

Best Practices for Optimal LAG/ECMP Component Link Utilization in Provider Backbone Networks

draft-krishnan-large-flow-load-balancing-01

- Ram Krishnan, Sanjay Khanna (Brocade Communications)
 - Bhumip Khasnabish (ZTE)
 - Anoop Ghanwani (Dell)

CURRENT ISSUES

- Hash based techniques for LAG/ECMP are flow unaware
 - Do not distinguish between large and small flows
 - Over-utilization/congestion of certain links
- LAGs of different member link speeds pose challenges
 - E.g. 2x100G and 20x10G LAG are not equal
- Some networks may lack end-to-end visibility e.g. transit networks
 - Global optimization techniques, e.g. SDN/Openflow, may not be feasible

SOLUTION

- “Targeted” local optimization is used for hotspots
 - Techniques are applied within a single network node
- Best practice techniques
 - Long-lived large flow definition
 - Long-lived large flow identification
 - sFlow/Netflow sampling
 - Automatic hardware identification
 - Egress link congestion notification
 - Long-lived large flow load-balancing
 - Manual (Operator driven), Automatic
 - No dedicated link(s) vs dedicated link(s) for large flows

NEXT STEPS

- Adopt as a work item in OPSAWG
 - Network Operator interest
 - Operators are facing this problem today
 - Vendor interest

ADDITIONAL WORK ITEMS

- Standardized data model
 - Router policy configuration for long-lived large flows detection and load-balancing
 - Moving detected long-lived large flows from router to central entity
- Applying these local optimization techniques for Data Centers
 - Use case
 - Switches/routers from different vendors; mix of 10G/100G
 - Global optimization techniques may not be feasible
 - Switch/router programming interface
 - Openflow besides normal Ethernet switching