Proportional Differentiated Service: Delay Differentiation and Packet Scheduling

Speaker: ***Yi-Neng Lin***
Author: Constantinos Dovrolis, Dimitrios Stiliadis
Advisor: Ying-Dar Lin

Department of Computer and Information Science
National Chiao Tung University
Hsinchu, Taiwan
2000,11,24

Outline

- Introduction
- The Proportional Differentiation Model
- Other relative differentiation models
- The Dynamics and Feasibility of Proportional Delay Differentiation
- Two Packet Schedulers for Relative Delay Differentiation
- Simulation study
- Conclusion

Introduction(1/2)

- Growing demand: same-service-to-all => classified
- Integrated Service: features & weakness
  - End-to-End guarantee
  - Interdomain policy & pricing
  - Complexity in deploying RSVP
- DiffServ: configurable in PHB (Per Hub Behavior)
  - absolute DS: tradeoff between "service assurance" and "net utilization"
  - relative DS: class (i) >= class (i-1); spacing between each class

Introduction(2/2)

- Tuning knobs for adjusting the quality space of each class
  - Predictable: the differentiation should be consistent
  - Controllable: according to user-selected criteria
- Two agreeable definitions of "better service"
  - lower likelihood of packet loss => TCP
  - lower queuing delay => IP-telephony & video conferencing

The Proportional differentiation model(1/3)

- Constrains for all pairs of classes

$$\frac{q_i}{q_j} = \frac{c_i}{c_j}$$

q_i : Performance measure for class i

c_i : quality differentiation parameter

where $c_1 < c_2 < \cdots < c_N$

While actual quality of each class varies with class loads,

the delay ratio remains fixed.

The Proportional Differentiation model(2/3)

- Proportional Delay Differentiation model (PDD)

$$\frac{\overline{d}_i}{\overline{d}_j} = \frac{\delta_i}{\delta_j}$$

$\overline{d}_i$: average queuing delay of the class-i packets

δ_i : Delay Differentiation Parameter

where $\delta_1 > \delta_2 > \cdots > \delta_N > 0$

While actual quality of each class varies with class loads,

the delay ratio remains fixed.

The Proportional differentiation model(3/3)

- Long-term average is not precise enough
- Unpredictable result between high and low class when bursty
- Hard to derive feasibility in short timescales
- Goals: identify and evaluate scheduling mechanisms that can approximate the proportional differentiation model in short timescales.

Other relative differentiation models (1/2)

- Strict Prioritization:
  - highest backlogged class first (delay aspect)
  - Drawback: Uncontrollable
- Price Differentiation:
  - higher price => fewer customers
- Capacity Differentiation:
  - Higher class, higher bandwidth and buffer
  - Drawback: controllable in bandwidth, not in delay

Other relative differentiation models (2/2)

Additive Differentiation

$$\overline{d}_i - \overline{d}_j = D_{i,j} > 0 \quad (j > i)$$

Some priority scheduler can approximate this kind of behavior at long-term sufficient heavy load

Example:

$$p_i(t) = w_i(t) + s_i$$

$w_i(t)$: waitting time of the packet at time t

s_i : scheduler differentiation parameters

Dynamics and Feasibility of PDD (1/3)

$$\frac{\overline{d}_i}{\overline{d}_j} = \frac{\delta_i}{\delta_j}$$

Not valid when delay of the class is less than delay in a FCFS with only one queue

δ_i : Delay Differentiation Parameter(**DDP**)

+

$$\sum_{i=1}^{N} \lambda_i d_i = \lambda \overline{d}(\lambda)$$

Conservation law

(Little's formula: $L = \lambda d$)

||

$$\overline{d}_i = \frac{\delta_i \overline{d}(\lambda)}{\delta_1 \frac{\lambda_1}{\lambda} + \delta_2 \frac{\lambda_{21}}{\lambda} + \cdots + \delta_N \frac{\lambda_N}{\lambda}}$$

Four properties for the dynamics of the proportional delay differentiation

Dynamics and Feasibility of PDD (2/3)

1. $\bar{d}_i$ increases with λ_j

2. Increasing the load of a higher class has more impact in delay

3. $\delta_i \uparrow$ => $d_j \downarrow$ and $d_i \uparrow$

4. Some load class i switches to class j

$$\begin{cases} i < j \Rightarrow d_k \uparrow \\ i > j \Rightarrow d_k \downarrow \end{cases}$$

Dynamics and Feasibility of PDD (3/3)

General traffic assumption:

$\{ d_i \}$ is feasible

➔ $$\sum_{i \in \pi} \lambda_i \bar{d}_i \geq \left(\sum_{i \in \pi} \lambda_i\right) \bar{d}\left(\sum_{i \in \pi} \lambda_i\right)$$ For all $\pi \in \{1, 2 \ldots N\}$

$\bar{d}(\sum_{i \in \pi} \lambda_i)$ queuing delay of aggregate traffic in a FCFS server

Backlog-Proportional Rate (BPR) scheduler

Dynamically adjust the service rate to the backlog of the class

$$\frac{r_i(t)}{r_j(t)} = \frac{s_i}{s_j}\frac{q_i(t)}{q_j(t)}$$

$$(s_1 < s_2 \cdots < s_N)$$

$q_i(t)$: backlog of queue i at time t

s_j : Scheduler Differential Parameter (SDP)

$$\sum_{i=1}^{N} r_i(t) = R$$

$r_i(t)$: service rate of queue i at time t

R : Link capacity

$$\frac{\overline{d}_i}{\overline{d}_j} \rightarrow \frac{\delta_i}{\delta_j} = \frac{s_j}{s_i}$$

DDP ratios tending to the inverse of the corresponding SDP ratios

Waiting-Time Priority (WTP) scheduler

The priority of a packet increases proportionally with its waiting-time

$$p_i(t) = w_i(t) s_i$$

$$(s_1 < s_2 \cdots < s_N)$$

$p_i(t)$: priority of the packet at time t

$w_i(t)$: waiting-time of the packet at time t

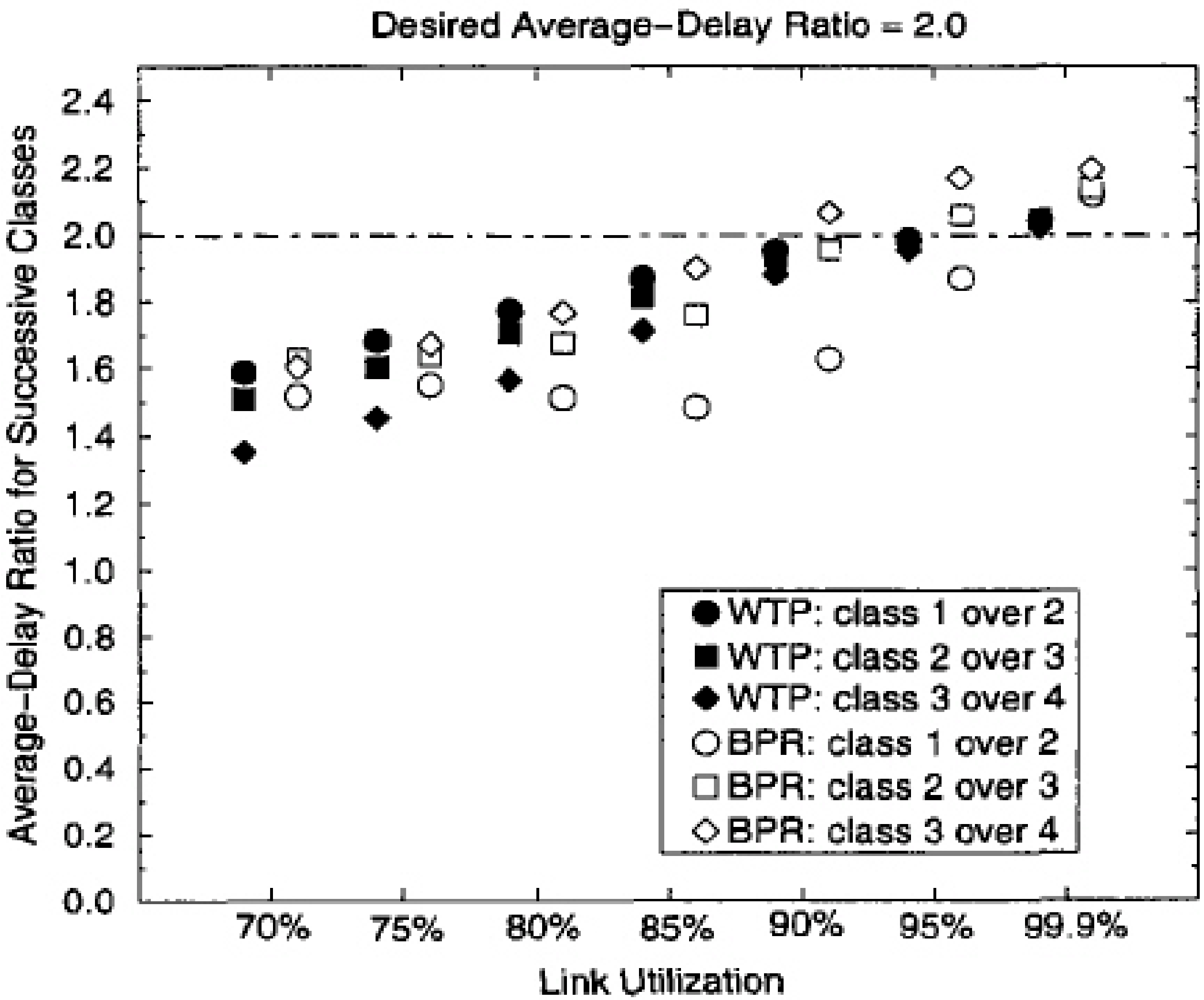
s_j : Scheduler Differential Parameter (SDP)

DDP ratios tending to the inverse of the corresponding SDP ratios

Reflect load in recent past of a queue more precisely than BPR by the waiting-time of head packet

★ May suffer from short-term starvation

Simulation Study:Evaluation of BPR and WTP (1/7)



(a) $s_1 = 1, s_2 = 2, s_3 = 4, s_4 = 8$

Effect of different link utilization

Load distribution

Class-1: 40%

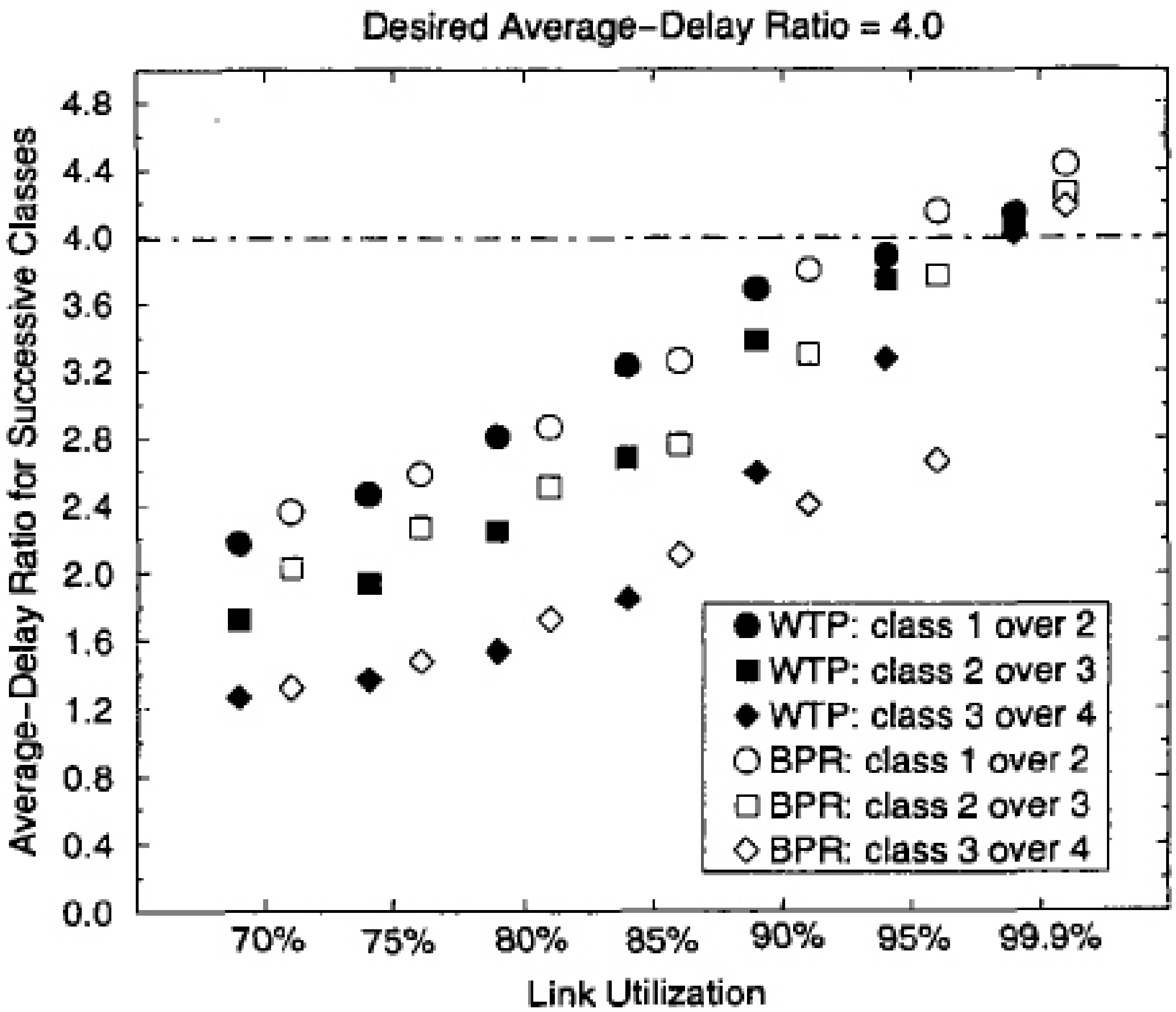
Class-2: 30%

Class-3: 20%

Class-4: 10%

The ratios of average-delays between successive classes

Simulation Study:Evaluation of BPR and WTP (2/7)



(b) $s_1 = 1, s_2 = 4, s_3 = 16, s_4 = 64$

Effect of different link utilization

Load distribution

Class-1: 40%

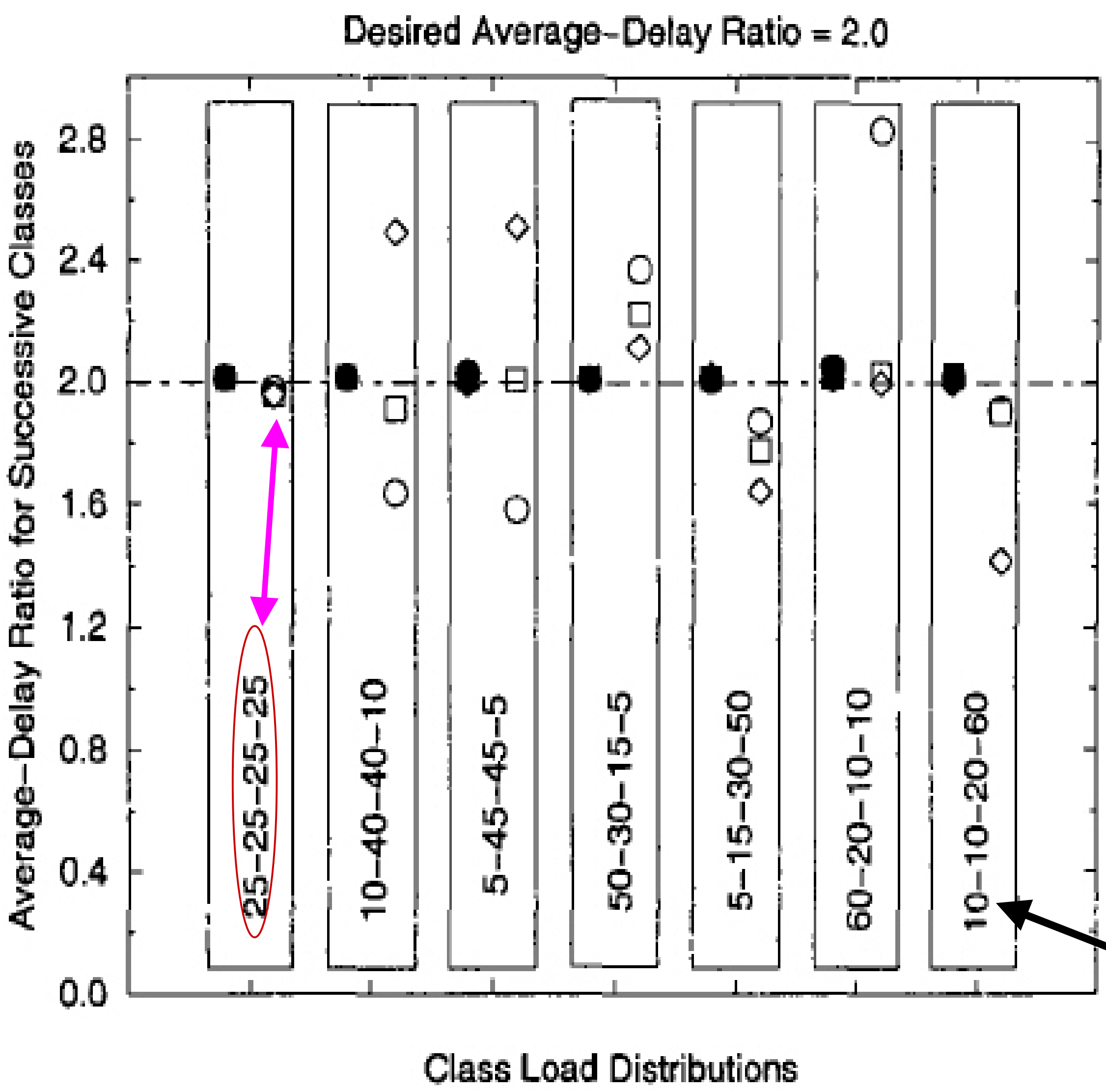
Class-2: 30%

Class-3: 20%

Class-4: 10%

The ratios of average-delays between successive classes

Simulation Study:Evaluation of BPR and WTP (3/7)



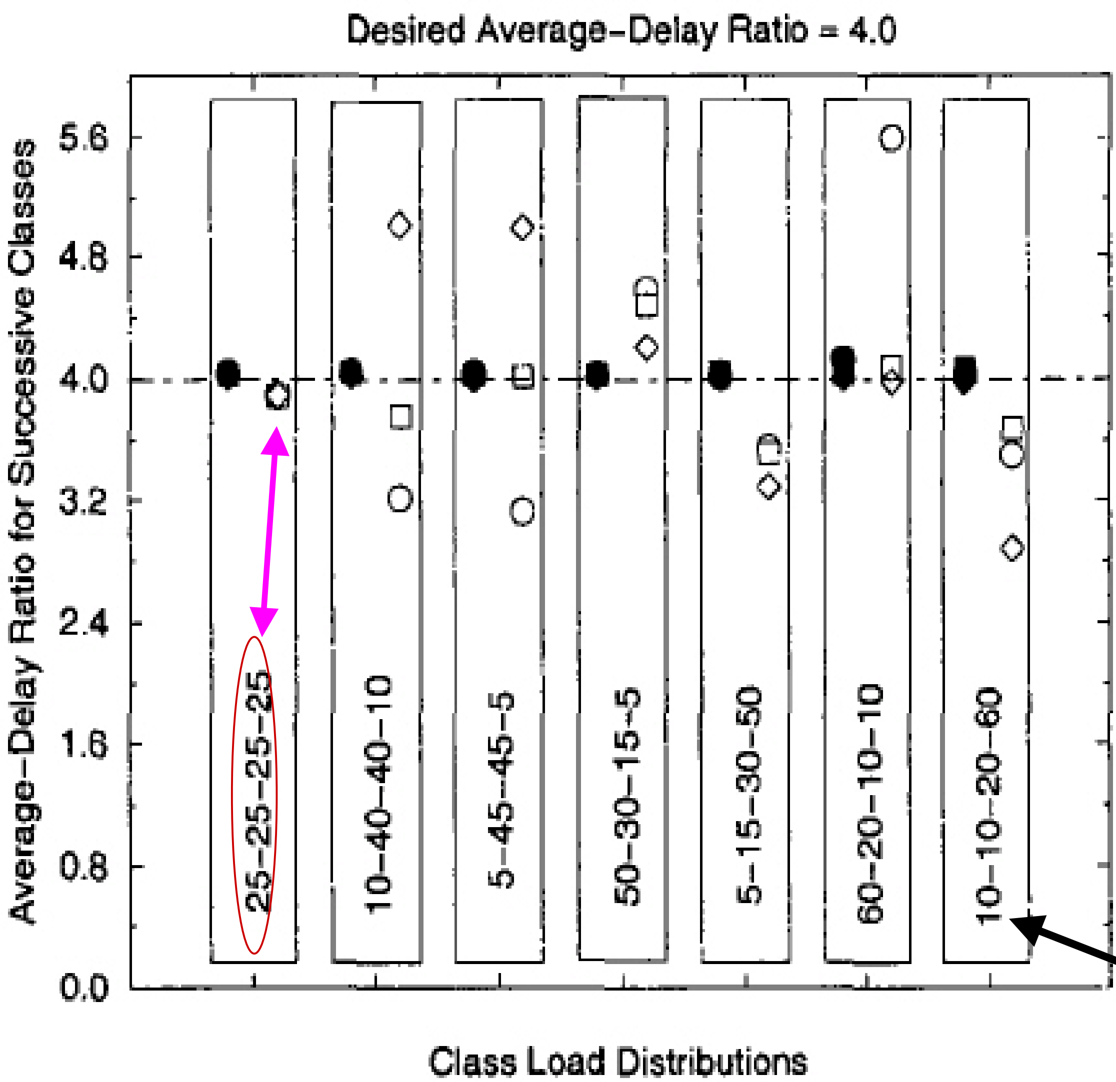
(a) $s_1 = 1, s_2 = 2, s_3 = 4, s_4 = 8$

The effect of the class load distribution

Utilization is 95% in all cases

Fraction of the four classes in the aggregate packet stream

Simulation Study:Evaluation of BPR and WTP (4/7)



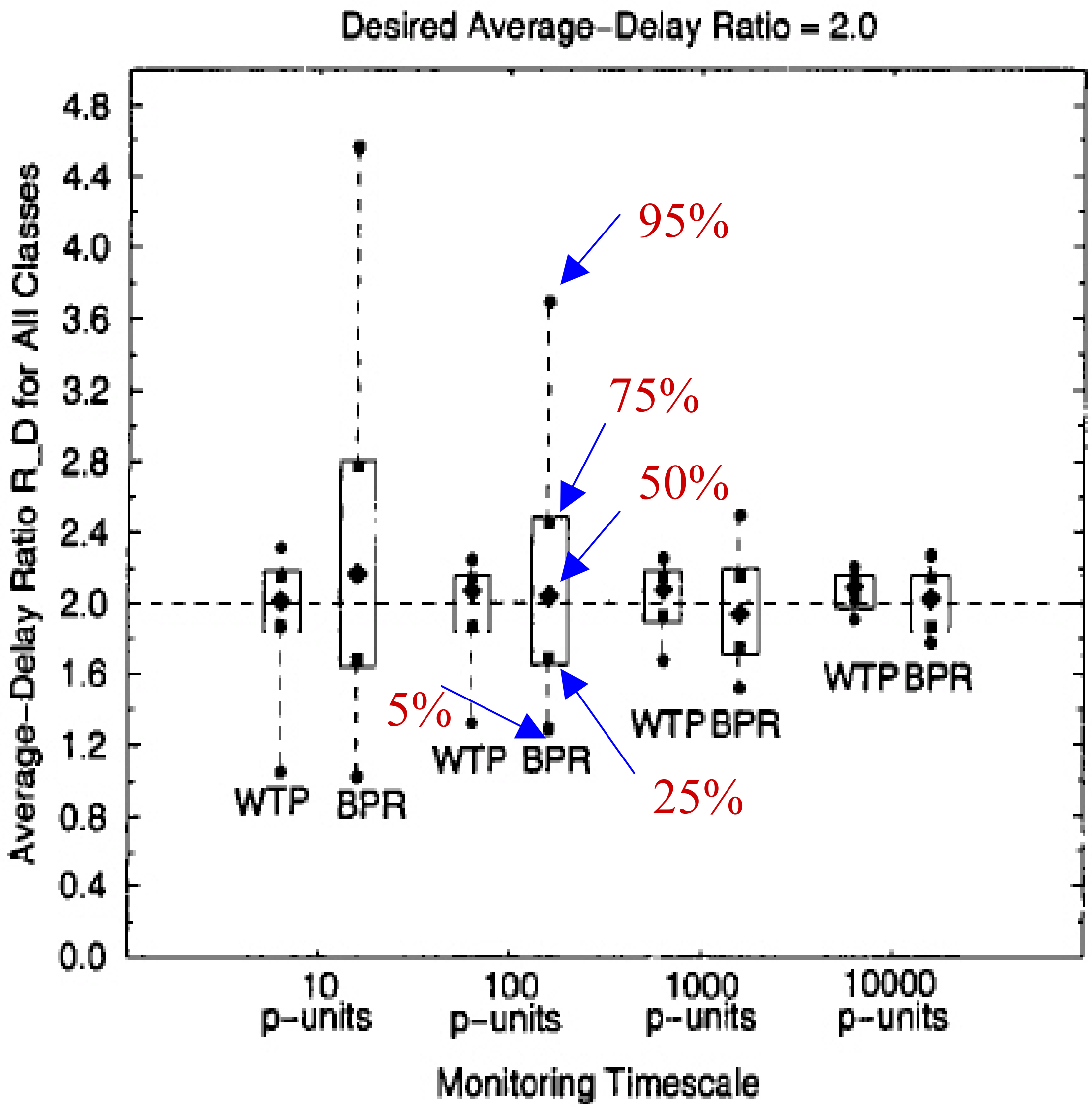
(b) $s_1 = 1, s_2 = 4, s_3 = 16, s_4 = 64$

The effect of the class load distribution

Utilization is 95% in all cases

Fraction of the four classes in the aggregate packet stream

Simulation Study:Evaluation of BPR and WTP (5/7)



(a) $s_1 = 1, s_2 = 2, s_3 = 4, s_4 = 8$

Five percentiles of the R_D measure

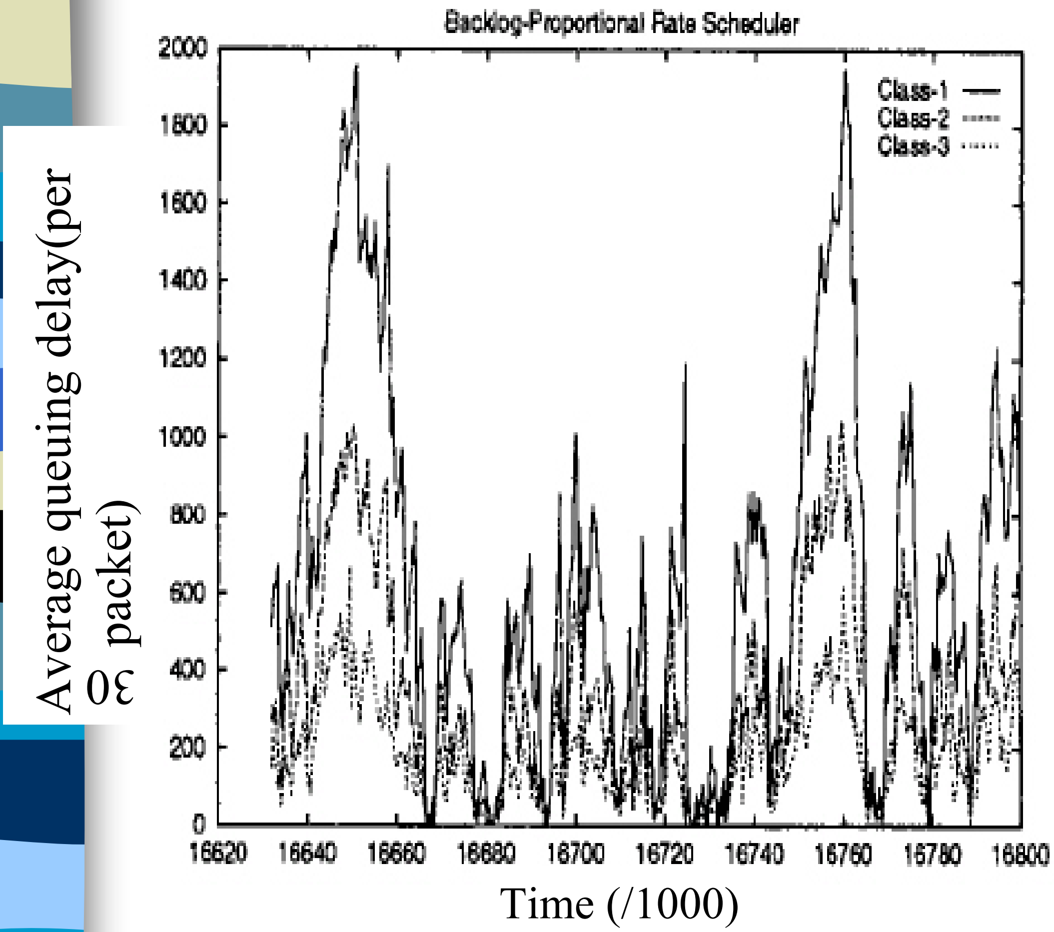
aggregate load: 95%

p-unit: average packet transmission time

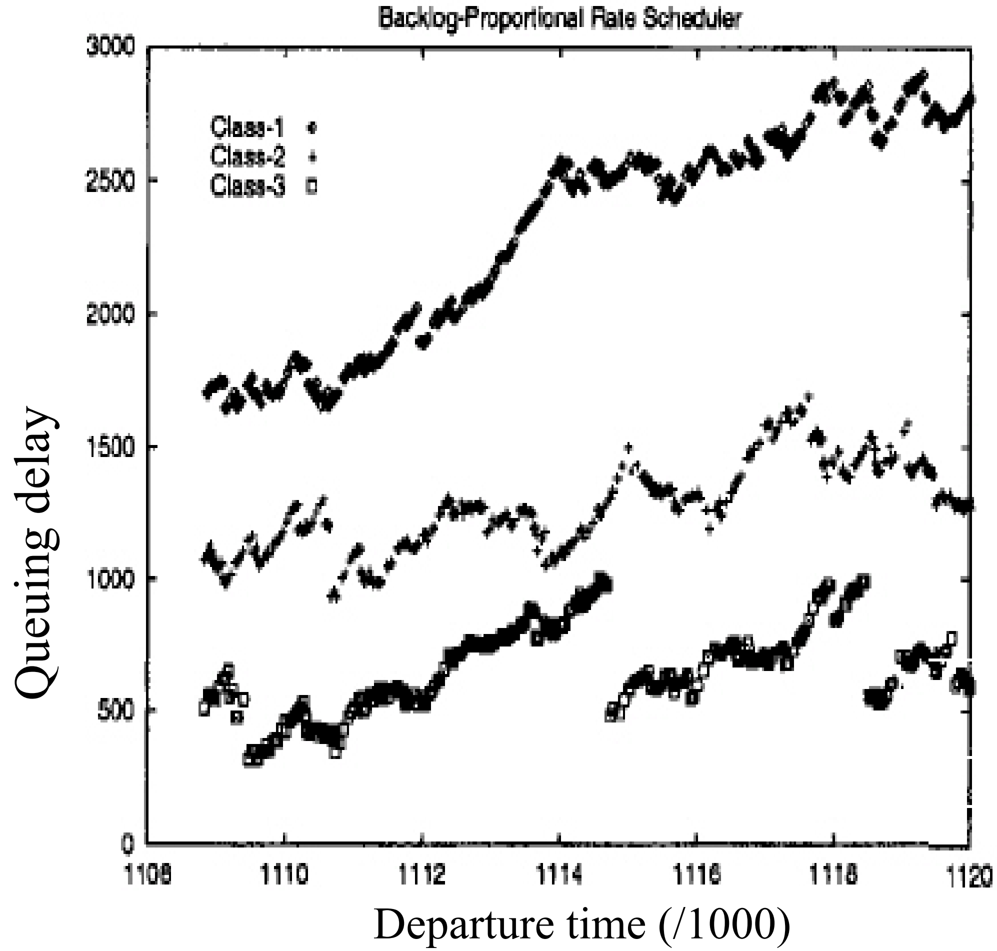
R_D : average of all ratios over all pairs of classes

★ WTP behaves well even in short time-scale

Simulation Study:Evaluation of BPR and WTP (6/7)



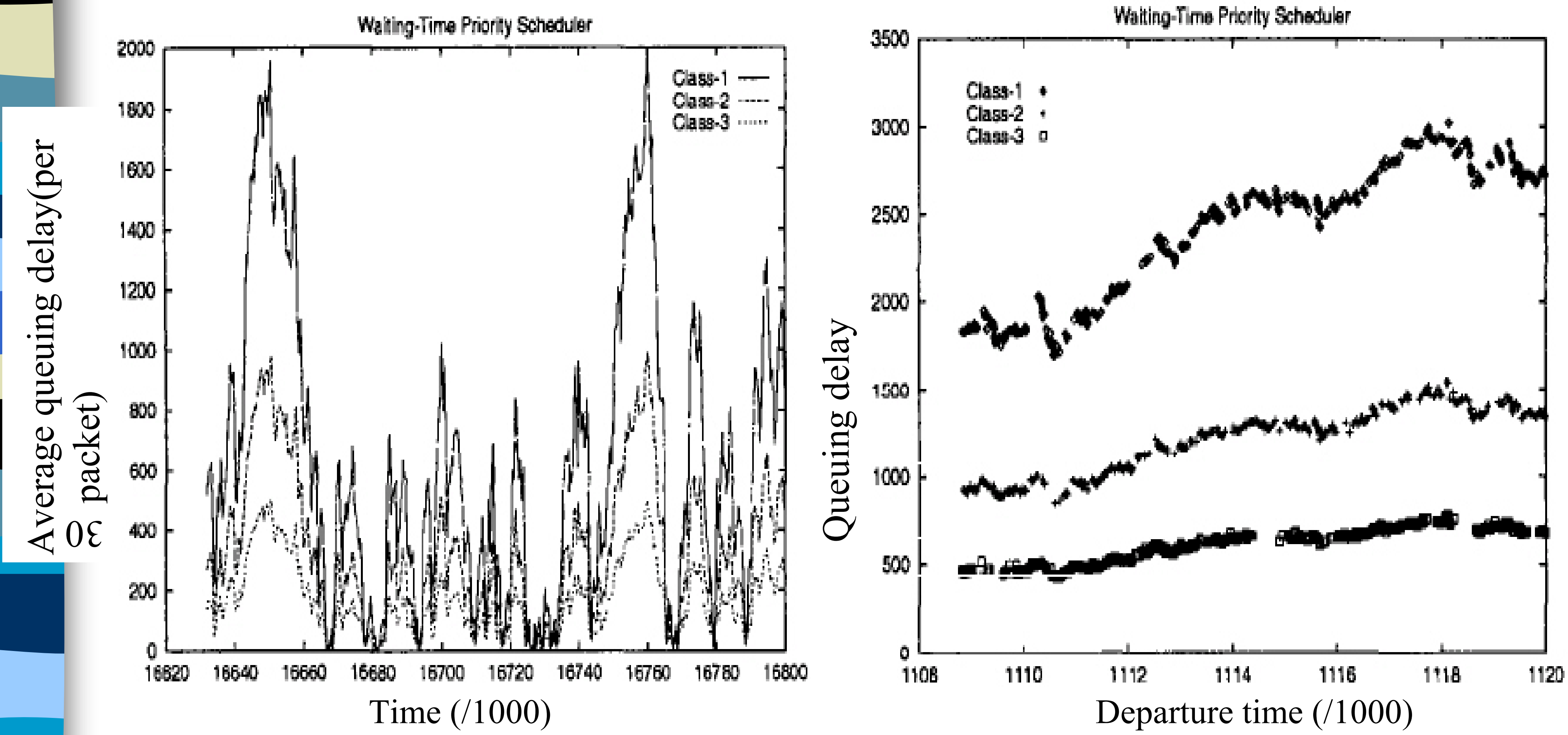
(a) Microscopic view I



(b) Microscopic view II

Queuing delays with the BPR scheduler when s1=1, s2=2, s3=4

Simulation Study:Evaluation of BPR and WTP (7/7)



(a) Microscopic view I

(b) Microscopic view II

Queuing delays with the WTP scheduler when s1=1, s2=2, s3=4

Simulation Study: The User's Perspective (1)

- Concern about the end-to-end performance of the packet flows
- Local class-based relative differentiation

  => end-to-end flow-based relative differentiation
- Simulation settings
  - Simulator: NS-v2
  - Cross-traffic is loaded in the network path
  - Propagation and transmission delays are ignored

Simulation Study: The User's Perspective (2)

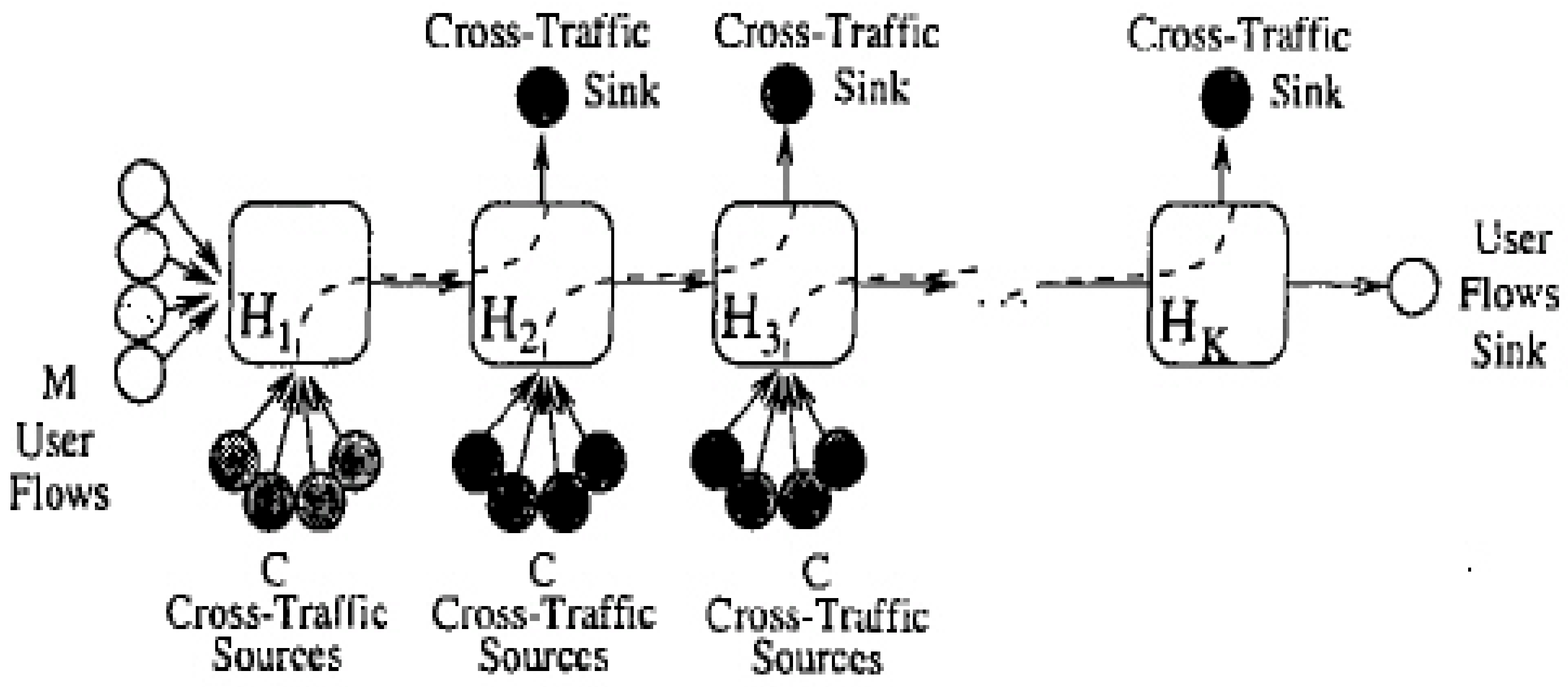


Figure 6: The multi-hop traffic configuration in this simulation study.

Simulation Study: The User's Perspective (3)

$\overline{R}_d$	F=10 R_u=50	F=10 R_u=200	F=100 R_u=50	F=100 R_u=200
K=4 ρ=85%	2.3	2.2	2.2	2.1
K=4 ρ=95%	2.1	2.1	2.1	2.0
K=8 ρ=85%	2.0	2.0	2.0	2.0
K=8 ρ=95%	2.0	2.0	2.0	2.0

Table 1: The metric $\bar{R}_D$: Ideally, it should be 2.00 in all cases. These results have been consistent in five simulation runs with different random seeds.

F : utilization of network link

R_u : rate of user flows

ρ : utilization of network link

K : number of hops in the path

$\overline{R}_d$ end-to-end delay ratio averaged over the N-1 pairs of successive classes

Simulation Study: The User's Perspective (4)

- No inconsistent delay differentiation in the runs
  - Inconsistent: higher class experience larger delay
- Increasing number of hops results in better $\overline{R_d}$

Conclusion and Open Issues

- Propose proportional differentiation model for controllable and predictable relative differentiation
- Identify and evaluate two schedulers in heavy load and short timescale
- What about the Additive differentiation model?
- Optimal scheduler of coupled delay and loss?
- Derive DDPs for various kind of network link by experiment
- Typical “profile” of feasible DDPs for network operators