The War Between Mice and Elephants

Liang Guo and Ibrahim Matta
Boston University
ICNP 2001

Presented By

Eric Wang

Outline

- Introduction
- Analyzing Short TCP Flow Performance
- Architecture And Mechanism
- Simulation
- Discussion
- Conclusion and Future Work

Short TCP Flows vs. Long TCP Flows

A real life example:

Mice and Elephants

- Elephants:

  Most traffic(80%) is carried out by only a small number of connections.

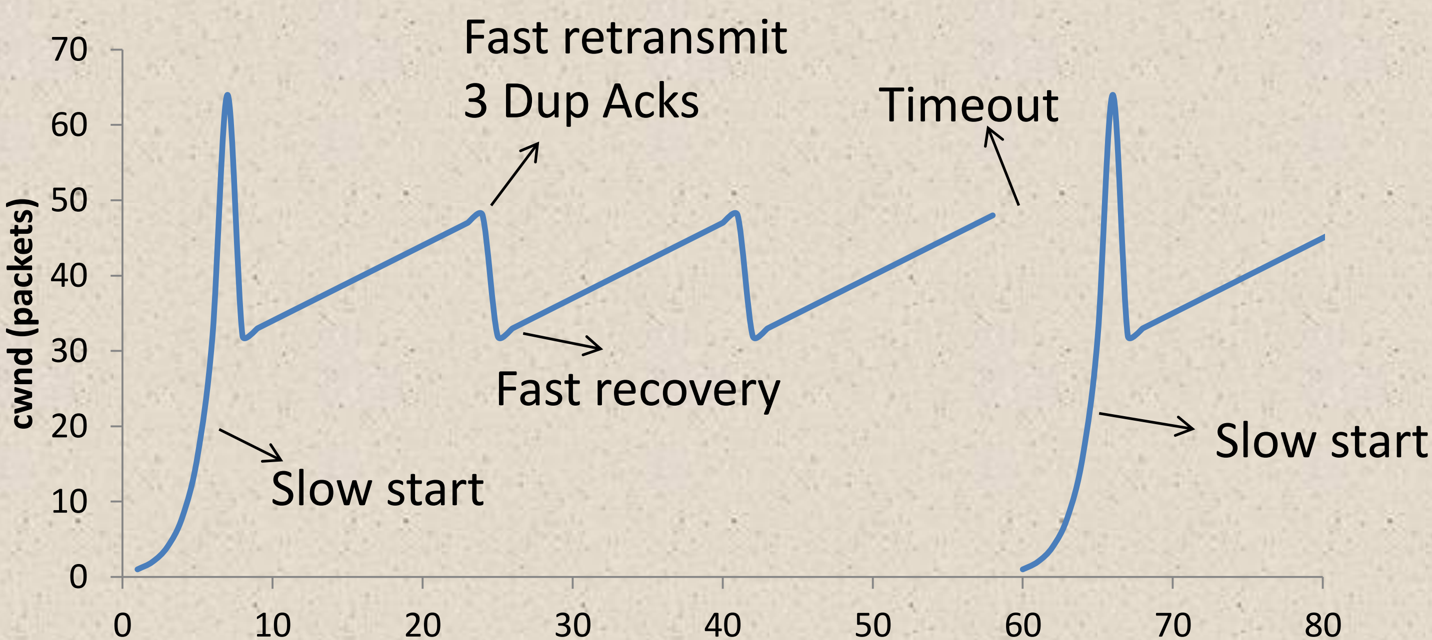
- Mice

  The remaining large amount of connections are very small in size or lifetime.

Is this really fair?

Short TCP Flows vs. Long TCP Flows

- In a fair network
  - Short connections expect relatively fast service compared to long connections
- Sometimes this is not the case with Internet



Unfair for Short flows Due to TCP Nature

- **TCP slow start**
  Sending window is initiated at minimum value regardless of what is available in the network.

- **Packet Loss detected by timeout or duplicate ACK**
  Sending window is initiated at minimum value regardless of what is available in the network.

- **ITO as initial value for RTO**
  For the first control packets and first data packets, TCP has to use ITO value as RTO, losing these packets can have disastrous effect on short connection performance.

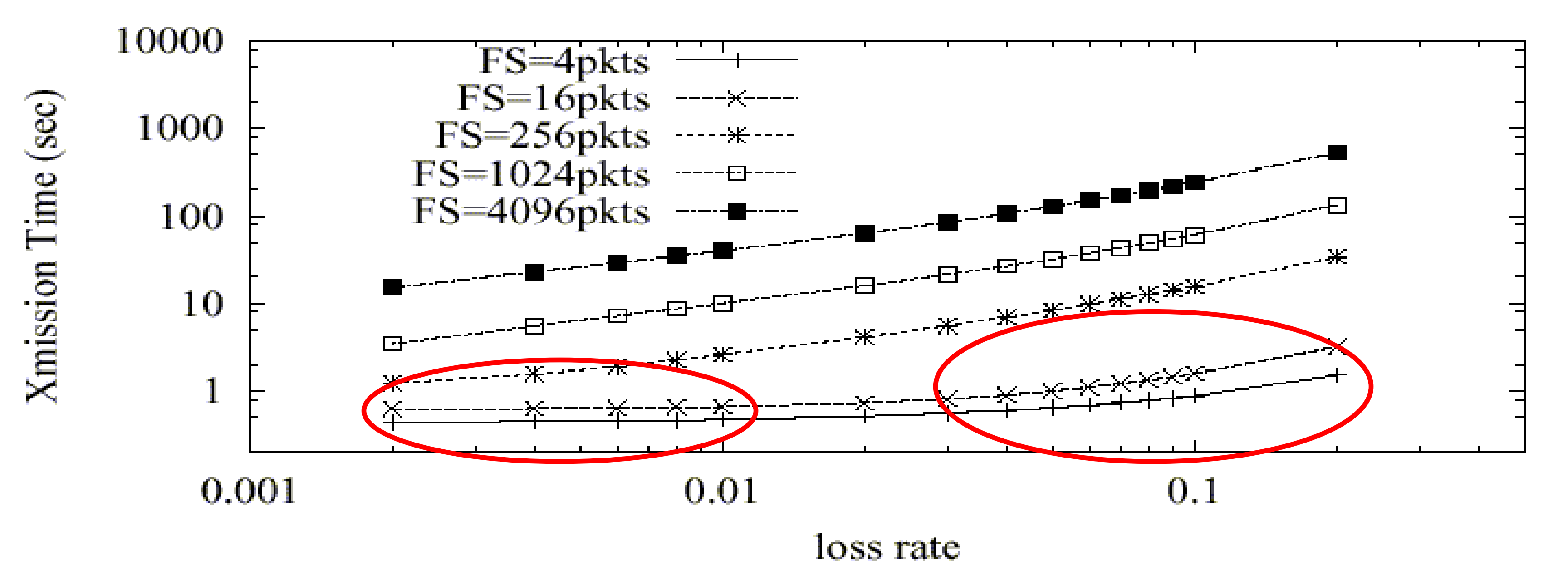
Proposed solution:

Active Queue Management + Differential Services(Diffserv)

Outline

- Introduction
- Analyzing Short TCP Flow Performance
- Architecture And Mechanism
- Simulation
- Discussion
- Conclusion and Future Work

Sensitivity Analysis of TCP flows to Loss Rate



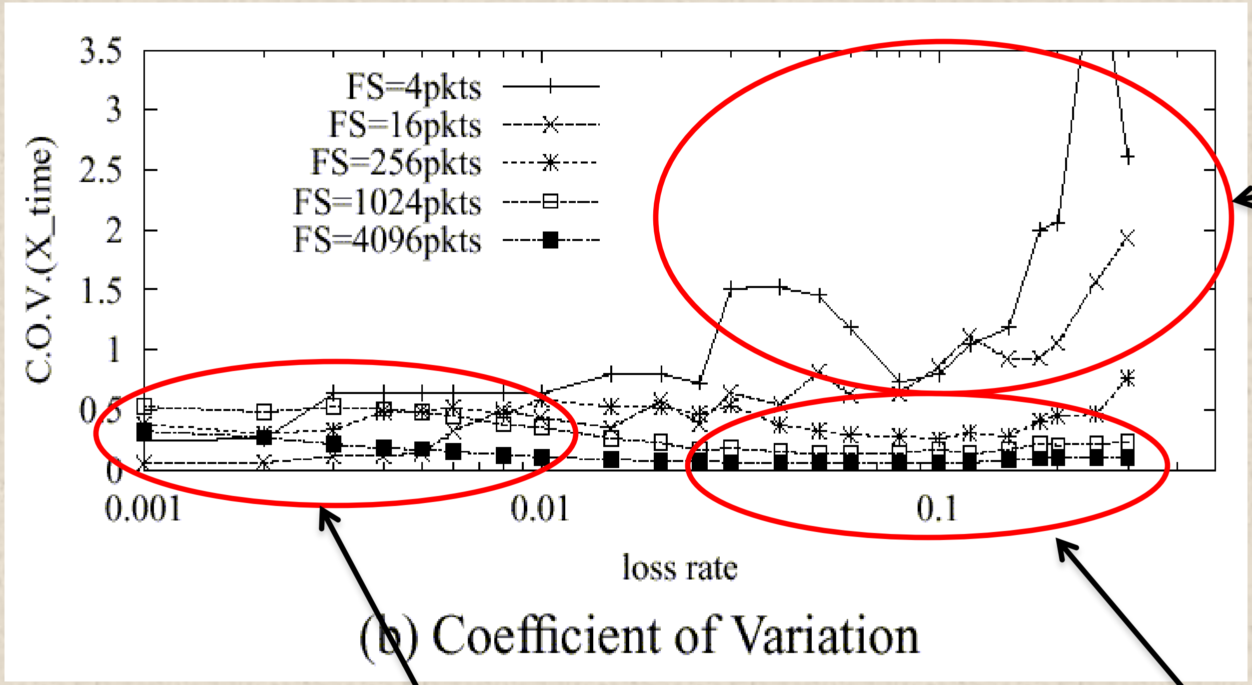
(a) Average Transmission Time

Average transmission time of Short TCP flows are not very sensitive to loss rate when loss rate is relatively small.

But it increase drastically as loss rate becomes larger (persistent congestion).

Variance of Transmission Times

COV = Standard deviation/mean



(b) Coefficient of Variation

Variability in short flows
Due to 1.
Law of large numbers

Variability in long flows
Due to 2.
Loss in slow start or congestion avoidance

Less variability in long flows
Loss in both slow start and congestion avoidance

Conclusions

- Short flows are more sensitive to increase of loss rate than long flows.
- For short flows, variability of transmission time is more sensitive to increase of loss rate

Preferential Treatment to Short flows

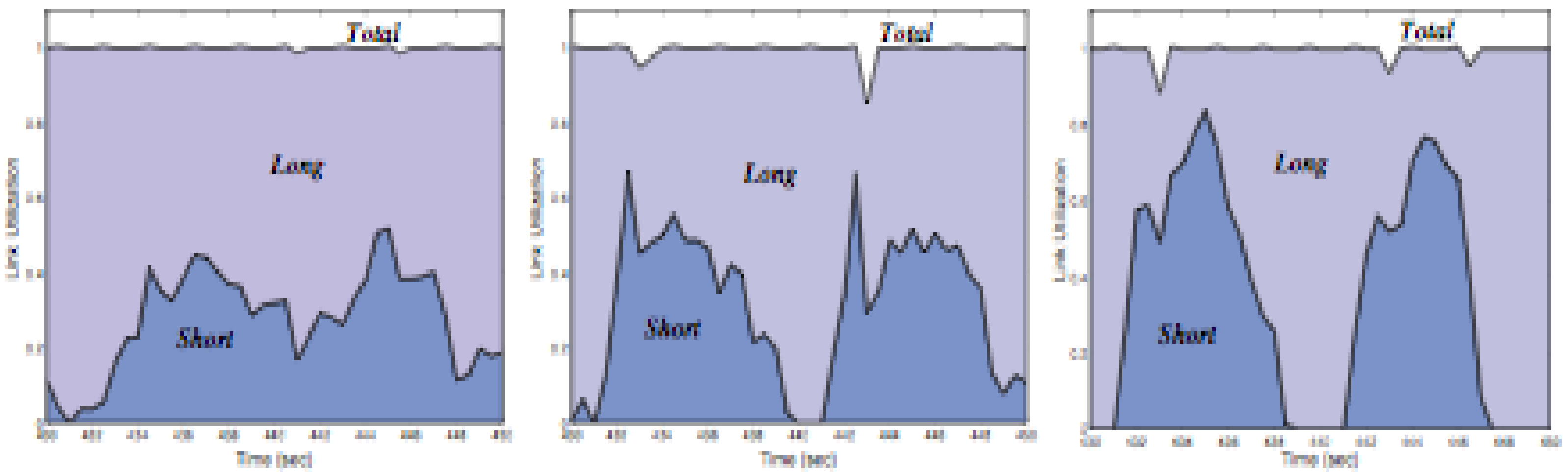


Fig. 3. Impact of Preferential Treatment— Link utilization under Drop Tail (left), RED (middle), and RIO-PS (right)

Drop Tail fails to give fair treatment to short TCP flows

RED gives almost fair treatment to all flows

RIO favors short flows by giving more than their fair share

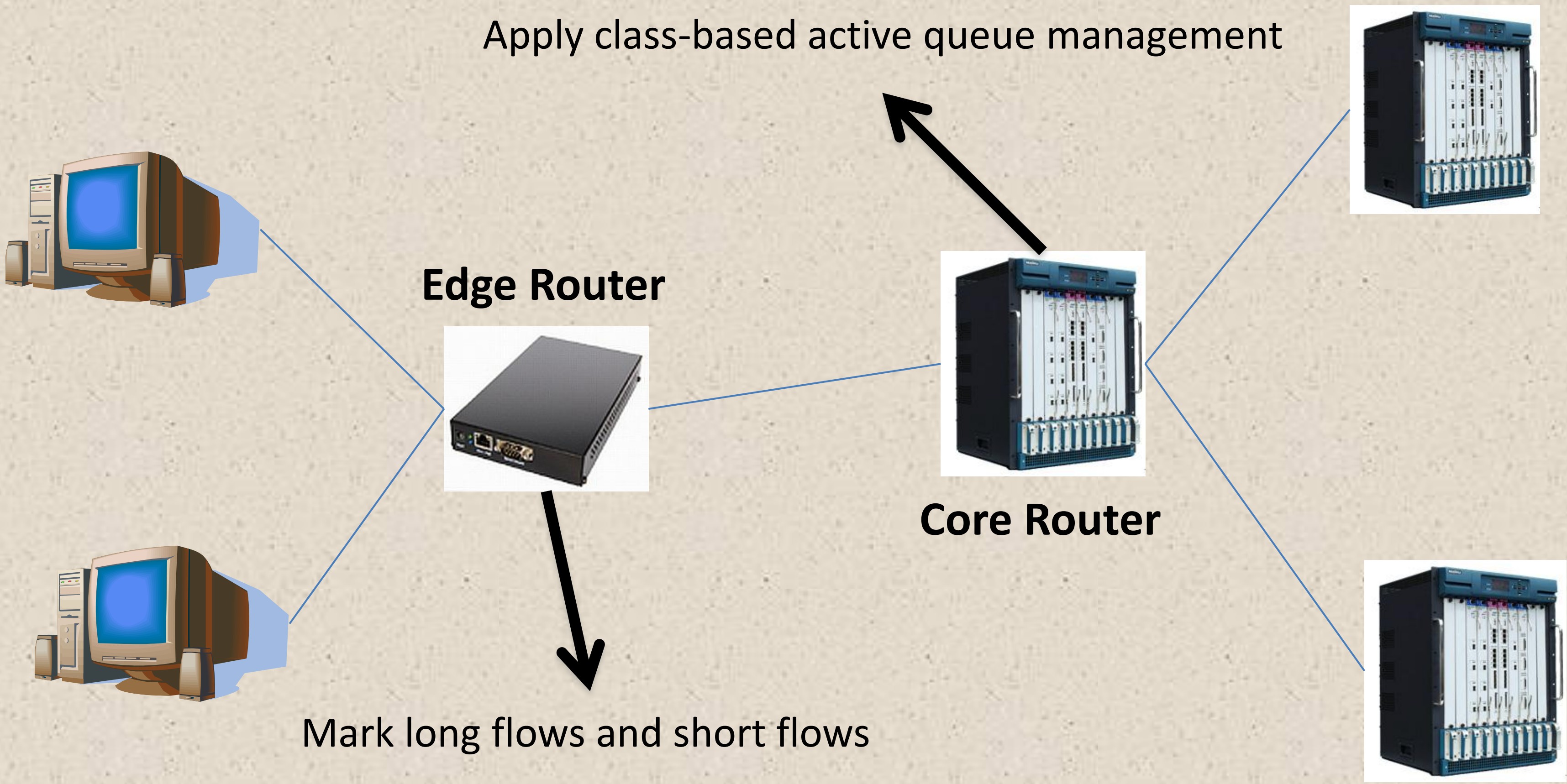
Why Using RIO for short flows?

- Short flows ends earlier, giving back resources to long flows.
- May even enhance long flows since they are less disturbed by short flows.
- Faster response time and better fairness for short flows, thus enhance the overall performance.

Outline

- Introduction
- Analyzing Short TCP Flow Performance
- Architecture And Mechanism
- Simulation
- Discussion
- Conclusion and Future Work

Proposed Architecture



Edge Routers

- Marking packets as from long flow and short flow
  - Setting a counter for each flow and a threshold L_t
  - When counter exceeds L_t , mark packets as from long flow, otherwise from short flow

- Maintaining per-flow state information
  - A flow hash table is updated every T_u time units.

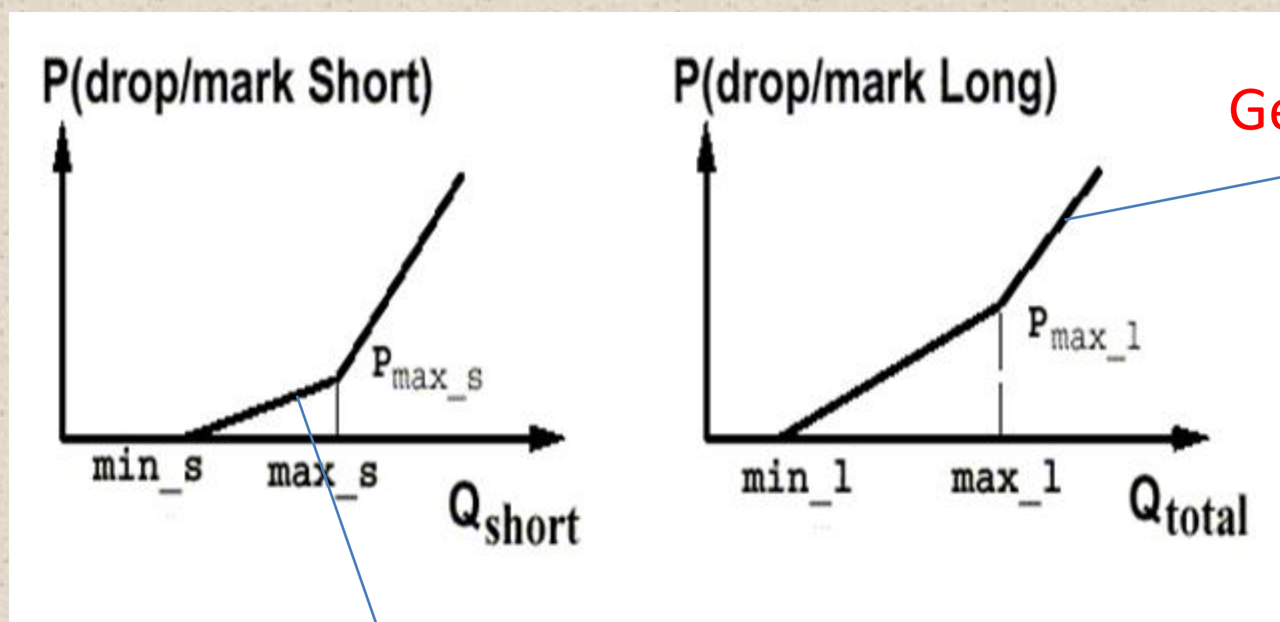
- Dynamically adjusting L_t to maintain SLR
  - SLR (Short-to-long-Ratio)
  - Maintain SLR by doing additive increase/decrease to L_t

Core Router – RIO-PS

- RIO - RED with In (Short) and Out (Long)
- Preferential treatment to short flows
  - Short flows
    - Packet dropping probability computed based on the average backlog of short packets only (Q_{short})
  - Long flows
    - Packet dropping probability computed based on the total average queue size (Q_{total})

RIO-PS

Two separate sets of RED parameters for each flow class



Gentle RED

Less Packet dropping probability for short flows

Features of RIO-PS

- Single FIFO queue is used for all packets
  - Packet reordering will not happen
- Inherits all properties of RED
  - Protection of bursty flows
  - Fairness within each class of traffic
  - Detection of incipient congestion

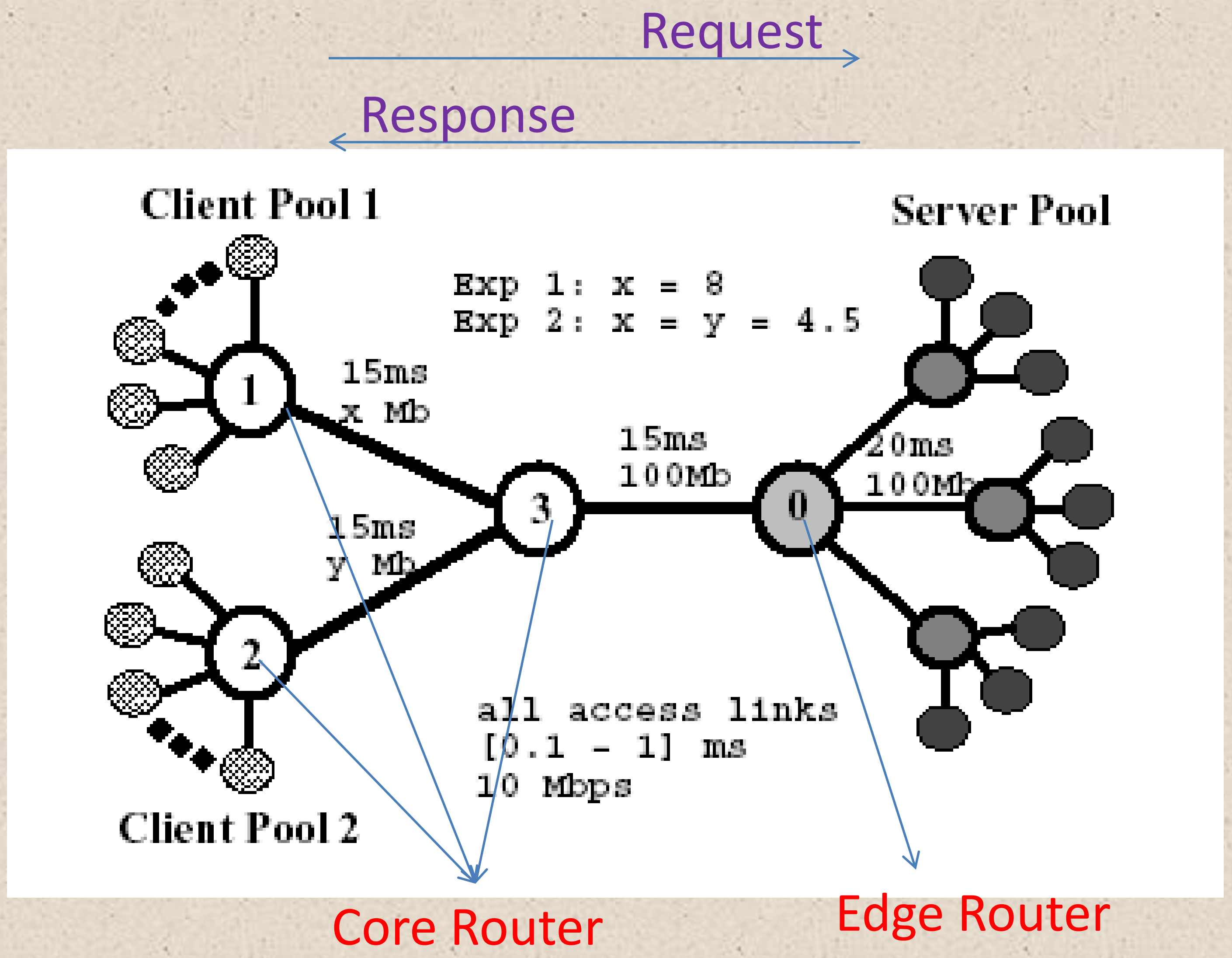
Outline

- Introduction
- Analyzing Short TCP Flow Performance
- Architecture And Mechanism
- Simulation
- Discussion
- Conclusion and Future Work

Simulations setup

- ns-2 simulations
- Web traffic model
  - HTTP 1.0
  - Exponential inter-page arrival (mean 9.5 sec)
  - Exponential inter-object arrival (mean 0.05 sec)
  - Uniform distribution of objects per page (min 2 max 7)
  - Object size; bounded Pareto distribution (min = 4 bytes, max = 200 KB, shape = 1.2)
  - Each object retrieved using a TCP connection

Simulation topology



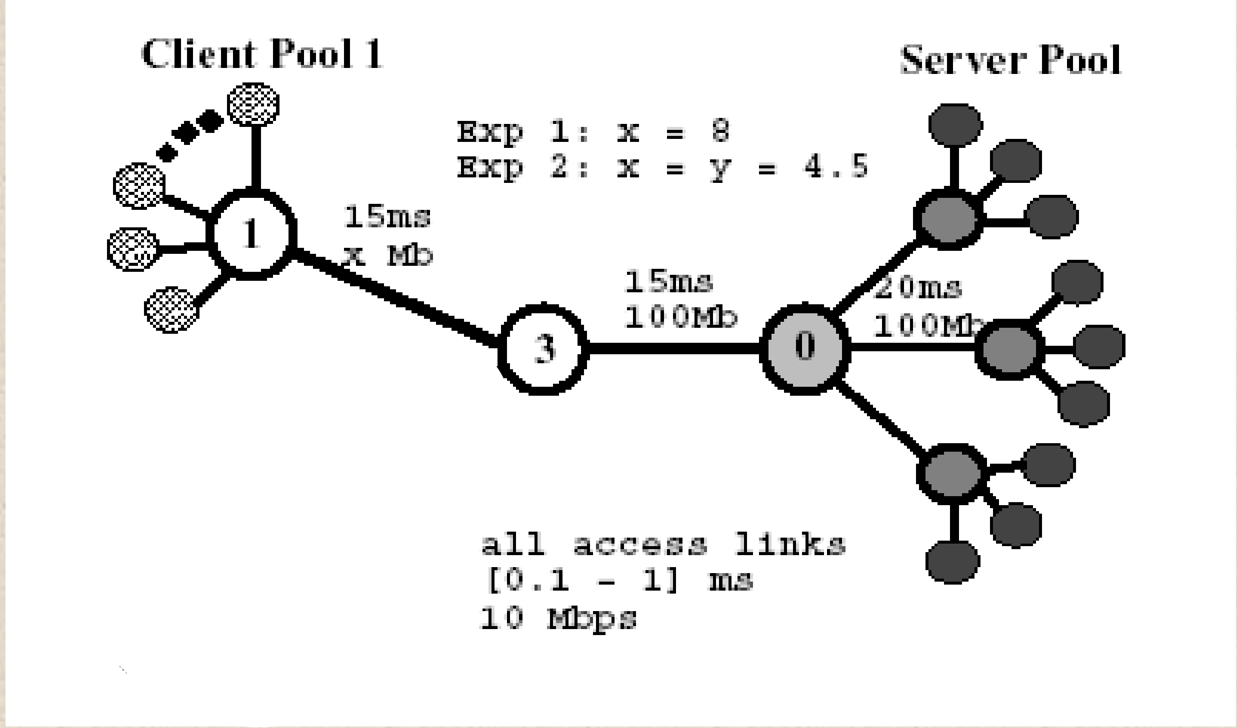
Network configuration

Description	Value
Packet Size	500 bytes
Maximum Window	128 packets
TCP version	Newreno
TCP timeout Granularity	0.1 seconds
Initial Retransmission Timer	3.0 seconds
B/W delay product (BDP)	≈ 200 pkts (Exp1) ≈ 120 pkts (Exp2)
Bottleneck Buffer Size (B)	DropTail: 1.5× BDP RED/RIO-PS: 2.5×BDP
Q. Parameters	$(min_{th}, max_{th}, P_{max}, w_q)$
RED	(0.15B, 0.5B, 1/10, 1/512)
RIO-PS short	(0.15B, 0.35B, 1/20, 1/512)
RIO-PS long	(0.15B, 0.5B, 1/10, 1/512)
RED & RIO-PS	ecn_on, wait_on, gentle_on
Edge Router	$SLR = 3, T_u = 1\ sec, T_c = 10\ sec$
Foreground Traffic	
(Src, Dest)	(Server Pool, Client Pool)
Long Connection Size	1000 packets
Short Connection Size	10 packets

Simulations details

- The load is carefully tuned to be close to the bottleneck link capacity
- RIO parameters
  - Short TCP flows are guaranteed around 75% of the total bandwidth in times of congestion
- Experiments run 4000 seconds with a 2000 second warm-up period

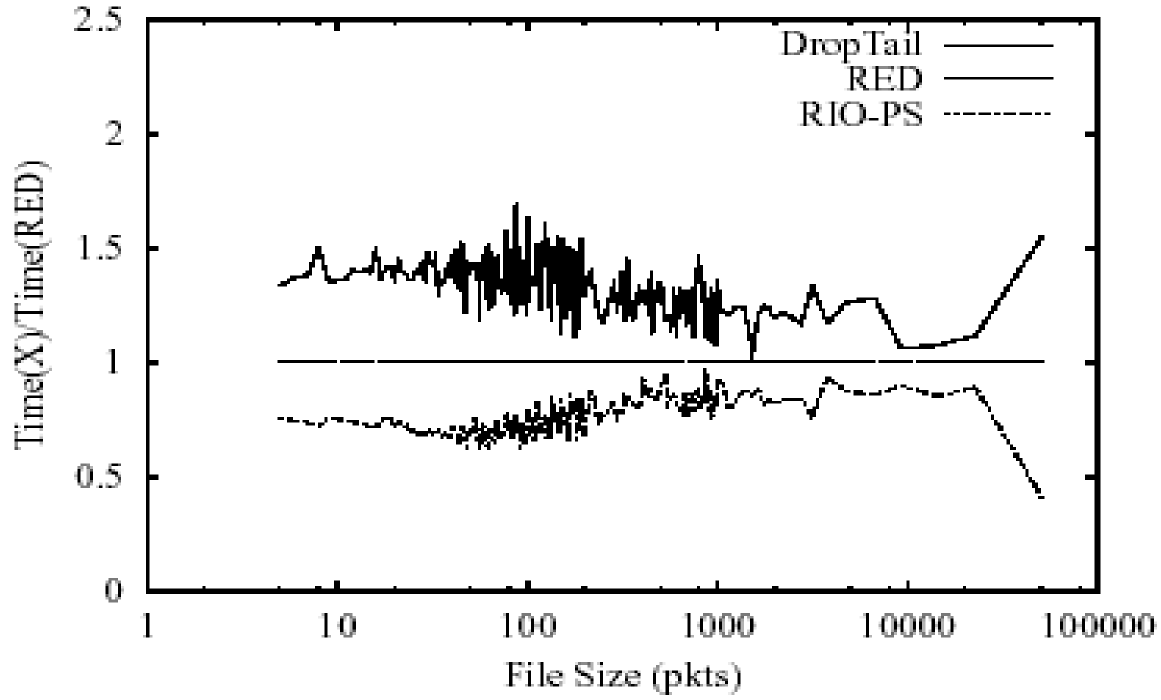
Experiment 1: Single Client Set



In this experiment, there is only one set of clients involved (client pool 1). Therefore, the traffic seen at the core router 1 is the same as that at edge router 0.

Average Response Time for Different sized objects

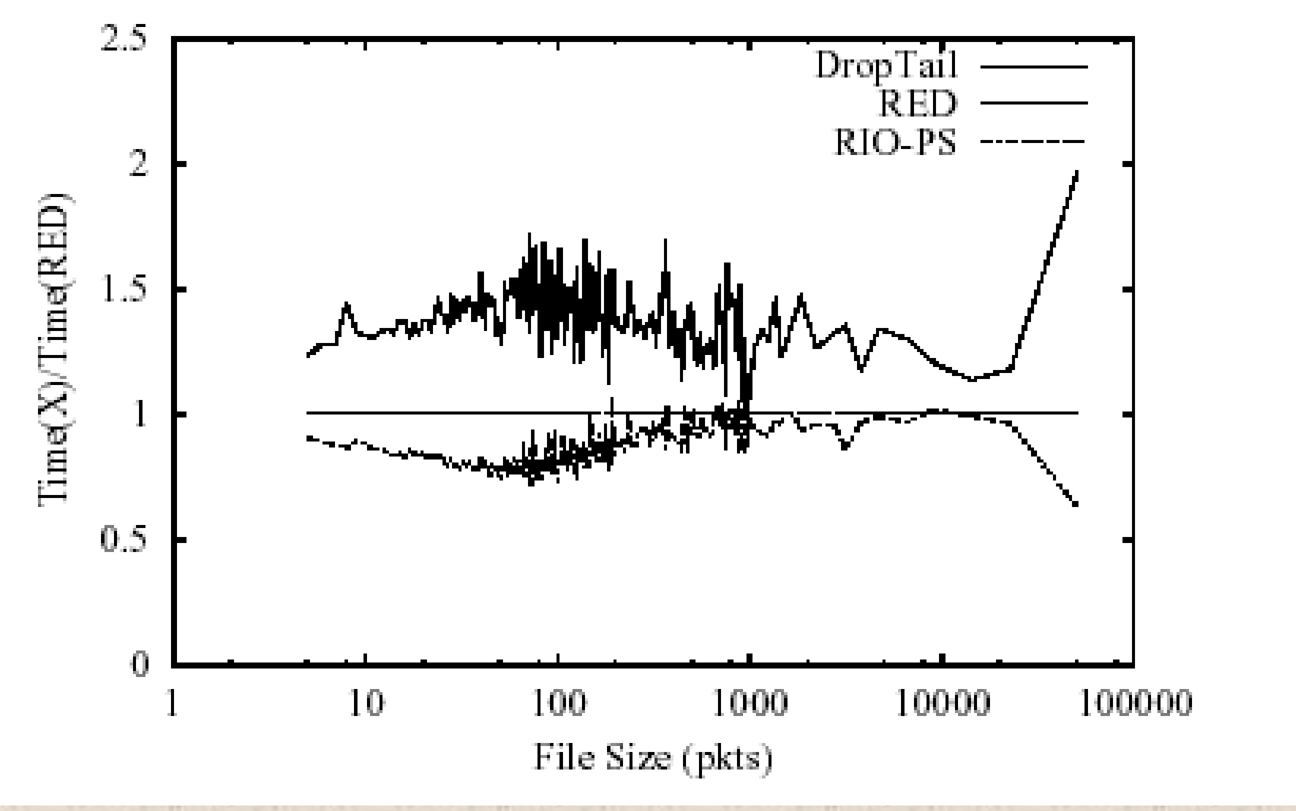
ITO = 3 sec



Preferential treatment can cut the average response time for short and medium sized files significantly (25-30 %)

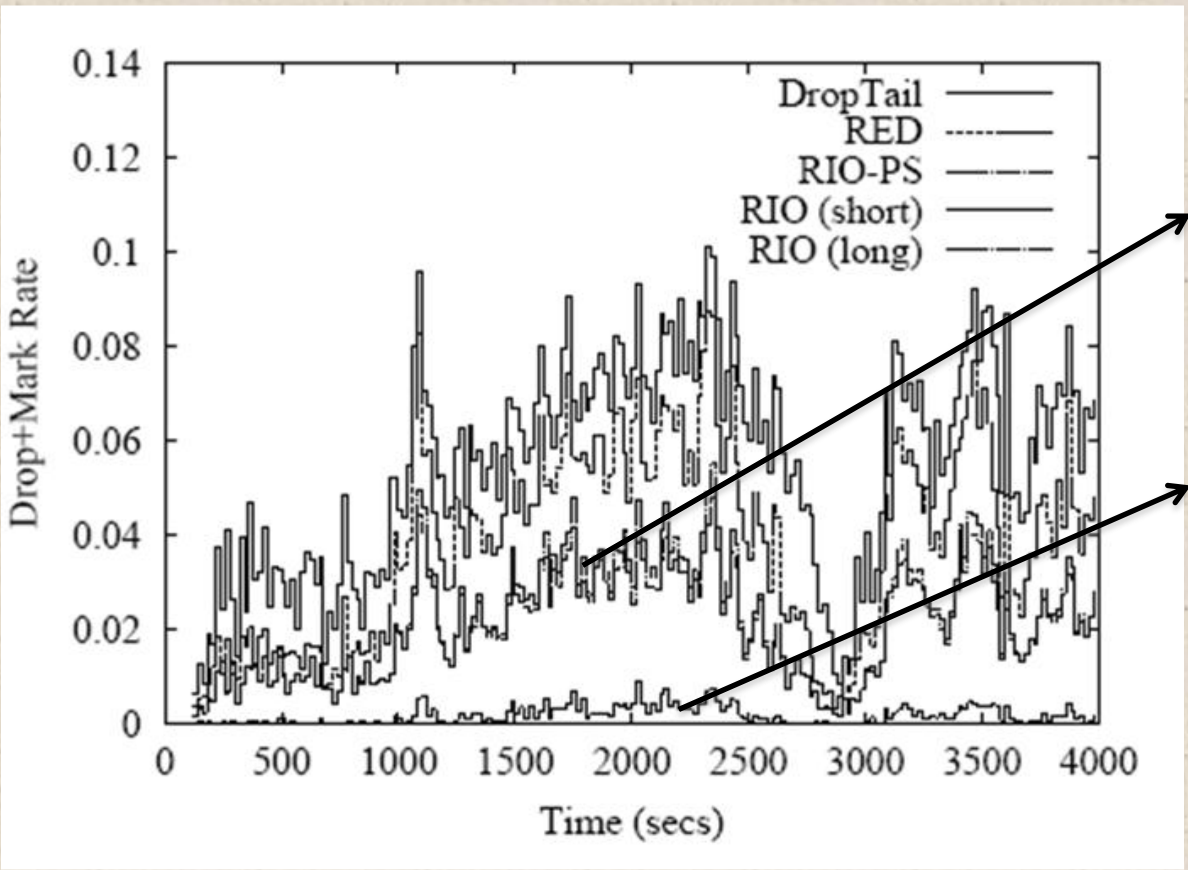
Average Response Time for Different sized objects

ITO = 1 sec



1. Significantly reducing the gap between RED and proposed scheme
2. Still large improvements with RIO-PS for medium sized connections(15%-25%).

Instantaneous Drop/Mark rate



RIO-PS reduces the overall drop/mark probability

Comes from the fact that short flows rarely experience loss

Also, Short TCP flows are not responsible for controlling congestion because of the time scale at which they operate.

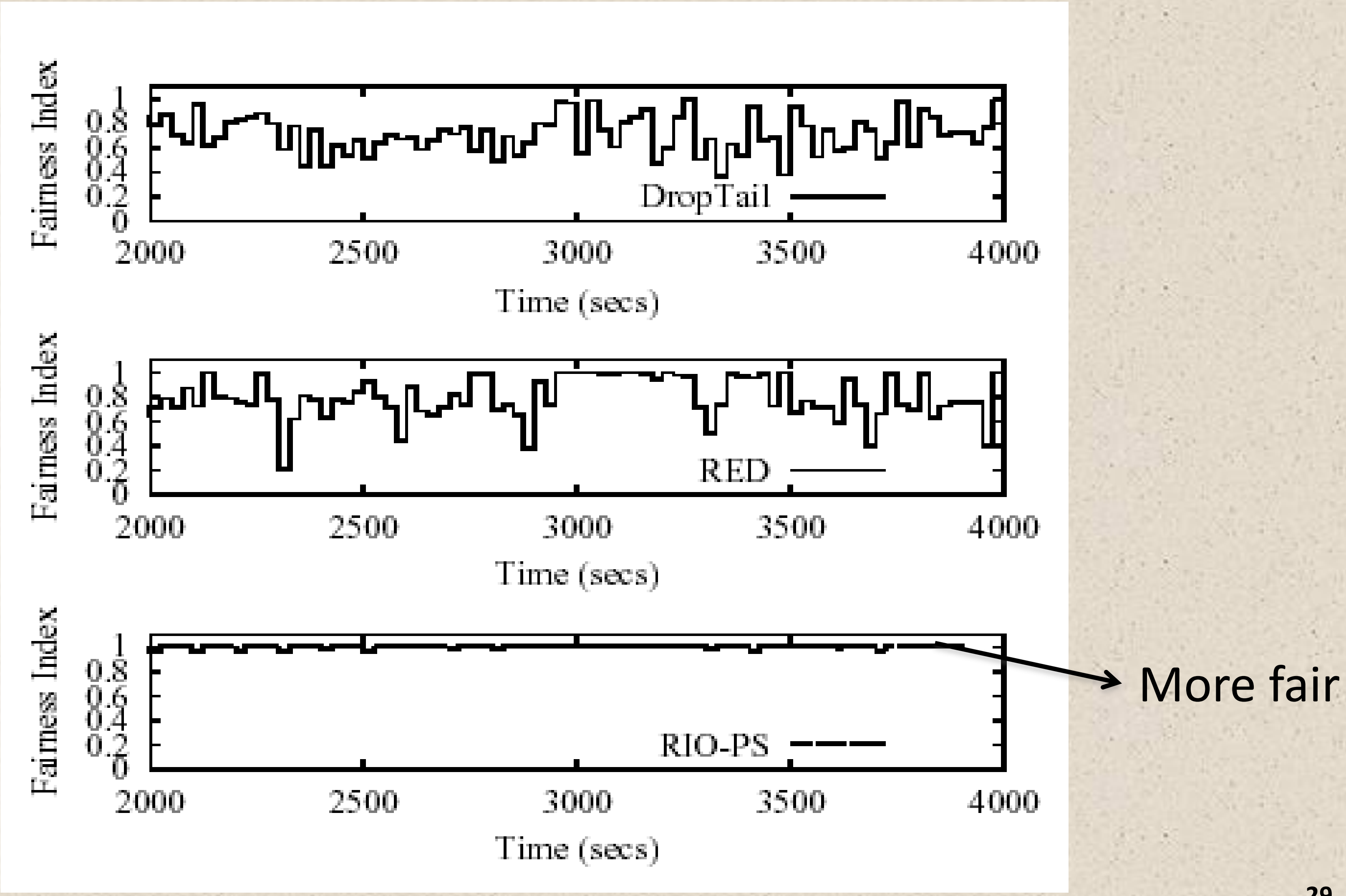
Preferential treatment to short flows does not hurt the network

Study of foreground traffic

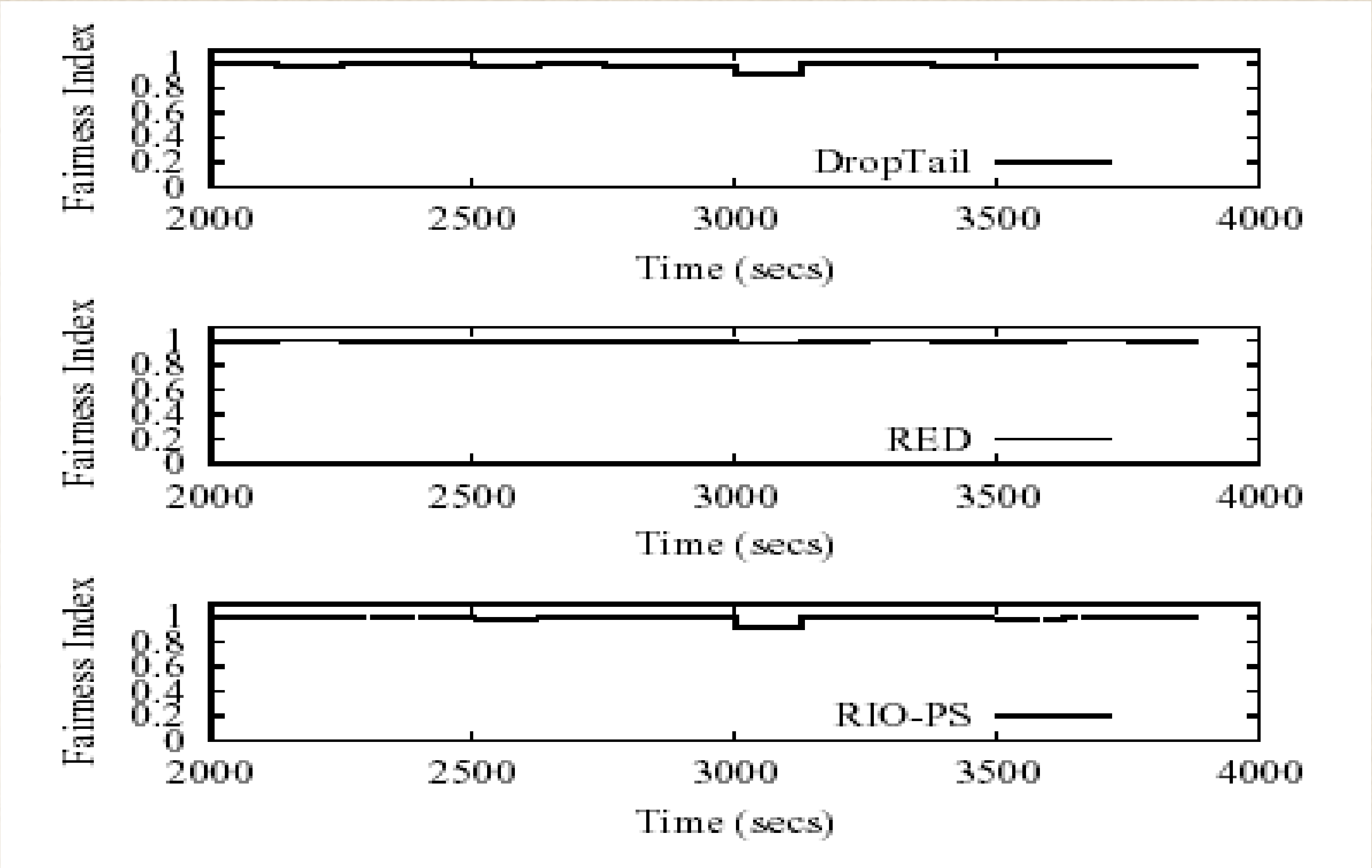
- Periodically inject 10 short flows (every 25 seconds) and 10 long flows (every 125 seconds) as foreground TCP connections and record the response time for i_{th} connection
- Fairness index
  - For any give set of response times (x_1, .., x_n), the fairness index is:

$$\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i^2}$$

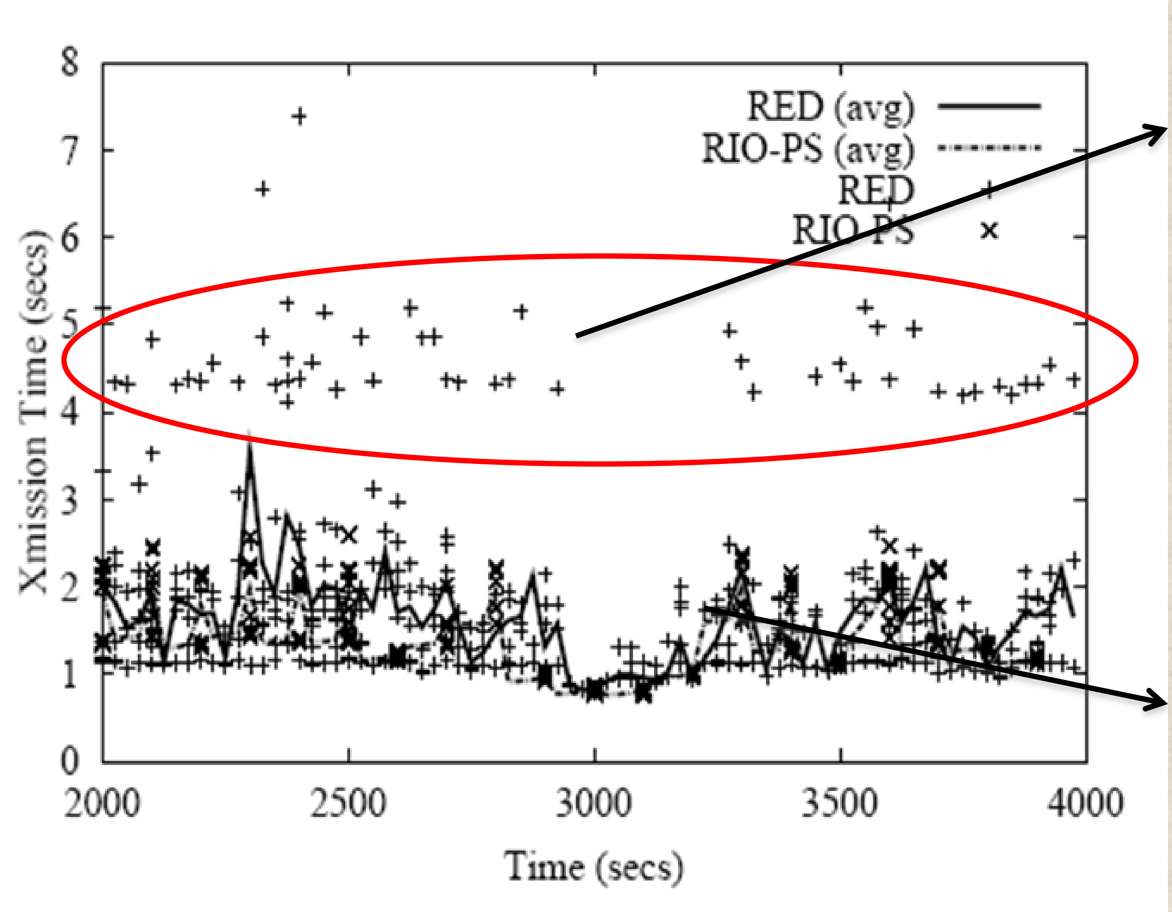
Fairness Index – Short Connections



Fairness Index – Long Connections



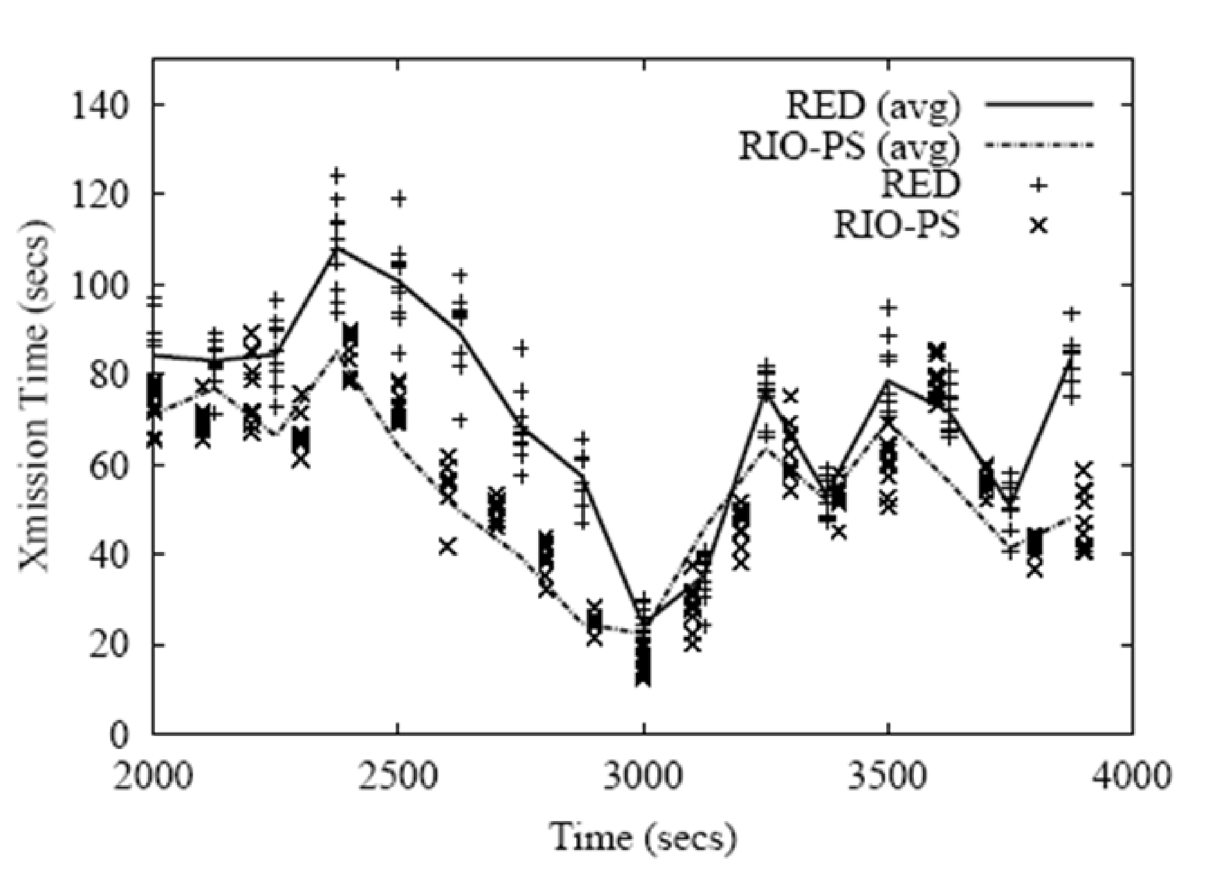
Transmission time – short connections



-Even with RED queues, many short flows experience loss
-Some lost first packet and hence timeout (3 sec)

RIO-PS
much less drops

Transmission time – long connections



RIO-PS does not hurt long flow performance

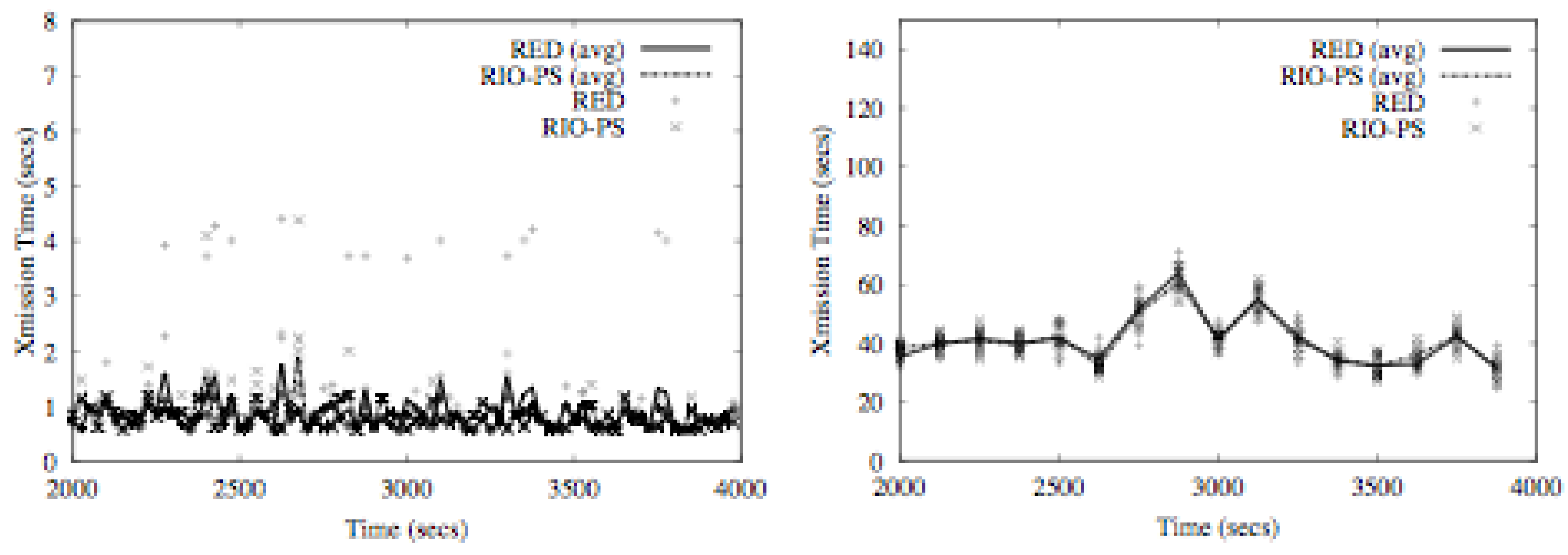
Goodput

Scheme	DropTail	RED	RIO-PS
Exp1 (ITO=3sec)	4207841	4264890	4255711
Exp1 (ITO=1sec)	4234309	4254291	4244158
Exp2 (ITO=3sec)	4718311	4730029	4723774

RIO-PS does not hurt overall goodput

Slightly improves over DropTail

Experiment 2: Unbalanced Request



(c) Transmission Time of Short Connections

(d) Transmission Time of Long Connections

When router is dominated by one class of flows (short or long), the proposed method reduces to traditional unclassified traffic plus RED queue policy.

Outline

- Introduction
- Analyzing Short TCP Flow Performance
- Architecture And Mechanism
- Simulation
- Discussion
- Conclusion and Future Work

Discussion

- Deployment Issues
- Flow Classification
- Controller Design

Outline

- Introduction
- Analyzing Short TCP Flow Performance
- Architecture And Mechanism
- Simulation
- Discussion
- Conclusion and Future Work

Conclusion

- TCP major traffic in the Internet
- Proposed Scheme is a Diffserv like architecture
  - Edge routers classifies TCP flow as long or short
  - Core routers implements RIO-PS
- Advantages
  - Short flow performance improved in terms of fairness and response time.
  - Long flow performance is also improved or minimally affected since short flows are rapidly served.
  - System overall goodput is improved
  - Flexible Architecture, can be tuned largely at edge routers