

The War Between Mice and Elephants

(by Liang Guo and Ibrahim Matta)

Treating Short Connections fairly against Long Connections when they compete for Bandwidth.

**Advanced Computer Networks
CS577 – Fall 2013
WPI, Worcester.**

**Presented by Pankaj Didwania
Sep.24th, 2013**

Outline

- Introduction
- Analyzing Short TCP Flow Performance
 - Sensitivity Analysis
 - Preferential Treatment
- Proposed Architecture and Mechanism
 - The Architecture
 - Edge Router – Packet Classification and State Maintenance
 - Core Router : Preferential Treatment to Short Flows

Contd...



Outline - Contd...

- Simulation
 - Simulation Setup
 - Experiment 1 : Single Client Set
 - Experiment 2 : Unbalanced Requests
- Discussion
 - Comments on Simulation Model
 - The Queue Management Policy
 - Deployment Issues
 - Flow Classification
 - Controller Design
 - Malicious Users
- Conclusion and Future Work



Introduction

- This paper highlights and resolves the 80-20 rule as applicable to Internet traffic.
- 80% of the traffic is actually carried by a small number of connections
 - the Elephants.
- And only the remaining 20%, large number of connections are very small in size or lifetime
 - the Mice.



Introduction contd...

- Short TCP Flows vs. Long TCP Flows, an example.
- In a Fair Network: the Short Connections expect faster service in comparison with their Long counterparts.
- However this is not true for Internet scenarios.
- Let's see why and what the authors recommend.



TCP characteristics

- TCP was originally designed for elephants.
- TCP slow start: Sending windows gets initiated at a minimum value without considering available network resources.
- TCP couples error control with congestion control.
- TCP depends upon timeout (vs. duplicate ACK mechanism) to detect packet loss for Short connection.



TCP characteristics

- TCP relies on its own packet samples to estimate an retransmission timeout (RTO) value.
- TCP uses conservatively estimated initial timeout (ITO) for the first control and data packets.
- This causes TCP flows to be more conservative for short connections and tend to get less than their fair share.

Approach

- Preferential treatment to ensure prompt responses to short TCP flows.
- Threshold based classification method.
- Active Queue Management (AQM) – RIO at core routers.
- Differentiated Services (Diffserv) architecture at the edge of networks.
- This approach achieves better goodput than traditional Drop Tail or RED policies.
- RIO guarantees ordered delivery of packets.



Related Work

- Authors :
 - study interaction between long and short flows.
 - propose to isolate long and short flows.
 - discover that ‘class based flow isolation’ in combination with ‘threshold based classification’ at the edge cause packet reordering and severely degrade TCP performance.
 - propose to push the bandwidth(load) control to the edges of the network.



Analyzing Short TCP Flow Performance

- Relationship between loss rate and TCP flow transmission.
- Sensitivity Analysis for Short and Long TCP flows.
- Preferential Treatment to Short TCP Flows.

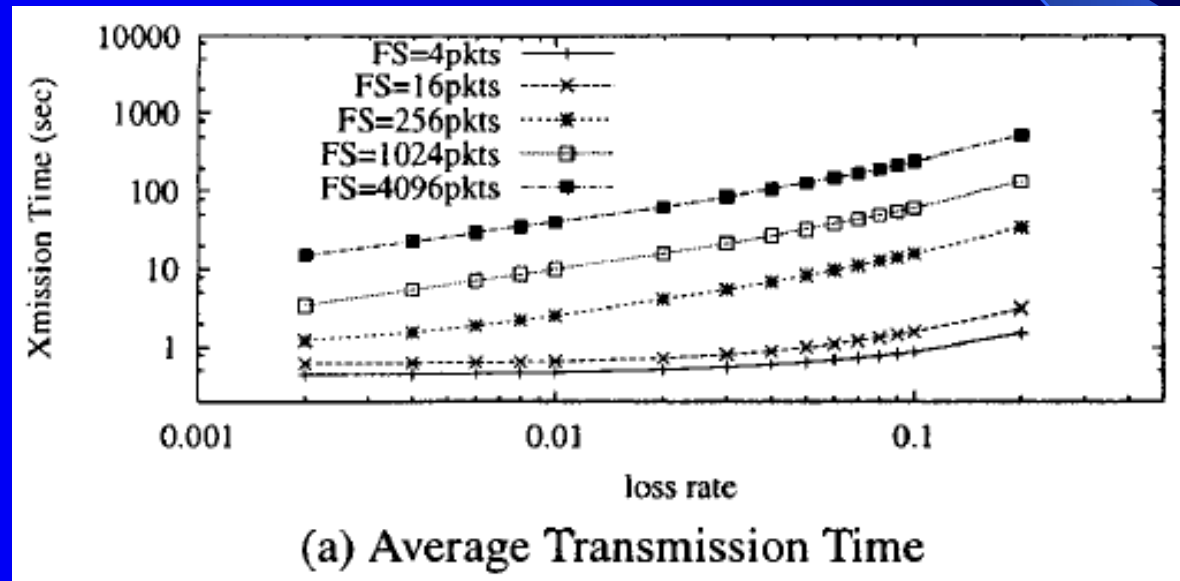


Sensitivity Analysis for Short and Long TCP flows

- In this section authors provide the analytical results on the transmission time for TCP flows of different sizes.
- It is observed that the average transmission time of short flows is not very sensitive to loss when the loss rate is relatively small. But it increases drastically as loss rate becomes larger (when persistent congestion happens).

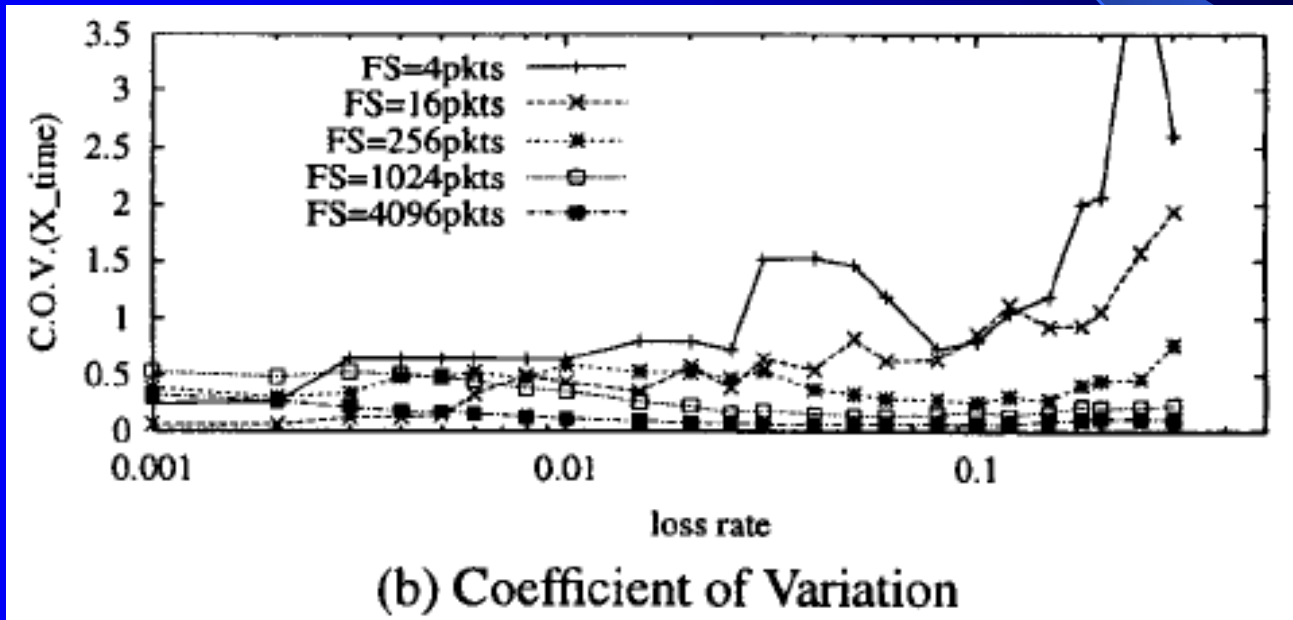
Sensitivity Analysis...

- Figure gives in a log-log plot the average total latency with avg RTT = 0.1 second, avg. RTO = 4 x RTT and the default initial retransmission timer ITO = 3 seconds, for a TCP flow of a fixed size FS for various loss rates.



Sensitivity Analysis...

- Figure plots C.O.V. against loss rate. Notice the trend - for small size TCP flows, increasing the loss probability can lead to increased variability, while for long TCP flows, large loss rate reduces the variability of transmission times.



Sensitivity Analysis...

- When Loss rate is high, TCP congestion control is more likely to enter the exponential back-off phase.
- When Loss rate is low, depending on when the packet loss occurs TCP can either transmit a significant amount of packets in slow-start phase or have to transmit them in the less aggressive congestion avoidance phase.
- Since the first source of variability is on individual packets of a flow, the law of large numbers indicates that its impact is more significant on short flows.
- The second source of variability is more pronounced for long flows since most short flows finish their transmission in slow-start phase.
- Authors thus conclude that reducing the loss probability is more critical to help short TCP flows experience less variations in transmission(response) time.



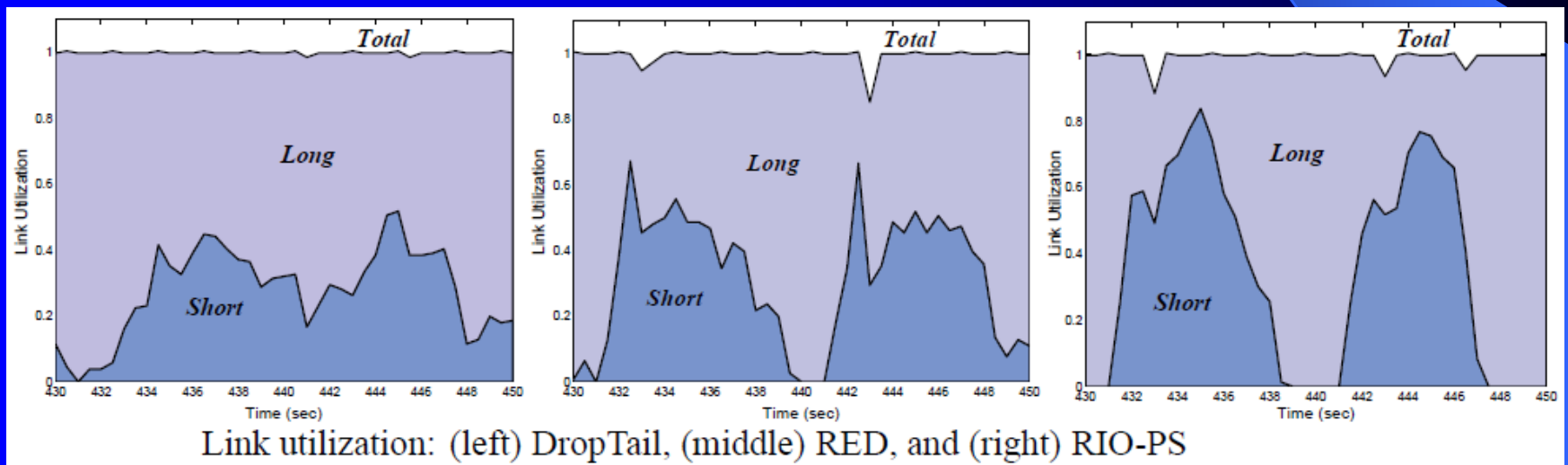
Preferential Treatment to Short TCP Flows

- Authors simulate the following scenario
 - Using ns Simulator
 - 10 Long(10000 packet) TCP-Newreno flows
 - 10 Short(100-packet) TCP-Newreno flows
 - Competing for bandwidth over a 1.25 Mbps link
 - Authors then vary the queue management policy at the bottleneck link and measure the instantaneous portion of bandwidth taken by each class of flows to show the effect of preferential treatment.
 - The results of Drop Tail Queue, RED Queue and the proposed RIO-PS(RIO with preferential treatment to Short flows) in the plot.
(left to right on the next slide).



Preferential Treatment to Short TCP Flows...

-Drop Tail Queue, RED Queue and the proposed RIO-PS (left to right).



Preferential Treatment to Short TCP Flows...

- Table below gives measured network goodput over the 500 seconds simulation period.
- The table also shows the measured goodput for a less loaded network with bottleneck link bandwidth of 1.5 Mbps

Link B/W	Flows	DropTail	RED	RIO-PS
1.25Mbps	All	153479	154269	154486
	Short	40973	49897	49945
	Long	112506	104372	104541
1.5Mbps	All	185650	184315	183154
	Short	43854	49990	49990
	Long	141796	134325	133164

TABLE I

NETWORK GOODPUT UNDER DIFFERENT SCHEMES

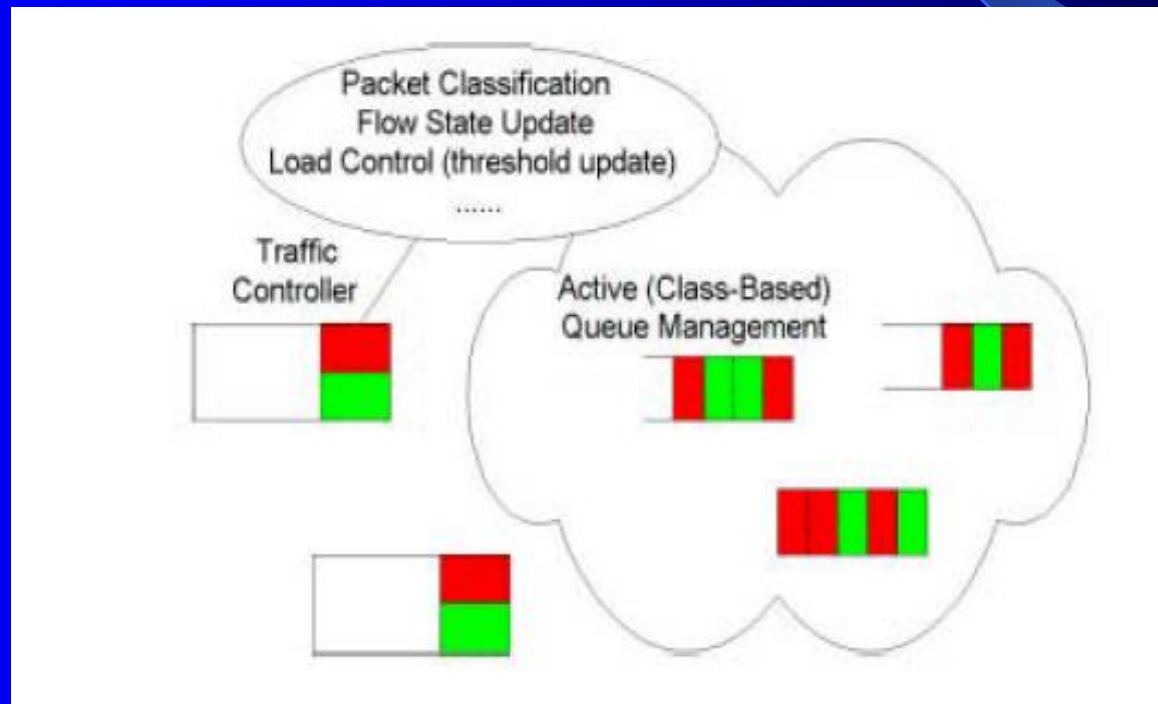
Proposed Scheme: Architecture & Mechanisms

- Architecture
- Edge Router: Packet Classification and State Maintenance
- Core Router: Preferential Treatment to Short Flows.



Architecture

This section covers the detailed implementation of the proposed scheme including the network architecture and the supporting mechanisms required to differentiate between short and long flows.



Edge Router : Packet Classification and State Maintenance

- ERs determine whether the packet is coming from a long or short flow.
- A threshold(L_t) based approximation method is used to mark them short vs. long.
- The per-flow state information are (softly) maintained to detect the termination of flow. The flow hash table is updated periodically every T_u time units.
- ER adjusts threshold dynamically using Short-to-Long Ratio(SLR), a ratio between the number of active short and long flows.

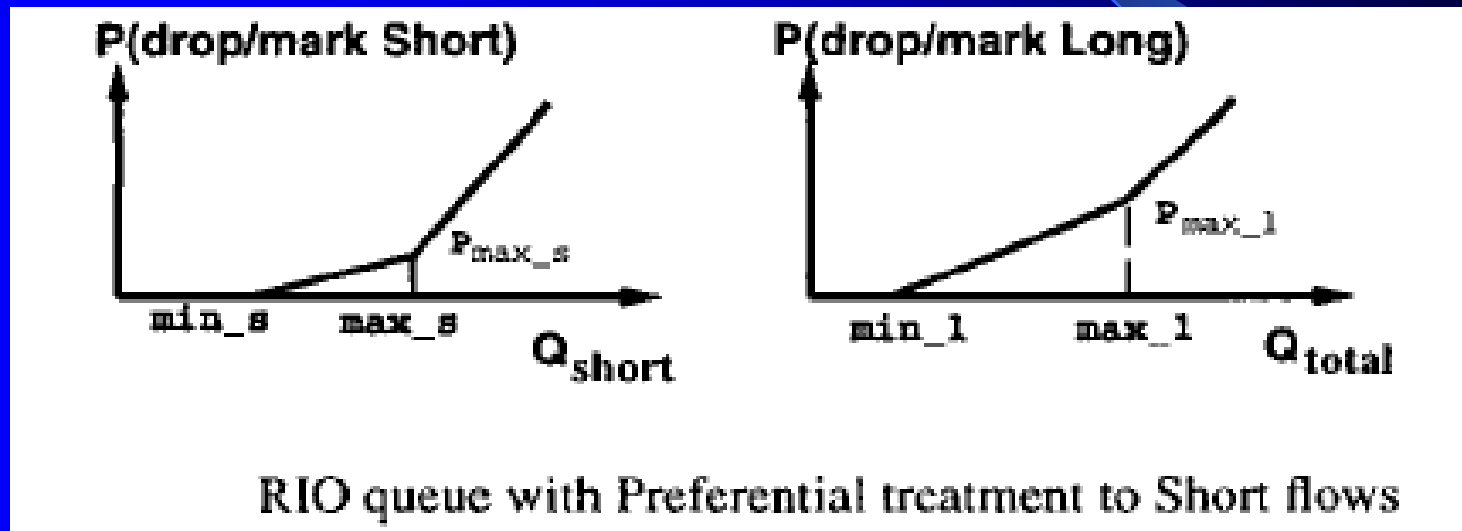


Core Router: Preferential Treatment to Short Flows

- Authors choose RIO (RED with In & Out) policy.
- RIO conforms to the DiffServ Specification.
- Only a single FIFO queue is used for all packets.
- RIO inherits all features of RED including protection of bursty flows.
- RIO performs soft prioritization, keeping benefits from statistical multiplexing.



Early dropping/marking function of an RIO queue



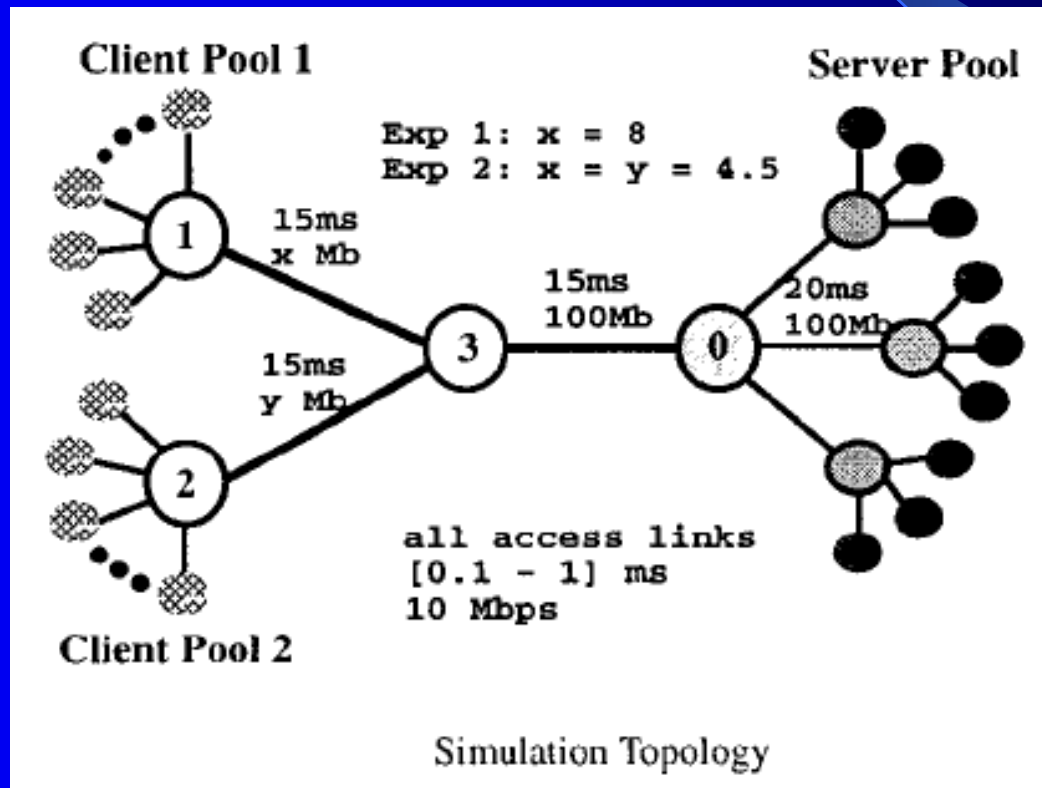
Simulation

- Simulation Setup
- Experiment 1: Single Client Set
- Experiment 2: .Unbalanced Requests



Simulation

- Topology : 0 = Edge Router; 1,2,3 = Core Routers.



Simulation

- Distribution of inter-page and inter-object time (in seconds), page size and object size(in packets).

Name	inter-page	objs/page	inter-obj	obj size
Exp1 client 1	Exponential mean 9.5	Uniform min 2 max 7	Exponential mean 0.05	Bounded Pareto [4,200000] shape 1.2
Exp2 client 1	Exponential mean 9.5	Uniform min 2 max 7	Exponential mean 0.05	Bounded Pareto [4,500] shape 1.2
client 2	Exponential mean 192	Uniform min 1 max 3	Exponential mean 1.5	Bounded Pareto [400,200000] shape 1.2

WEB TRAFFIC CONFIGURATION

Simulation

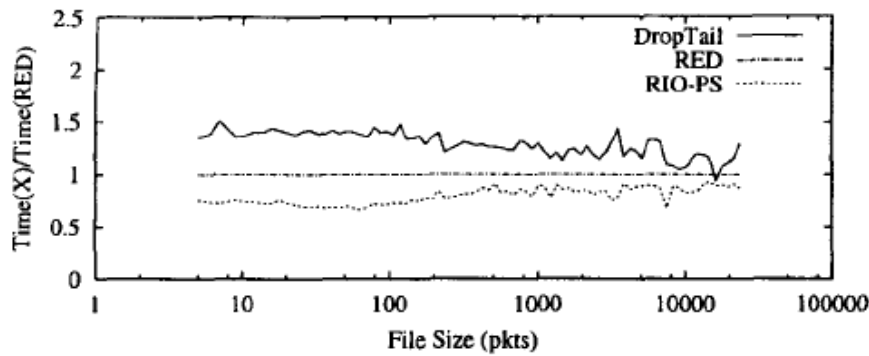
- Detailed Simulation Configuration :

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: $1.5 \times \text{BDP}$ RED/RIO-PS: $2.5 \times \text{BDP}$
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	$SLR = 3, T_u = 1 \text{ sec}, T_c = 10 \text{ sec}$
Foreground Traffic	
(Src, Dest)	(Server Pool, Client Pool)
Long Connection Size	1000 packets
Short Connection Size	10 packets

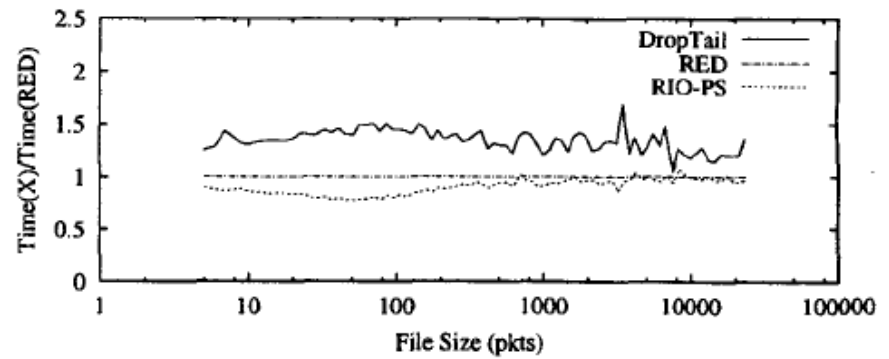
SIMULATION CONFIGURATION

Experiment 1

- Single Client Set



(a) Initial Retransmission Timer 3 seconds

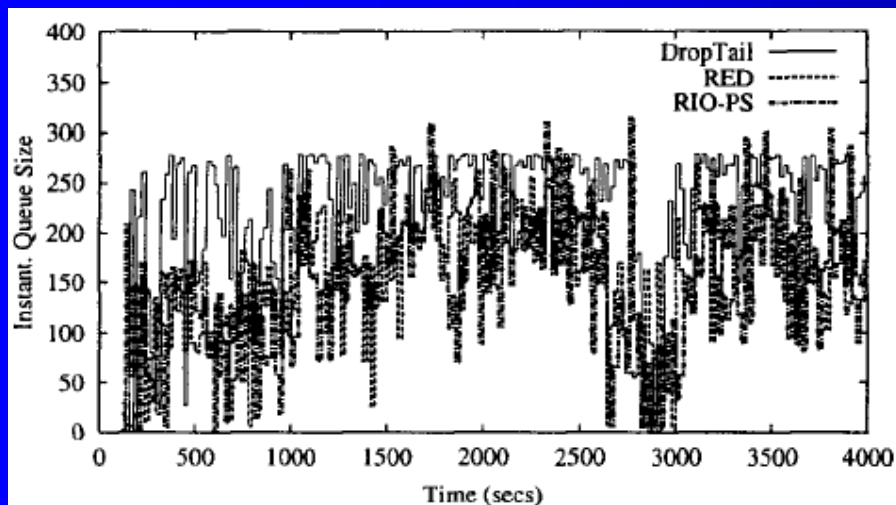


(b) Initial Retransmission Timer 1 second

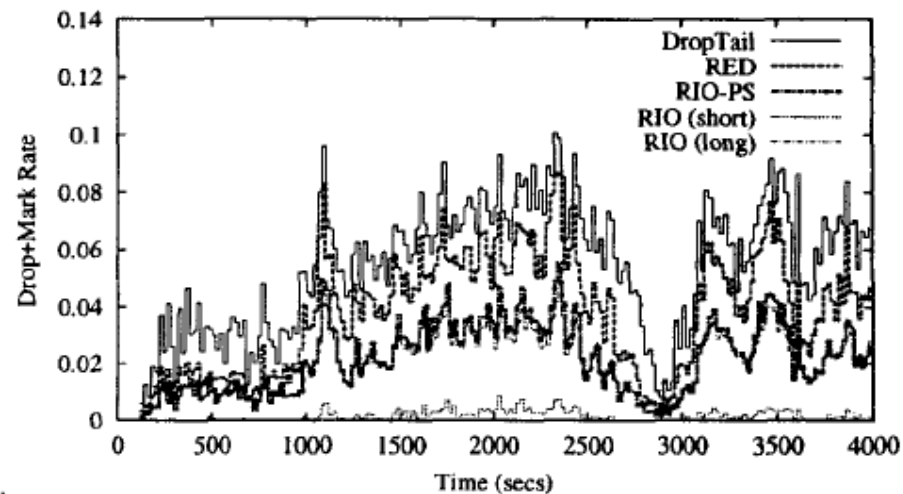
Average response time relative to RED

Experiment 1 ...

- Instantaneous queue size and drop rate in the last 20 seconds for the case of 3-seconds ITO



(a) Instantaneous Queue Size

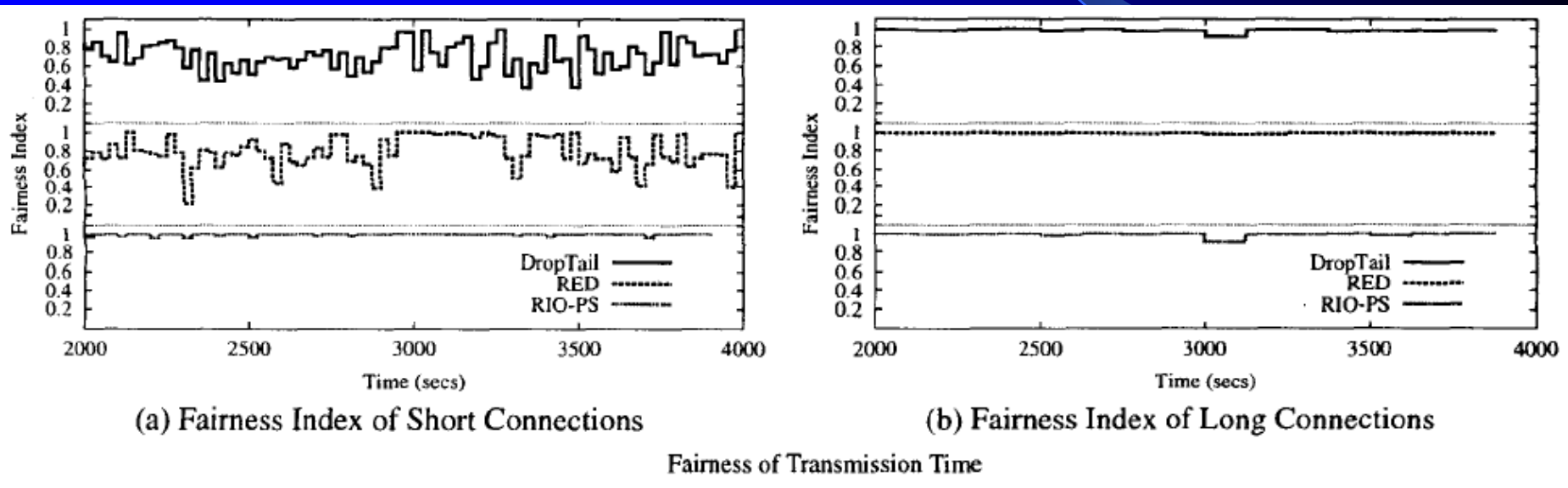


(b) Instantaneous Drop/Mark Rate

Revealing the secret of better performance

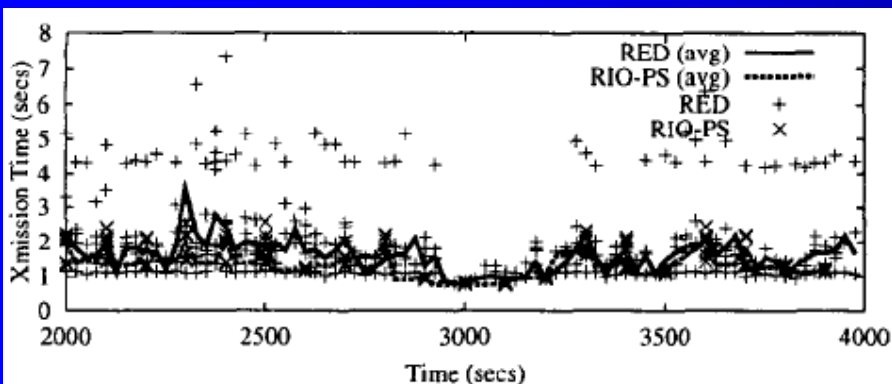
Experiment 1 ...

- Fairness Index of response time

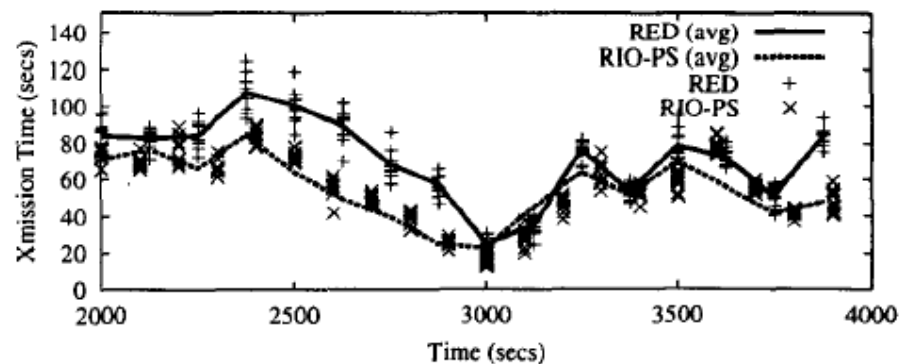


Experiment 1 ...

- Transmission time for each individual connection and their ensemble average



(a) Transmission Time of Short Connections



(b) Transmission Time of Long Connections

Transmission Time of Foreground Traffic

Experiment 1 ...

- Network Goodput

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

NETWORK GOODPUT OVER THE LAST 2000 SECONDS

Discussion

- Comments on Simulation Model
- The Queue Management Policy
- Deployment Issues
- Flow Classification
- Controller Design
- Malicious Users

Discussion

- The simulation presented in this paper uses 'DumbBell and DanceHall' (one-way traffic) and all TCP connections have similar end-to-end propagation delays.

Discussion

- To be conformant to existing DiffServ implementations authors chose RIO like AQM policy to be used at core routers.



Discussion

- The proposed scheme requires edge devices to be able to perform per-flow state maintenance and per-packet processing.

Discussion

- The proposed scheme involves Controller design issues at different places and timescales.

Discussion

- One concern regarding the proposed scheme may be that users are then encouraged to break long transmissions into small pieces so that they can enjoy faster services.

Conclusions and Future Work

- Performance of majority of TCP flow is improved.
- The performance of few TCP long flows is also enhanced.
- The overall Goodput of the system is improved.
- The proposed architecture is extremely flexible and can be largely tuned at the edge routers.
- Authors currently investigating an approach that integrates size-aware traffic management at both the network and transport layers.



Questions & Class Discussion

- Questions
- Suggestions
- Professor Comments
- Others
- Thanks!!



References:

- Paper by Liang Guo, Ibrahim Matta.
- Prof. Kinicki – WPI CSFQ paper.
- Review document : Preeti Phadnis

