



On Generalized Processor Sharing with Regulated Traffic for MPLS Traffic Engineering

Shivendra S. Panwar

New York State Center for Advanced Technology in Telecommunications (CATT)

Department of Electrical and Computer Engineering

Polytechnic University, Brooklyn, NY

<http://catt.poly.edu/CATT/panwar.html>

GPS Background



- Two important QoS mechanisms
 - Deterministic traffic shaping: leaky bucket
 - Generalized processor sharing (GPS)
 - Available in most commercial router/switches
- GPS
 - A work conserving scheduling discipline: efficient
 - Each class is guaranteed a minimum rate: isolating the classes
 - Assumes fluid traffic: many discrete approximations available



Related Work on GPS Queues



- Deterministic analysis [Parekh'93, Parekh'94]
 - Leaky bucket regulated sources: **easy to enforce**
 - Worst case analysis: **low utilization**
- Stochastic analysis:
 - [Zhang'95]: Exponential bounded burstiness sources
 - [Lo Presti'96]: Markov Modulated Fluid Processes
 - [Borst'99]: Long-tailed traffic sources
 - [Pereira'01]: Long-range dependent (LRD) traffic sources
 - [Mannersalo'02]: Gaussian traffic sources
 - **Efficient in utilizing network resources**
 - But the traffic models are **hard to measure or enforce**
- Effective envelope [Boorstyn'00]



MPLS TE Queues for QoS Routing



- Joint work with Yihan Li and C.J. (Charlie) Liu (AT&T Laboratories)
- Outline
 - MPLS Traffic Engineering (TE)
 - Motivation
 - MPLS TE queues
 - Performance analysis with GPS model

MPLS Traffic Engineering

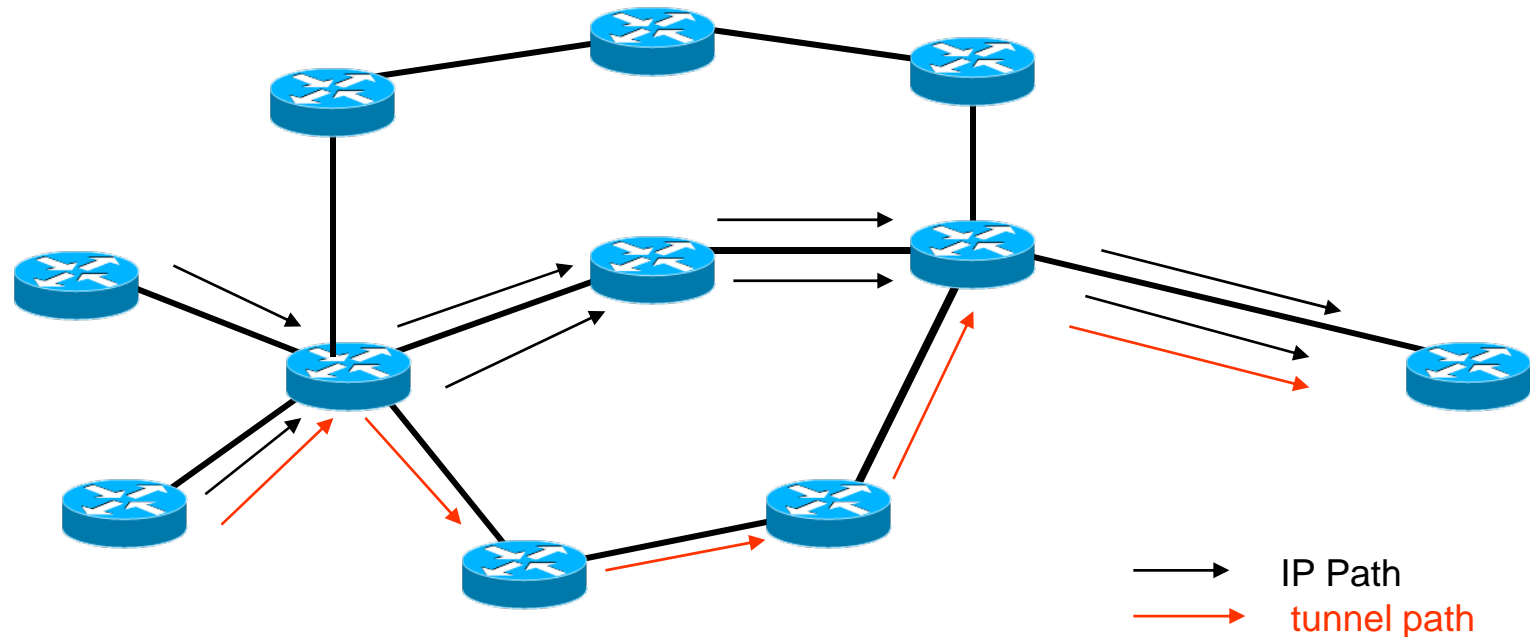


- Multi-Protocol Label Switching (MPLS)
 - assigns short labels to network packets that describe how to forward them through the network.
 - Is independent of any routing protocol.
 - provides a mechanism for engineering network traffic patterns.
- MPLS Traffic Engineering (TE)
 - OSPF always chooses the shortest path, which may be over used and congested.
 - MPLS TE allows path selection without adjusting link OSPF cost, so that flows can be moved from congested links to alternate links with larger costs.

Deficiency in IP Routing



- IP traffic takes shortest path destination based routing.
- Shortest path may not be the only path, or the best path.
- Alternate paths may be under-utilized while the shortest path is over-utilized.



Motivation

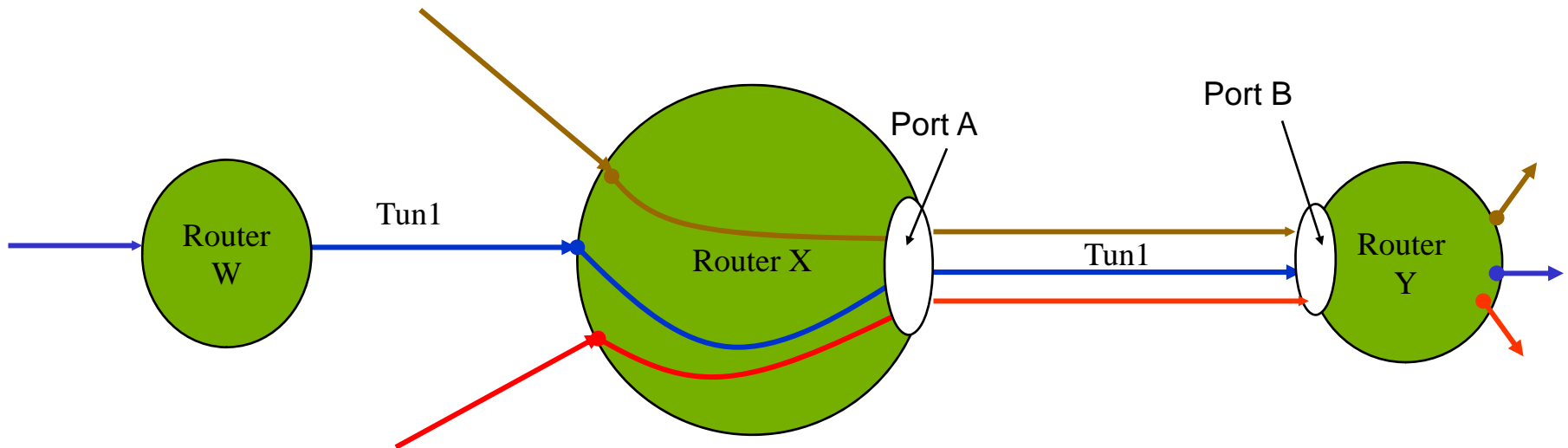


- Admission control mechanism is available only at tunnel setup time, but not at packet sending time.
 - Bandwidth reservation is policed only at tunnel setup time to limit the number of tunnels traversing a given link.
 - At packet sending time, tunnel traffic has to compete with all other traffic either in another tunnel or non- tunnel traffic.

MPLS TE Queue (1)



- Tun1 traffic competes with non-tunnel traffic for BW available in link X-Y.
- If RSVP reserved BW does not receive preferential treatment over other traffic, mixing of tunnel traffic and non tunnel traffic.
- **If there is preferential treatment,**
 - Router X recognizes tunnel packets by label associated with the packets.
 - Create MPLS TE Queue for tunnel traffic.
 - Packets in MPLS TE queue take priority over packets in any other non- TE queues.

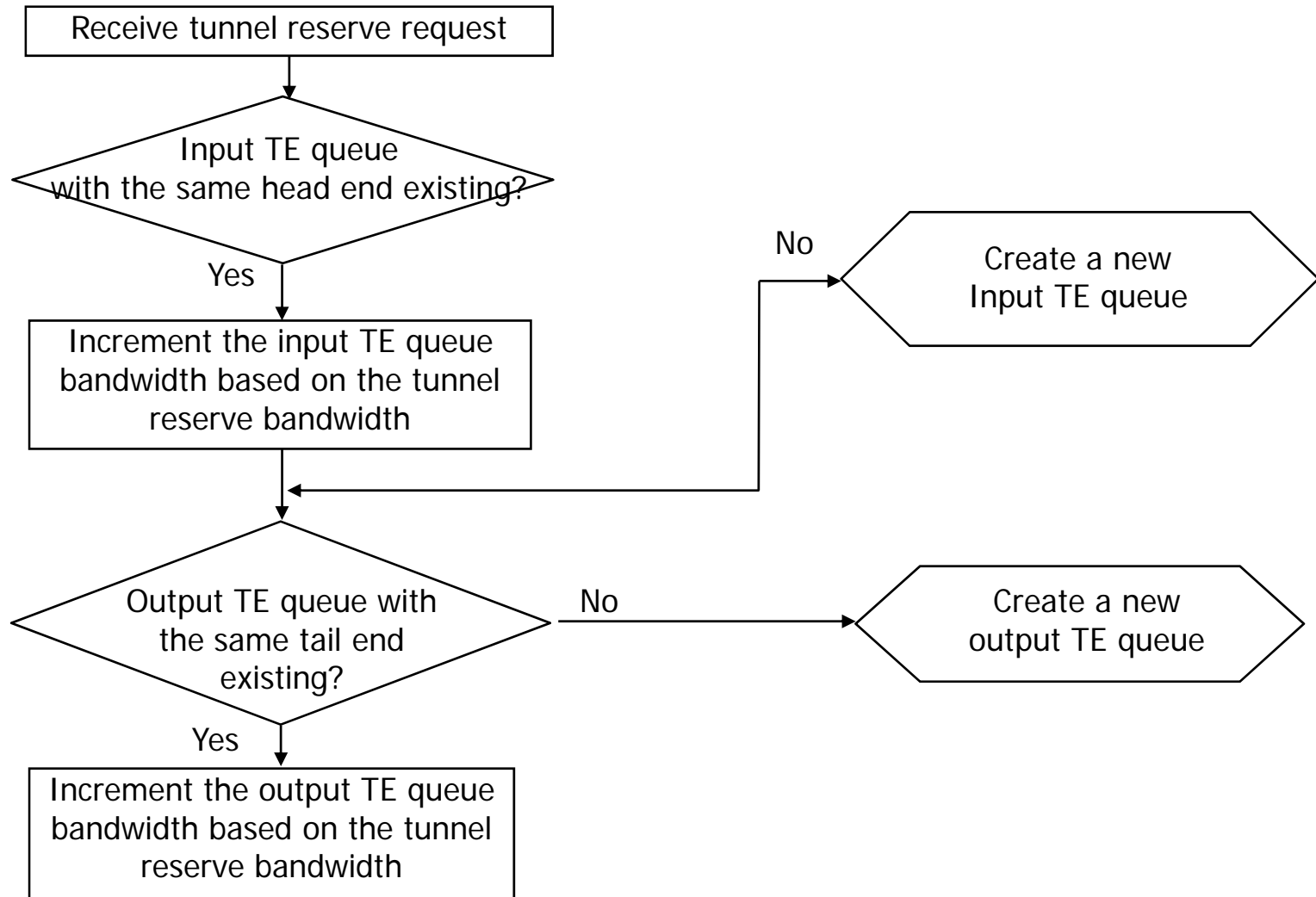


MPLS TE Queue (2)

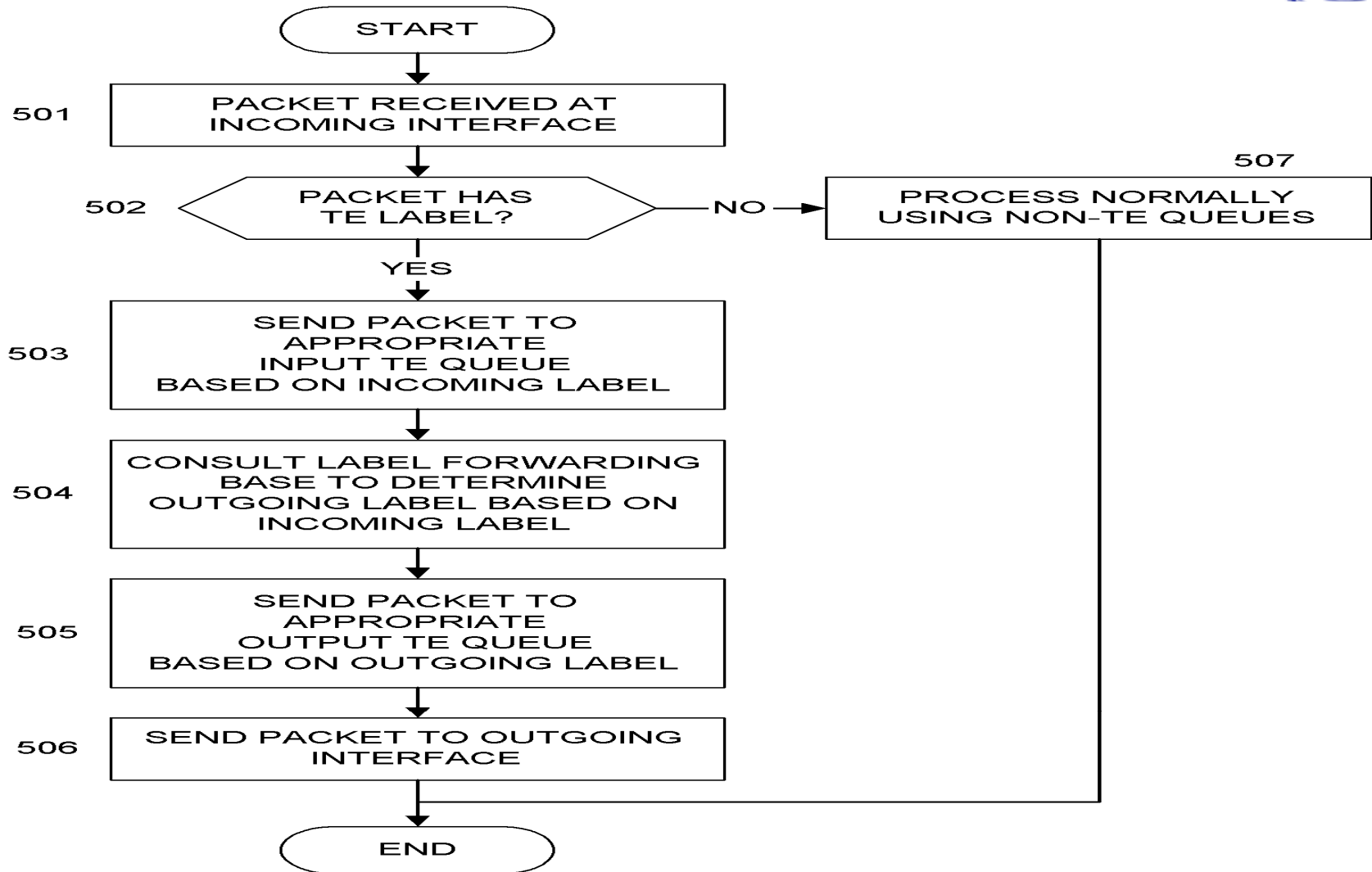


- MPLS TE Queues are created to ensure tunnel traffic's priority over non-tunnel traffic.
 - **Only real time traffic will be sent into MPLS TE tunnels via policy routing.**
 - **Input TE queues are shared by tunnels with the same head end.**
 - **Output TE queues are shared by tunnels with the same tail end.**
- Tunnel packets are identified by the label associated with the packets, and sent to a TE queue based on its label.
- This is a new idea to enable scalable MPLS TE Tunnels deployment with QoS guarantee for real time traffic in Service Provider's network.

MPLS TE Queues Creation



Switching Process for Tunnel Packets



Performance Analysis

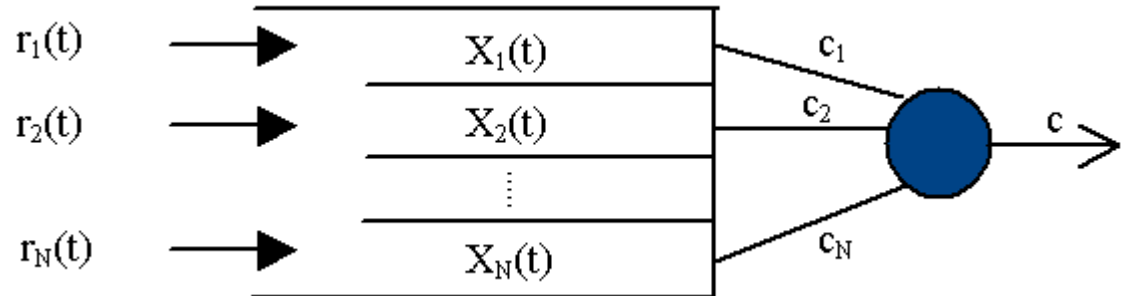


- An output switch with output TE queues is considered.
- We analyzed the process from the time packets enter output TE queues to the time they are forwarded to the next hop.
- Assume that each traffic source can be modeled as a **continuous-time Markov process**.
- The system is analyzed and simulated as a **Generalized Processor Sharing (GPS)** system.
 - S. Mao and S. Panwar, "The Effective Bandwidth of Markov Modulated Fluid Process Sources with a Generalized Processor Sharing Server," Globecom 2001, pp. 2341-2346.
- Using TE queues leads to lower overflow probabilities for TE tunnel traffic.

The System Model



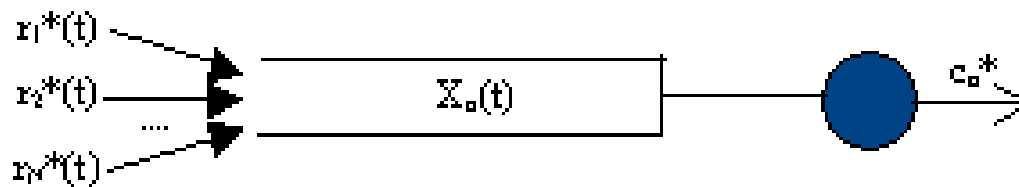
- Assume that each input maintains N (output) TE queues and K non-TE queues.
- All TE queues have the same priority, which is higher than the priorities of non-TE queues.
- When all TE queues are served, the residual service is distributed to non-TE queues.
- The buffer is infinite for each TE queue. Need to find out **the overflow probability** with threshold B .
- Each queue is modeled as a **Markov Modulated Fluid Process (MMFP)**.



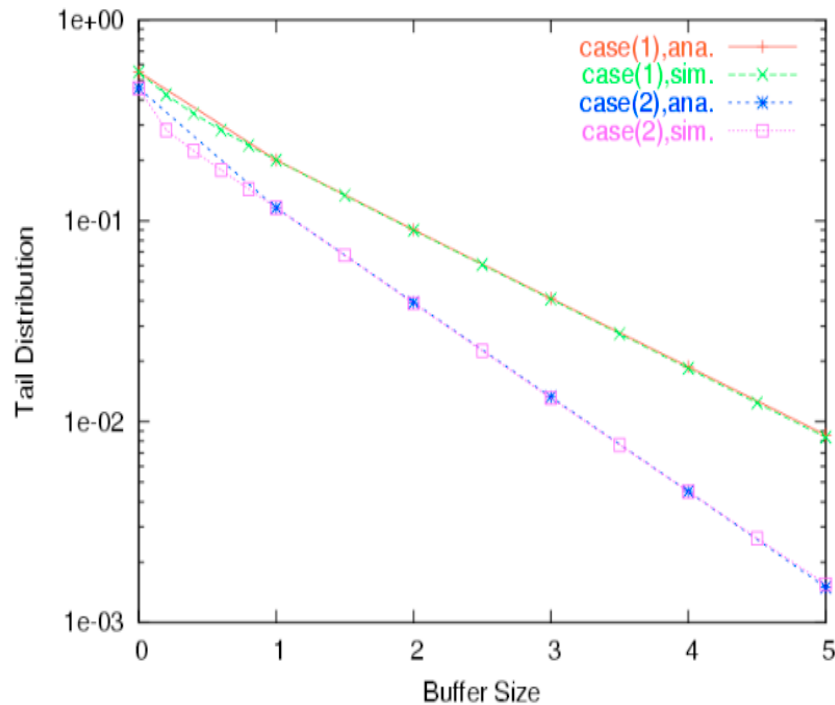
Analyze a Queue with GPS Scheduling



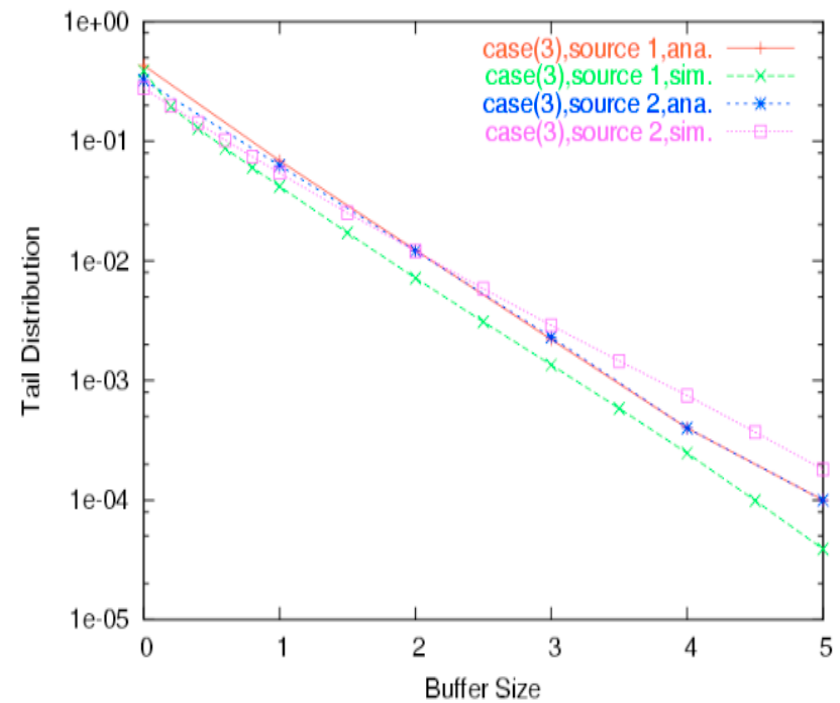
- When one queue is analyzed, the system can be simplified as below.
- Resolve the problem with **Fluid-Flow Model**.
- Assume that there are three on-off sources,
 - two for TE traffic, and
 - one for non-TE traffic.
- Three cases are considered:
 - One queue: all traffic share one queue.
 - Two queues: all TE traffic share one TE queue and non-TE traffic goes to the non-TE queue
 - Three queues: each TE source traffic goes to its own TE queue and non-TE traffic goes to the non-TE queue



Analysis and Simulation Results



Tail distributions of case 1 and 2.



Tail distributions of case 3.

- With the selected system parameters,
 - using TE queue leads to lower overflow probabilities for TE tunnel traffic, and
 - using multiple TE queues can further improve the service of TE tunnel traffic.

Number of MPLS TE Queues Per Router



- How many queues can a router effectively support?
 - For a large network with full meshed TE tunnels, the number of tunnels can easily go to many thousands.
- Limit MPLS TE tunnels to backbone only
 - For a typical IP core network with 36 backbone routers, there will be 1260 tunnels.
 - Assumes each tunnel, on the average, will traverse five routers, including its head end and tail end.
 - Each router, on the average, will have to accommodate 175 tunnels.
 - head end of 35 tunnels, tail end of 35 tunnels, and at mid point of 105 tunnels.
- More complicated analysis may be needed

Generalized Processor Sharing with Regulated Multimedia Traffic



- Joint work with Chaiwat Oottamakorn and Shiwen Mao
- Outline
 - Background
 - The system model
 - Backlog and delay bounds for each class
 - Numerical results
 - Conclusions



Background



- Multimedia traffic: an increasing portion of today's network traffic
 - Music/video streaming
 - Video teleconferencing
 - IP telephony
 - Distance learning
 - ...
- QoS provisioning
 - *Hard* guarantees: e.g., telemedicine, distributed computing
 - *Soft* guarantees: e.g., video, audio



Observations



- The traffic model should be easy to police at network edge
- Parsimonious models are preferred
 - Reduces *state-explosion* problems
 - Need to handle a large number of flows
- Statistical QoS guarantees
 - Most multimedia applications can tolerate a certain amount of loss or delay violations
 - Efficient in utilizing network resources



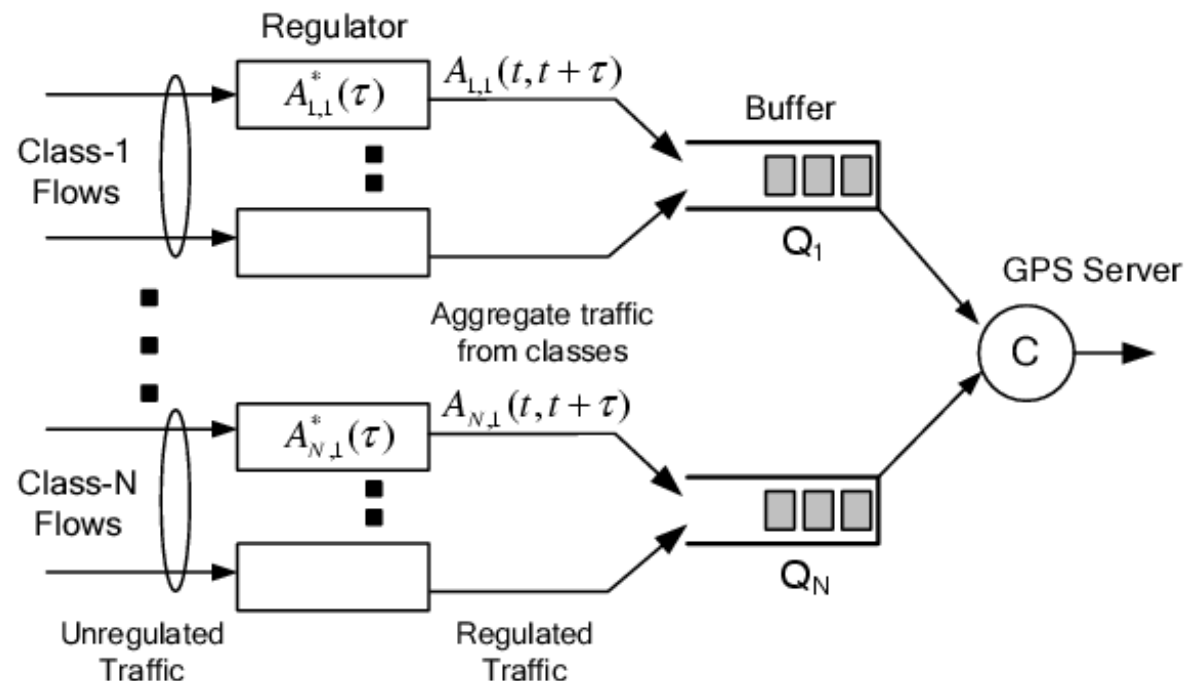
System Model



- A GPS queue with multiple classes flows, each being regulated **deterministically**
- Objective: **statistical bounds on backlog and delay**

- Assumptions:

- Subadditive envelopes
- Additivity
- Stationarity
- Independence



System Model (contd.)



- GPS Weights: $\{\phi_1, \phi_2, \dots, \phi_N\}$

- Minimum service rate for Class i :

$$g_i = \frac{\phi_i}{\sum_{j=1}^N \phi_j} C$$

- Feasible ordering

$$\rho_i < \frac{\phi_i}{\sum_{j=i}^N \phi_j} \left(C - \sum_{j=1}^{i-1} \rho_j \right), \quad 1 \leq i \leq N$$

- Deterministic sub-additive envelope

$$A_{i,k}(t, t + \tau) \leq A_{i,k}^*(\tau), \quad \forall t \geq 0, \forall \tau \geq 0$$



Bounding the Cumulative Traffic



- Moment generating function of cumulative traffic

- A single source: $M_{i,k}(\theta, \tau) := E \left[e^{\theta A_{i,k}(t, t+\tau)} \right]$
- A class (due to the independence assumption):

$$M_i(\theta, \tau) := E \left[e^{\theta A_i(t, t+\tau)} \right] = \prod_{k=1}^{C_i} M_{i,k}(\theta, \tau)$$

- From Theorem 1 in [Boorstyn'00]:

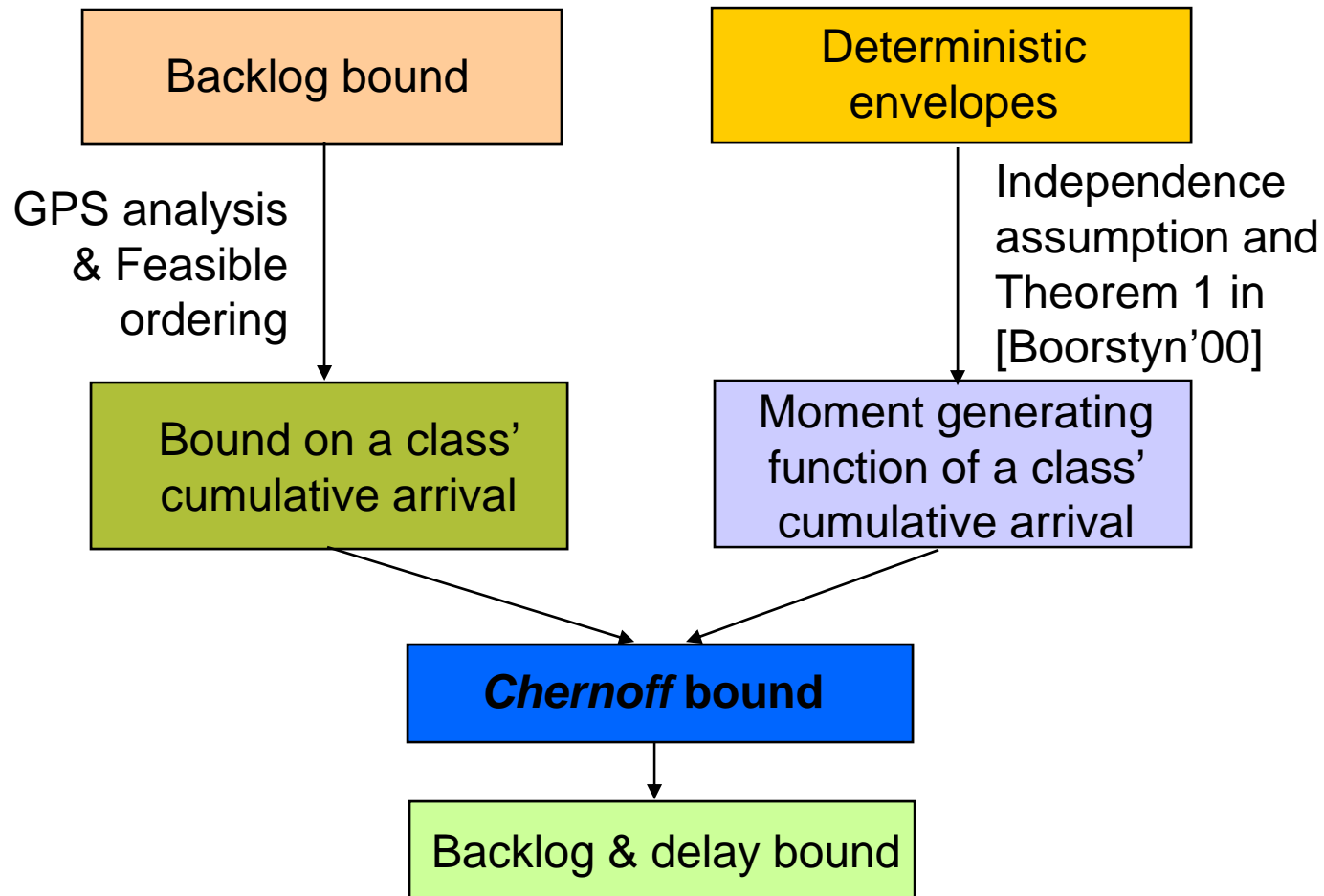
$$M_i(\theta, \tau) \leq \prod_{k=1}^{C_i} \left[1 + \frac{\rho_{i,k} \tau}{A_{i,k}^*(\tau)} \left(e^{\theta A_{i,k}^*(\tau)} - 1 \right) \right]$$

- Chernoff bound: for a random variable Y

$$Pr[Y \geq y] \leq e^{-\theta y} E \left[e^{\theta Y} \right], \quad \forall \theta \geq 0$$



Putting All Pieces Together



The Backlog Bound for a Class



Theorem 1: Given a GPS server with a transmission rate C serving N incoming traffic classes $\{A_1, A_2, \dots, A_N\}$ satisfying (P1)-(P4). Assume that $\sum_{i=1}^N \rho_i < C$ and a feasible ordering with respect to $\{\phi_1, \phi_2, \dots, \phi_N\}$ and $\{\rho_1, \rho_2, \dots, \rho_N\}$. For any time instance τ in a busy period and any $q_i \geq 0$,

$$Pr[Q_i(\tau) \geq q_i] \leq \inf_{\theta \geq 0} \left\{ e^{-\theta(q_i + \psi_i C \tau)} \prod_{k=1}^{C_i} \left[1 + \frac{\rho_{i,k} \tau}{A_{i,k}^*(\tau)} \left(e^{\theta A_{i,k}^*(\tau)} - 1 \right) \right] \prod_{j=1}^{i-1} \prod_{k=1}^{C_j} \left[1 + \frac{\rho_{j,k} \tau}{A_{j,k}^*(\tau)} \left(e^{\psi \theta A_{j,k}^*(\tau)} - 1 \right) \right] \right\}$$



The Delay Bound for a Class



- For FCFS queues, delay of a traffic unit enqueued at time τ :

$$D_i(\tau) = \min\{z : z \geq 0 \text{ and } A_i(0, \tau) \leq S_i(0, \tau + z)\}$$

- Delay requirement is violated when:

$$A_i(0, \tau) \geq S_i(0, \tau + d_i)$$

- So we have:

$$Pr[D_i(\tau) \geq d_i] = Pr[A_i(0, \tau) \geq S_i(0, \tau + d_i)]$$



The Delay Bound (contd.)



Theorem 2: Given a GPS server with a transmission rate C serving N incoming traffic classes: A_1, A_2, \dots, A_N satisfying properties (P1)-(P4). Assume that $\sum_{i=1}^N \rho_i < C$ and a feasible class ordering is achieved with respect to $\{\phi_1, \phi_2, \dots, \phi_N\}$. For any time τ in a busy period and any $d_i \geq 0$,

$$\Pr[D_i(\tau) \geq d_i] \leq \inf_{\theta \geq 0} \left\{ e^{-\theta \psi_i C(\tau + d_i)} \prod_{k=1}^{C_i} \left(1 + \frac{\rho_{i,k} \tau}{A_{i,k}^*(\tau)} (e^{\theta A_{i,k}^*(\tau)} - 1) \right) \prod_{j=1}^{i-1} \prod_{k=1}^{C_j} \left(1 + \frac{\rho_{j,k}(\tau + d_i)}{A_{j,k}^*(\tau + d_i)} (e^{\psi \theta A_{j,k}^*(\tau + d_i)} - 1) \right) \right\}.$$



Numerical Results



- Settings
 - C: 0.8 ~ 1.0 Gb/s
 - Four traffic classes:

TRAFFIC SOURCE PARAMETERS

-	Avg. Rate (Kbps)	Peak Rate (Kbps)	GPS Assignment ϕ_i
Class 1: on-off (Exp.)	230	1600	280
Class 2: on-off (Pareto)	520	2220	500
Class 3: Video Trace 1	773	2687	800
Class 4: Video Trace 2	1053	3150	1000

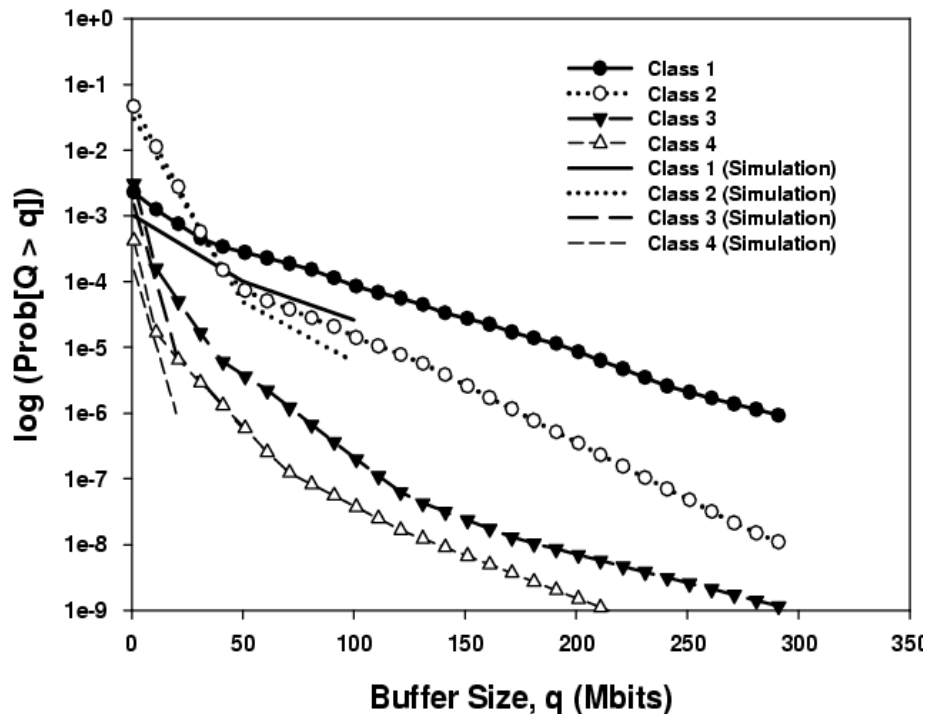
- Comparison of analysis and simulation results (OPNET)



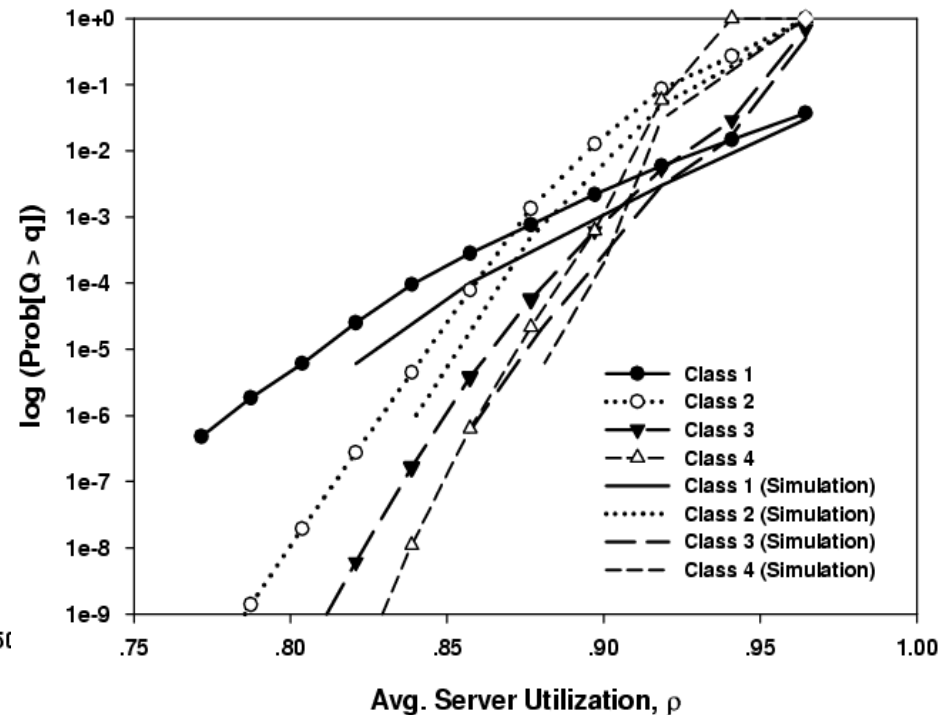
Numerical Results - I



- Backlog distribution: (a) fixed load (300~400 sources for each class); (b) fixed buffer size



(a) Tail distribution as a function of buffer size q with fixed average server utilization ($\rho = 85.7\%$)



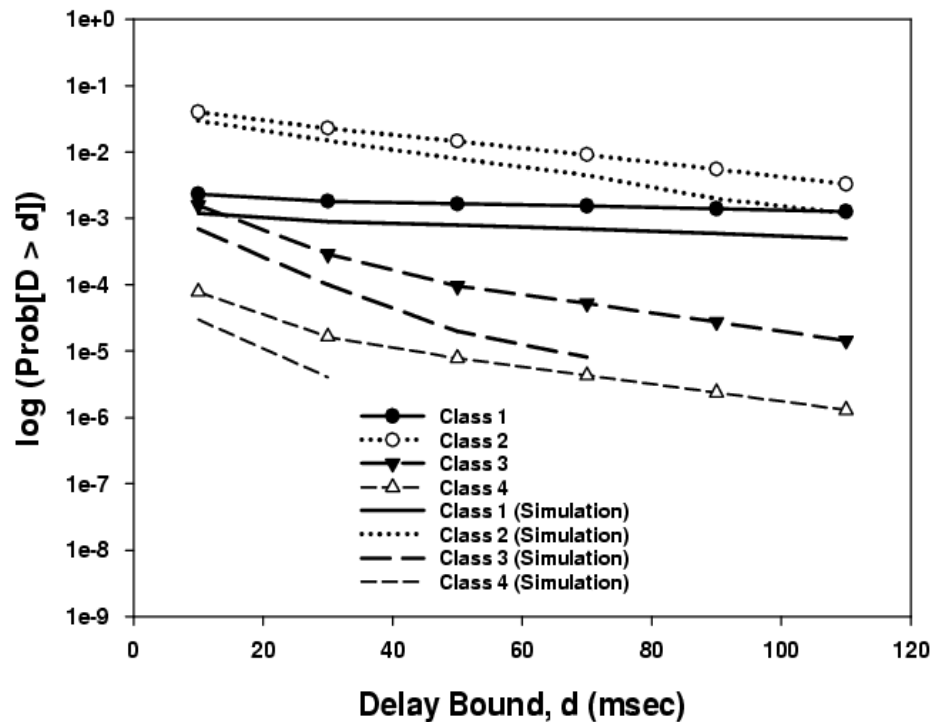
(b) Tail distribution as a function of average server utilization ρ with fixed buffer size ($q = 50$ Mbits).



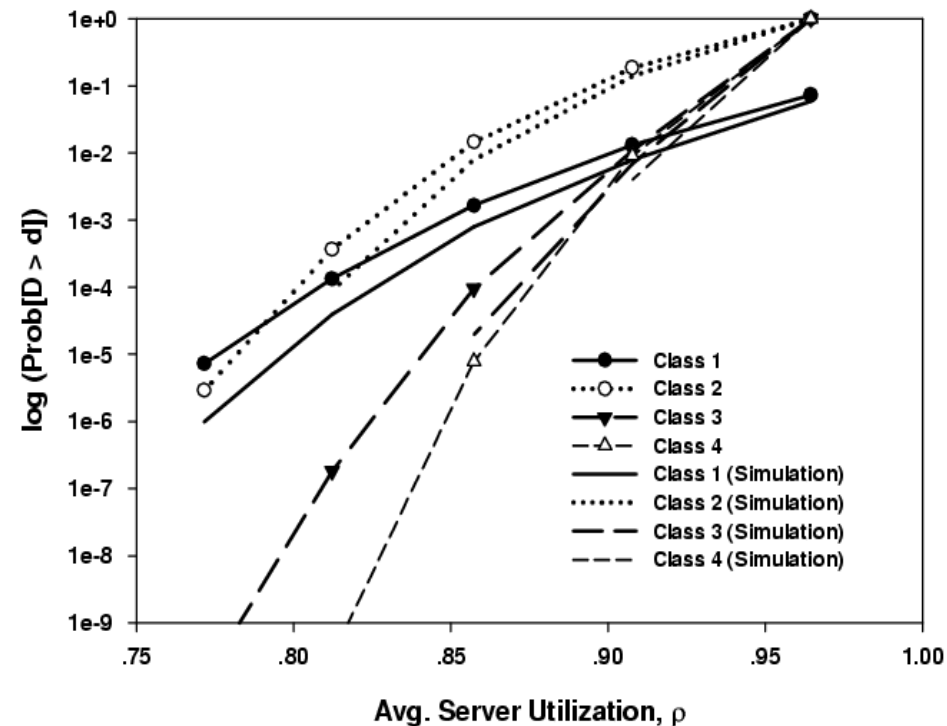
Numerical Results - II



- Delay distribution: (a) fixed load; (b) fixed delay requirement



(a) Tail distribution as a function of the delay bound d with fixed average server utilization ($\rho = 85.7\%$)



(b) Tail distribution as a function of average server utilization ρ with fixed delay bound ($d = 50$ msec).



Numerical Results - III



■ Admissible region:

- $C=900$ Mb/s
- Fixed number of Class
4 flows: 300

□ Class 1:

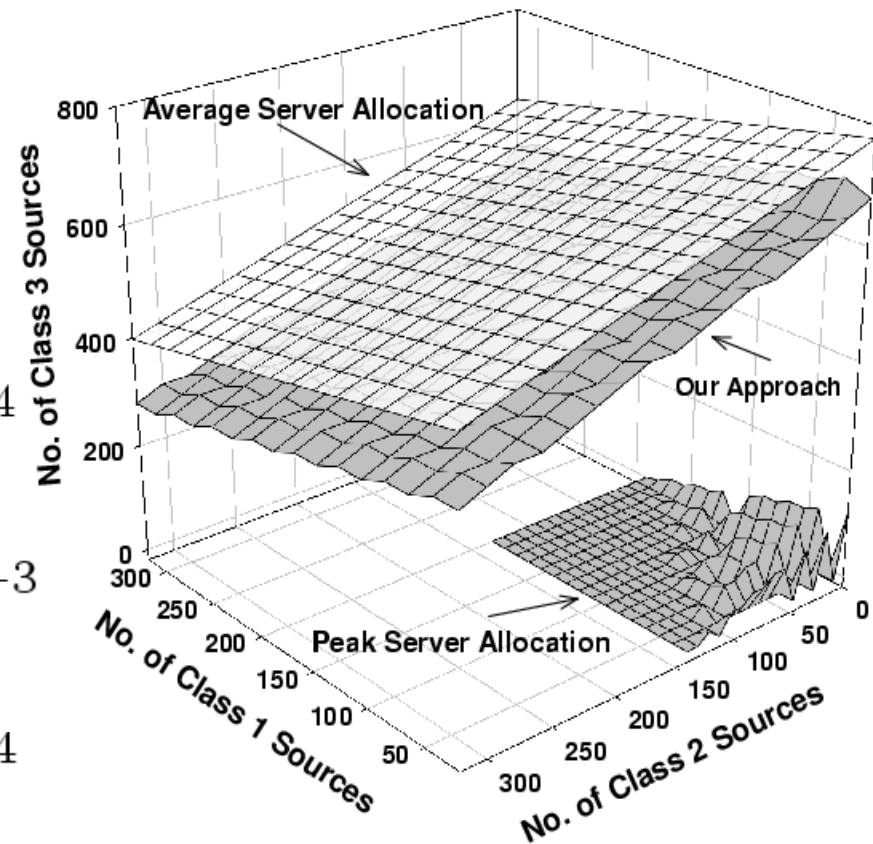
$$Pr[Q_1 \geq 50Mbits] = 10^{-4}$$

□ Class 2:

$$Pr[D_2 \geq 30msec] = 10^{-3}$$

□ Class 3:

$$Pr[D_3 \geq 50msec] = 10^{-4}$$



Conclusions



- Studied the GPS system with deterministically regulated multimedia flows
 - Easy for regulation and monitoring, while maintaining high efficiency in utilizing network resources
 - Leaky bucket and GPS scheduler:
 - Available in most commercial routers
- Derived bounds on backlog and delay
 - Computationally efficient
 - Can handle a large number of flows
 - Can handle LRD video, and other self-similar sources
 - Analytical bounds match simulation results

