IEEE 802.1Qau Reaction Point Tag: Issues & Questions

Bruce Kwan & Ashvin Lakshmikantha

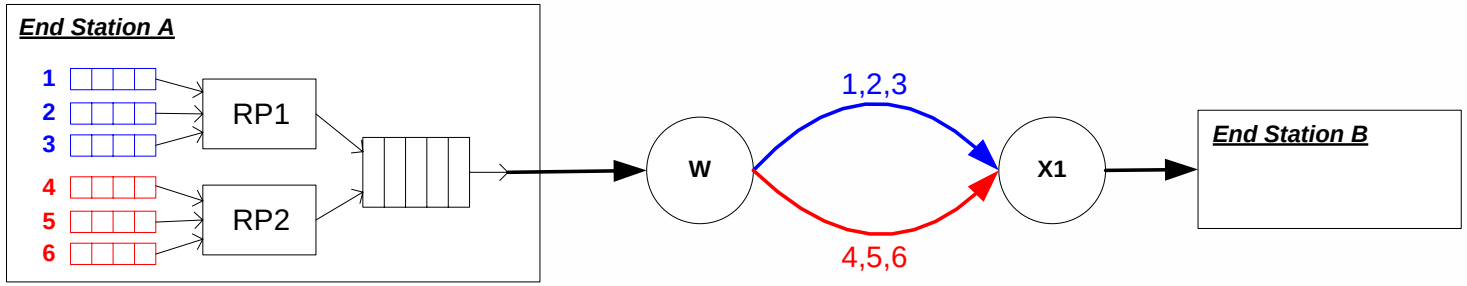
IEEE Plenary Meeting (Denver)

July 15, 2008

Overview

- LAG and EoNECMP
- Link Aggregated NICs
- Conclusion

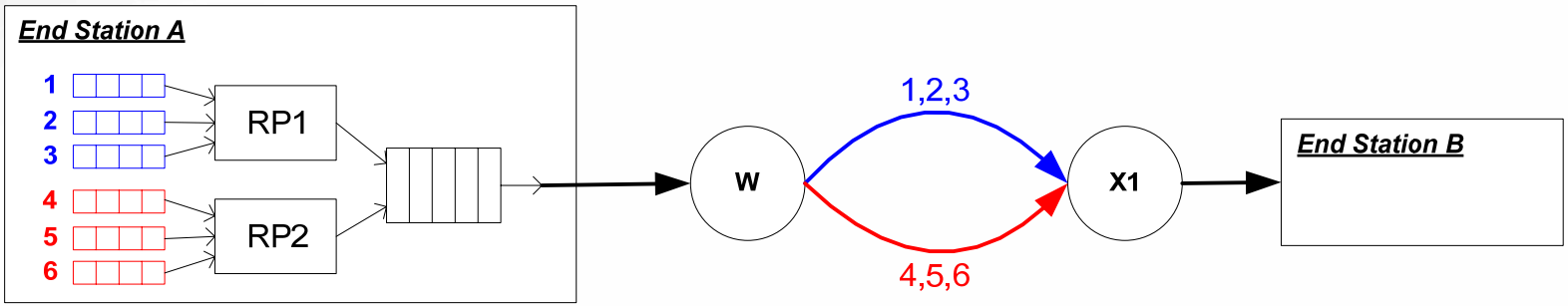
Solution Overview for the LAG/EoNECMP Issue



- Goal is to coordinate Flow to RP selection and Flow to Path selection to limit fate sharing
- Every RP is assigned a locally unique ID which is transmitted as a tag (RPID) along with every packet leaving the NIC from that RP
- LAG resolution is performed using the RPID
- Only RPs that have flows on the congested path will be slowed down

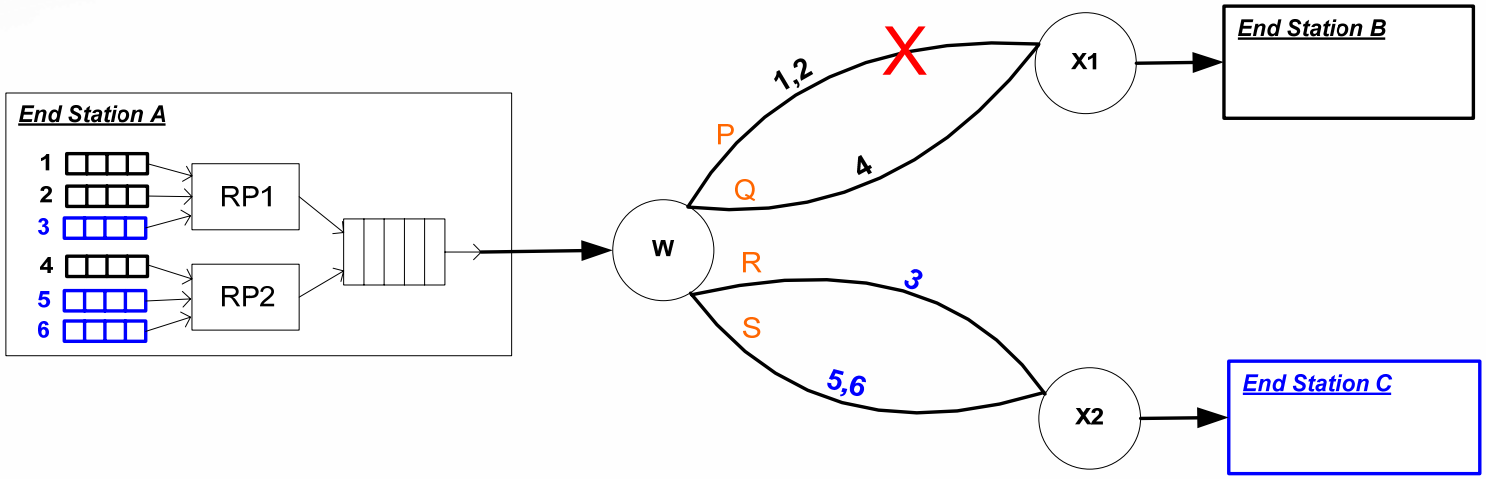
BROADCOM.
Connecting everything

Bridge Behavior: Open Questions



- Defining Bridge Load Balancing Behavior
  - What is being proposed for the bridge behavior? Need to define it now and rather than leaving it undefined.
  - At present, the standard does not dictate the bridge load balancing algorithm. Doing so would limit vendor differentiation.

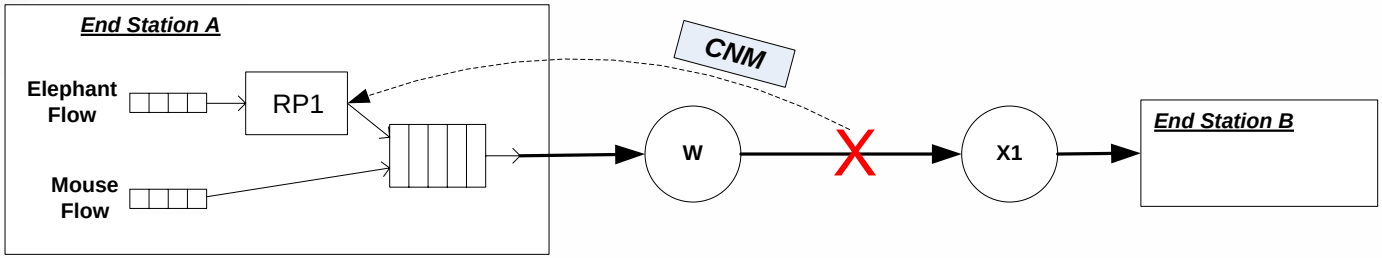
Fate Sharing Issue Remains



When congestion occurs on path P, the rate of flows associated with RP1 will be slowed down. Innocent flow 3 will be impacted.

Fate sharing is not addressed with the RPID under this common scenario where RP's contain flows with different destinations.

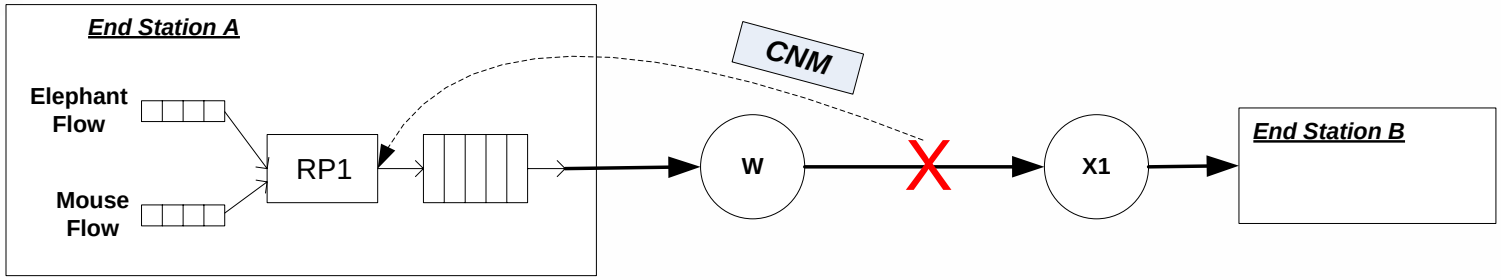
QCN and Fast Delivery of Mice Flows



- QCN (& BCN) Design Philosophy
  - Control the elephants
  - Allow the mice to zip through the network
- Achieving Fast Delivery of Mice Flows
  - QCN-Sampling behavior is designed so that statistically elephant flows are more likely to be sampled (and consequently receive a CN Message)
  - When a new flow starts, it is allowed to burst at line rate
    - Mice with a few packets to transmit will zip through the network since it's transmission rate is high
  - Results in high utilization of the network

BROADCOM.
Connecting everything

RPID, QCN and Mice Flows Fate Sharing Degradation



- Impact of RPID on QCN and Mice Flows
  - RPID assigned to all incoming flows
  - If an RP is being congestion managed, any mice flows mapped to that RP will result in fate sharing and ***slow delivery of mice flows***
  - Reduced utilization of the network

Load Balancing Degradations

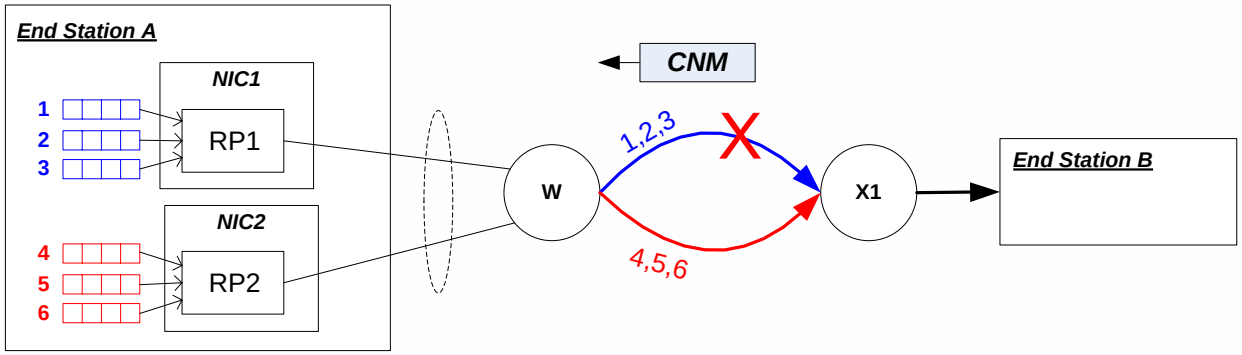
- Assumption
  - To achieve desired behavior, one approach would be to perform hash based load balancing based only on the RPID
- Performance Concerns
  - Load balancing microflows can yield even load balancing across paths
  - Load distribution based on the coarse-grained definition of a flow (RPID) can lead to degraded load balancing behavior

BROADCOM.
Connecting everything®

Overview

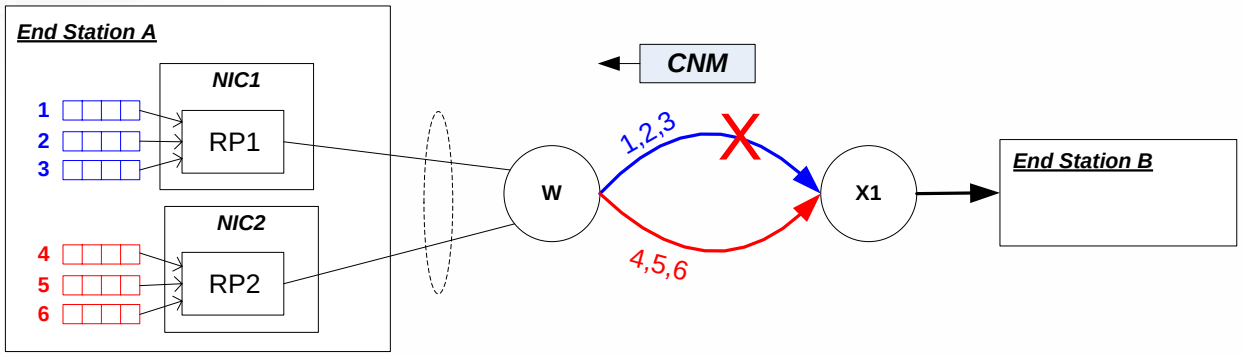
- LAG and EoNECMP
- Link Aggregated NICs
- Conclusion

Solution Overview for Link Aggregated NICs



- Every CNM message will include the RP-ID tag associated with the sampled packet
- The Bridge uses the RP-ID within the CNM to identify the correct egress port to send the CNM

Bridge Behavior: Open Questions



- Bridge Behavior
  - What is the impact on the bridge?
- Support May Lead to Increased Cost/Complexity
  - RPID to Port Mapping Table
    - To achieve stated goal, edge bridge connected to NICs requires a mapping table from RPID to port
    - When a CNM message arrives, the mapping table can be used to resolve how to direct the CNM to the correct NIC
  - Populating the Mapping Table
    - Manual Configuration, or
    - Protocol definition needed to “learn” the binding between RPID and port

BROADCOM.
Connecting everything®

Overview

- LAG and EoNECMP
- Link Aggregated NICs
- Conclusion

Conclusion

- Understand What is (& what is not) Being Solved
- Clarify Solution
  - Bridge load balancing behavior for LAGs
  - Bridge behavior for Link Aggregated NICs
  - Needs to be defined now and not later to insure this is solving the stated problems
- Understand the Compromises
  - Limiting bridge vendor differentiation in terms of load balancing
  - Fate sharing remains
  - Slowed Mice flow delivery
  - Load Balancing Degradations
  - Increased cost/complexity

BROADCOM.
Connecting everything®