MPLS-Based Hierarchical SDN for Hyper-Scale DC/Cloud
draft-fang-mpls-hsdn-for-hsdc-04

Luyuan Fang, Microsoft
Deepak Bansal, Microsoft
Fabio Chiussi
Chandra Ramachandran, Juniper
Ebben Aries, Facebook
Shahram Davari, Broadcom
Barak Gafni, Mellanox
Daniel Voyer, Bell Canada
Nabil Bitar, Alcatel Lucent

IETF 94, MPLS WG, November 4, 2015

Status

- HSDN architecture has been widely supported from both industry and academia
- It created the foundation architecture for scalability to tens of millions endpoints, which was not previously achievable
- It is being adopted and applied to different forwarding technologies (L3, L2, L1, IPv4, IPv6, MPLS, SR)
- The key concept of partition and hierarchy is being applied to the control plane

Reference: "Hierarchical SDN for the Hyper-Scale, Hyper-Elastic Data Center and Cloud," *Proc. ACM SIGCOMM Symposium on SDN Research*, June 2015, http://dl.acm.org/citation.cfm?id=2775009

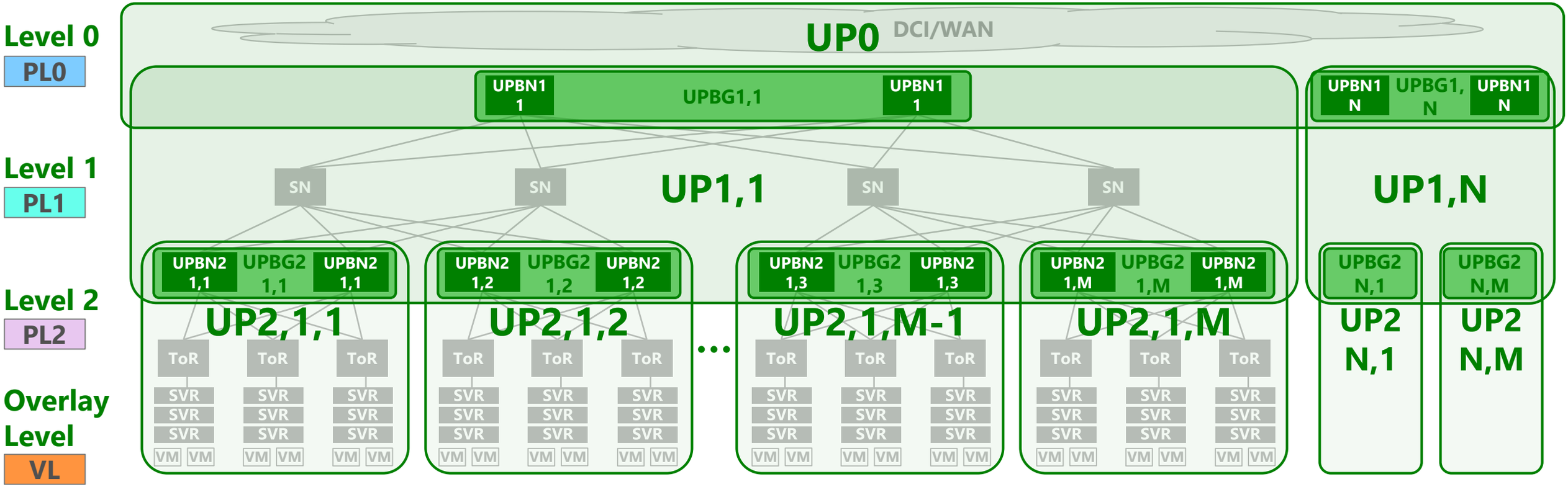
- Contains the LFIB computation details with ECMP and TE, scalability analysis, and performance data

Benefit of Hierarchical SDN (HSDN)

- Partitioning is crucial for scaling to 10's of millions endpoints
- HSDN is the architecture for partitioning the DC and DCI
  - The principle applies to any forwarding: MPLS, SR, IPv4, and IPv6, L2 or even L1
  - The control plane can be implemented with full SDN approach or using BGP-LU for label distribution (draft-fang-idr-bgplu-for-hsdn-01)
- Two game-changing properties of HSDN
  - All paths in the network can be pre-established in the LFIBs (with small LFIBs)
  - Labels can identify paths, not just destinations

All Paths are set: support End-to-End Any-to-Any TE and ECMP concurrently

HSDN: Hierarchical Underlay Partitioning

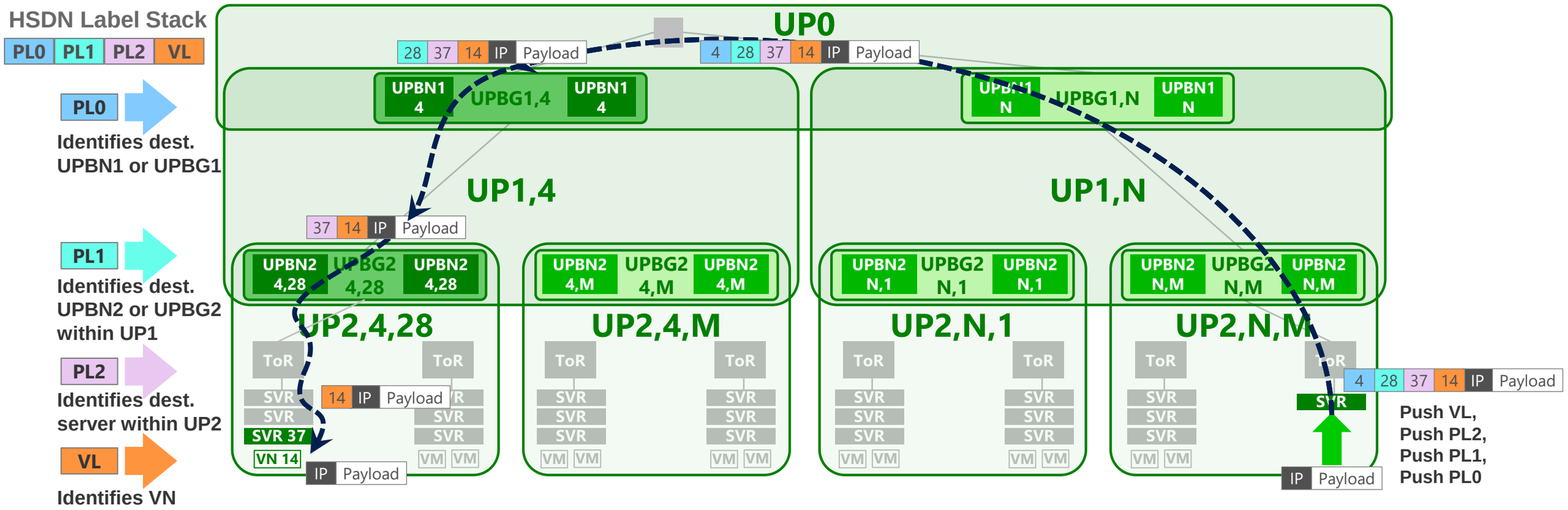


- One path label per level of underlay partition, plus one VN label
- Labels are “static,” globally unique within each partition

Example:

- UP0 = DCI; UP1s = DCs; UP2s = Clusters ➔ With 3 levels, easily scale to 10’s of millions of endpoints

HSDN Forwarding: The Life of a Packet



- Route optimization
  - Forward a packet from any source to any destination using the same (or less) number of hops as in a flat architecture and without introducing any additional latency
  - "Turn Around" entry to optimize label usage

Next Steps

- Issue -05 draft adding new co-authors and contributors
- Request for WG adoption