I2RS Large Flow Use Case

draft-krishnan-i2rs-large-flow-use-case-00

IETF 88

Ram Krishnan (Brocade Communications)

Anoop Ghanwani (Dell)

Sriganesh Kini (Ericsson)

Dave Mcdysan (Verizon)

PROBLEM STATEMENT

- Current flow based LAG/ECMP load balancing techniques treat all flows as equal; they make inefficient use of the network bandwidth in the presence of long-lived large flows (Note 1) such as file transfers.

- This use-case aims to improve the network bandwidth efficiency under such conditions and aims to drive requirements for the I2RS WG.

Note 1: Terminology -- Large flow(s): long-lived large flow(s)

LARGE FLOW RECOGNITION/SIGNALLNG

- Network-based Recognition of Large Flows
  - Automatic hardware-based recognition, e.g. IPFIX, NetFlow
  - Packet sampling, e.g. sFlow, IPFIX

- Application-based Signaling of Large Flows
  - Large flows signaled by application to the management entity capable of performing flow rebalancing
  - This communication is outside the scope of I2RS

FLOW REBALANCING (1)

LOCAL REBALANCING

- The utilization of the component links that are part of the LAG or ECMP are monitored
- Flows are redistributed among the member links to ensure optimal load balancing across all of the component links (Note 2)
- Works for IP and MPLS networks

Note 2: Details --
http://tools.ietf.org/html/draft-ietf-opsawg-large-flow-load-balancing-05

FLOW REBALANCING (2)

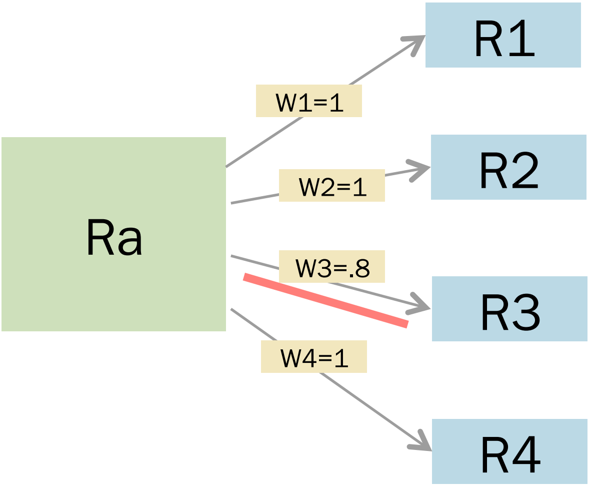
GLOBAL REBALANCING – IP NETWORKS

- Program a globally optimal path for the large flow using hop-by-hop PBR rules
- The weights of the ECMP table for different nexthops should be adjusted to factor the large flows

FLOW REBALANCING (2)

GLOBAL REBALANCING – IP NETWORKS

- Simple Illustrative Example:
  - Consider a 4 way ECMP at node n1 with IP nexthops n11, n12, n13, n14 using links l1, l2, l3, l4 each of capacity 10 Gbps.
  - Say, a long-lived large flow of average bandwidth 2 Gbps is admitted to one of the links l3.
  - The ECMP nexthop table will be programmed as w1*n11, w2*n12, w3*n13, w4*n14 where w1=w2=w4=1 and w3=0.8;



FLOW REBALANCING (4)

GLOBAL REBALANCING – MPLS NETWORKS

- Have multiple LSPs between ingress and egress edge routers. Program PBR entry at the edge LSR that forwards the large flow to a specific LSP known to have the necessary bandwidth is needed.
- Program a new LSP for a given large flow.

I2RS INFORMATION MODEL REQUIREMENTS SUMMARY

- IP Networks
  - Hop-by-hop PBR entries with IP nexthop
  - Identify ECMP entries and associate weights that can be programmed for each of the components. Useful to have the notion of an ECMP group that is used by multiple routes.
- MPLS Networks
  - PBR entries at the edge LSR with LSP tunnel nexthop
  - Program new LSPs in the network
- Most requirements are PBR related
  - Aligned with PBR information model work
- Other requirements
  - The ability to address individual ports in a router is desirable. I2RS topology to be aware of LAG members; ability of routers to accept route or PBR entries that map to a specific member port within a LAG.

NEXT STEPS

- Interesting Comments/Thoughts so far from Diego Lopez (off the list)
  - User based signaling mechanisms
  - Global rebalancing using PCE and/or ALTO

- Request – please read the draft and send comments !