Enhanced Flow control in Ethernet Networks

Agenda

- Review the proposal to support the communication of congestion information

Some Facts …

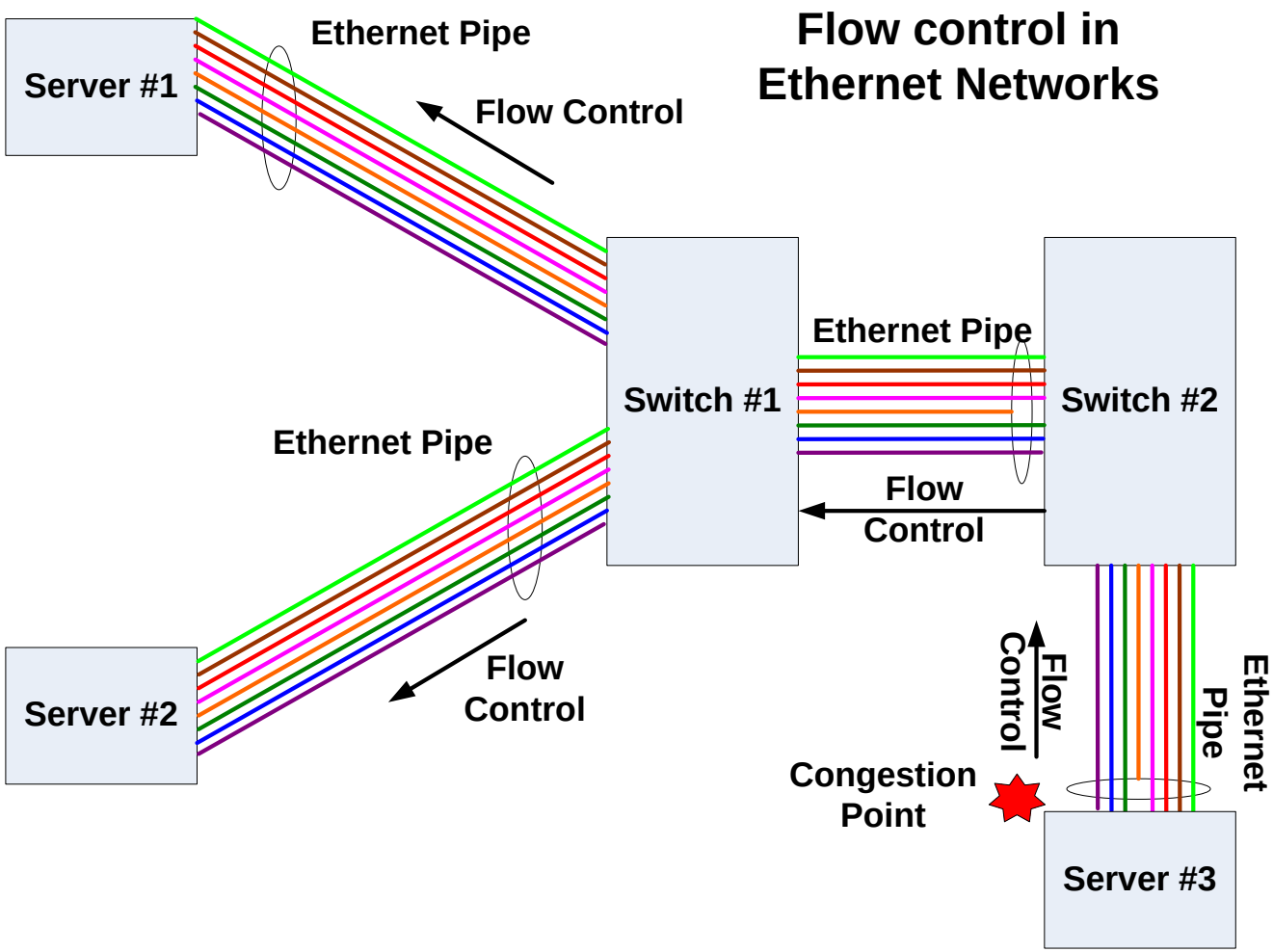
- Ethernet has fulfilled the expectations of applications with best effort delivery model for past 3 decades
- Throwing more bandwidth at the problem does not solve congestion problem since queuing delays are more than wire delays
- Convergence of mission critical and non-mission critical applications over shared network infrastructure require traffic differentiation
- Is it time to enhance ethernet?

Ethernet : Is it time to enhance it?

- Businesses rely heavily on predictable performance of IT and network infrastructure
- Businesses needs more intelligent and adaptable network infrastructure to meet business needs
- Commoditization of compute and sharing of same network infrastructure between mission critical and non-mission critical applications require
  - Intelligent network infrastructure with prioritization and differentiation of traffic
  - Guaranteed performance per traffic type to meet business SLA

- Ethernet requires enhancements not disruptive changes

802.1p and Flow control in Ethernet Networks



802.1p supports prioritization of the traffic

Flow control

- Limitations
  - One traffic class affect another in true loss-less fabric implementation
    - Flow control (802.3X) does not have prioritization
    - One VLAN traffic can block another VLAN traffic

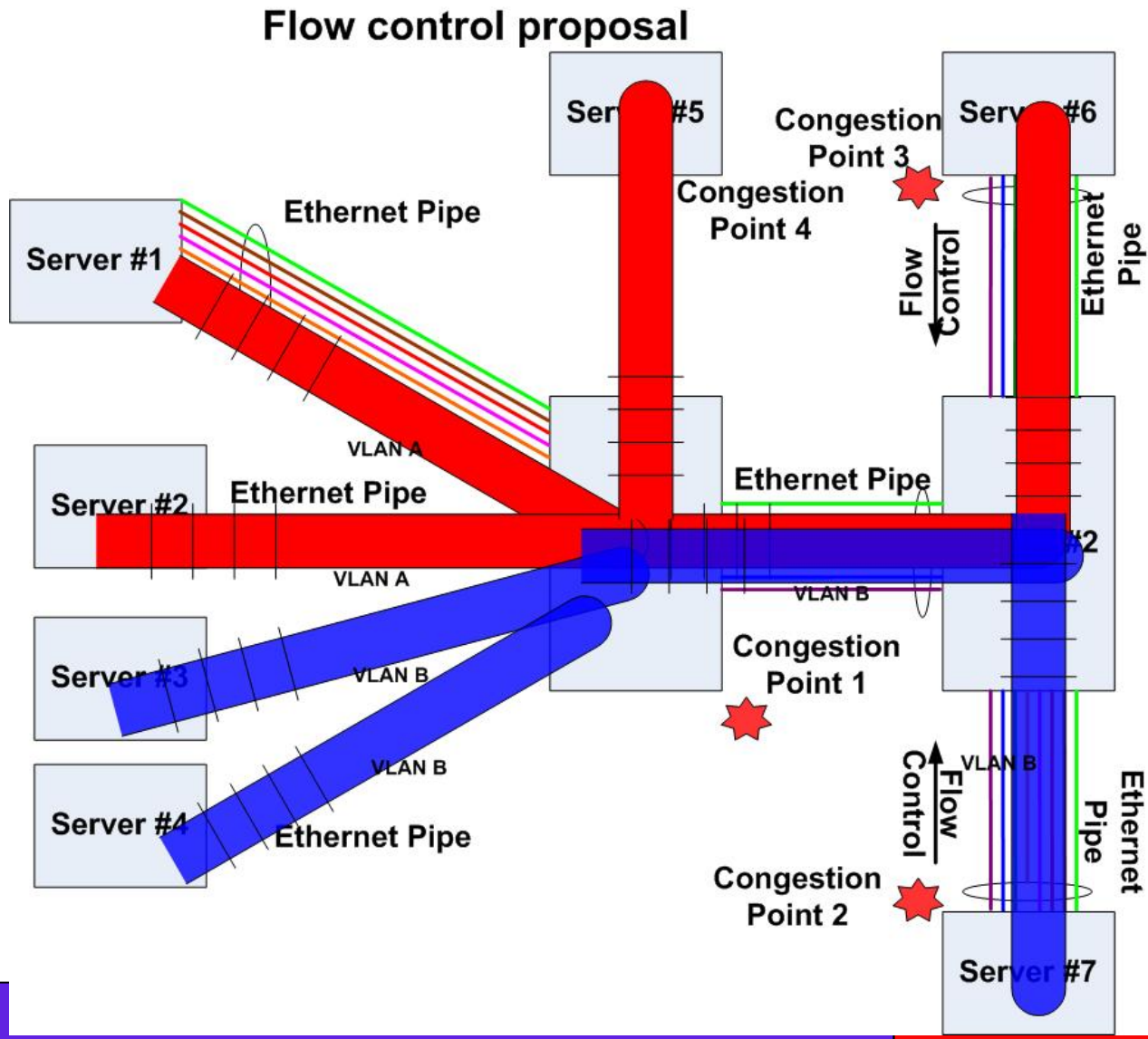
Flow control Requirements

- Traffic type differentiation
  - Congestion should be reported based on traffic types, E.g. priority, {VLAN,Pri}, other finer level flows are possible, not relevant in this group
  - Avoid head-of-line blocking between traffic types
- Fairness
  - Congestion information should be reported to all the sources who can potentially send traffic to congested destination
- Keep Flow control traffic as low as possible

Flow control Proposal

- Flow control per VLAN per priority
- Flow control packet per VLAN can be broadcasted back to all the sources to turnoff traffic destined to a MAC address (End node) or group of MAC addresses going through congested switch port
- This is in addition to link level flow control defined today

Flow control Example



Congestion Point #1

- **Switch #1's output port connected to switch #2 is congested**
- **This port carry traffic for server #6, VLAN A and server #7, VLANB**
- **Ingress side of this port sends two broadcast flow control packets**
- **One on VLAN A**
  - **Packet includes DMAC 6, VLAN A, Priority and time to stop etc …**
- **One on VLAN B**
  - **Packet includes DMAC 7, VLAN B, Priority and time to stop etc …**
- **Traffic towards server #5 will not be affected**

Congestion Point #2

- Server #7 is congested
- Server #7 sends broadcast flow control packet on VLAN B to all sources belong to same VLAN
  - Packet includes DMAC 7, VLAN B, Priority and time to stop
- Traffic towards any other server except #7 will not be affected!

Congestion Point #3

- Server #6 is congested
- Server #6 sends broadcast flow control packet on VLAN B to all sources belong to same VLAN
  - Packet includes DMAC 6, VLAN B, Priority and time to stop
- Traffic towards any other server except #6 will not be affected!

Congestion Point #4

- Server #5 is congested
- Server #5 sends broadcast flow control packet on VLAN A to all sources belong to same VLAN
  - Packet includes DMAC 5, VLAN A, Priority and time to stop
- Traffic towards any other server except #5 will not be affected!

Summary

- Ethernet needs enhancement in congestion reporting mechanism back to sources
  - The mechanism has to be fair to all the sources
- Use VLAN based broadcast support to report congestion information back to sources with fairness
- This proposal allows to differentiate various traffic types on ethernet networks with fairness and guaranteed service

Thank You