

“Difficulties in Simulating the Internet”

Sally Floyd, Van Paxson
ACM/IEEE TON, 9(4) August 2001

Techniques for Networking Research

Measurement

V. Paxson. "End-to-end Internet packet dynamics,"

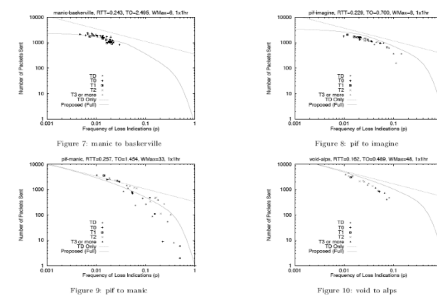
J. Padhye, V. Firoiu, D. Towsley, and J. Kurose "Modeling TCP Throughput: A Simple Model and its Empirical Validation,"

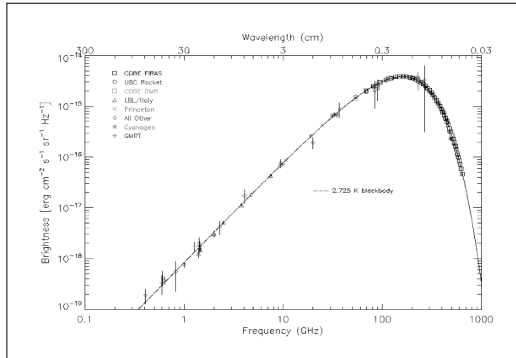
“Reality Check”

Are our assumptions reasonable? Is our mathematical model a good estimation of the real world?

e.g., from Paxson’s study

1. packet losses are bursty
2. $OTT \neq RTT/2$





Experimentation

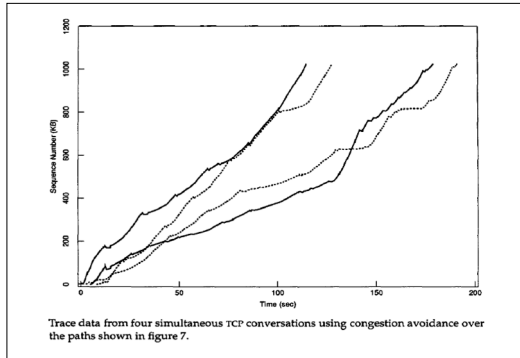
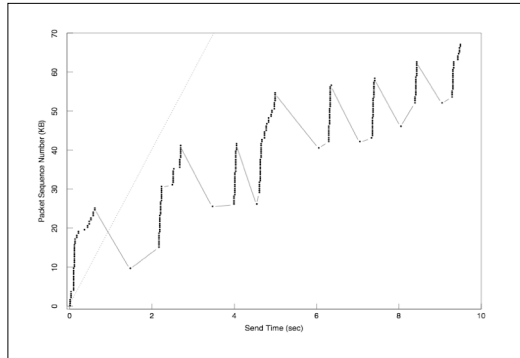
e.g., V. Jacobson. "Congestion Control and Avoidance"

Deal with implementation issues

Sometimes unforeseen complexities (e.g. own research experience in Unreliable TCP)

Understand the Behavior of Systems

Some systems are too complex to understand with "thought experiments" alone.



Analysis

D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks,"

J. Padhye, V. Firoiu, D. Towsley, and J. Kurose "Modeling TCP Throughput: A Simple Model and its Empirical Validation,"

Explore with Complete Control

We can understand the basic forces that affect the system.
e.g. TCP throughput is inversely proportional to $\sqrt{\rho}$

Simplify complex systems

If too simplified, important behavior could be missed
(TCP throughput without timeout)

Simulation

K. Fall and S. Floyd, "Simulation-based comparison of Tahoe, Reno, and SACK TCP,"

S. Floyd, K. Fall, "Promoting the Use of End-to-End Congestion Control in the Internet,"

S. Floyd, V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance,"

Check Correctness of Analysis

If simulation uses the same assumptions/model as the analysis, this simply verify the correctness of the mathematical derivations.

Check Correctness of Analysis

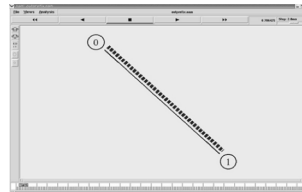
Simulation can relax some assumptions, use more complex models, etc. to test the limits of analysis.

(Real measurement/experiments still needed to check the usefulness of analysis results)

Explore Complex Systems

Some systems are too difficult/impossible to analyze
e.g. Internet

Helps Develop Intuition



Measurement
Experimentation } Real World

Analysis
Simulation } Abstract Model

Why is Internet hard to simulate?

!
Internet is diverse

End-hosts: phones, desktops,
servers, iPod, Wii

Links: Ethernet, WiFi, Satellite, Dial-up, 3G

Transport: TCP variants, UDP, DCCP

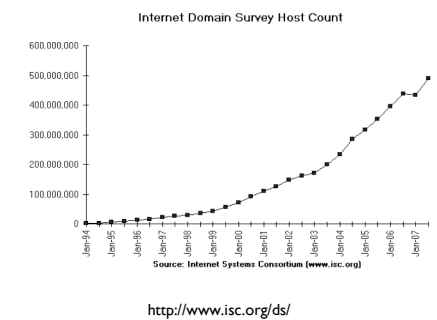
Applications: games, videos, web, ftp, bittorrent

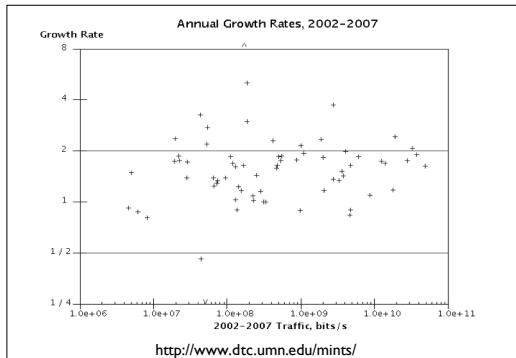
2

Internet is huge

3

Internet is changing





Time	Median File Transfer Size
March 1998	10.9 kB
December 1998	5.6 kB
December 1999	10.9 kB
June 2000	62 kB
November 2000	10 kB

Measurement at LBNL: Statistical property of Internet changes as well.

Why is Internet hard to simulate?

1. Heterogeneous
2. Huge
3. Changing

Suppose you come up with the greatest BitTorrent improvement ever..

You want to simulate it to make sure it works before you release it (and call the press)

What Internet topology should you use in your simulation?

How end hosts are connected? What are the properties of the links?

Topology changes constantly
Companies keep info secrets
Routes may change
Routes may be asymmetric

You will need to simulate
over a wide range of
connectivity and link
properties

Suppose you come up
with the greatest TCP
optimization ever..

You want to know if it is
fair to existing TCP
versions before you write
your SIGCOMM paper..

Which TCP versions to
compare with?

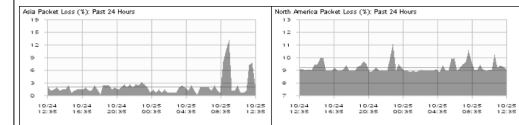
Using “fingerprinting”,
831 different TCP
implementations and
versions are identified.

Which to use? Which to ignore?

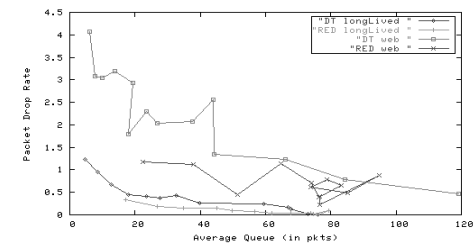
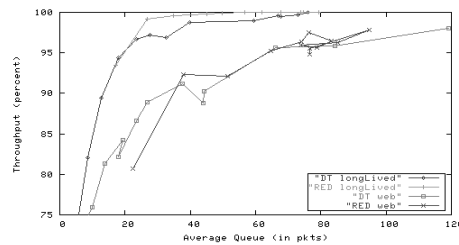
What applications to run?
What type of traffic to generate?

Telnet? FTP? Web?
BitTorrent? Skype?

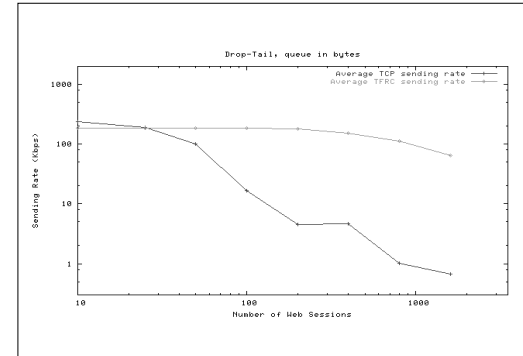
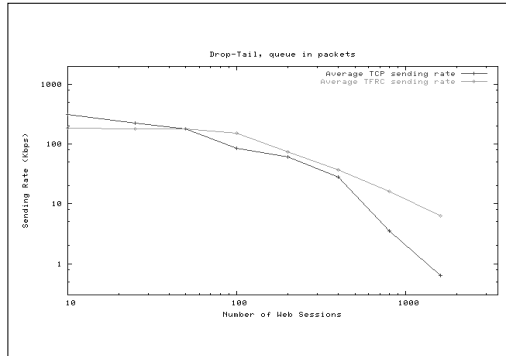
How congested should the network be?



Example from Sally Floyd:
RED vs DropTail



Example from Sally Floyd: Using TFRC for VoIP



We can focus our simulation
on dominant
technology/application
today..

TCP: NewReno SACKS
OS: Windows Linux
Applications: Web, FTP

What about tomorrow?

WiMax?
Sensors?
Virtual World?
DCCP?

10 years ago, you came up
with a router mechanism
to improve TCP Reno..

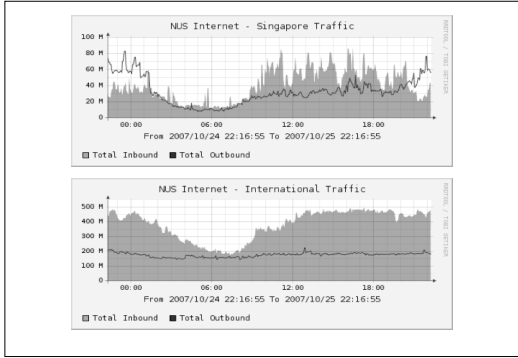
No one cares today.

How to verify the
simulator itself?

So, how?

Looking for Invariants

I. Diurnal Patterns



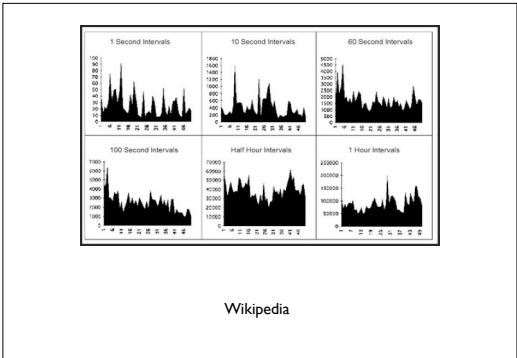
```

hour #constrained
-----
00 139 2.5% -----X
01 144 2.6% -----X
02 146 2.6% -----X
03 140 2.5% -----X
04 119 2.1% -----X
05 89 1.6% -----X
06 69 1.2% -----X
07 55 1.0% -----X
08 45 0.8% -----X
09 40 0.7% -----X
10 40 0.7% -----X
11 42 0.8% -----X
12 51 0.9% -----X
13 57 1.0% -----X
14 68 1.2% -----X
15 75 1.3% -----X
16 77 1.4% -----X
17 92 1.6% -----X
18 98 1.8% -----X
19 105 1.9% -----X
20 108 1.9% -----X
21 113 2.0% -----X
22 124 2.2% -----X
23 134 2.4% -----X
  
```

U Waterloo Data 24 Oct 2007

2. Self-Similar Traffic

The traffic is bursty regardless of time scale



3. Poisson Session Arrival

$$f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Remote logins, starting
FTP, beginning of web
surfing etc.

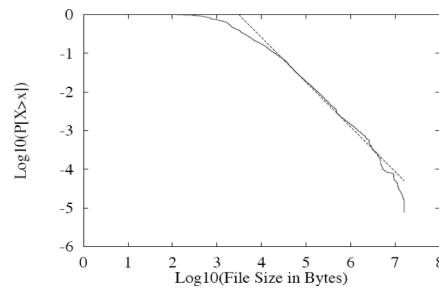
(so are dead light bulbs,
spelling mistakes, etc.)

4. Log-normal Duration

$$f(x; \mu, \sigma) = \frac{e^{-(\ln x - \mu)^2 / (2\sigma^2)}}{x\sigma\sqrt{2\pi}}$$

5. Heavy Tail Distributions

$$P[X > x] \sim x^{-\alpha},$$



Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes, by Mark E. Crovella and Azer Bestavros

I. Looking for Invariants

2. Explore Parameter Space

Change one parameter,
fix the rest

Explore a wide range of
values

3. Use Traces

e.g. collects traces of web
sessions, video files, VoIP
traffic

Use it to simulate the
traffic source

But must be careful about traffic shaping and user/application adaptation.

e.g. traces collected during non-congested time should not be used to simulate congested networks.

4. publish simulator script for others to verify

Conclusion

Simulation is useful but needs to be done properly

Be careful about your simulation model: you want it to be as simple as possible, but not simpler.

Be careful about your conclusion: “A is 13.5% better than B” is probably useless.

“A is 13.5% better than B under these environment” is better but not general

Not really for quantitative results, but more for

understanding the dynamics,
illustrate a point,
explore unexpected behavior.