Internet2 Network Research Workshop

QoS Amplification Experiments

Kihong Park
Network Systems Lab
Dept. of Computer Sciences
Purdue University
park@cs.purdue.edu

http://www.cs.purdue.edu/nsl

Network Systems Lab

Overview

Goal Achieve QoS amplification over imperfect network service substrate

→ end-to-end control

- End-to-end QoS amplification techniques
  - Multiple time scale traffic control
  - Adaptive redundancy control
  - Adaptive label control
- Internet2 experiments
  - Local environment
  - Internet2 requirements
  - Proposed WAN experiments and collaborations

Outline

Network Systems Lab

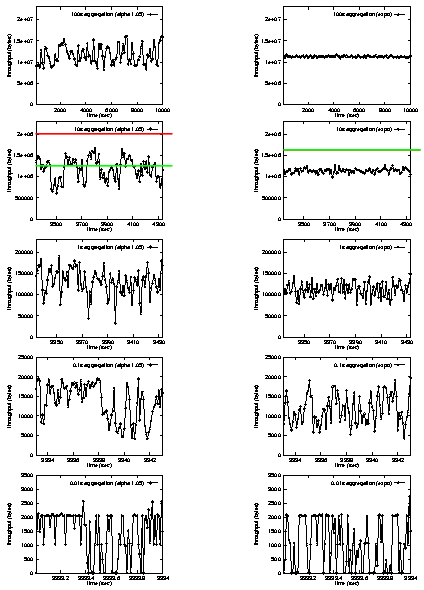
Multiple Time Scale Traffic Control

Self-similar Network Traffic

- Data traffic is fundamentally different from telephony traffic (Leland *et al.* '93)
  - → self-similar or long-range dependent
- Causality
- Performance Impact
- Control

Self-similar Network Traffic and Performance Evaluation, Park and Willinger (eds.), Wiley-Interscience, 2000

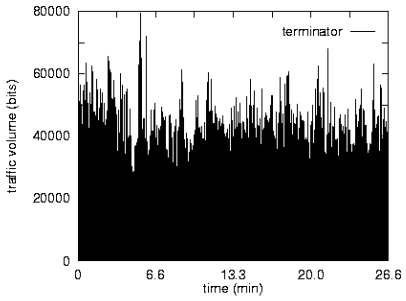
Network Systems Lab



Network Systems Lab

Multiple Time Scale (cont.)

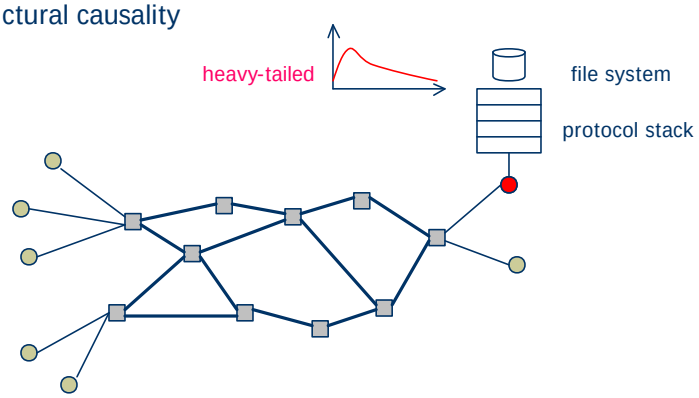
- Causality
  - Single-source causality (e.g., MPEG video)



Network Systems Lab

Multiple Time Scale (cont.)

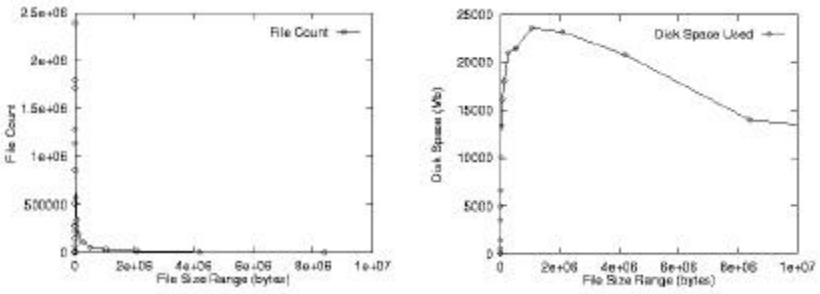
- Structural causality



Network Systems Lab

Multiple Time Scale (cont.)

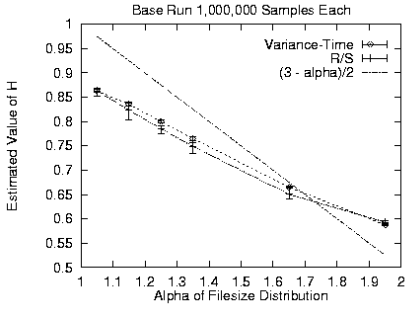
- Structual causality (cont.)



→ UNIX file system (G. Irlam)

Multiple Time Scale (cont.)

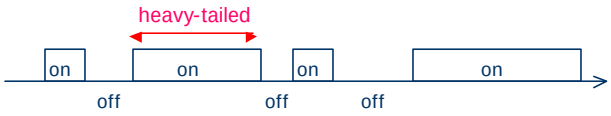
- Structural causality (cont.)



→ impervious to “details”

Multiple Time Scale (cont.)

- Structural causality (cont.)



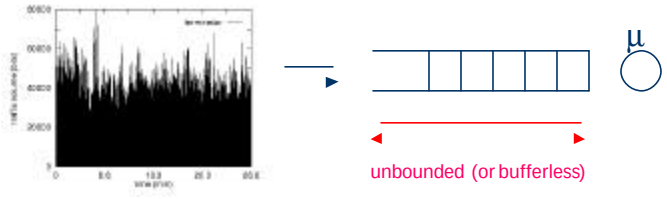
  - on/off traffic (0/1 reward renewal process)
  - asymptotic second-order self-similarity

- Two principal traits
  - Invariant correlation structure across multiple time scales
  - Correlation at a distance (long-range dependence)

Network Systems Lab

Multiple Time Scale (cont.)

- Detrimental performance impact: queueing

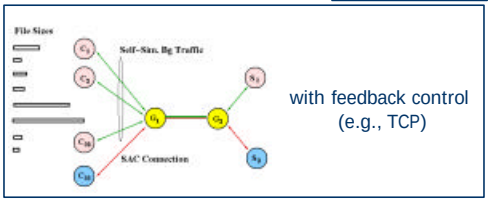


- polynomial (vs. exponential) queue length distribution
- infinite memory/asymptotic analysis

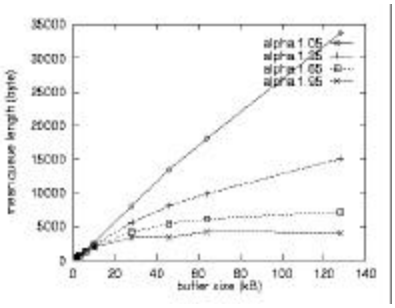
Network Systems Lab

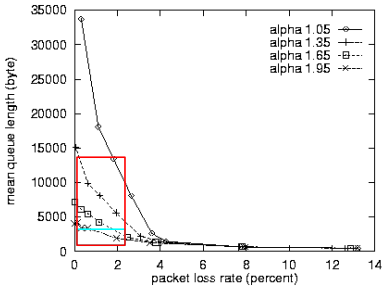
Multiple Time Scale (cont.)

- Empirical validation



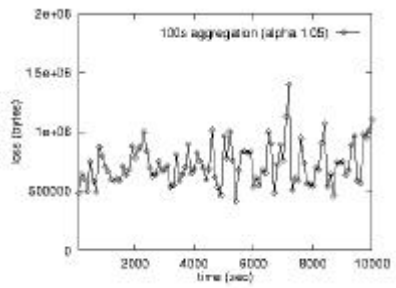
with feedback control (e.g., TCP)

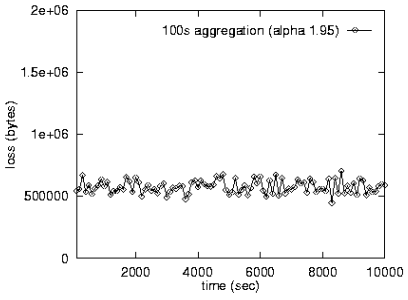




Multiple Time Scale (cont.)

- Importance of second-order performance measures
  - → e.g., jitter





- concentrated periods of over- and under-utilization
- bufferless queueing does not help

Multiple Time Scale (cont.)

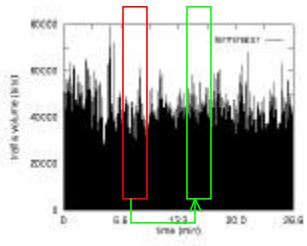
- Impact of long-range structure can be curtailed
  - extreme: bufferless queueing
  - time horizon implied by finite memory
  - short-range correlation can dominate
- Small buffer/large bandwidth resource provisioning policy
  - statistical multiplexing
  - central limit theorem

Network Systems Lab

Multiple Time Scale Traffic Control (cont.)

Traffic Control

- Premise: exploit long-range correlation for traffic control
  - correlation/predictability structure at large time scales

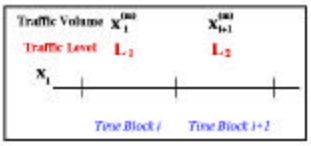


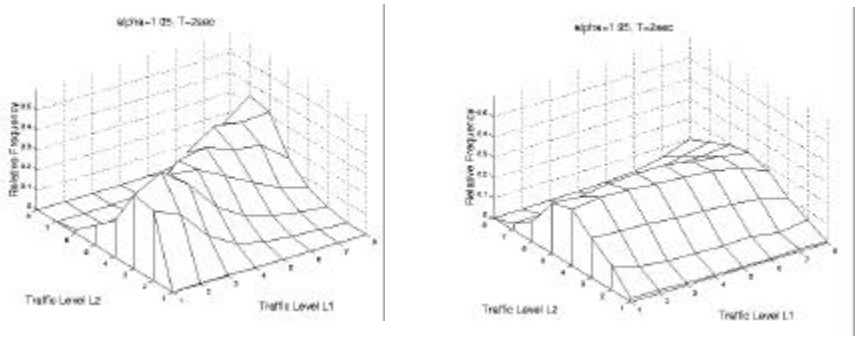
  - relevant in broadband WANs with high delay-bandwidth product

Network Systems Lab

Multiple Time ScaleTraffic Control (cont.)

Large time scale predictability:

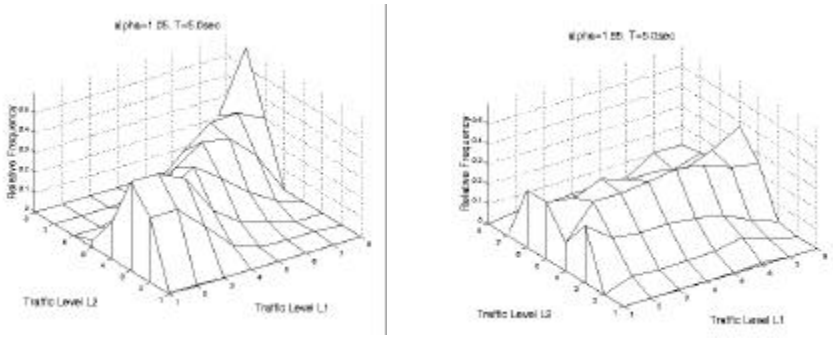




Network Systems Lab

Multiple Time Scale Traffic Control (cont.)

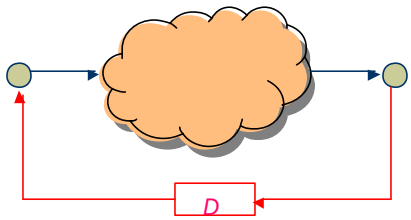
Large time scale predictability (5 sec):

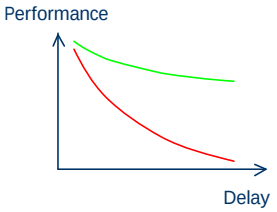


Network Systems Lab

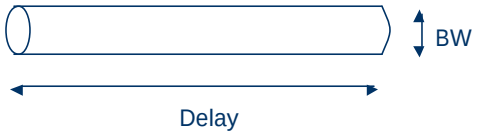
Multiple Time Scale Traffic Control (cont.)

- Implications: mitigate reactive cost of feedback control





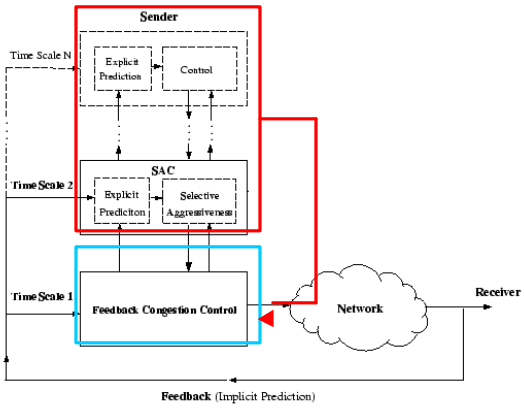
LRD time scale » RTT



Network Systems Lab

Multiple Time Scale Traffic Control (cont.)

Multiple time scale traffic control:



Network Systems Lab

Multiple Time Scale Traffic Control (cont.)

Application domains:

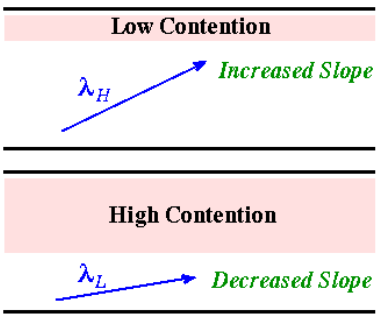
- Bulk data transport – congestion control
  - → throughput maximization (TCP-MT)
- Real-time data transport – adaptive redundancy control
  - → end-to-end QoS (AFEC-MT)

Network Systems Lab

Multiple Time Scale Traffic Control (cont.)

Congestion control: TCP and rate-based

Idea:

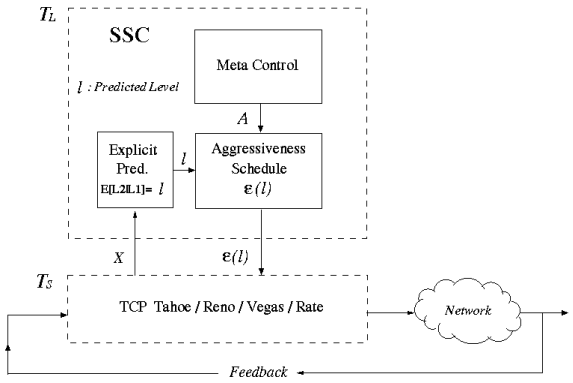


→ modulate slope of linear increase phase in AIMD

Network Systems Lab

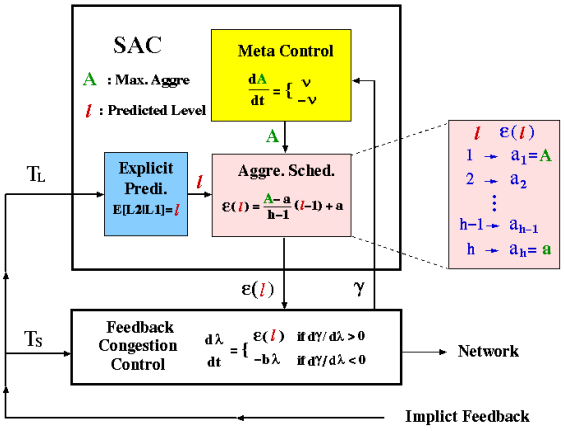
Multiple Time Scale Traffic Control (cont.)

- Multiple time scale TCP (TCP-MT):



Multiple Time Scale Traffic Control (cont.)

- Multiple time scale rate-based congestion control:

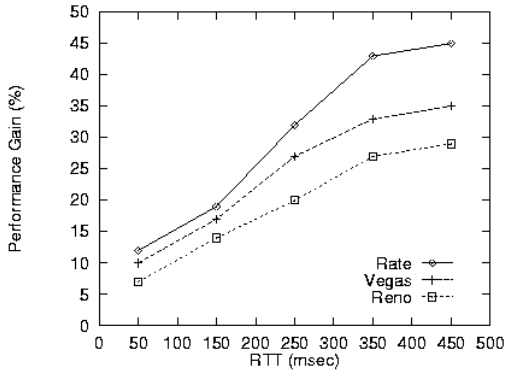


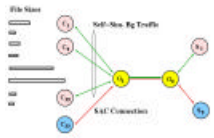
Multiple Time Scale Traffic Control (cont.)

- TCP-MT: performance gain as function of RTT

TCP-MT

TCP

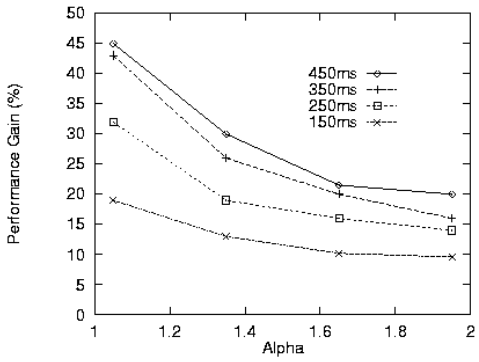




Network Systems Lab

Multiple Time Scale Traffic Control (cont.)

- TCP-MT: performance gain as function of self-similarity



Network Systems Lab

Adaptive Redundancy Control

Real-time traffic transport

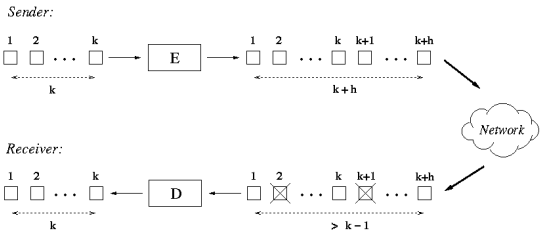
- Achieve invariant end-to-end QoS
- User-specified QoS
- ARQ infeasible (RTT & timeliness)
- Packet-level FEC
  - → proactive QoS protection
- Purely end-to-end (black box network)
- MPEG video/audio implementation (UDP)

Network Systems Lab

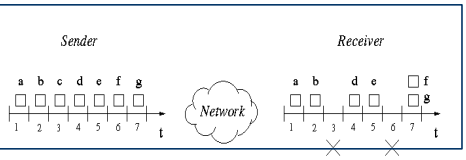
Adaptive Redundancy Control (cont.)

Adaptive redundancy control (AFEC):

FEC:



hit rate:

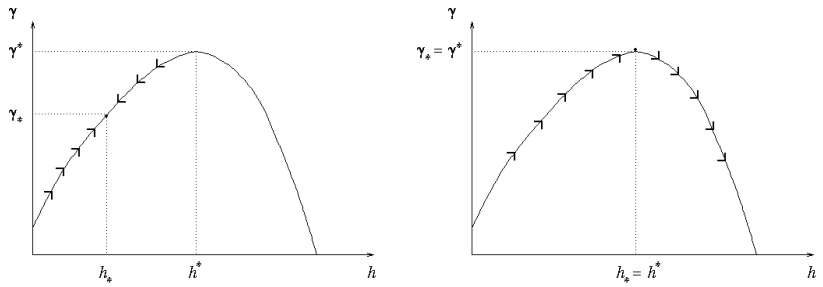


$0 \le \gamma \le 1$

Network Systems Lab

Adaptive Redundancy Control (cont.)

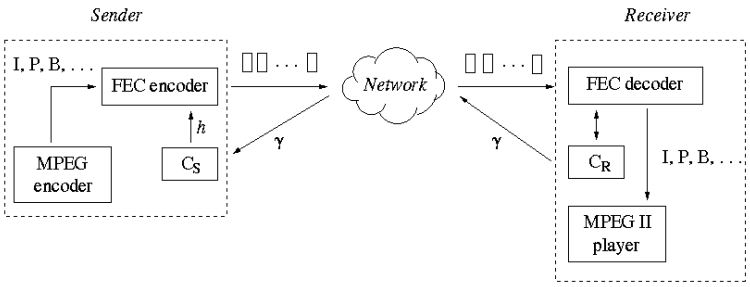
- Redundancy-recovery relation:



→ stability & optimality

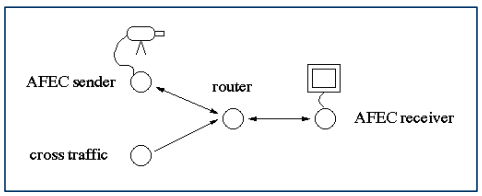
Adaptive Redundancy Control (cont.)

- AFEC structure:



Adaptive Redundancy Control (cont.)

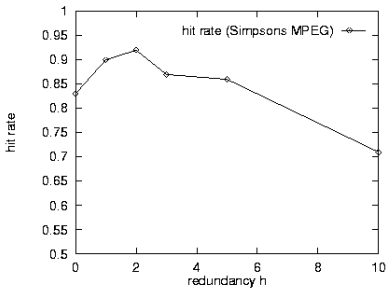
- Experimental set-up:

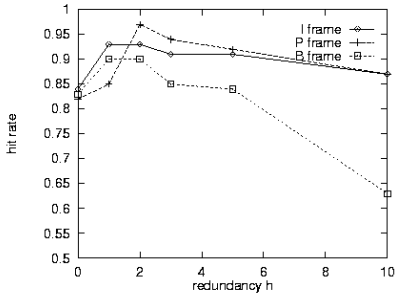


  - UltraSparc 1 & 2, SGI, x86
  - Solaris UNIX, Windows NT
  - Optibase, Futuretel MPEG I & II compression boards
  - Sony DCR-VX 1000, Panasonic F250

Adaptive Redundancy Control (cont.)

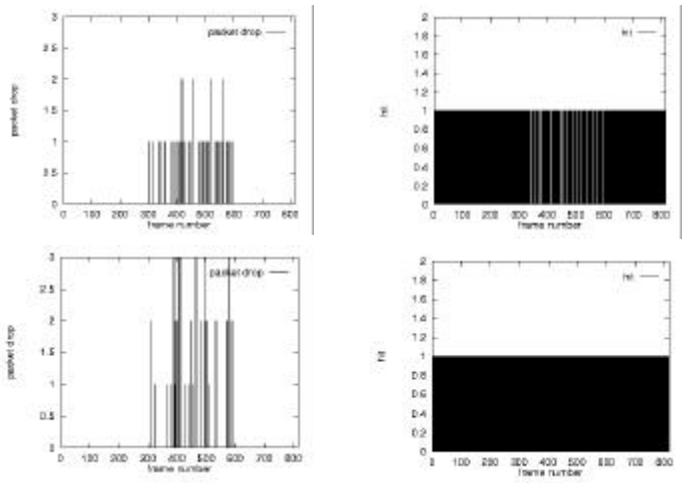
- Impact of redundancy: Static FEC





Adaptive Redundancy Control (cont.)

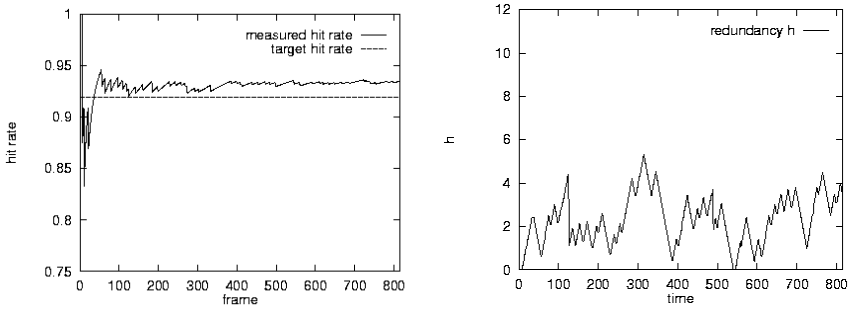
- Adaptive FEC vs. static FEC



Network Systems Lab

Adaptive Redundancy Control (cont.)

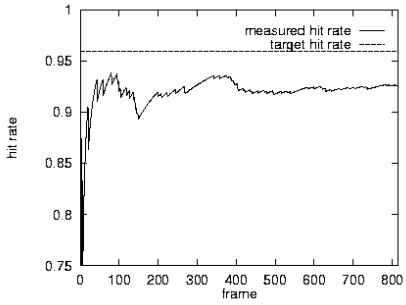
- Stable target QoS: symmetric control

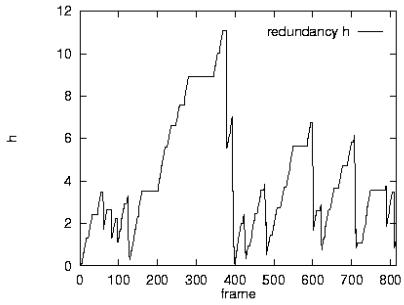


Network Systems Lab

Adaptive Redundancy Control (cont.)

- Unstable target QoS: asymmetric control

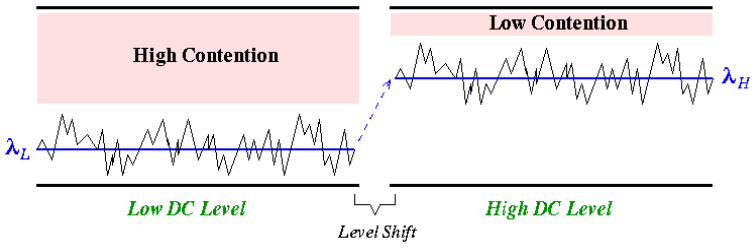




Network Systems Lab

Adaptive Redundancy Control (cont.)

- Multiple time scale redundancy control

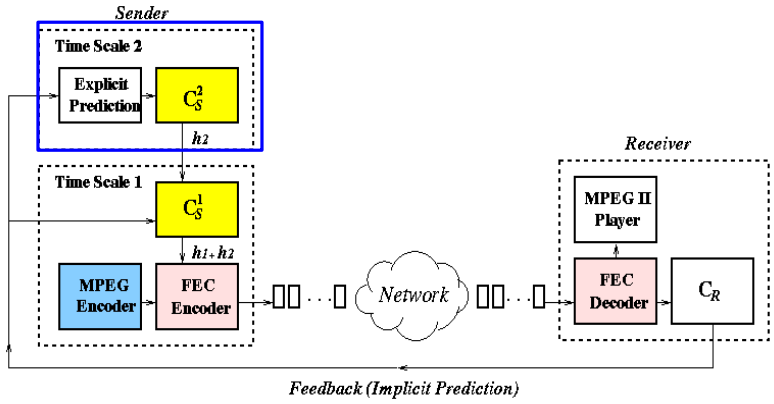


→ level control

Network Systems Lab

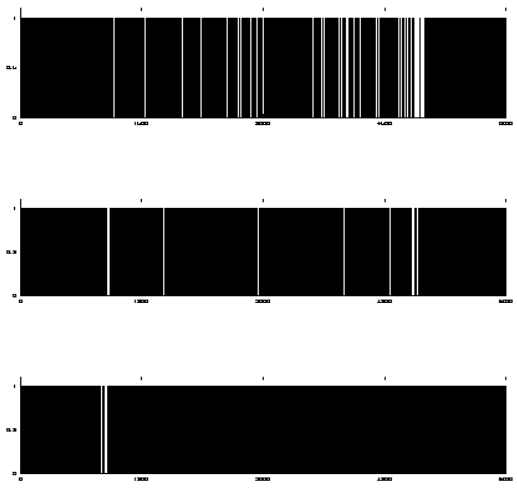
Adaptive Redundancy Control (cont.)

- AFEC-MT structure:



Adaptive Redundancy Control (cont.)

- AFEC-MT:



Adaptive Redundancy Control (cont.)

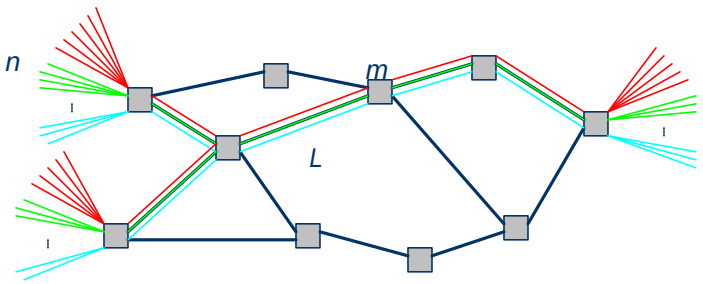
- Principal performance effect:
  - impart proactivity above and beyond AFEC
  - proactivity of reactive control in broadband WANs
  - mitigate reactive cost

predictability at time scales exceeding RTT imparts timeliness

Network Systems Lab

Adaptive Label Control

Differentiated services network:



n users » *L* labels (colors) ≥ *m* classes

Network Systems Lab

Adaptive Label Control (cont.)

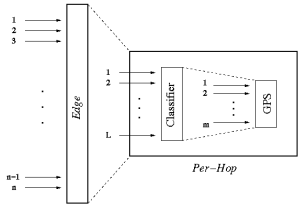
Questions:

- What is a “good” (optimal) per-hop control?
  - → optimal aggregate-flow per-hop behavior

- What is a “good” (optimal) edge control?

Adaptive Label Control (cont.)

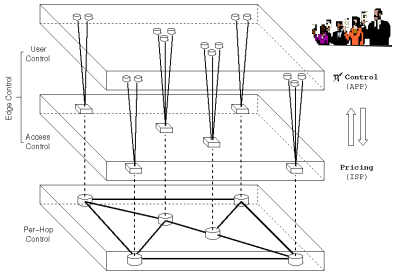
- What is the loss of power due to aggregation?
  - → $n \gg L \geq m$
  - → loss of resolution vis-à-vis per-flow switching



- What is the impact of finite, discrete label set {1, 2, ..., L}?
  - → $\eta \in \mathbf{Z}_+, \mathbf{R}_+, [0,1]$, or $\mathbf{R}_+^S$

Adaptive Label Control (cont.)

- What is the system dynamics when driven by selfish users?
  - → end-to-end label control
  - → stability (Nash equilibria) and efficiency (system optimality)



- What is the impact of selfish service provider (ISP)?

Network Systems Lab

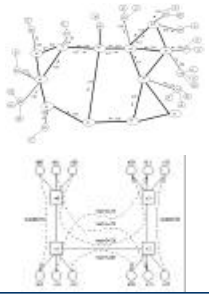
Adaptive Label Control (cont.)

Theory

- optimal PHB
  - – differentiation/shaping
  - – efficiency
- adaptive label control
- selfish users
- selfish service provider
- performance analysis

Simulation

QSim: WAN QoS Simulator



Implementation

Purdue Infobahn

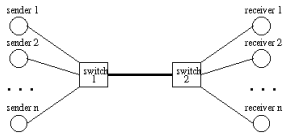
Cisco 7206 VXR IP-over-SONET QoS Testbed

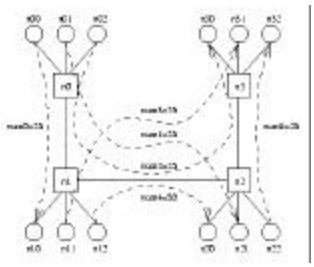
Network Systems Lab

Adaptive Label Control (cont.)

Set-up:

- QSim: *ns* based WAN QoS simulation environment

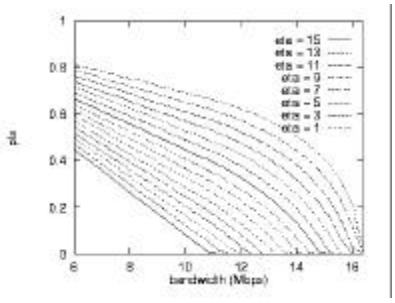


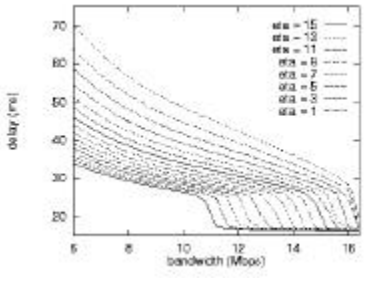


Network Systems Lab

Adaptive Label Control (cont.)

- Structural: bottleneck BW, L = 16 (m = 16)

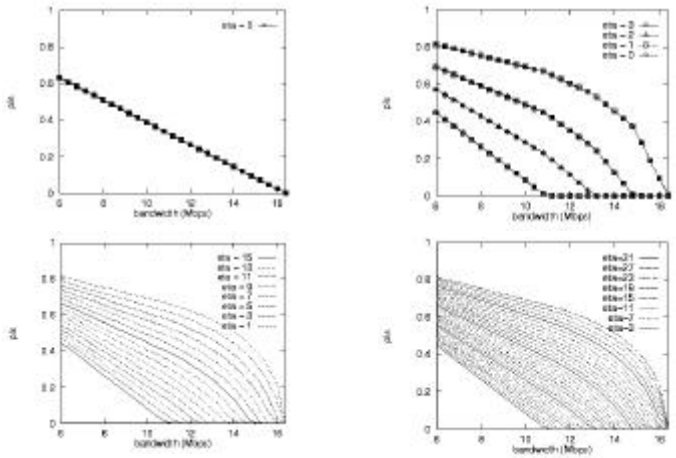




Network Systems Lab

Adaptive Label Control (cont.)

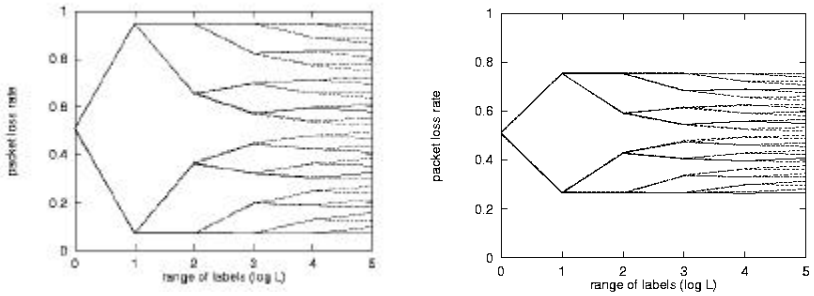
- Structural: L = 1, 4, 16, 32



Network Systems Lab

Adaptive Label Control (cont.)

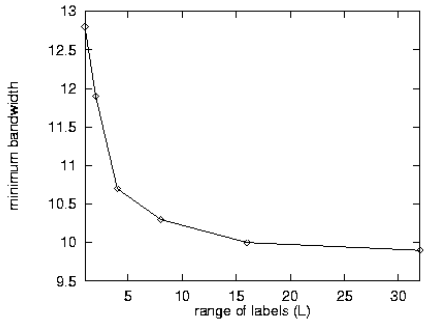
- Structural: log L = 0, 1, 2, 3, 4, 5 (bits)



Network Systems Lab

Adaptive Label Control (cont.)

- Structural: system optimal BW requirement

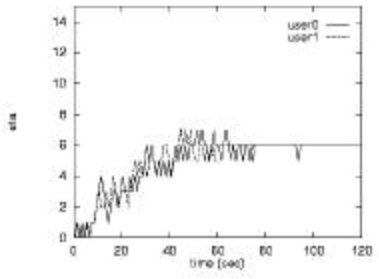


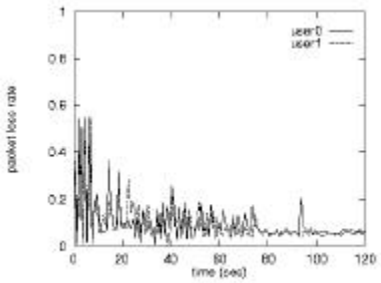
Network Systems Lab

Adaptive Label Control (cont.)

- Dynamical: adaptive label control (end-to-end)

  → reachability

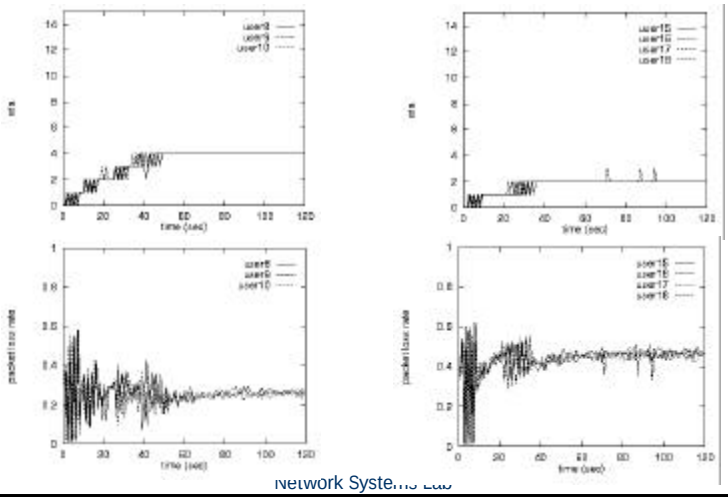




Network Systems Lab

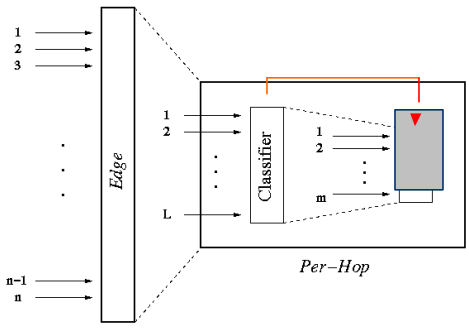
Adaptive Label Control (cont.)

- Dynamical: adaptive label control (cont.)



Adaptive Label Control (cont.)

Optimal aggregate-flow per-hop control:



$n \gg L \geq m$

→ n users, L labels, and m service classes

Adaptive Label Control (cont.)

- Of interest: $n \gg L \geq m$
- Special case: $n = m$
  - per-flow per-hop control
- Of special interest: $L = m$
  - as many service classes as label values

Optimality I: service differentiation/shaping

Network Systems Lab

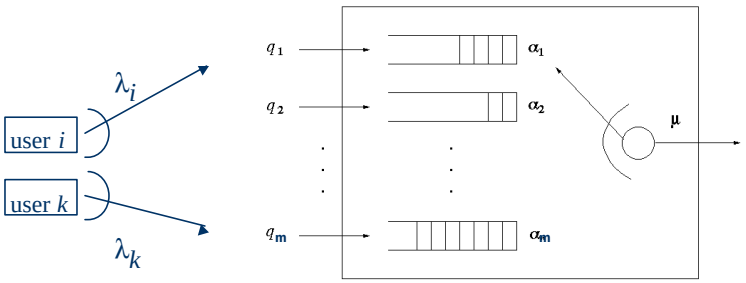
Adaptive Label Control (cont.)

- Per-flow Control ($n = m$):
  - Label value η viewed as "code" of user requirement
    - e.g., 1.5 Mbps, relative share of link bandwidth, etc.
  - If infinite resources, then no interaction/coupling
    - e.g., INDEX
  - In resource-bounded systems, $\exists$ coupling (externality)

Network Systems Lab

Adaptive Label Control (cont.)

- Illustration of coupling in simple single switch case:



GPS switch

Adaptive Label Control (cont.)

- INDEX (Varaiya et al.)

Platinum Service	BW_1	$Price_1$
Gold Service	BW_2	$Price_2$
Silver Service	BW_3	$Price_3$
Bronze Service	BW_4	$Price_4$

→ service class: volume insensitive
→ infinite resoures
→ no externality

Adaptive Label Control (cont.)

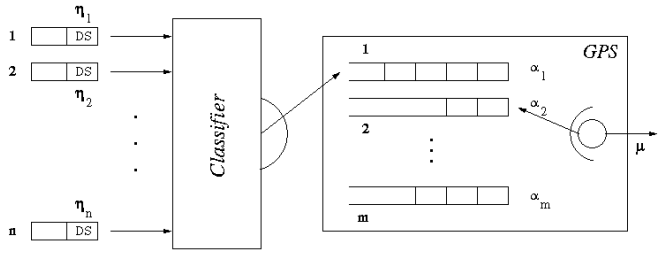
- Assume label set is metric space (totally ordered)
  - → e.g., Euclidean distance (L_2 norm)
  - → e.g., $\eta = 1 < 2 < \ldots < L$
- Mean square measure of goodness:

  Given η, find resource configuration v s.t.

$$\min_{v} \quad \sum_{i=1}^{n} (h_i - v_i)^2$$

Adaptive Label Control (cont.)

- GPS: $\varpi_i = \alpha_i / \lambda^i$



$\eta_i \in \{1,2,\ldots,L\}$; $\xi : \{1,\ldots,L\} \rightarrow \{1,\ldots,m\}$

Adaptive Label Control (cont.)

- Normalization: $\frac{h_i - h_{\min}}{h_{\max} - h_{\min}} \in [0,1]$
- Solution: $a_i = (1-u)\frac{l^i h^i}{\sum_k l^k h^k} + u\frac{l^i}{\sum_k l^k}$

Network Systems Lab

Adaptive Label Control (cont.)

- Optimal aggregate-flow classifier:

  Given η, find resource configuration v s.t.

  $$\min_{v} \sum_{i=1}^{n} (h_i - v_i)^2$$

- Optimal solution:

  Reduce to per-flow optimal solution

  → optimal clustering problem

Network Systems Lab

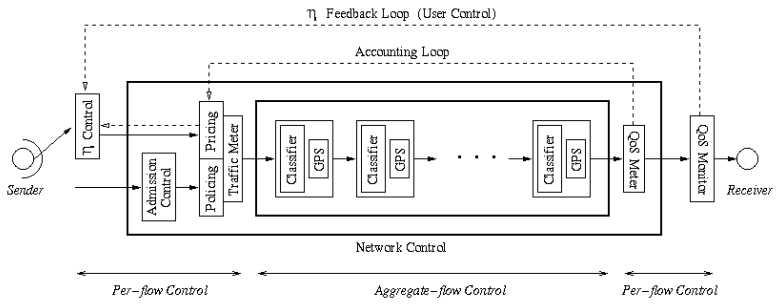
Adaptive Label Control (cont.)

- Properties (A1), (A2), and (B)
  - (A1) If η_i increases, then QoS of user i improves
  - (A2) If η_i increases, then QoS of user j deproves
  - (B) If $\eta_i \geq \eta_j$ then QoS of user i is better than QoS of user j
- Optimal per-flow classifier satisfies (A1), (A2), (B)
- Optimal aggregate-flow classifier with $L = m$ satisfies (A1), (A2), (B)

Network Systems Lab

Adaptive Label Control (cont.)

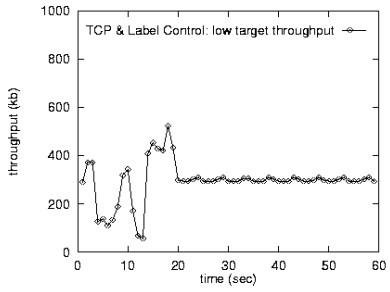
- End-to-end QoS control:

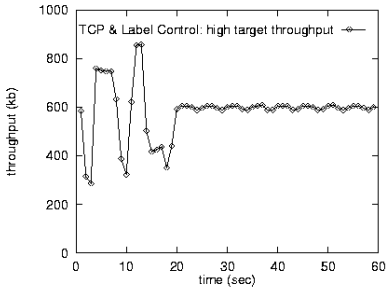


Network Systems Lab

Adaptive Label Control (cont.)

- Integrated QoS control:
  - e.g., TCP over adaptive label control





Network Systems Lab

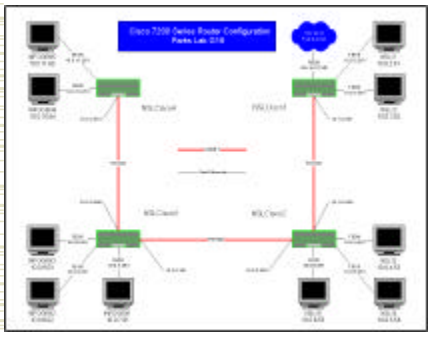
Local Environment: Network Systems Lab

- Purdue Infobahn QoS testbed: 4+5 Cisco 7206 VXR routers
  - IP-over-SONET backbone
  - custom classifier implementation in IOS
- NSF vBNS and Abilene connectivity (DS-3)
  - Purdue vBNS/Internet2 Advisory Committee
- Fore ATM, FastEthernet switches

Network Systems Lab

Local Environment: NSL (cont.)

Purdue Infobahn



Network Systems Lab

Local Environment: NSL (cont.)

- Real-time MPEG I & II video/audio compression engines
  - → Optibase, Futuretel (Windows NT)
- Video/audio capture equipment
- 35+ Sun/Intel/SGI workstations & PCs
- Prototype software systems: UNIX, Windows NT

Network Systems Lab

Proposed WAN Experiments

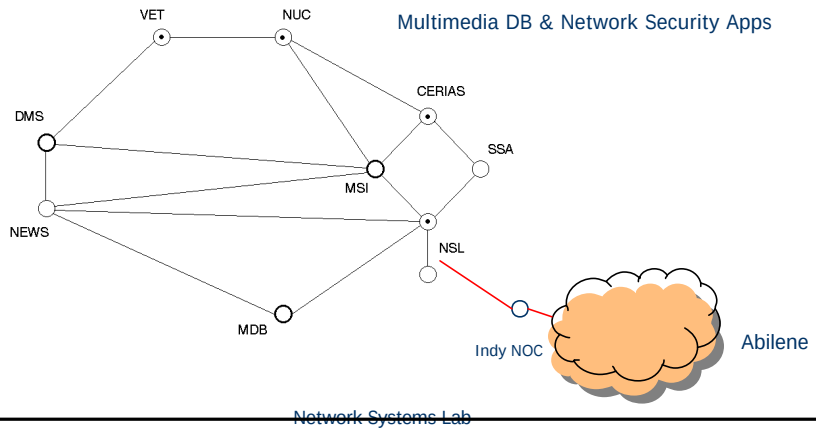
Performance Evaluation and Benchmarking

- Internet2 benchmarking of
  - Multiple time scale traffic control (TCP-MT, AFEC-MT)
  - Adaptive redundancy control (AFEC)
  - Adaptive label control (Diff-Serv router support)
    - → vBNS/Abilene
- Commodity Internet benchmarking
- Evaluate effectiveness of end-to-end QoS amplification
    - → model of future Internet (NGI)

Network Systems Lab

Proposed WAN Experiments (cont.)

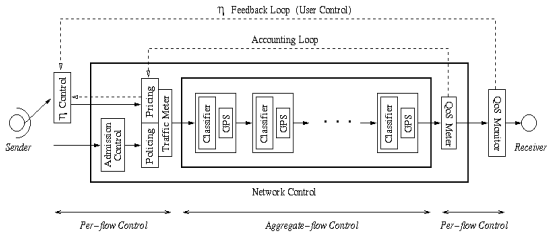
- Integration with Purdue Infobahn & QoS peering



Network Systems Lab

Proposed WAN Experiments (cont.)

- Pricing, accounting, and access control



- Incremental optimal aggregate-flow classifier deployment (Cisco IOS)
- IP-over-ATM, IP-over-SONET, IP-over-? issues

Collaborations

- Academic:
  - Boston Univ. (A. Bestavros)
  - Ohio State Univ. (J. Hou)
  - Santa Fe Institute (Fellow-at-Large)
  - Univ. of Wisconsin (P. Barford; WAWM)
  - Seoul National Univ. (S. Bahk)
- Industry/Research Labs:
  - AT&T Research (W. Willinger)
  - Cisco (F. Baker)
  - Sprint (K. Metzger)

Internet2 Requirements

- Multi-point channel set-up, resource reservation
- Bottleneck configuration
- Congestion susceptibility
  - → traffic generators and packet drops
- Diff-Serv support
- Network monitoring & management
- Modified router software deployment (partial)
- Member institution benchmark participation
  - → scale

Network Systems Lab

Acknowledgments & More Info

- Supported by:
  - NSF ANI-9714707, ANI-9875789 (CAREER), ESS-9806741, EIA-9972883; ANI-9729721 (vBNS)
  - Purdue Research Foundation
  - Santa Fe Institute
  - Sprint
  - CERIAS, SERC

- Research assistants & postdocs:
  - RAs: A. Balakrishnan, S. Chen, J. Cruz, G. Nalawade, H. Ren, M. Tripunitara, T. Tuan, W. Wang
  - Postdocs/visting scientists: S. Bahk, H. Lee, J. Park, W. Zhao
- Network Systems Lab
  - **http://www.cs.purdue.edu/nsl**

Network Systems Lab

Acknowledgments & More Info (cont.)

- Related publications:
  - Chen & Park. An architecture for noncooperative QoS provision in many-switch systems. In *Proc. IEEE INFOCOM*, 1999.
  - Cruz & Park. Towards performance-driven system support for distributed computing in clustered environments. *Journal of Parallel and Distributed Computing*, 1999.
  - Park & Tuan. Performance evaluation of multiple time scale TCP under self-similar traffic conditions. *ACM Trans. on Modeling and Computer Simulation*, 2000.
  - Park & Wang. QoS-sensitive transport of real-time MPEG video using adaptive forward error correction. In *Proc. IEEE Multimedia Systems*, 1999.
  - Park & Willinger. *Self-Similar Network Traffic and Performance Evaluation*. Wiley-Interscience, 2000.
  - Ren & Park. Toward a theory of differentiated services. In *Proc. IEEE/IFIP IWQoS*, 2000.
  - Ren & Park. Efficient shaping of user-specified QoS using aggregate-flow control. In *Proc. International Workshop QofIS*, Lectures Notes in Computer Science, 2000.
  - Tuan & Park. Multiple time scale congestion control for self-similar network traffic. *Performance Evaluation*, 1999.
  - Tuan & Park. Multiple time scale redundancy control for QoS-sensitive transport of real-time traffic. In *Proc. IEEE INFOCOM*, 2000

Network Systems Lab