Addressing Concerns with Closed Loop Congestion Management Protocols

Guenter Roeck, Teak Technologies

IEEE 802.1Qau
Stockholm Interim Meeting, September 2007

Rationale

- Several concerns have been raised against the use of a Closed Loop CM protocol
- List all concerns about Closed Loop CM protocols in a single place
- For each concern,
  - Determine if it is a real problem
  - Propose solutions if necessary

Definition

- Open Loop Protocols
  - CP->RP communication
  - Negative feedback only
  - Example
    - QCN
- Closed Loop Protocols
  - CP->RP communication for negative feedback
  - RP->CP/RfP->RP communication for positive feedback
  - Examples
    - Path probing
      - FECN, E2CM, (ECM-SP, QCN-SP, QCN-PP)
    - CP probing
      - (ECM-P, QCN-P)
    - Tagging
      - ECM

Advantages

- Open Loop Protocols
  - Simplicity
- Closed Loop Protocols
  - More accurate control loop

Concerns with Closed Loop Protocols

- CP probes
  - Wrong RP<->CP association may cause RP to be stuck in low data rates
  - Network re-configuration may cause RP to be stuck with CP which is no longer associated with rate limited flow(s)
- Path probes
  - Multi-path environment
    - May cause instability due to probes taking wrong path
  - Shared rate limiters have no well defined path
    - May cause instability
- All probe based protocols
  - Protocol packets sent directly to CP/switch

Concerns - continued

- CPID
  - CPID association with shared rate limiters or in multipath-scenarios causing false feedback
  - CPID Thrashing
  - CP loses anonymity due to existence of CPID
- All
  - Security: Fake probe messages
  - Increased complexity
  - Protocol might have impact on or require modifications of other L2 protocols

Addressing Concerns - CP Probes

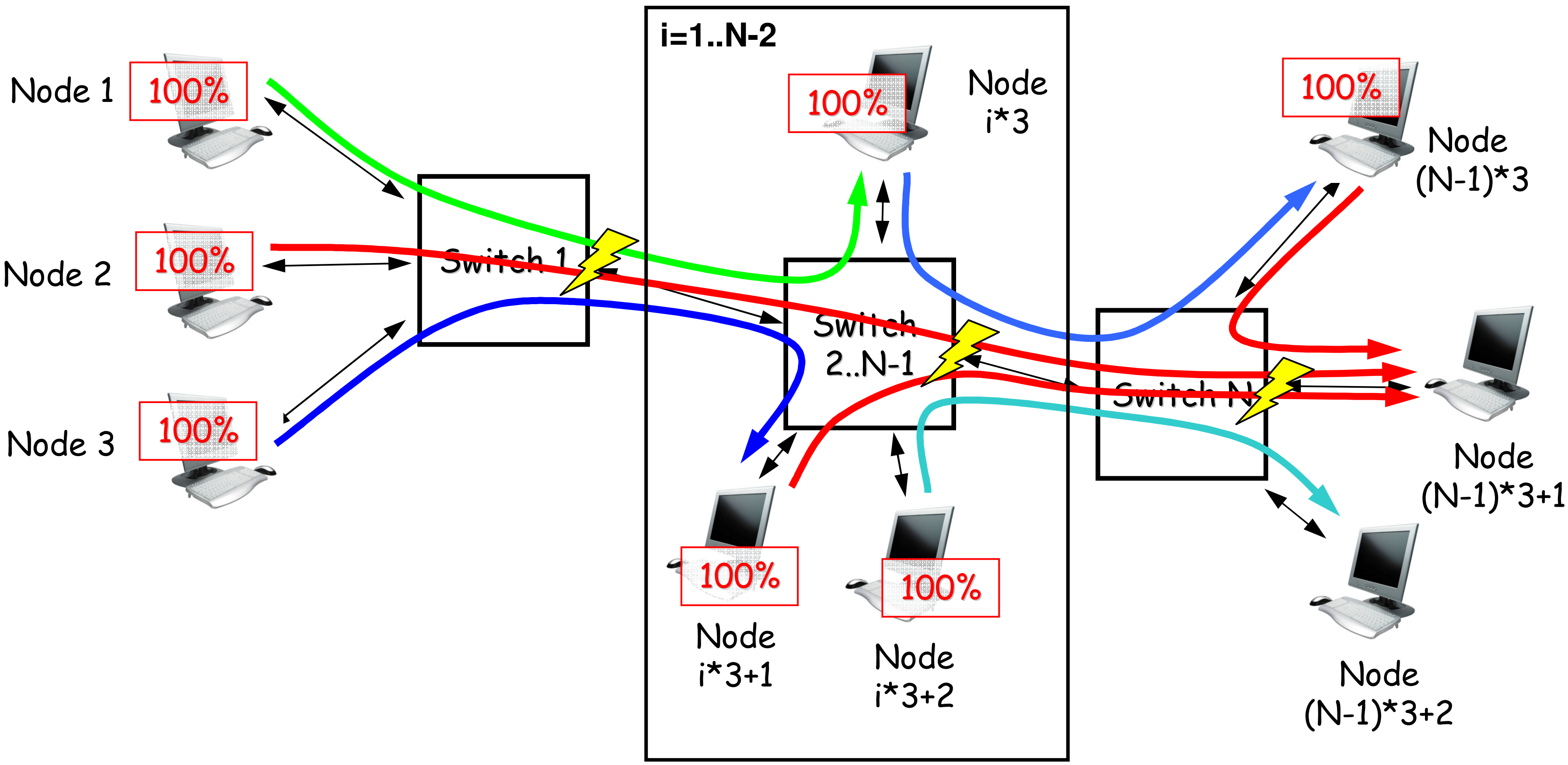
- RP stuck with low data rate
  - Use aggressive self-increase or a timeout if there is no positive feedback
    - Example: QCN-style self-increase
- Network re-configuration may cause RP to be stuck with CP which is no longer associated with rate limited flow(s)
  - Change CPID association whenever negative feedback is received
  - Use aggressive self-increase if there is no positive feedback

Addressing Concerns - Probes

- Probes taking wrong path
  - Problem does not apply to directed probes
  - Sub-path probes always provide as good or better results than directed probes, thus the problem does not apply to sub-path probes either
  - ➢ Use either directed or sub-path probes
- No well defined path for shared rate limiters
  - No real difference to open loop protocol behavior
  - Constantly changing CPID will ensure that lowest throughput CP will dominate
- Protocol packets addressed to CP/switch
  - Is this really a problem ?

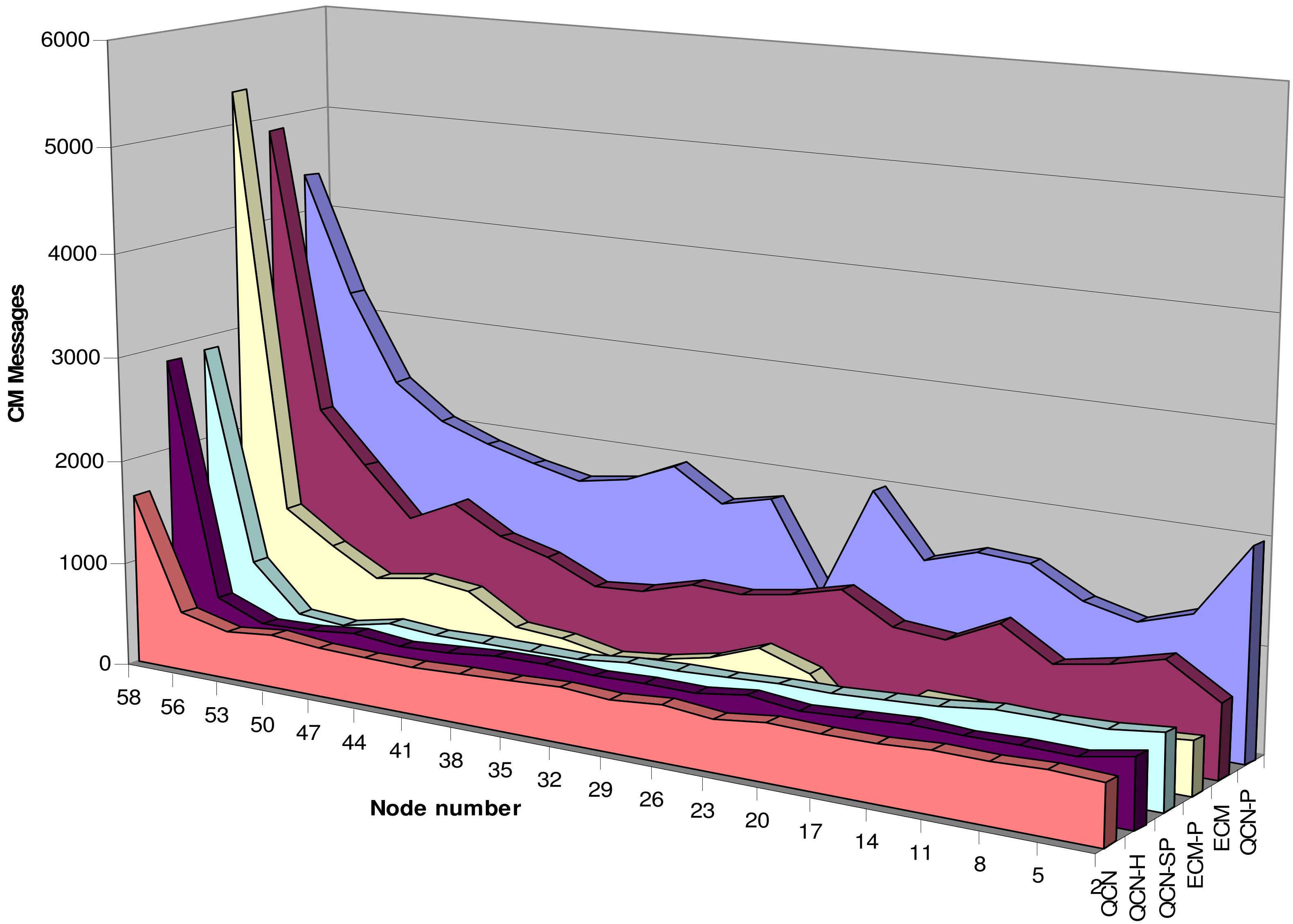
Addressing concerns:
CPID Thrashing

20-stage Hotspot

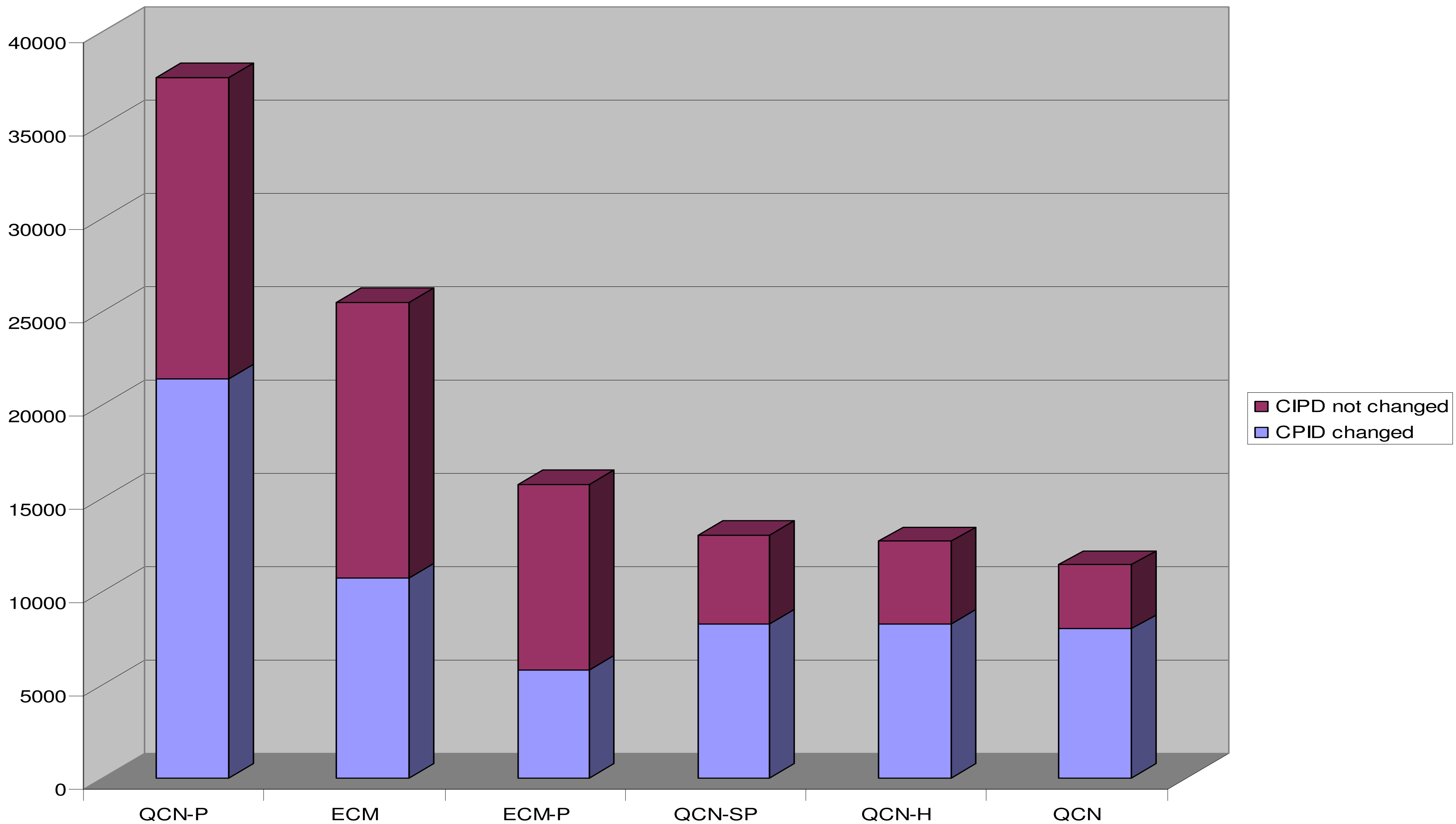


- N=18 switches; 3 hosts per switch
- Node <i> sends to node <i+3>; Node <i+1> sends to node (N-1)*3+1; node <i+2> sends to node <i+4>
- 100% load from all nodes
- Node (N-1)*3+1 receives traffic from <N> sources
- N hotspots

CM Packets Received by Nodes 2,5,8,...

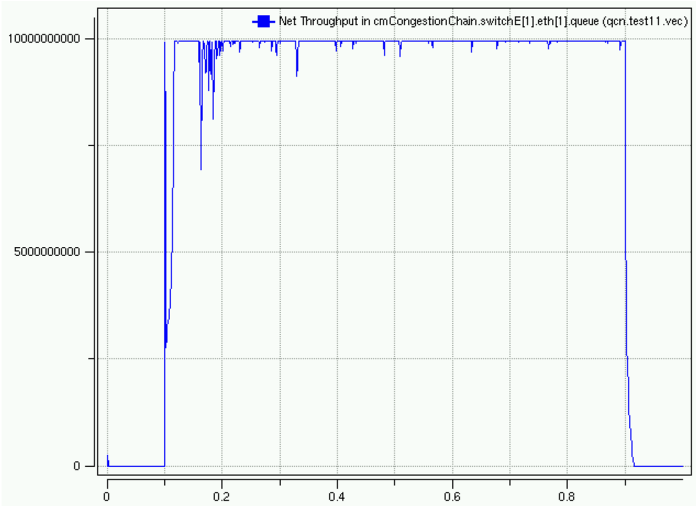


CM Messages per Protocol

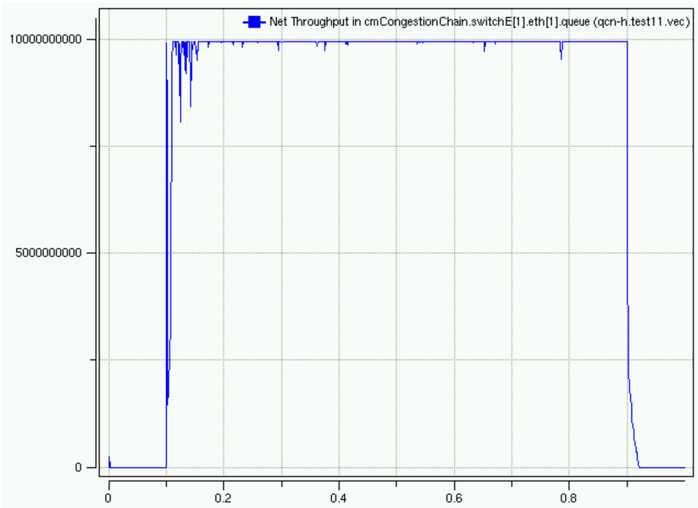


Throughput at Switch N CP: Open-Loop Protocols

QCN

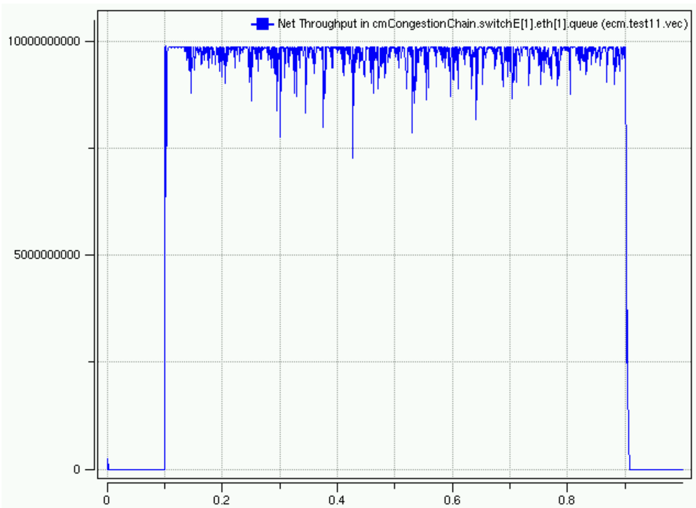


QCN-H

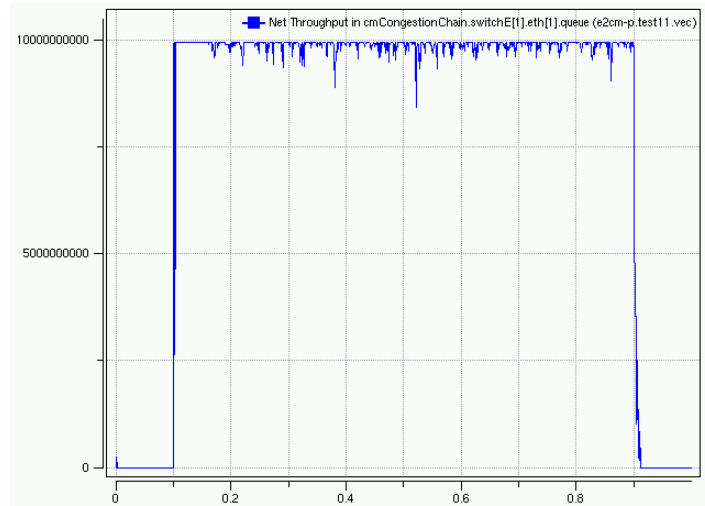


Throughput at Switch N CP: Closed-Loop Protocols

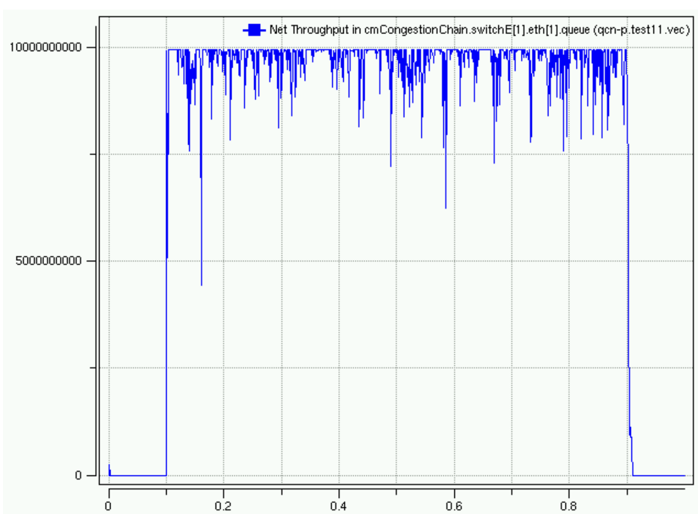
ECM



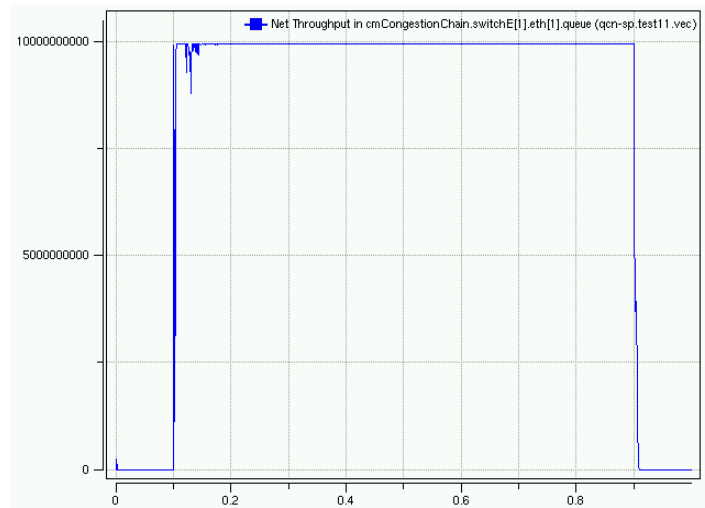
ECM-P



QCN-P



QCN-SP



Conclusions

- In multi-hotspot scenarios, **every** protocol changes its CP association all the time
  - ... even if such an association is not explicitly defined (QCN)
- No evidence that CPID Thrashing could be a problem
- Protocol stability **depends** on changing CPID association in multi-path and multi-hotspot operation

- Wrong CPID association with shared rate limiters or in multi-path scenarios
  - Update CPID association whenever a negative feedback message is received
    - If rate gets too high, another CP with higher congestion will take over
    - CP with lowest rate (highest level of congestion) will dominate
    - Similar to open loop protocols
  - If this is insufficient,
    - Do not use probes if rate limiters are shared
    - Use directed or sub-path probes instead of path probes
  - Need to verify in simulation

- Fake probe messages
  - Answer 1: Security is not commonly addressed in 802.1 protocols. Furthermore, every CM protocol has this problem. Why is it a concern here ?
  - Answer 2: What can happen ?
    - Fake probes sent to CP
      - CP only replies if feedback is positive
      - Worst case, the “offender”, i.e., the host referenced in fake probes, would get more bandwidth
        - Impact similar to the host simply increasing its rate or not caring about negative adjustment requests
    - Fake probes sent to RP
      - RP will reduce its data rate
        - Same impact for all protocols, independent of probe mechanism

Addressing concerns - Complexity

- Increased complexity
  - RP: Needs to send probes (or tags) and evaluate results
  - CP: Detect and evaluate probes/tags
- Looking into the code, this seems to be a minor issue
  - Most of the code to generate CM packets is already there anyway
    - Arguable, since simulation code and implementation may only be loosely coupled
- According to HW engineers, added complexity is not really a problem as long as probes/tags have a well defined (static) packet format
  - More concerned with complex calculations

Addressing concerns – Anonymity

- Loss of CP anonymity
  - Not really a problem
  - CP is not anonymous anyway
    - Always sends its MAC address with each CM message
  - Customers like the idea of knowing where they may have a problem in the network
    - Knowing where the problem is seems to have higher value than trying to automatically fix it

Impact on Other Protocols

- Protocol might have impact on or require modifications of other L2 protocols
  - This is a generic argument which can be used against any protocol
    - Does not have much practical value without substantiation
  - Can be addressed by stating that protocol must be independent of other protocols

Worst case scenarios

- CP switch disappeared
  - No probe replies; RL auto-increases data rate until full rate recovered, or until negative adjustment request received from another CP
    - No worse than QCN
- Path probes take wrong path
  - Use Sub-path or CP directed probes
  - No positive feedback if protocol designed correctly
    - No worse than QCN
- Data path changed
  - Only positive feedback received from CP
  - RL increases data rate until full rate recovered, or until negative rate adjustment request received from another CP
    - Better than QCN

Summary

Problem	Solution
Wrong CP-RP Association	✓
RP stuck in low rate	✓
Instability due to probes taking wrong path	✓
No well defined path with shared rate limiters	✓
Probes sent directly to switch/CP	?
CPID Thrashing	-
Loss of anonymity	-
Fake probe messages	- (✓)
Increased complexity	- (?)

Conclusions

- Even in worst case scenarios, directed or sub-path probes do not have a negative impact on protocol performance
- Significant performance gains in all other scenarios
- Improved performance outweighs increased complexity
- Protocol elegance and simplicity should not outweigh performance
- Good performance requires a closed loop protocol
  - Closed Loop protocol implies use of CPID to identify CP

Thank you

Questions ?

Backup slides

Probe algorithm overview and assumptions

- Probes sent to solicit **positive** feedback only
  - CP does not reply if feedback would be negative
  - Options
    - Directed probes
      - Probes sent to CP associated with RL
    - Sub-path probes
      - Probes sent to flow destination address, and reflected by “last” CP supporting switch in path
      - In-path CP removes probe from network if it is congested (Fb would be negative)
- RL associated with CP from which the most recent negative adjustment request was received
  - RP<->CP association will change each time a negative adjustment request is received from a different CP (for a given RL)
- RP<->CP association per RL queue
  - Deleted when a queue/RL is deleted
- RP<->CP context (per RL queue)
  - CPID
  - CP MAC address