =
[3.0.0, 3.0.1, 3.1.0, 3.1.1, ...,3.15.0, 3.15.1]
SAI_ARS_PATH_PROFILE_ATTR_PUB_PORT_LIST =
[1.0.0, 1.0.1, ..., 1.0.30, 1.0.31]
SAI_ARS_PATH_PROFILE_ATTR_REMOTE_PATH_ID_LIST =
[1.0.0.x, 1.0.1.x, ..., 1.0.30.x, 1.0.31.x]
SAI_ARS_PATH_PROFILE_ATTR_TYPE =
SAI_ARS_PATH_PROFILE_TYPE_BOTH

[L1G0, L1G2 …L1G1023]

SAI_NEXTHOP_GROUP_ATTR_ARS_PATH_PROFILE_LIST =
[path_profile_obj1, path_profile_obj2]

Path Characterization – 3 Tier Network ..contd



Node 3.0.0: Remote Quality of 1.x nodes

L1 ECMP groups are create with remote nexthop type for 1.x.x

POD 1:
L1G0 -> [1.0.0] . . .
L1G31 -> [1.0.31]
L2 ARS ECMP group
L2G0 -> [2.0.0, 2.0.1]
POD 2:
L1G32 -> [1.1.0] . . .
L1G63 -> [1.1.31]
L2 ARS ECMP group
L2G1 -> [2.1.0, 2.1.1]
. . .
POD 31:
L1G991 -> [1.31.0] . . .
L1G1023 -> [1.31.31]
L2 ARS ECMP group
L2G31 -> [2.31.0, 2.31.1]

Node:3.0.0, Monitoring and Publishing Ports:

All L2Gx member ports are monitored and published to all L2Gx member ports

path_profile_obj =

SAI_ARS_PATH_PROFILE_ATTR_MON_PORT_LIST =
[2.0.0, 2.0.1, 2.1.0, 2.1.1 . . . . . 2.31.0, 2.31.1]
SAI_ARS_PATH_PROFILE_ATTR_PUB_PORT_LIST =
[2.0.0, 2.0.1, 2.1.0, 2.1.1 . . . . . 2.31.0, 2.31.1]
SAI_ARS_PATH_PROFILE_ATTR_REMOTE_PATH_ID_LIST =
[2.0.0.x, 2.0.1.x, 2.1.0.x, 2.1.1.x . . . . . 2.31.0.x, 2.31.1.x]
SAI_ARS_PATH_PROFILE_ATTR_TYPE =
SAI_ARS_PATH_PROFILE_TYPE_BOTH

[L1G0, L1G2 …L1G1023] -> path_profile_obj

Call to Action

- SONiC HLD
  - Come join us to add more usecases and refine the HLD

- SAI Spec Enhancement
  - Come join us to define a SAI spec, write test cases and more

Thank you!