A Measurement Based Admission Control Algorithm for Integrated Services Packet Networks

Sugih Jamin, Peter B. Danzig, Scott Shenker, Lixia Zhang

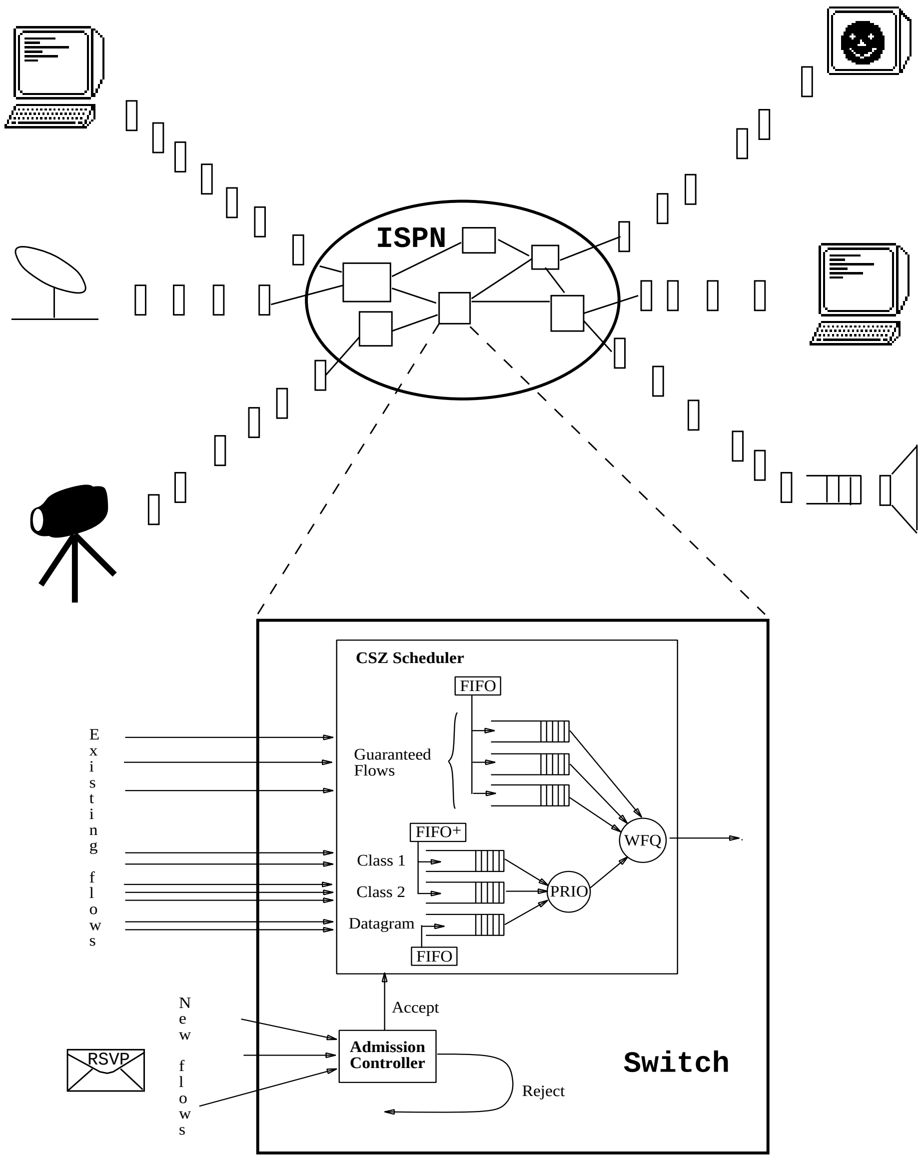
Presented by

Sugih Jamin
jamin@usc.edu
http://netweb.usc.edu/jamin/

ACM SIGCOMM 1995,
Cambridge, MA

August 30, 1995

Integrated Services Packet Networks (ISPN):



Examples of ISPN:
the Internet, ATM networks.

Framework:

- ISPN/ATM with limited bandwidth
- Accounting & Pricing, e.g. [Cocchi, et al., Herzog, et al.]
- Resource reservation protocol, e.g. [Zhang et al., Topolgic]
- Work conserving scheduler, e.g. [CSZ92]
- E2E delay: $\sum D_i, \prod p(D_i)$ [Yates]
- Adaptive Source Routing, e.g. [Breslau95]
- Token-bucket filter

Outline:

- Framework and Context
- Realtime services:
  Delay Bound vs. Utilization
- Equivalent token bucket filter and the Measurement-based admission control algorithm
- Measurement mechanism
- Simulation results: Order of magnitude increase in link utilization

Realtime Services:

Need *delay bound* guarantee.
Question: What *kind* of guarantee?
Answers give rise to **service models**:

- Unbounded delay:
  red-button service.
- Deterministic guarantee:
  $\sum$ resource requirements $<$ capacity.
- Probabilistic guarantee:
  Enough resources such that
  Pr(delay violation) $< \epsilon$.

  Disadvantage: Both require precise source models; with loose models resulting in low utilization.

- Predictive service.

Predictive Service:

- Exploits adaptive playback point.
- Does not require precise source model.
- *Requires* measurement based admission control.
- Provides reliable, but not guaranteed, delay bound. (CSZ provides *both* guaranteed and predictive services.)
- Costs less.
- Made up of multiple classes, each with its own bound.

Advantages of Predictive Service:

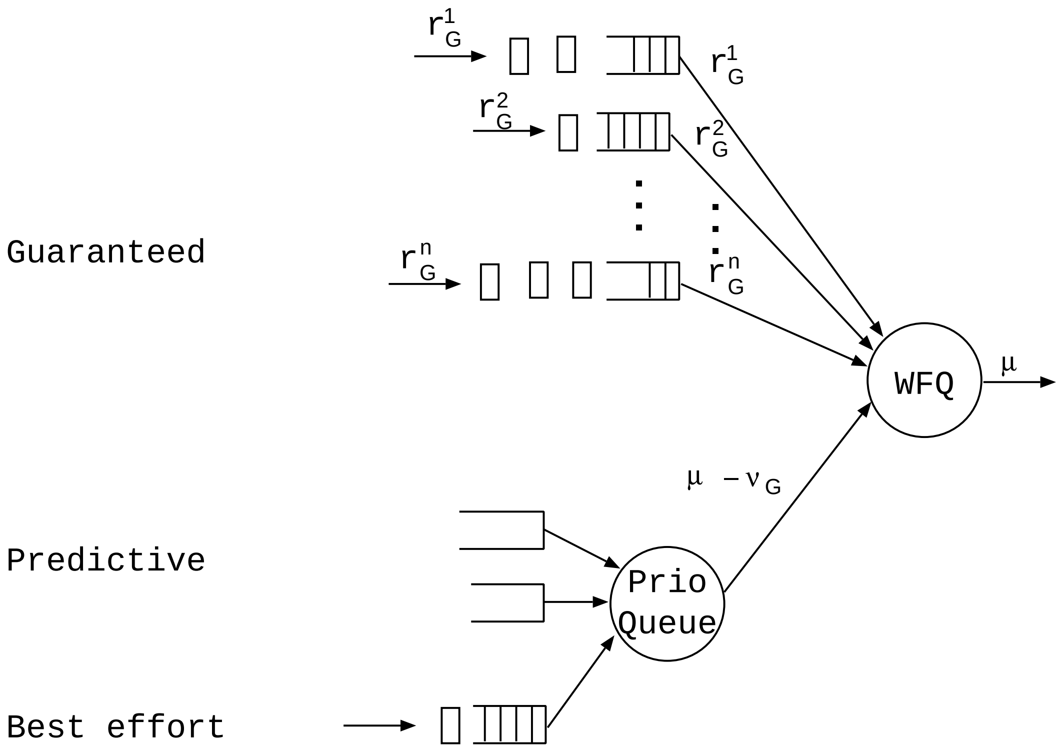
- Results in higher utilization
- Still gives bounded delay

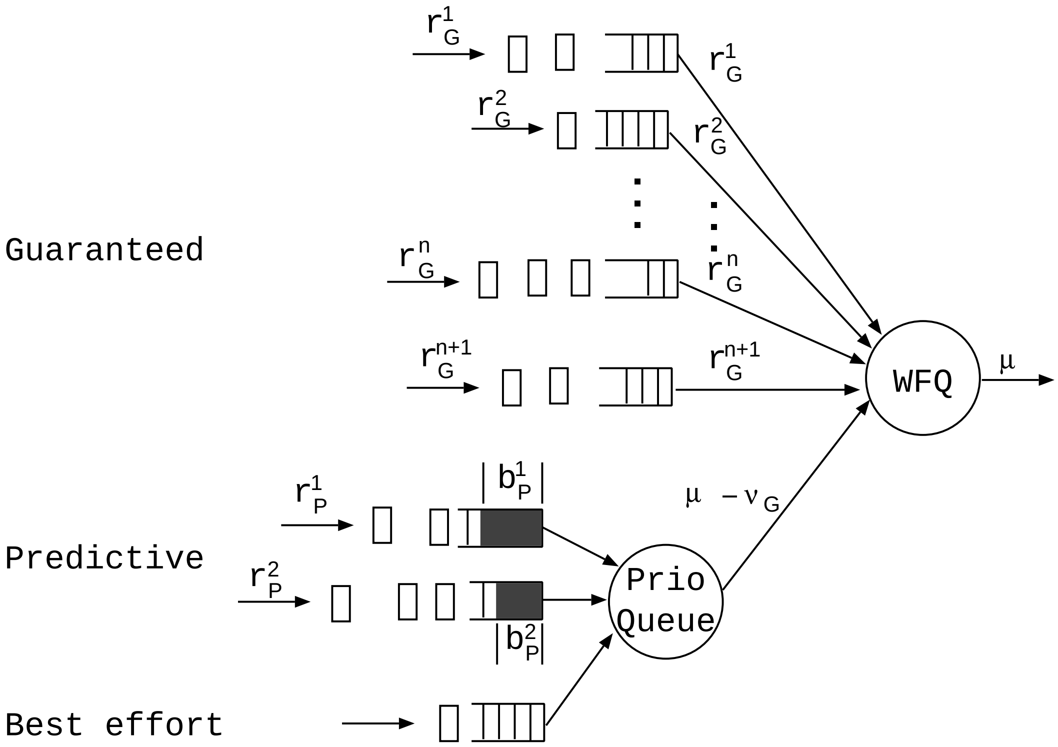
Admission Control Mechanism:

Need to know:

- Token bucket filter of new flow
- Effect on existing traffic

Effect of adding a new flow





Guaranteed
Predictive
Best effort
WFQ
Prio
Queue
r_G^1
r_G^2
r_G^n
r_G^{n+1}
r_P^1
r_P^2
$|b_P^1|$
$|b_P^2|$
$\mu - \nu_G$
μ

Equivalent Token Bucket Filter

Per flow maximal delay in a predictive class:

$$D_j^* = \frac{\sum_{i=1}^{j} b_i}{\mu - \nu_G - \sum_{i=1}^{j-1} r_i} \tag{1}$$

Equivalent token bucket filter:

- Treat aggregate traffic as one flow
- Use measured values to approximate token bucket filter

The Admission Control Algorithm

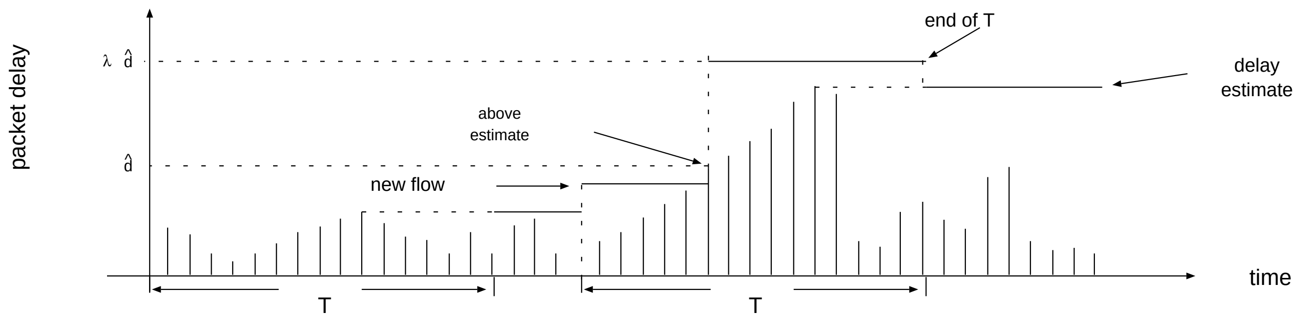
Incoming predictive flow:

- Enough bandwidth
- Same priority delay ok
- Lower priority delay ok

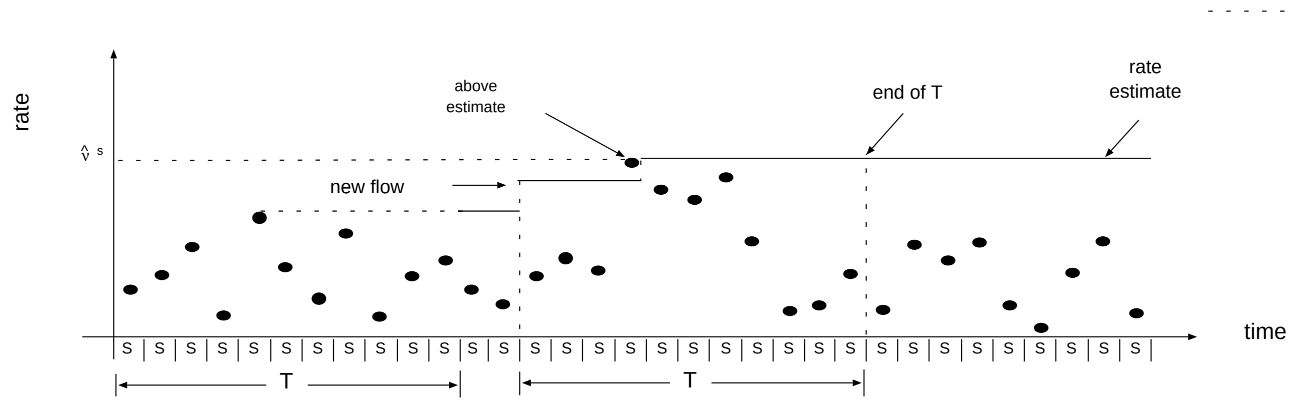
Incoming guaranteed flow:

- Enough bandwidth
- All predictive delay ok

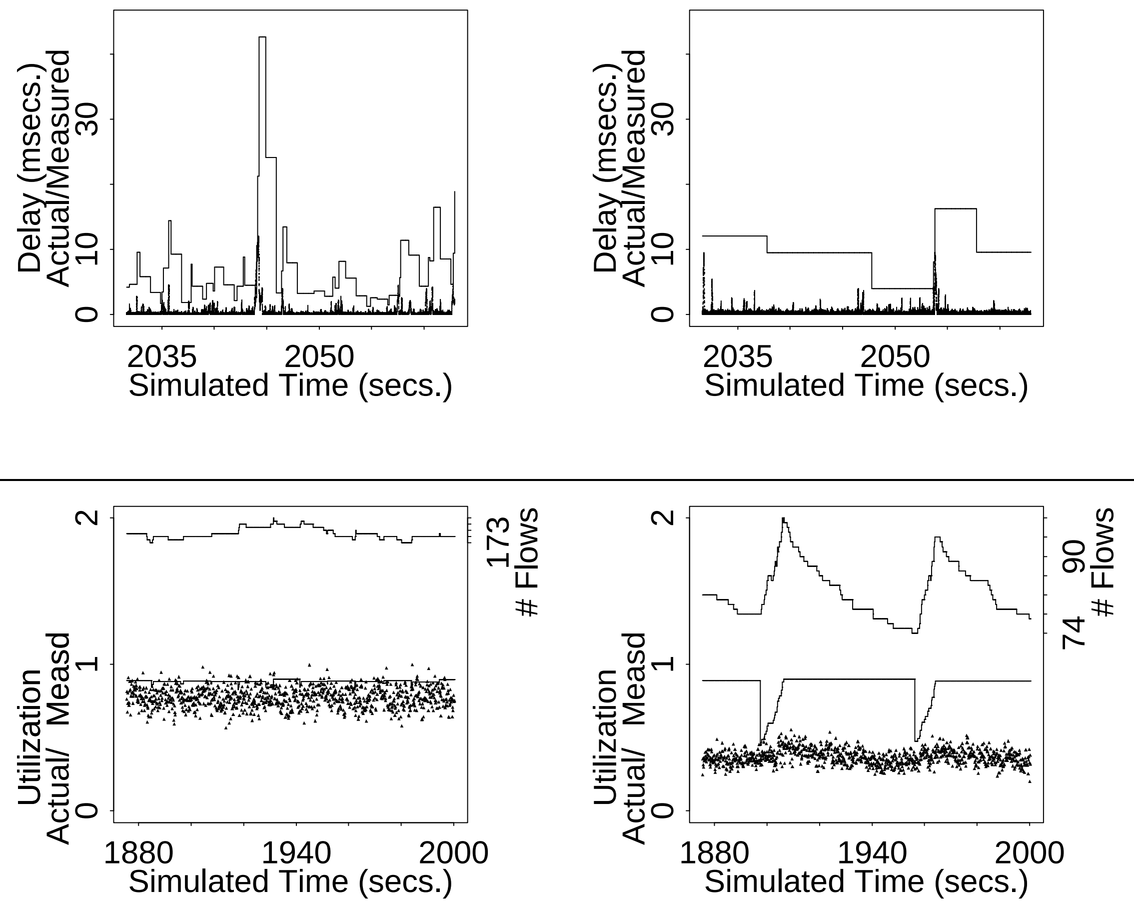
Delay measurement:



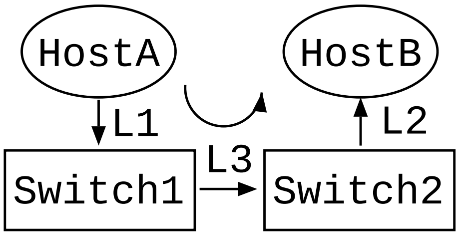
Rate measurement:



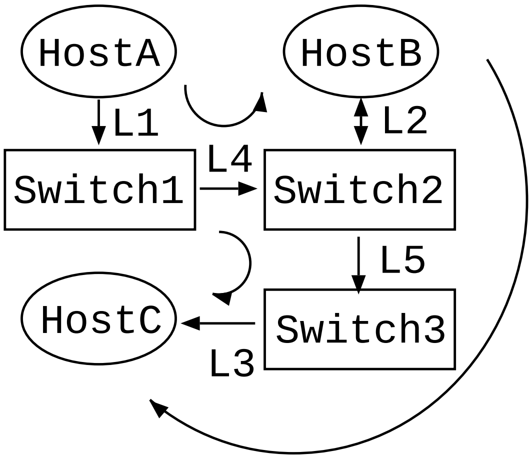
On the Appropriate Value of T



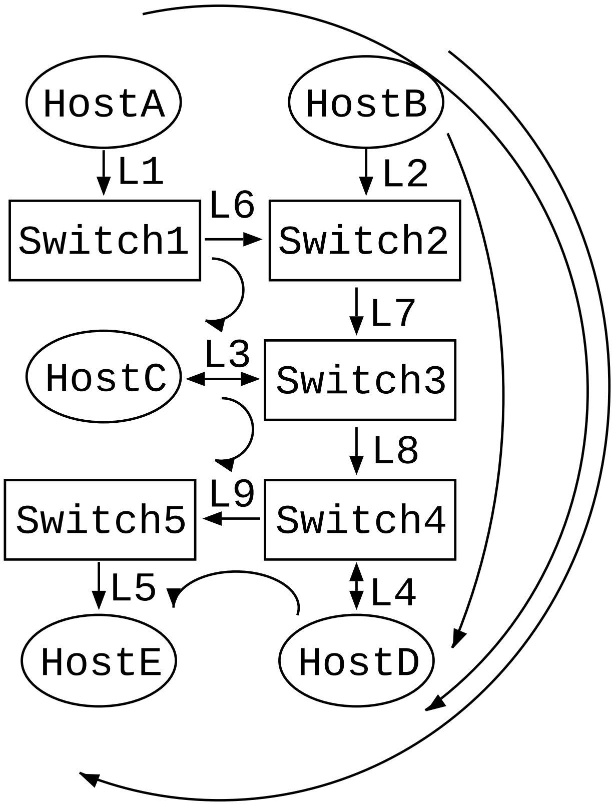
Simulated Topologies



(a) One-Link



(b) Two-Link



(c) Four-Link

Source Models

Two-state Markov-ON/OFF:
Exponentially distributed ON, OFF times.
Exponentially distributed duration.

Two long-range dependent (LRD) sources:

Two-state Pareto-ON/OFF:
Pareto distributed ON, OFF times
Lognormal distributed duration.

fARIMA:
fARIMA output as ON times.
Fixed length ON time interarrivals.

Fixed rate packet transmission.
Poisson flow arrivals.

Single-hop Homogeneous Sources Simulation Results

Model Name	Guaranteed %Util	Guaranteed #Actv	Predictive %Util	Predictive #Actv	$\lceil d_j \rceil$	$\overline{L}/T$	**Gain Factor**
EXP1	46	144	80	250	3	60	**1.4**
EXP2	28	28	76	75	42	300	**2.7**
EXP3	2	18	62	466	33	600	**31.0**
POO1	7	144	74	1637	5	60	**10.6**
POO2	3	38	64	951	8	60	**21.3**
fARIMA	55	9	81	13	72	60	**1.5**

Repeated 10 times, very tight confidence intervals.

Six Instantiations of the Three Source Models

Model Name	Model's Parameters				Token Bucket Parameters				Bound (ms)	
	p pkt/sec	I msec	N pkts	p/a	r tkn/sec	b tkns	cut rate	max qlen	D^*	D_j
EXP1	64	325	20	2	64	1	0	0	16	16
EXP2	1024	90	10	10	320	50	2.1e-3	17	160	160
EXP3	∞	684	9	∞	512	80	9.4e-5	1	160	160
				β						
POO1	64	2925	20	1.2	64	1	0	0	16	16
POO2	256	360	10	1.9	240	60	4.5e-5	220	256	160
				s						
fARIMA ({0.75}, 0.15, -)	∞	125	8	13	1024	100	1.1e-2	34	100	160

Single- and Multi-hop, Multiple Source Models,
All Services Link Utilization

Topology Name	Link Name	Guaranteed %Util	Guaranteed and Predictive %Util	$\lceil d_1 \rceil$	$\lceil d_2 \rceil$	**Gain Factor**
One-Link	L3	25	65	1	43	**2.6**
Two-Link	L4	26	75	3	72	**2.9**
	L5	22	74	2	48	**3.4**
Four-Link	L6	15	47	1	38	**3.1**
	L7	21	75	2	65	**3.6**
	L8	20	72	2	45	**3.5**
	L9	21	76	1	50	**3.6**

Summary

Four major points:

1. Delay Bound vs. Utilization
2. Delay Formula to EQTBF to Admission Control
3. Being Conservative vs. Aggressive
4. Simulation results that show
   - Measurement-based admission control can be *reliable*
   - Predictive service is *viable*