Feedback Output Queuing

A Novel Architecture for Efficient Switching Systems

Victor Firoiu

Performance Engineering Center / Advanced Technology

Nortel Networks

August 21, 2002

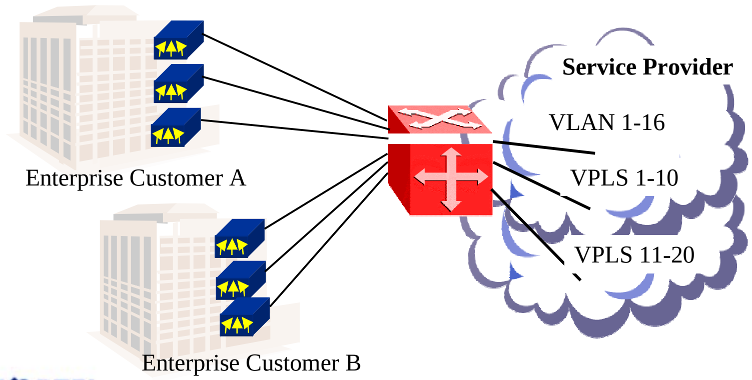
NORTEL NETWORKS

Outline

- Motivation
- FOQ architecture
- Performance
- Conclusion

Motivation: Application Addressed

Edge router providing services to Enterprise customers



NORTEL NETWORKS

Motivation: Switch Requirements

- **Port speeds: 10Mb/s to 10Gb/s**
- **Port count: 32x10G, 320x1G, etc**
- **QoS guarantees: similar to DiffServ**
  - Premium service: zero loss, minimal delay and delay variation
  - Tiered service: min bw guarantee (CIR), overdraft allowed (PIR) (similar to Frame Relay)
- **Fine-grained QoS: multiple flows per port**
  - Per-flow guarantee (VLAN, VPLS)
  - Multiple guarantees for each Enterprise customer
  - Potentially many flows per port (100s..1000s)

NORTEL NETWORKS

Motivation: Current architectures

Architectures to provide fine-grained QoS

- **Emulation of OQ with CIOQ, VOQ and Matching**
  [Stoica,Zhang'98, Chuang et al.'99, Charny et al.'98]
  - Matching: *$O(N^2)$* at each cell time, N=total number of ports
  - Or matching *$O(N)$*, space *$O(P)$*, time *$O(logP)$*, *P*=total no pkts in system
  - QoS guarantees highly accurate: O(pkt transmission time)<1us
  - High complexity and cost of implementation
- **Envelope switching [Kar et al'00]**
  - Packets grouped in envelopes
  - Matching *$O(N^2)$* at each envelope time
  - Lower accuracy, lower computational complexity
  - Complex 3-level scheduler, complex envelope management
- **All require policed inputs, no support for CIR/PIR tiered service**

NORTEL NETWORKS

FOQ - 5

Ideas

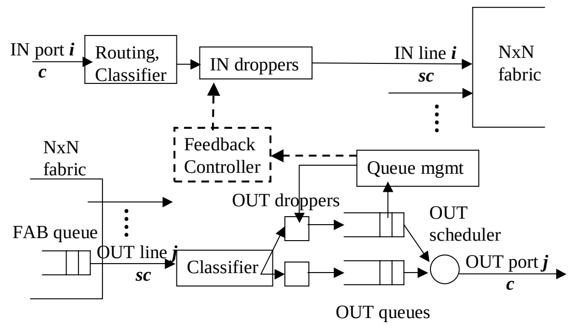
- **Accuracy of QoS guarantees needed by applications << provided by current architectures**
  - E.g.: 1-10ms for VoIP >> packet transmission time (1us or smaller)
- **No need to spend fabric resources to forward packets that get dropped at output**
- **Average congestion is likely to be "low" (<20%)**
  - Typical TCP loss rates below 10%

NORTEL NETWORKS

FOQ - 6

FOQ Switch Architecture

- **Switch: output buffered, small external speedup**
- **Fabric: internal speedup=N; minimal QoS support**
- **Per-flow OUT queues**



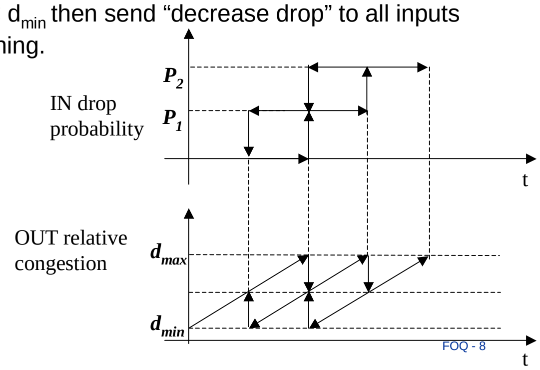
NORTEL NETWORKS

The Gear-box control algorithm

- Goal: keep fabric uncongested, no drops in fabric
- Define for each queue: RelCong(T) = 1-OutBytes(T)/InBytes(T)
- Observe: if RelCong<1-1/s for all queues, then fabric is not congested (s=fabric speedup)

Gearbox Algorihm:

- At each OUT queue, compute **relative congestion** over time T
- If RelCong > d_{max} then send "increase drop" to all inputs
- If RelCong < d_{min} then send "decrease drop" to all inputs
- Else, do nothing.



NORTEL NETWORKS

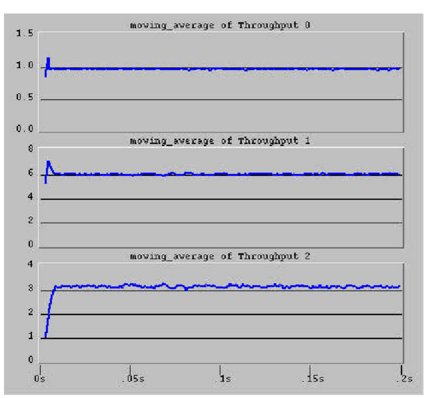
Performance: CBR Traffic

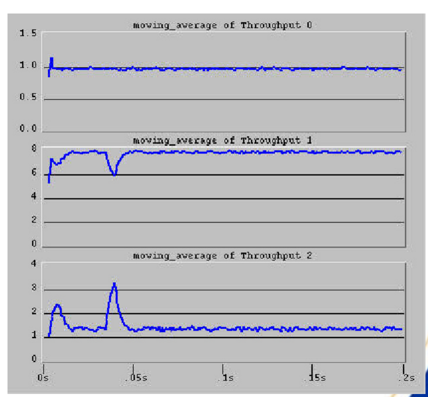
- **Description of experiment**
  - 16x10Gb/s switch
  - Fabric: memory 5MB, external speedup=1.28
  - Out queues: 2MB, Drop-tail
  - FOQ sampling rate T=1ms
  - dmin=0.02, dmax=0.17
- **QoS goals**
  - Offered Load: F0=0.952Gb/s, F1,F2=9.52Gb/s, converging to 10G port
  - F0: Premium service, F1>7.75Gb/s, F2>1.3Gb/s (6:1 ratio)

  Similar to Network World test of 2001

Performance: CBR Results

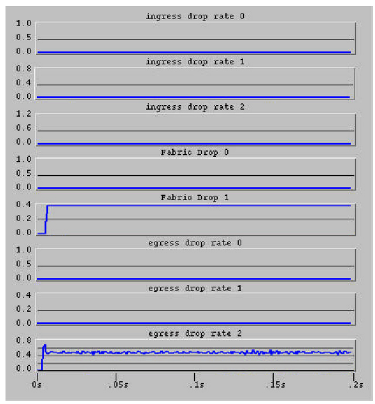
Throughput without FOQ **with FOQ**

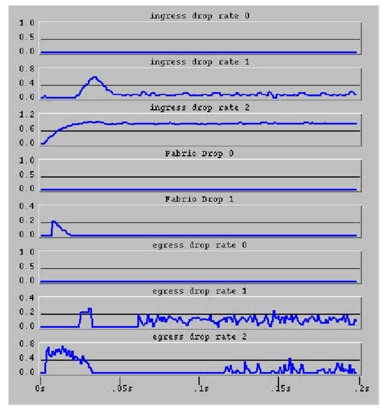




Performance: CBR Results

Drop rates without FOQ **with FOQ**





NORTEL NETWORKS

FOQ - 11

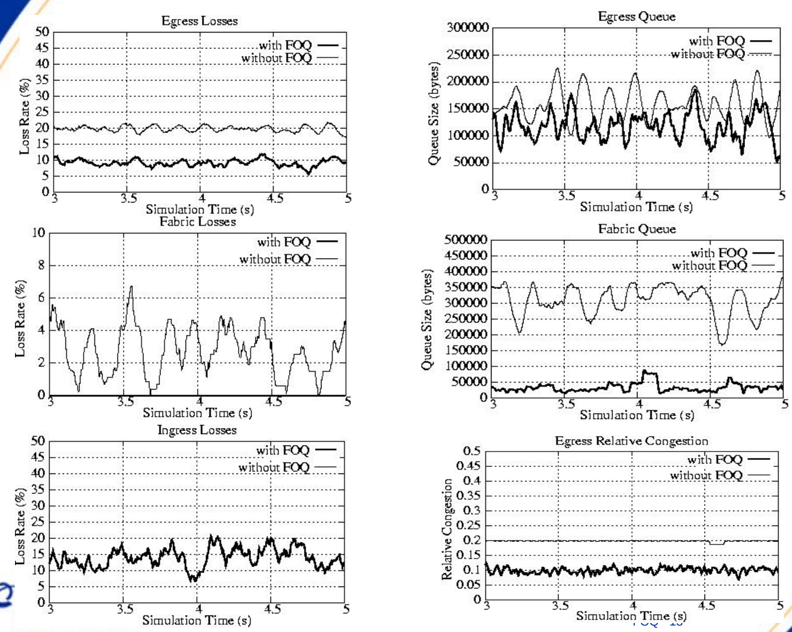
Performance: TCP traffic

- **Description of experiments**
  - 10Gb/s port
  - 45000 TCP micro-flows (Reno), one QoS flow (class)
  - 20ms propagation delay each way
  - Internal speedup s=1.25
  - Feedback interval T=1ms
  - dmin=0.02, dmax=0.17
  - Fabric buffer 400KB
  - Output buffer 320KB, running RED with min_th=80KB, max_th=240KB, max_P=0.4, w_q=0.003.
- **FOQ goal: no fabric drops, all QoS at ingress and egress queue**

NORTEL NETWORKS

FOQ - 12

Performance: TCP results



Complexity

- **Computational complexity**

  O(MN), N=ports, M=flows per port

  At each time interval T (typically 1ms)

- **Feedback communication bandwidth**

  B = MNF/T, F=bits per feedback indication

  Example: B=64Mb/s over entire system for M=1000, N=32, F=2, T=1ms.

Conclusion

- Simple architecture
- Low complexity, low implementation cost
- Can adapt to any traffic condition, no need for strict policing
- Provides fine-grained QoS guarantees
- Provides tiered service (CIR/PIR)
- Accuracy of QoS guarantees sufficient for any known QoS application

NORTEL NETWORKS

Acknowledements

- **Coauthors**

  Xiaohui Zhang, Emre Gunduzhan
- **Collaboration**

  Eric Haversat, Tom Holtey

  Nicolas Cristin, Univ. of Virginia, student intern

NORTEL NETWORKS