Proposed Technical Approach Congestion Isolation

IEEE 802.1 Interim

Oslo

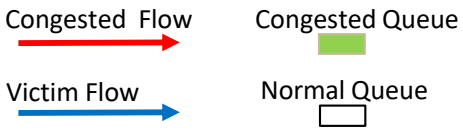
September 2018

Paul Congdon (Huawei/Tallac)

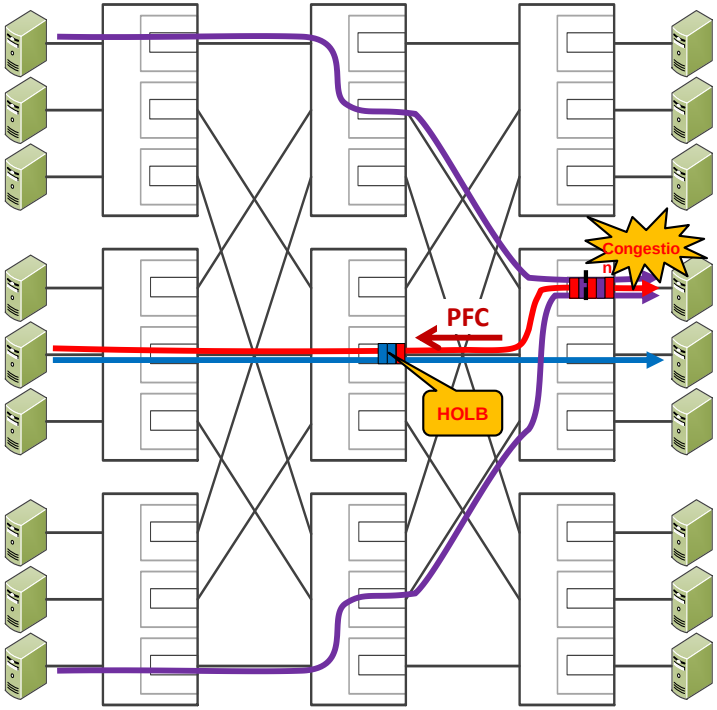
Previous Presentation References

- Technical overview of CI
  - http://www.ieee802.org/1/files/public/docs2018/cz-congdon-congestion-isolation-review-0418-v1.pdf
- Possible changes to 802.1Q
  - http://www.ieee802.org/1/files/public/docs2018/cz-congdon-ci-Q-changes-0618-v1.pdf
- Objectives Discussion
  - http://www.ieee802.org/1/files/public/docs2018/new-dcb-congdon-ci-objectives-0118-v02.pdf

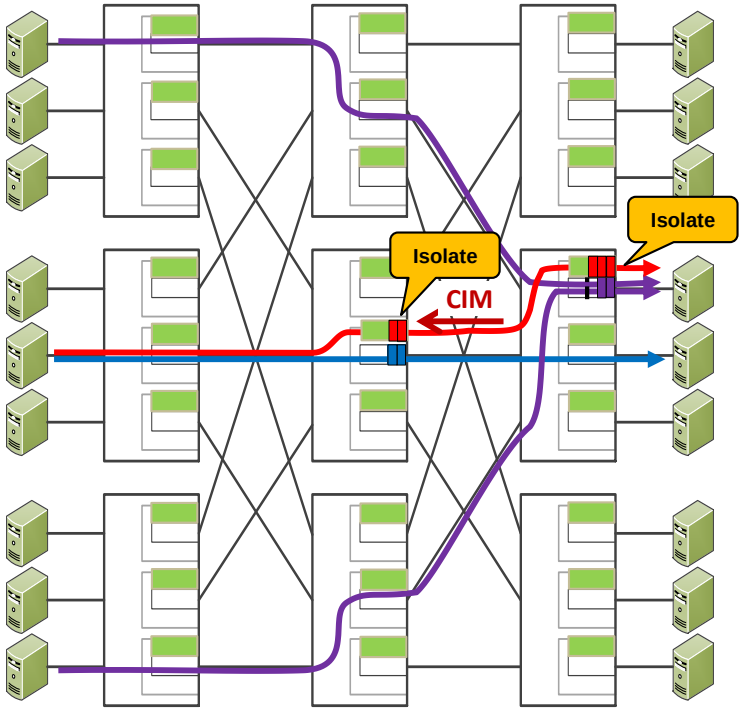
One Slide Congestion Isolation Refresh



Today – Without Congestion Isolation



Congestion Isolation

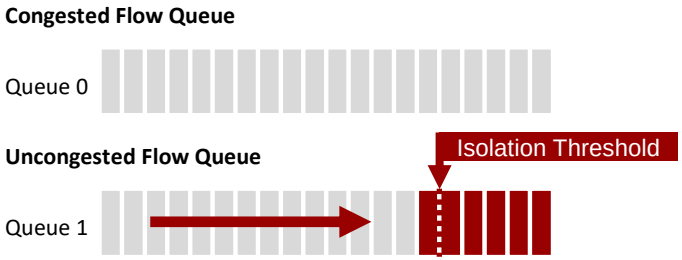


Congestion Isolation Critical Processes

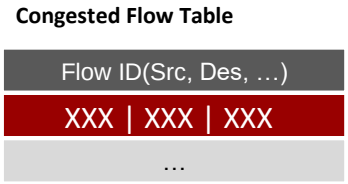
1. Identifying flows causing congestion
2. Creating flows in the congested flow table
3. Signaling congested flow identify to neighbors
4. Isolating congested flows without ordering issues
5. Interaction with PFC generation
6. Identifying when congested flows are no longer congested
7. Removing entries from congested flow table
8. Signaling congested to non-congested flow transitions to neighbors
9. Un-isolating previously congested flows without ordering issues

Congestion Isolation - Illustrations

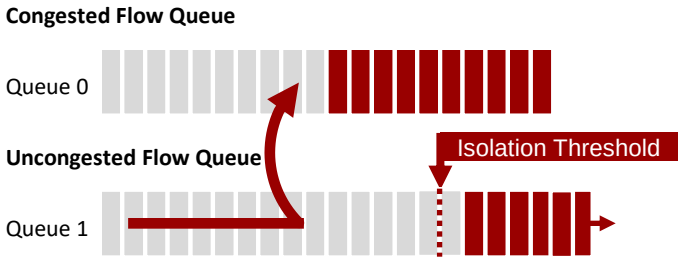
1. Identify flows causing congestion



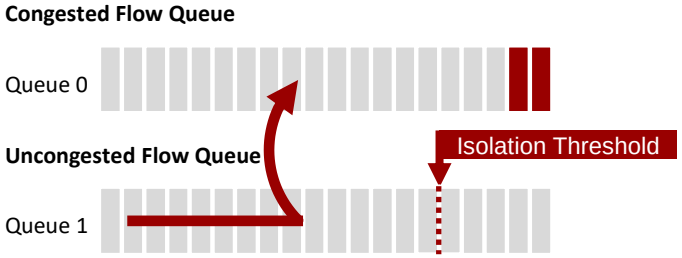
2. Create entries in congested flow table



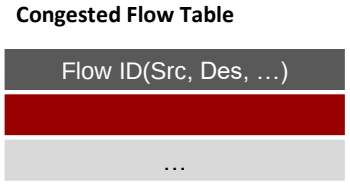
4. Isolate congested flows



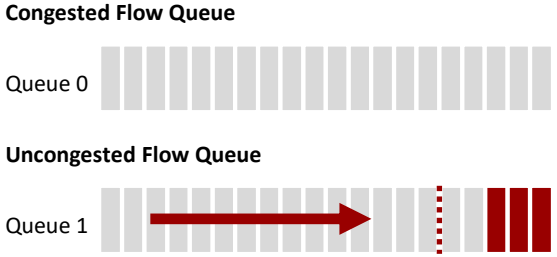
6. Identify when flows are no longer congested.



7. Remove entries in the congested flow table

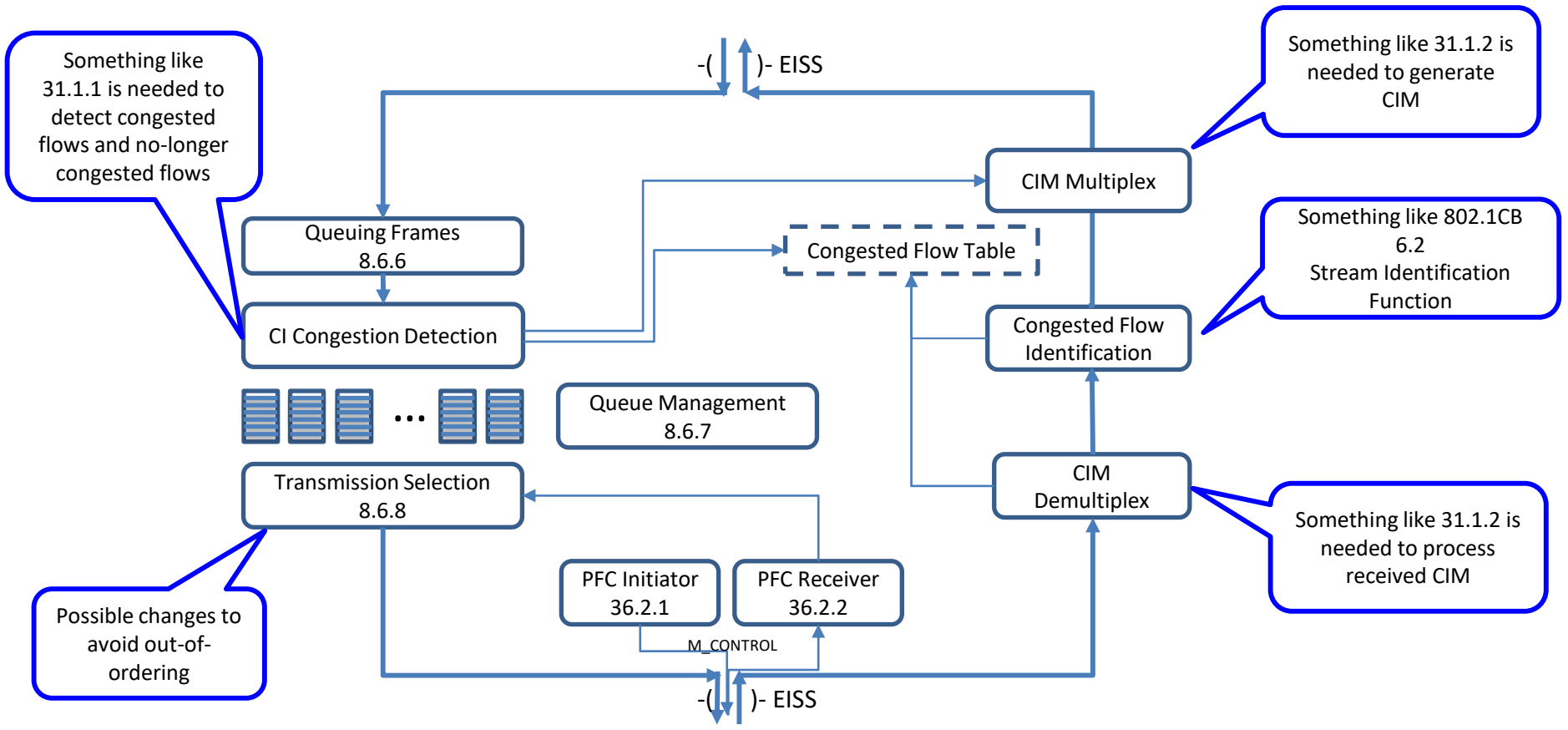


9. Forward to uncongested flow queue



Proposed Reference Diagram

(without the callouts)



Objective Categories

- Functionality
- Compatibility
- Performance
- Scale
- Implementation (Cost/Complexity)
- Manageability

Functional Objectives

- With high probability, identifies flows that are causing congestion
- Quickly adjusts transmission scheduling of offending flows
- Avoids head-of-line blocking by signaling to upstream neighbor to also adjust transmission scheduling.
- Reduces frequency of PFC usage to create lossless environments

Compatibility Objectives

- Works in legacy environments
  - Signaling is not enabled unless peer bridges are compatible
  - Does not require network wide upgrade
- Works with existing PFC deployments
  - Does not require additional traffic classes
- Works in conjunction with end-to-end congestion control schemes (e.g. ECN, BBR, RoCEv2 CNM, QCN)
- Coexists with existing scheduling paradigms in other traffic classes
- Works with load balancing techniques

Performance Objectives

- Metrics to measure performance gains
  - Average flow completion time (mice vs elephants)
  - Reduction in pause time if PFC is enabled
  - Reduction in frame loss if PFC is not enabled
  - Reduction in number of victim flows from HoLB
  - Reduction in overall congestion signaling
  - Increased link utilization

Correctness Objectives

- Does not introduce packets re-ordering within a flow
- Does not introduce deadlock vulnerabilities
- Avoids starvation
- Resilient to loss of control messages

Scale Objectives

- Works in arbitrary data-center topologies with a mix of link speeds
- Limits messaging overhead
  - Does not require message propagation beyond hop-by-hop
  - Does not increase frequency of messages over existing approaches (e.g. QCN)
- Limits flow table size requirements
  - Flow entries are aged
  - Only offending flows are required to be stored
  - Limit amount of state per-flow required

Implementation Objectives

- Limits impact on traffic selection implementations
- Benefit is achieved without additional buffer requirement
- Can be implemented using existing traffic classes
- Limited flow table size requirements
  - Can be implemented by only registering offending flows in flow table

Management Objectives

- Requires only a small set of configuration parameters which are consistent across deployments
- Impossible to configure a inoperable environment (stretch?).
- Limits configuration requirements
  - Does not require additional tuning
- Provides auto discovery of peer capability
  - LLDP CI Discovery TLV
  - No new Hello or auto-configuration protocols