Investigating Forms of Simulating Web Traffic

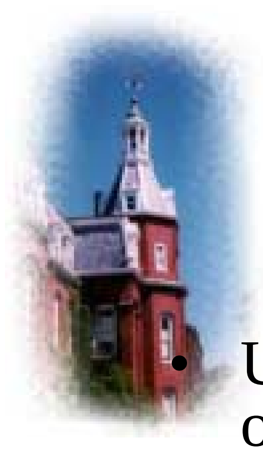
Yixin Hua
Eswin Anzueto
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA

WPI

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

WPI



Introduction

- Understanding the nature of network traffic is critical in order to properly design and implement computer networks services like congestions control. Web Traffic simulations provide the ability of exploring complicated scenarios that would be either difficult or impossible to analyze. More specific, the complexities of Internet topologies and traffic, and the central role of adaptive congestion control, make simulation the most promising tool for addressing many of the questions about Internet traffic dynamics. This paper present an analysis on the characteristics of real web traffic, i.e. the typical life of web objects, most commonly used TCP protocol, etc. The goal of this paper is to provide a framework for future web traffic simulations.

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

WPI

Web Traffic Characteristics

- The traffic measured in the Internet is known to be self-similar, which is generated in simulations with power-tail distributed random variables for the file size distribution.
- TCP Tahoe and Reno implementations are no longer the dominant families of TCP congestion control in the Internet and have been replaced by NewReno.
- IP Traffic: TCP accounts for 95 percent or more of the bytes, 85-95 percent of the packets, and 75-85 percent of the flows. TCP flows average fewer than 20 packets, about 7 Kbytes, and under 20s in duration. UDP makes up most of the remaining IP traffic, and ICMP packets account for less that 1 percent of all packets.
- Web traffic dominates as the single largest Internet application, with client/server traffic accounting for more than half the bytes (65-80 percent), packets (55-75 percent), and flows (65-75 percent). Web server traffic averages 10 Kbytes/flow, 15 packets/flow, and 13 s in duration per flow.

WPI

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

WPI

Web Traffic Simulations Found in Different Papers.

- In "Tuning RED for Web Traffic"
  - Used Mah model to write Web-traffic generating programs using socket system calls provided in FreeBSD.
  - Mah's model is application-level description of critical elements that characterize how HTTP1.0 protocols are used
- "Core-stateless Fair Queuein: Achieving Approximately Fair Bandwidth Allocation in High Speed Networks"
  - web traffic is simulated by using 60 ON-OFF TCP source, whose inter-arrival times are exponentially distributed with a mean of 0.05 ms, and the length of each transfer is drawn from a Pareto distribution with a mean of 20 packets with packet size 1KB and a shape parameter of 1.06

Web Traffic Simulations Found in Different Papers.

- In “The War Between Mice and Elephants” and “Dynamics of IP traffic: A study of the role of variability and the impact of control”
  – Web traffic is simulated by randomly selected clients initiate sessions by surfing several web pages of different size with randomly chosen website. Each page may contain several objects, each of which requires a TCP connection for delivery (implies HTTP 1.0 protocol). The client sends requests, and server responds with an acknowledgment and then starts to transmit the web object requested by client. They used shape parameter 1.2 for the Pareto ON-OFF sources.

WPI

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

Analysis

- There are two models used when simulating web traffic. The first model uses an abstract process which tries to capture the statistical traffic properties, independently of how the traffic is generated
- The second model use a hierarchical architecture that offers a more significant way to fully describe the intricacy of the web traffic. The main advantage of this model is that its parameters have a physical meaning, consequently those parameters can be change more easily to reflect changes on network conditions.

Analysis

These models normally consist of the following levels

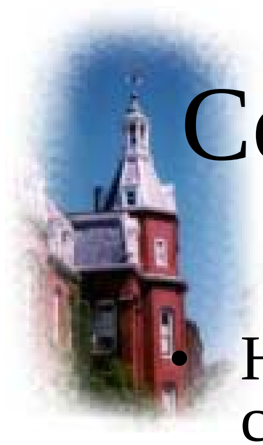
- Session Level: It describe the number of number of web sessions per day (or week), and the distribution of the sessions along the day or otherwise, the time between two consecutive sessions.
- Page Level: This parameter describes the number of pages per session and the statistical distribution of time between pages; this is associated to the average reading time of the users.
- Connection Level: A web page consists of a bunch of objects, which are conveyed through one or more TCP connection. The size of Web objects follows a heavy tailed distribution; this fact offers an explanation to the existence of self similarity in Internet traffic.

WPI

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

WPI



Considerations when Modeling Web Traffic

- HTTP1.0 protocol implies the use of a new TCP connection for each request/response pair. This protocol is gradually being replaced by the more efficient HTTP1.1 protocol which allows multiple and pipelined request to reuse TCP connections
- User behavior is very important when simulating web traffic, for example wireless links are becoming a more popular method for how millions of users access the network.
- A new “Killer application” comes along. While Web traffic dominates today, it is vital not to make the easy assumption that it will continue to do so tomorrow. There are many possible applications that could take its place, and these could greatly alter how the network tends to be use, and consequently the modeling will change

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

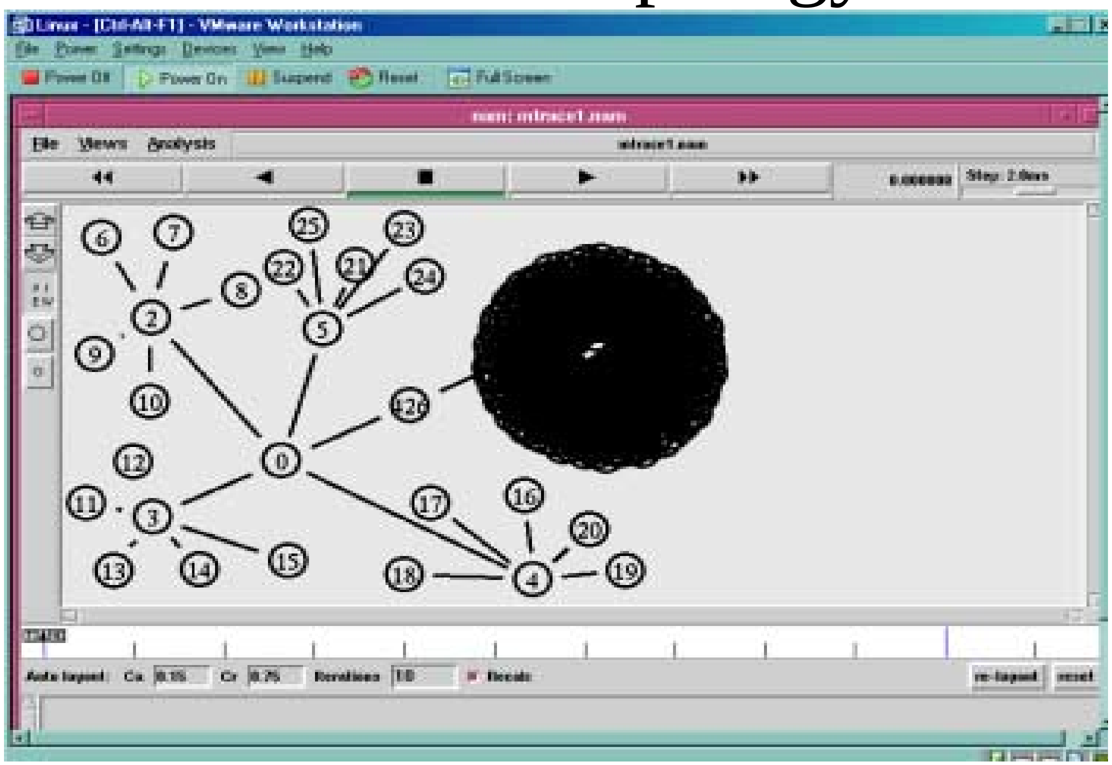
WPI

Simulations

- We carry out our simulations using the ns2 simulator, following a hierarchical architecture (model 2).
- We use 400 browser nodes, which can simulate a large group of users in an organization connected to internet simultaneously.
- It is been study that the differences between HTTP 1.0 and 1.1 does not impose notable differences on HTTP traffic, therefore we use HTTP1.0 (one connection per object) to simplified our work.

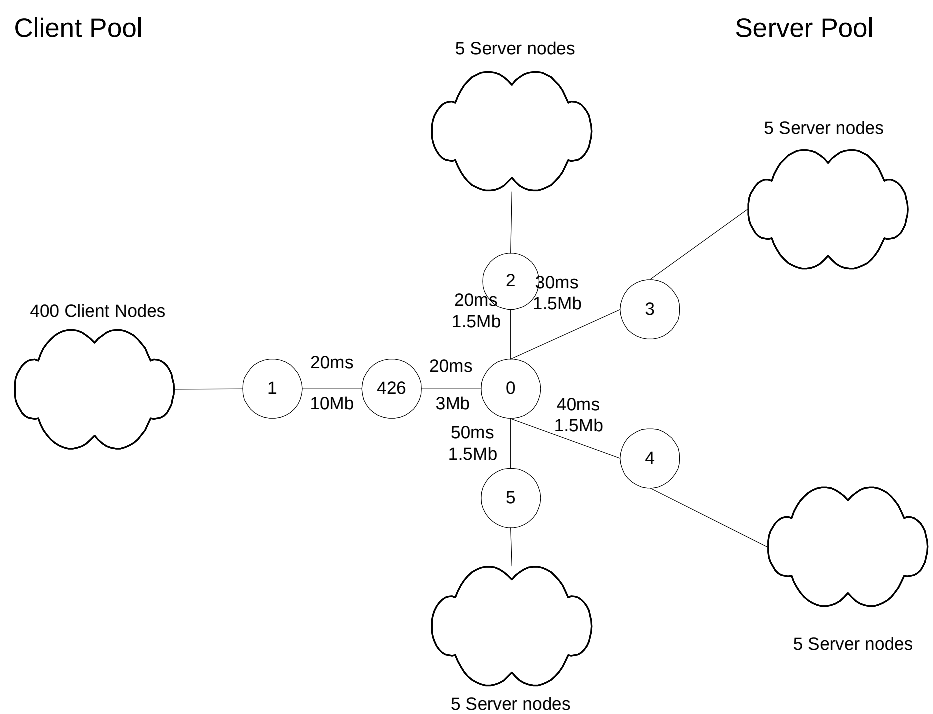
Simulations (cont)

- Simulation network topology



Simulations (cont)

- Abstractive simulation network topology

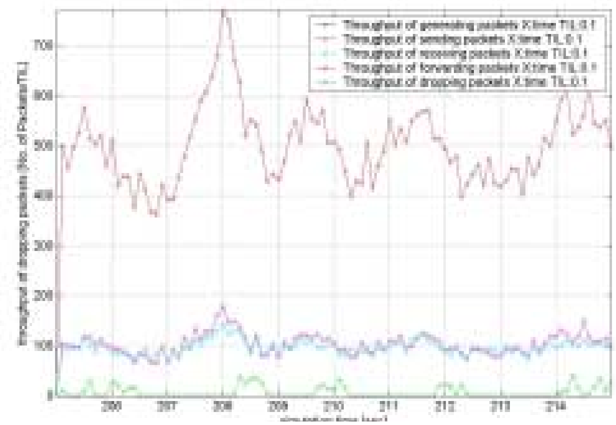


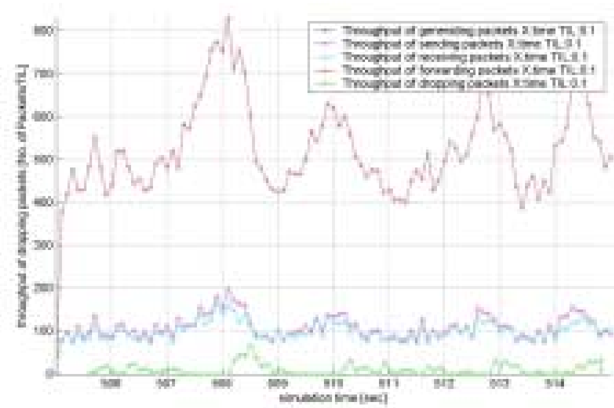
Experiment 1

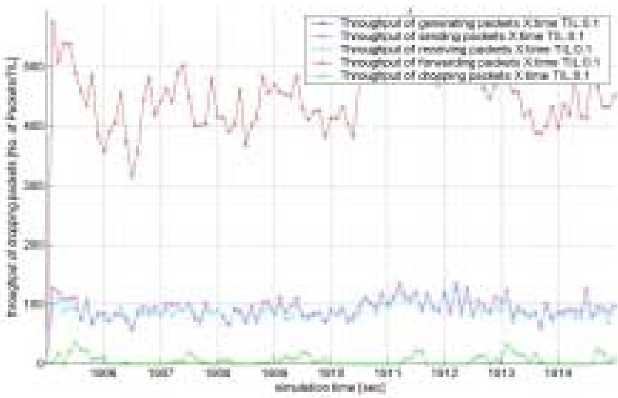
- Experiment 1 parameter setup:

Time	interPage Exponential	pageSize Pareto II	interObject Exponential	ObjSize Pareto II	objectSize Shape
205	15 Sec.	3 Objects	0.01 Sec.	12 packet	1.2
505	15 Sec.	3 Objects	0.01 Sec.	12 packet	1.2
1905	15 Sec.	3 Objects	0.01 Sec.	12 packet	1.2

Experiment 1 (Overall traffic pattern)

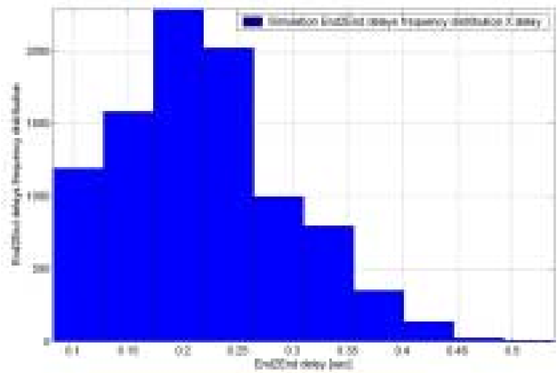


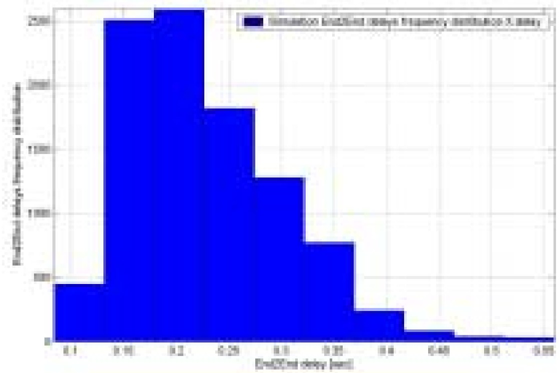


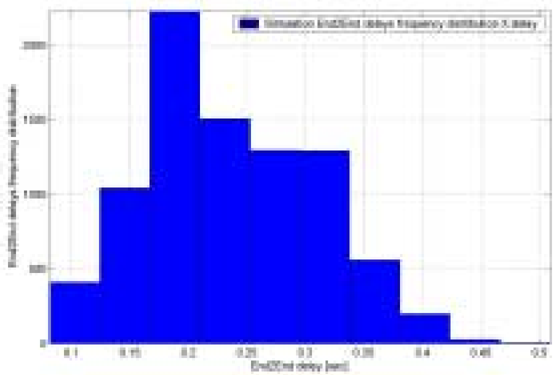


205 sec.(Top Left)
505 sec.(Top Right)
1905 sec.(Lower left)

Experiment 1 (End-to-End delay)

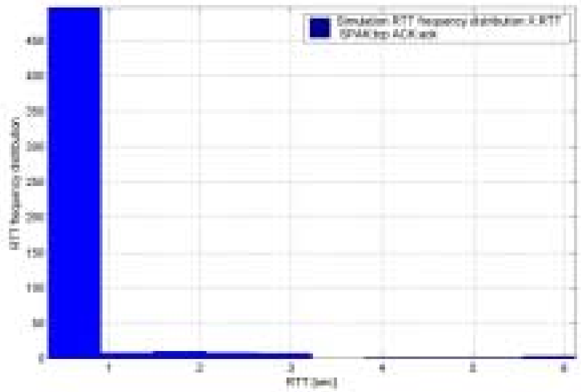


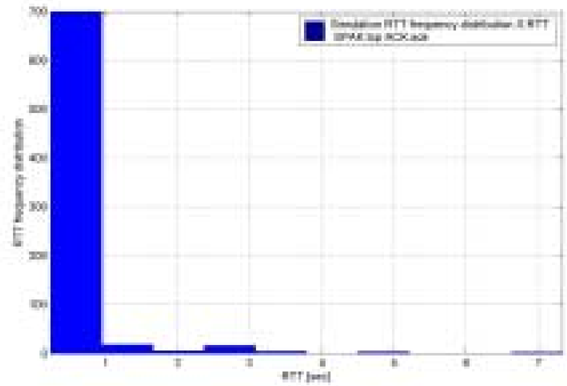


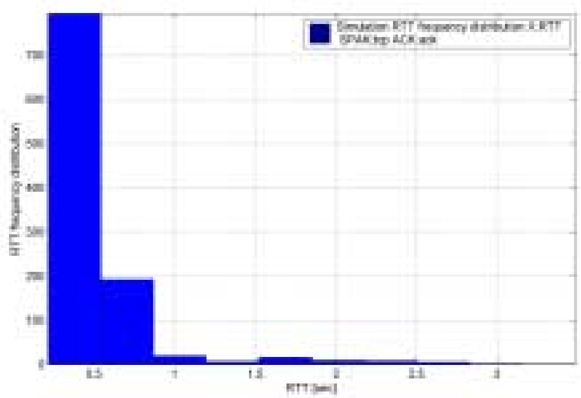


205 sec.(Top Left)
505 sec.(Top Right)
1905 sec.(Lower left)

Experiment 1 (RTT)







205 sec.(Top Left)
505 sec.(Top Right)
1905 sec.(Lower left)

WPI

Experiment 1(Observation)

- Overall traffic pattern evolves. Self similarity?
- Traffic pattern smoothes up during its lifetime
- Traffic pattern displays an increasing latency after long run
- Question: Do we need to consider to run longer simulation for research?? Caution, there are many web traffics start anytime in real life. What’s the best way to simulate the web traffic?

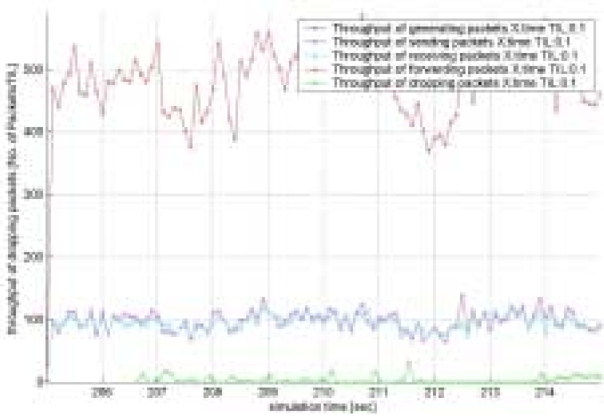
Experiment 2

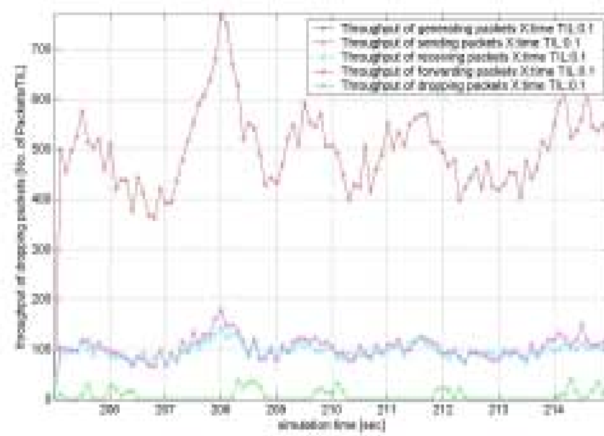
- Experiment 2 parameter setup:

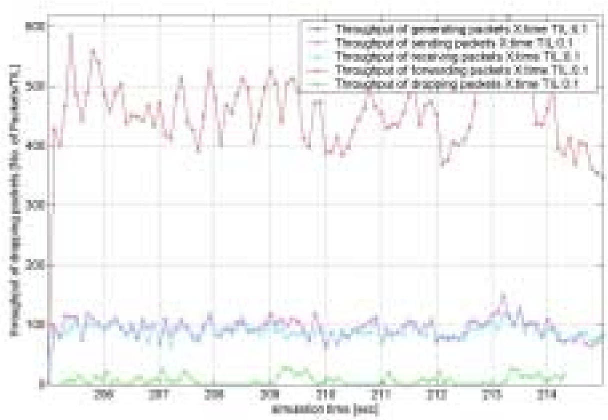
Optimize Time	interPage Exponential	pageSize Pareto II	interObject Exponential	ObjSize Pareto II	objectSize Shape
205	15 Sec.	3 Objects	0.01 Sec.	12 packet	1.1
205	15 Sec.	3 Objects	0.01 Sec.	12 packet	1.2
205	15 Sec.	3 Objects	0.01 Sec.	12 packet	1.5

WPI

Experiment 2 (Overall traffic pattern)



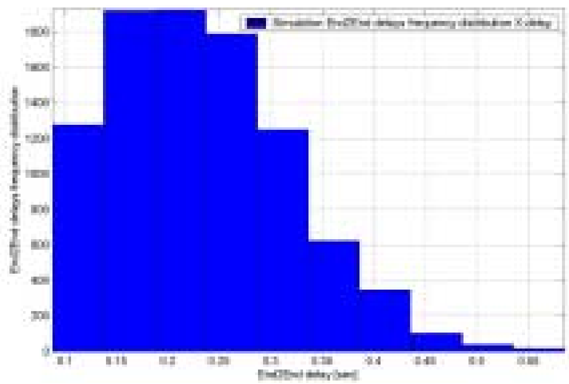


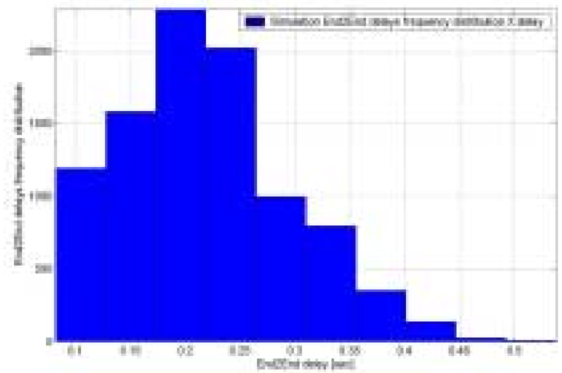


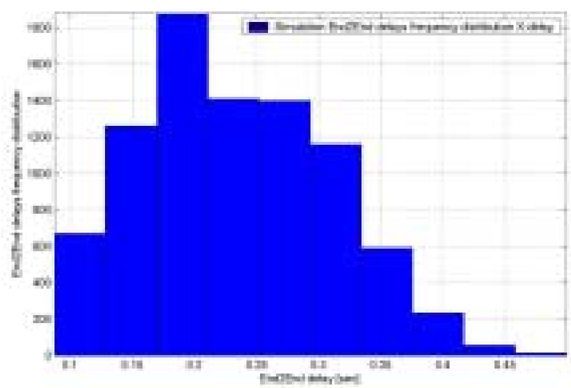
Shape =
1.1 (Top Left)
1.2 (Top Right)
1.5 (Lower left)

WPI

Experiment 2 (End-to-End delay)



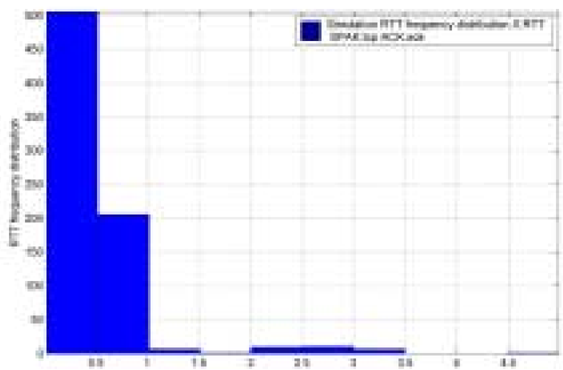


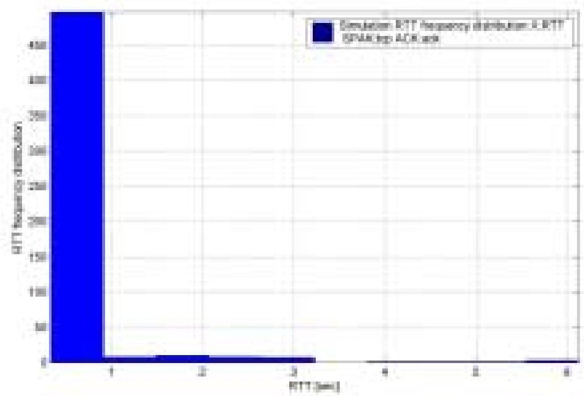


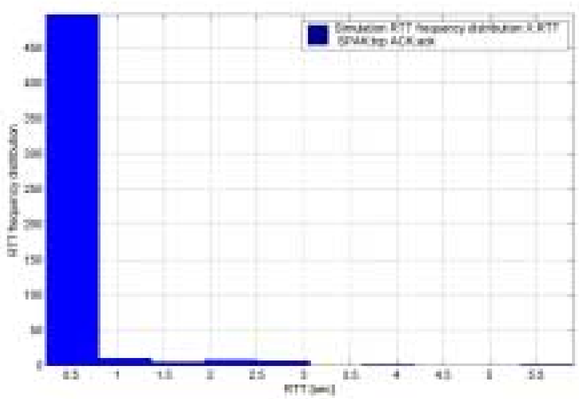
Shape =
1.1 (Top Left)
1.2 (Top Right)
1.5 (Lower left)

WPI

Experiment 2 (RTT)







Shape =
1.1 (Top Left)
1.2 (Top Right)
1.5 (Lower left)

Experiment 2(Observation)

- No clear traffic changing trend we can perceive
- In End-to-End delay and RTT charts, traffic with shape 1.2 are close to traffic with shape 1.5
- Traffic with shape 1.1 and traffic with shape 1.5 are contradicting with each other
- Traffic burstiness changes
- Question: Need longer simulation to reach steady state for comparison??

WPI

Outline

- Introduction
- Web Traffic Characteristics
- Web Traffic Simulations Found in Different Papers.
- Analysis
- Considerations when Modeling Web Traffic
- Simulations
- Conclusions

Conclusions

- Understanding the nature of network traffic is critical when simulating the internet, however unpredictable changes make it difficult to accomplish such simulations with accuracy.
- User behavior is very important when modeling web traffic.
- Simulation time (data needs to stabilize)
- Another import aspect worth mentioning is how important is the selection of the shape parameter used in the Pareto II distribution when simulating web traffic [13].

WPI