Thoughts on Modeling From a Measurement Perspective

Vern Paxson

AT&T Center for Internet Research at ICSI (ACIRI)

International Computer Science Institute

Berkeley, CA USA

vern@aciri.org

April 25, 2001

The power of modeling to prejudice our thinking:

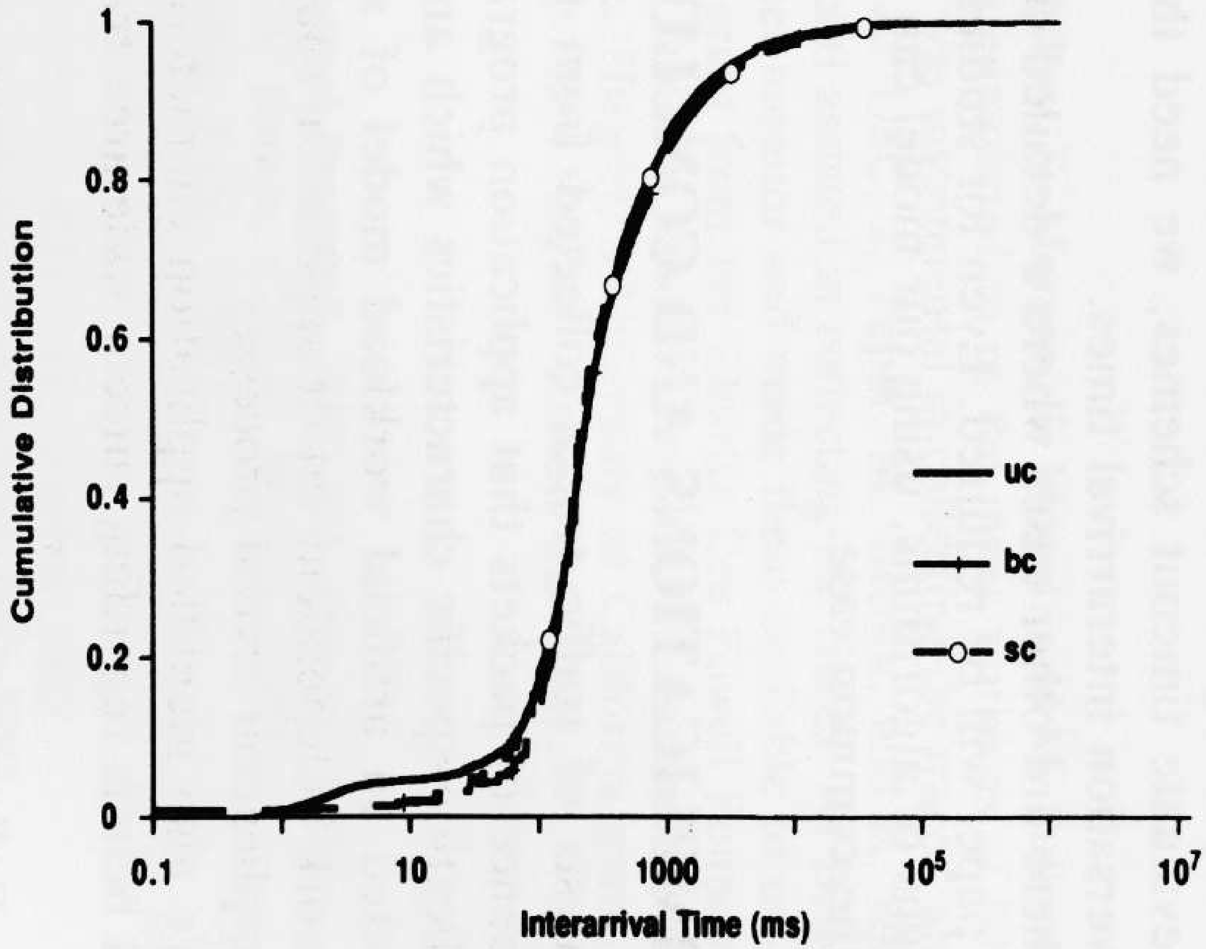


Figure 19. Distribution of TELNET packet interarrivals

"... interarrival times for interactive applications should be modeled by a uniform plus exponential random time ..."

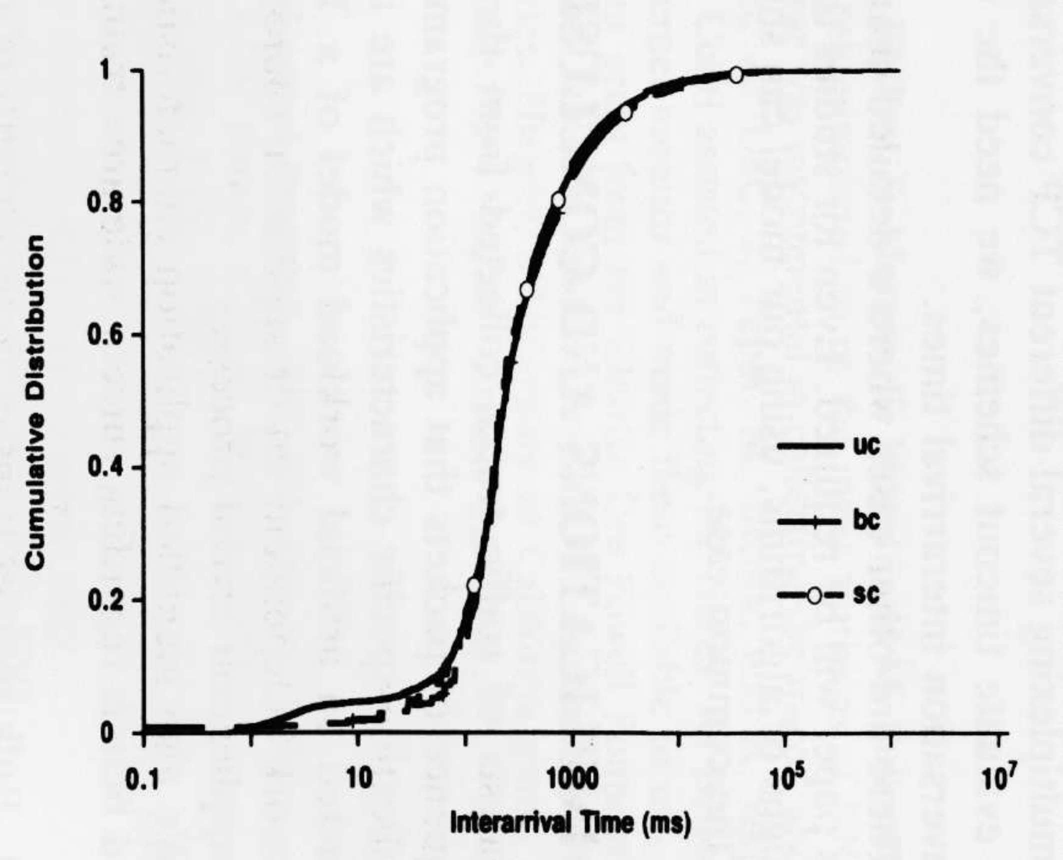
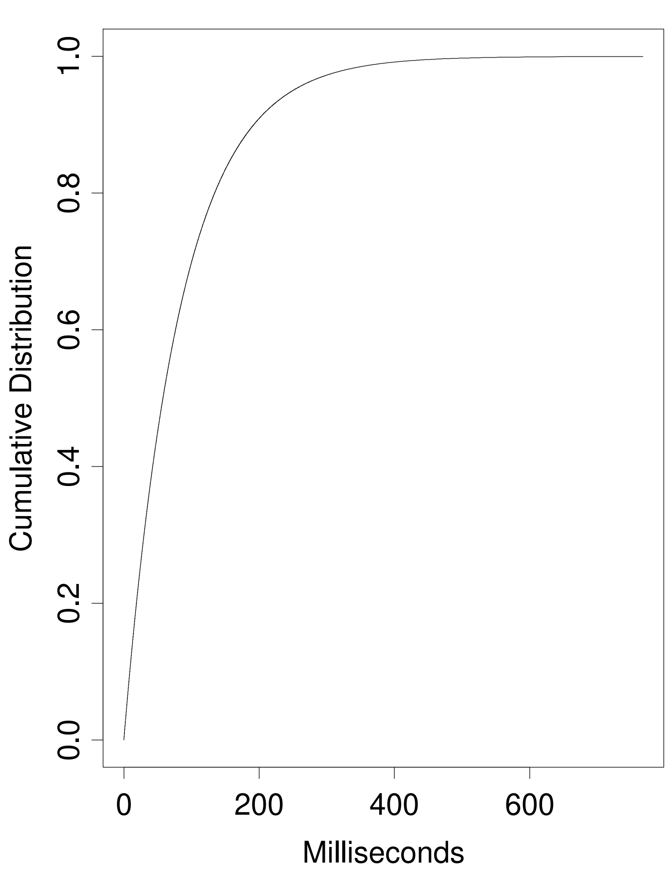


Figure 19. Distribution of TELNET packet interarrivals



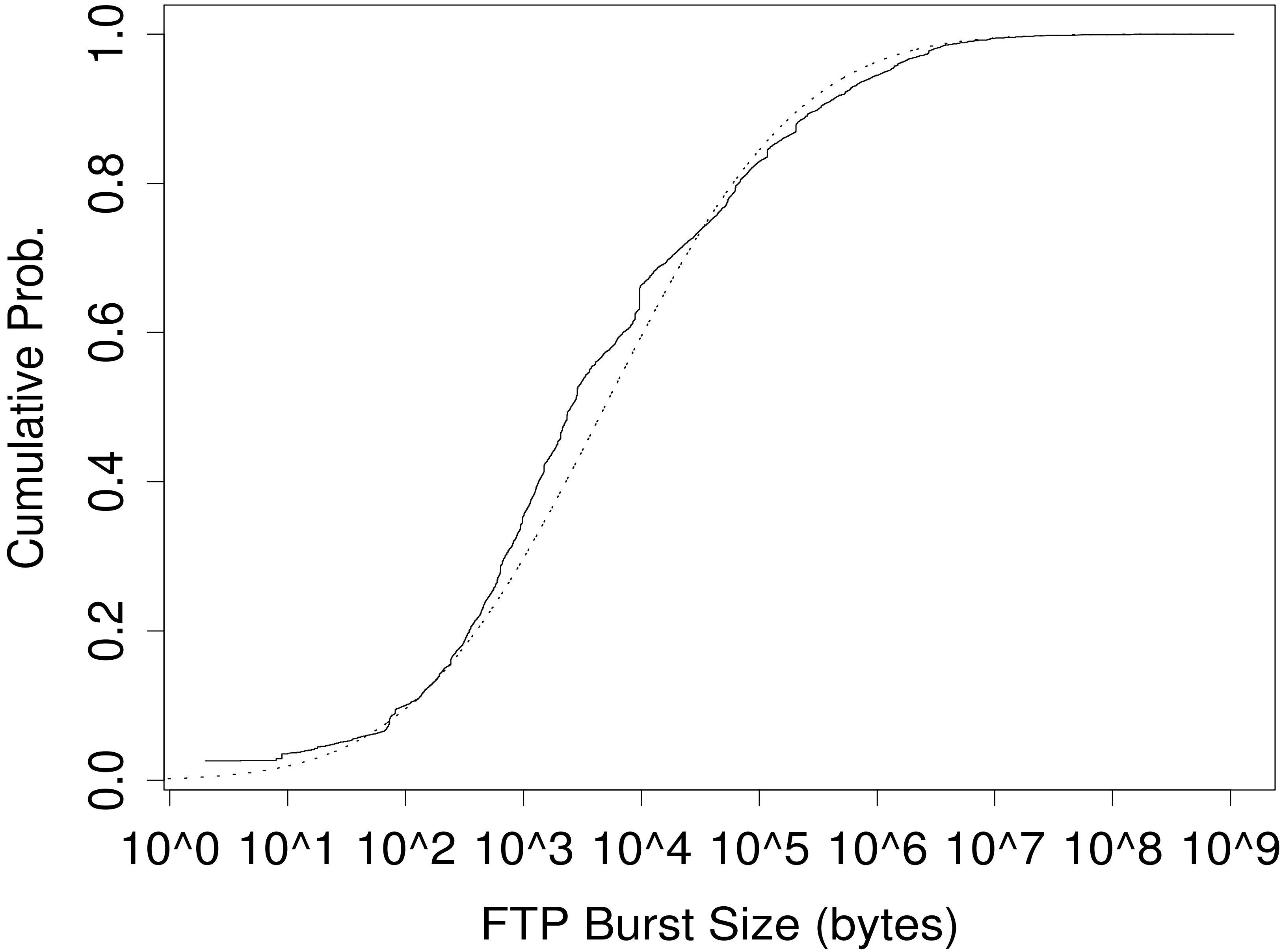
The power of modeling to open our eyes:

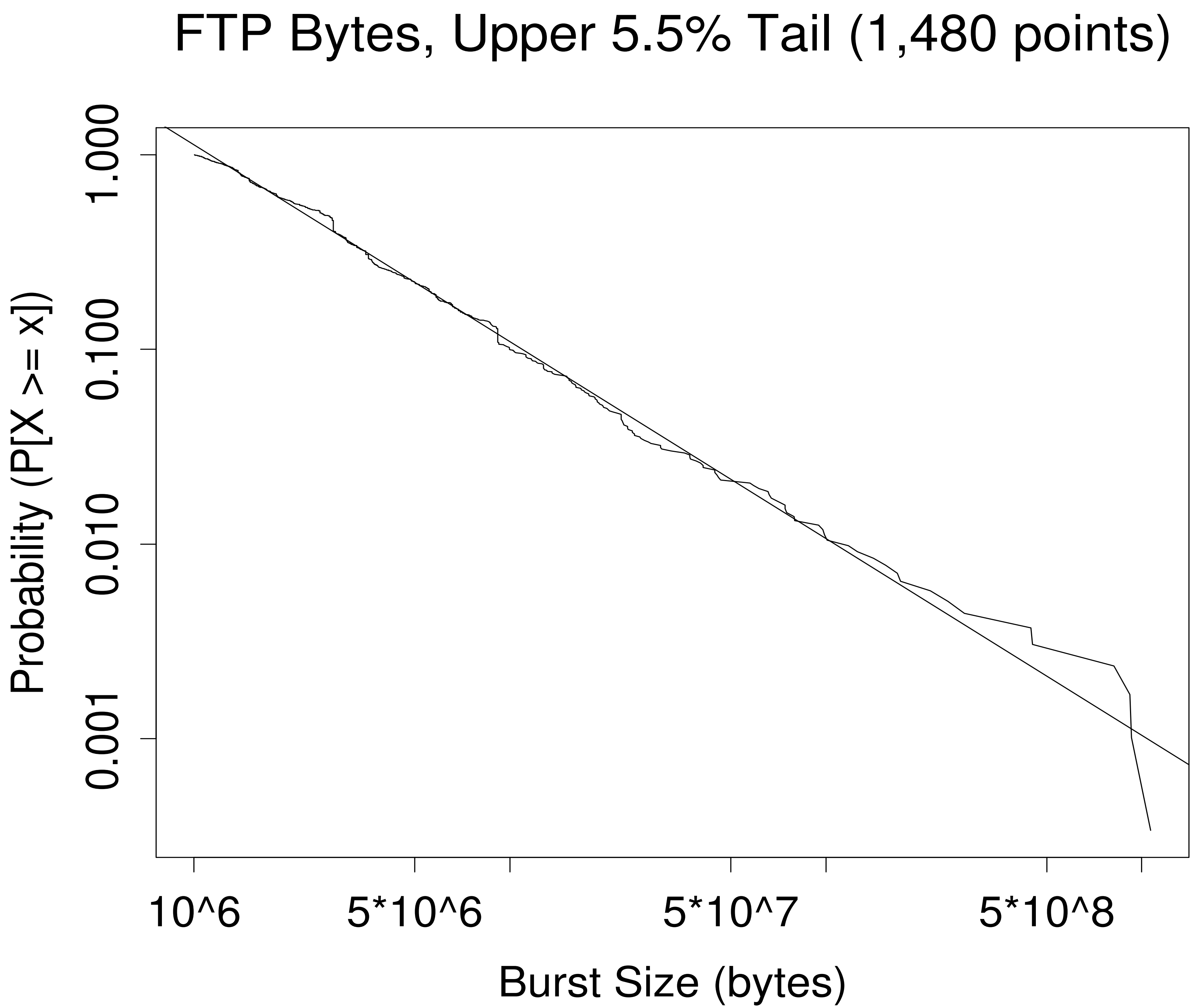
[LTWW93] developed the case for *self-similarity* as a parsimonious model of Ethernet traffic . . .

"*We did so without first studying and modeling the behavior of individual Ethernet users (sources).*"

They speculated: self-similarity in aggregates could be rooted in heavy-tailed behavior by (independent) individuals.

$\Rightarrow$ "Go look for heavy tails."


Cumulative Prob.
0.0
0.2
0.4
0.6
0.8
1.0
10^0 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9
FTP Burst Size (bytes)


FTP Bytes, Upper 5.5% Tail (1,480 points)
Probability (P[X >= x])
1.000
0.100
0.010
0.001
10^6
5*10^6
5*10^7
5*10^8
Burst Size (bytes)

There Is No Such Thing As Typical:

Major modeling/measurement pitfall is to underappreciate the Internet's vast dynamic range.

Immense heterogeneity:

- across different sites
- at a single site over (not much) time
- in range of conditions encountered and perf. achieved

It keeps changing:

- sustained exponential growth for two decades
- repeated arrival of explosively popular new apps

There Is No Such Thing As Typical, con't:

Observation #1:

If you fail to take this into account,

nothing tells you that you're wrong.

Observation #2:

We are desperate for parsimonious models . . .

. . . because we have no clue how to deal with

a lot of parameters in a meaningful way.

Dealing with a sea of change — searching for invariants:

- Diurnal patterns of activity.
- Self-similar correlational structure.
- Poisson session arrivals.
- Log-normal connection sizes.
- Heavy-tailed distributions.
- Invariant (heavy-tailed) distribution of keystroke timing.
- Speed-of-light constraints.
- Topology power laws (?).