Improving our Evaluation of Transport Protocols

Sally Floyd
Hamilton Institute

July 29, 2005

“Computer System Performance Modeling and Durable Nonsense”

- “A disconcertingly large portion of the literature on modeling the performance of complex systems, such as computer networks, satisfies Rosanoff's definition of durable nonsense.”

- "THE FIRST PRINCIPLE OF NONSENSE:
  For every durable item of nonsense, there exists an irrelevant frame of reference in which the item is sensible."

- "THE SECOND PRINCIPLE OF NONSENSE:
  Rigorous argument from inapplicable assumptions produces the world's most durable nonsense."

- "THE THIRD PRINCIPLE OF NONSENSE:
  The roots of most nonsense are found in the fact that people are more specialized than problems"

The quote is 25 years old!

- John Spragins, "Computer System Performance Modeling and Durable Nonsense", January 1979.

- R. A. Rosanoff, "A Survey of Modern Nonsense as Applied to Matrix Computations", April 1969.

Outline of Talk:

- Metrics for evaluating congestion control.
- Models for use in simulations, experiments, and analysis.
- Examples:
  - HighSpeed TCP
  - Quick-Start
  - Congestion control for VoIP

Metrics for evaluating congestion control: throughput, delay, and drop rates

- Throughput:
  - Router-based metric: link throughput.
  - User-based metrics:
    - per-connection throughput or file transfer times.
    - Throughput after a sudden change in the app's demand (e.g., for voice and video).
    - Fast startup.
- Delay:
  - Router-based metric: queueing delay
  - User-based metrics: per-packet delay (average or worst-case?)
- Drop rates.

Throughput, delay, and drop rates:

- Tradeoffs between throughput, delay, and drop rates.
- The space of possibilities depends on:
  - the traffic mix;
  - the range of RTTs;
  - the traffic on the reverse path;
  - the queue management at routers;
  - …

Metrics for evaluating congestion control: response times and minimizing oscillations.

- Response to sudden congestion:
  - from other traffic;
  - from routing or bandwidth changes.
- Concern: slowly-responding congestion control:
  - Tradeoffs between responsiveness, smoothness, and aggressiveness.
- Minimizing oscillations in aggregate delay or throughput:
  - Of particular interest to control theorists.
- Tradeoffs between responsiveness and minimizing oscillations.

Metrics for evaluating congestion control: fairness and convergence

- Fairness between flows using the same protocol:
  - Which fairness metric?
  - Fairness between flows with different RTTs?
  - Fairness between flows with different packet sizes?
- Fairness with TCP
- Convergence times:
  - Of particular concern with high bandwidth flows.

Robustness to failures and misbehavior:

- Within a connection:
  - Receivers that “lie” to senders.
  - Senders that “lie” to routers.
- Between connections:
  - Flows that don’t obey congestion control.
- Ease of diagnosing failures.

Metrics for evaluating congestion control: robustness for specific environments

- Robustness to:
  - Corruption-based losses;
  - Variable bandwidth;
  - Packet reordering;
  - Asymmetric routing;
  - Route changes;
  - …
- Metric: energy consumption for mobile nodes
- Metric: goodput over wireless links
- Other metrics?

Metrics for evaluating congestion control: metrics for special classes of transport

- Below best-effort traffic.
- QoS-enabled traffic

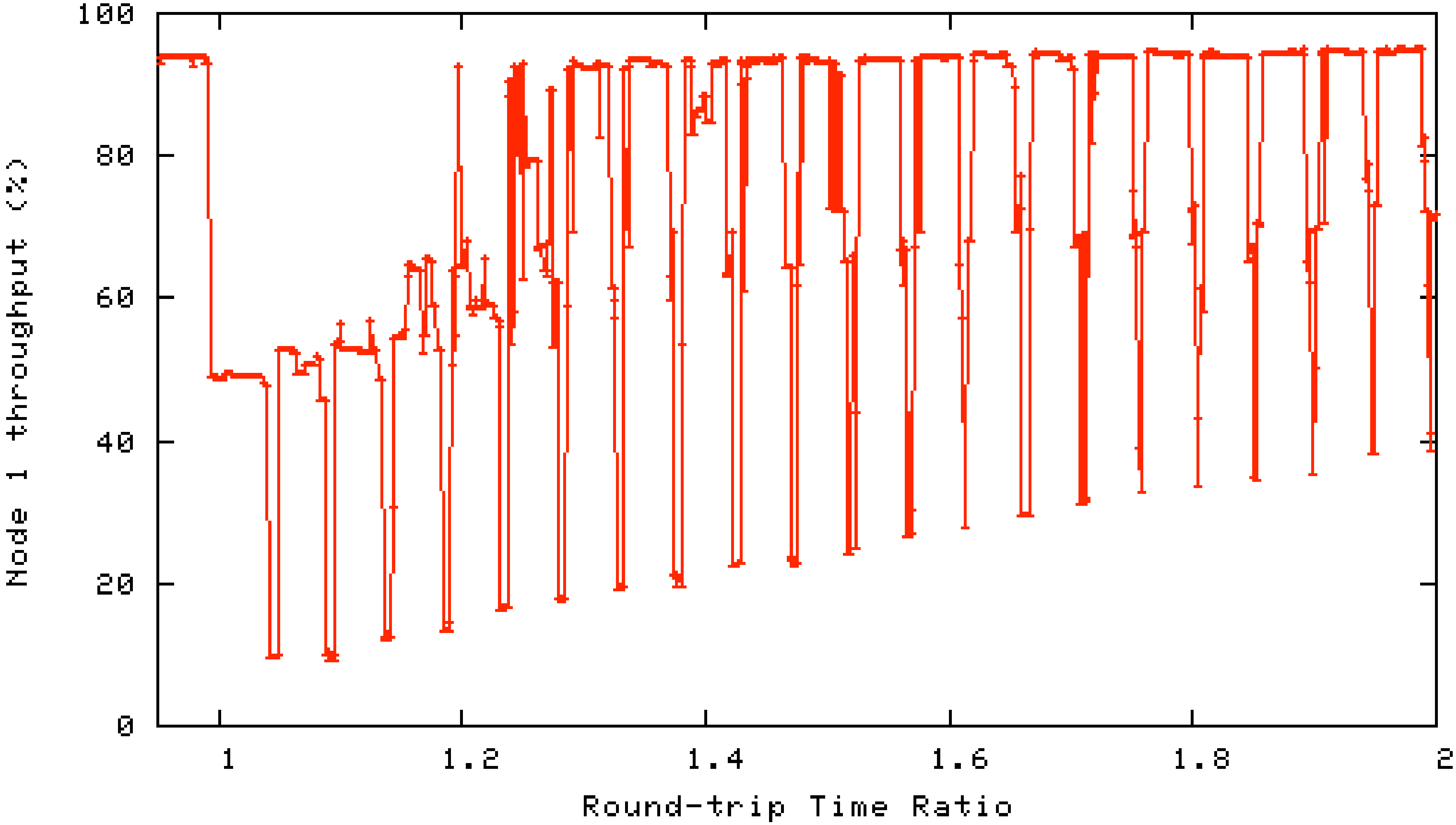
Metrics for evaluating congestion control: Deployability

- Is it deployable in the Internet?

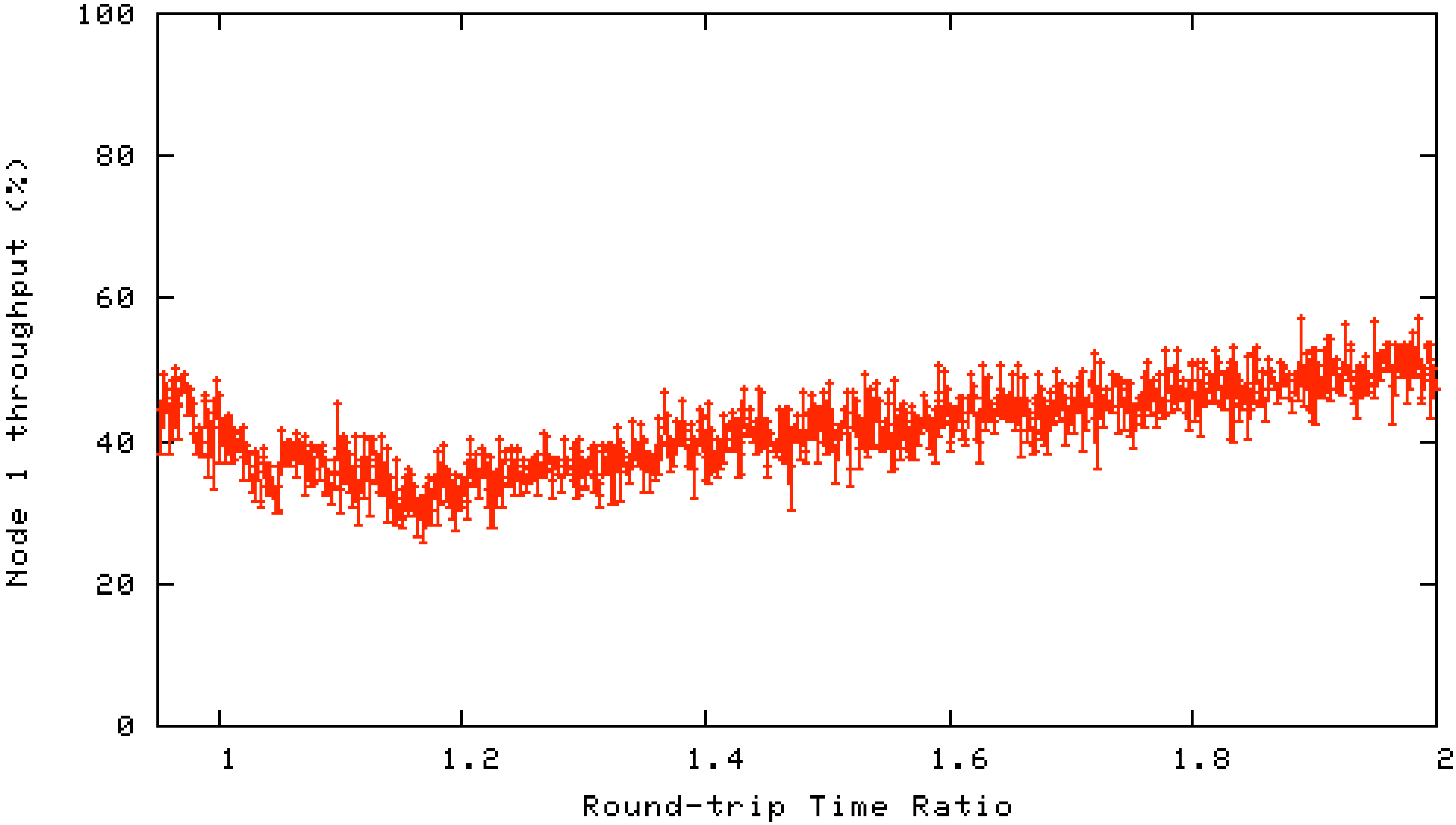
Internet research needs better models!

- What models do we use for evaluating transport protocols?
  - In simulations, experiments, and analysis.
- The simpliest model sufficient, but no simplier!
  - A simple topology with one-way traffic of long-lived flows all with the same RTT?
  - A complex topology aiming for full realism of the global Internet?
  - Or something in between…

Simulation with Two Long-lived Flows:



Two Long-lived Flows, with Telnet and Reverse-path Traffic:



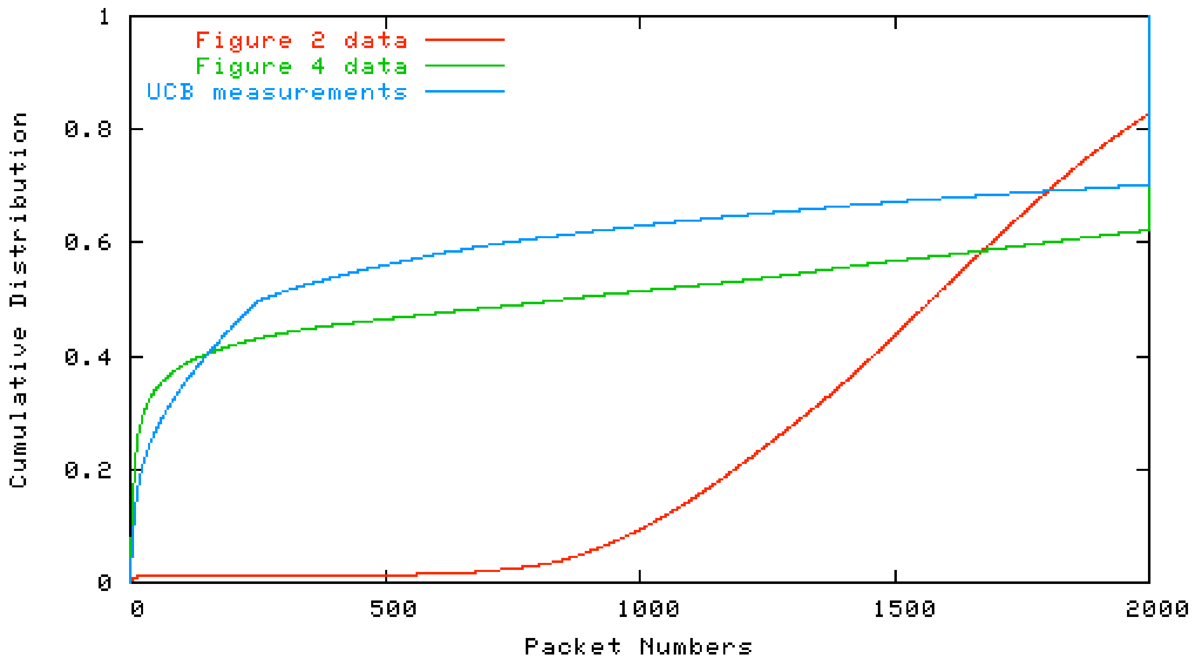
Use a range of scenarios!

- A range of:
  - Link bandwidths;
  - Levels of congestion;
  - Levels of statistical multiplexing
- Look for weaknesses as well as strengths!
- Look for the space of possible tradeoffs.

Characterizing scenarios:

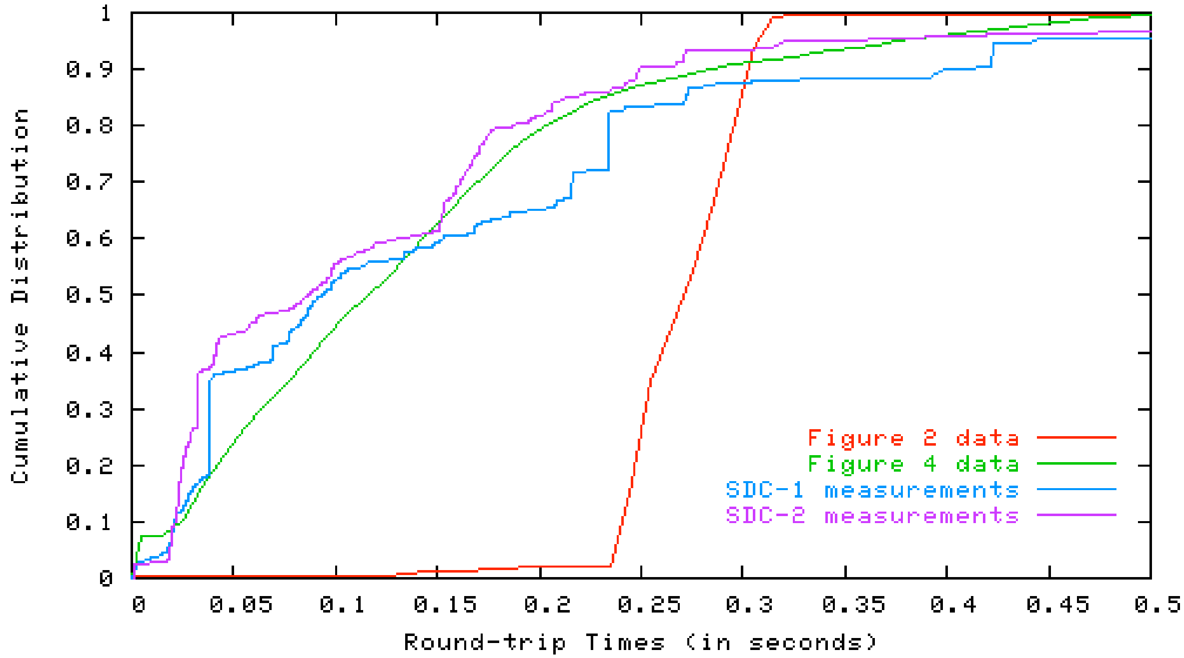
- The distribution of RTTs:
  - Measured by per-packet RTTs.
  - Affects throughput, delay, etc.
- The distribution of connection sizes:
  - Measured by packet sequence numbers.
  - Affects throughput, delay, much else.
  - Medium-sized flows slow-starting.

Distribution of Flow Sizes



- Distributions of packet numbers on the congested link over the second half of two simulations, with data measured on the Internet for comparison.

Distribution of RTTs



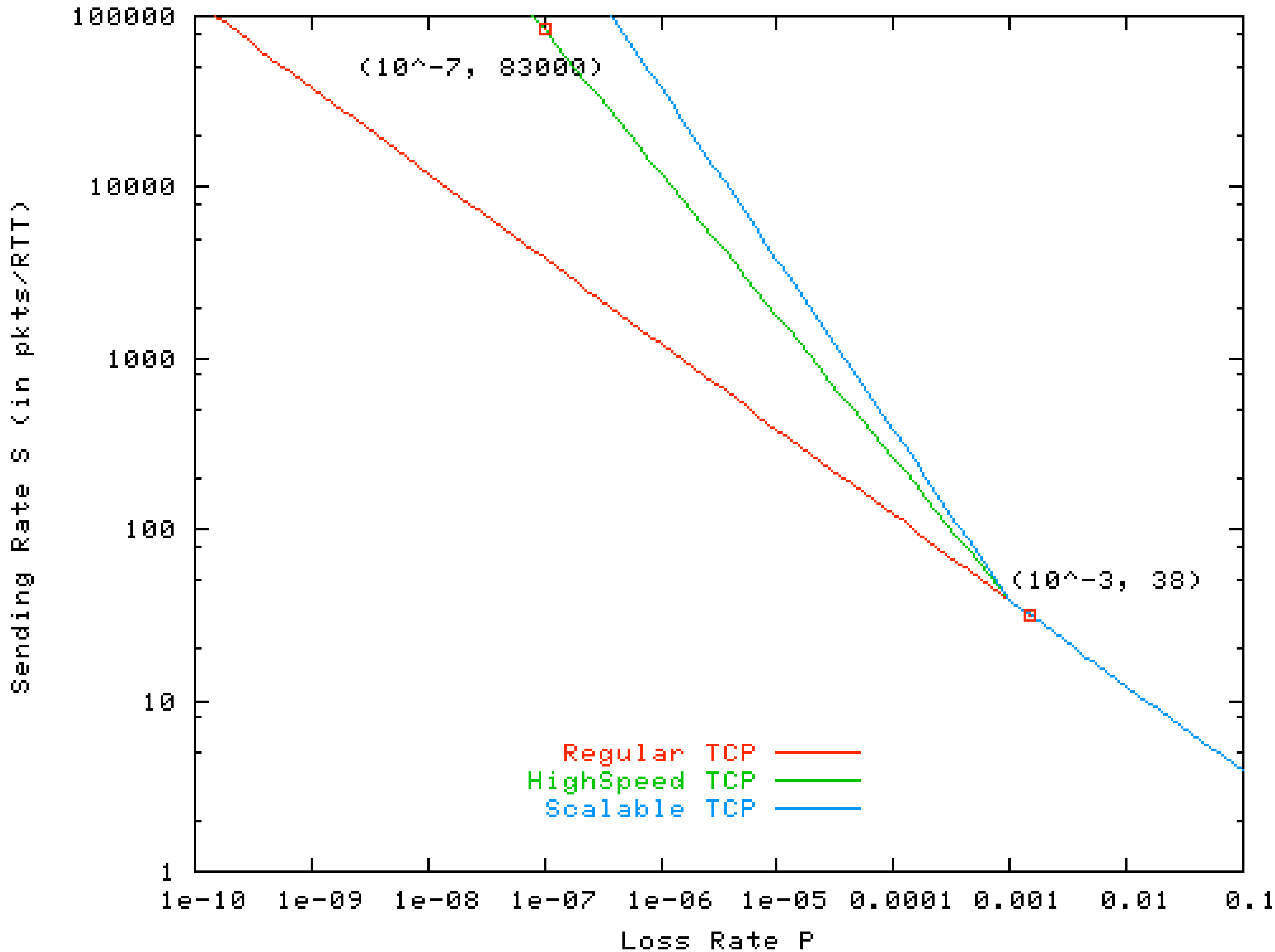
- Distributions of packet round-trip times on the congested link of two simulations, with data measured on the Internet for comparison.

Characterizing scenarios, continued.

- The degree of synchronization between flows:
  - Measured by synchronization between flow pairs.
  - Affects convergence times.
- The effect of burstiness:
  - Measured by congestion response to bursts within flows.
  - Affects throughput, delay, etc.
- Other characterizations?

Example: HighSpeed TCP

- The problem with standard TCP::
  - achieving 10 Gbps requires a window of 83,000 packets,
  - and at most one loss every 1 2/3 hours,
  - for 1500-byte packets,100 ms RTT
- The answer: more aggressive forms of TCP
  - HighSpeed TCP (HSTCP), Scalable TCP, HTCP, FAST TCP, BIC TCP, XCP, etc.
- HighSpeed TCP:
  - With higher congestion windows,
    - increase faster than one packet per RTT,
    - decrease less than halving the window.



100000
10000
1000
100
10
1
1e-10
1e-09
1e-08
1e-07
1e-06
1e-05
0.0001
0.001
0.01
0.1
Sending Rate S (in pkts/RTT)
Loss Rate P
(10^-7, 83000)
(10^-3, 38)
Regular TCP
HighSpeed TCP
Scalable TCP

Concerns raised by HighSpeed TCP:

- Two key metrics:
  - Fairness with standard TCP.
  - Convergence times.
- Different models give different convergence times!
  - Model #1: DropTail queues, global synchronization for loss events.
  - Model #2: Drop Tail queues, some synchronization, (depending on traffic mix).
  - Model #3: RED queues, some synchronization, (depending on traffic mix).
  - Model #4: RED queues, no synchronization
- Which model is the best fit for the current and future Internet?

Synchronization for HighSpeed TCP:

- What to we know about synchronization on high-bandwidth paths?
  - Has it been measured?
  - Is there a rich traffic mix at the congested router?
  - Do the congested routers use AQM?
- Are future high-bandwidth paths likely to be similar to current ones?

Example: Quick-Start

- Quick-Start (QS):
  - A proposal for end nodes to ask permission from routers to send at a high rate.
    - Uses an IP option.
    - Routers approve request by decrementing a counter.
    - Approval only by underutilized routers.
- Metrics for evaluation?
  - Effective use of bandwidth in underutilized paths?
  - Incremental deployment?
  - Robustness against competition and attacks?

Evaluation of QuickStart:

- Possible problems:
  - Senders asking for too much QuickStart bandwidth.
  - QuickStart requests denied downstream.
  - Malicious QuickStart requests.
- The result:
  - No QuickStart bandwidth available for others.
- The partial fix:
  - Routers keeping history of sender's behaviors.

Evaluation of QuickStart, continued:

- Lessons:
  - Each mechanism has its own strengths and weaknesses, that need to be discovered and explored.
- For evaluating QuickStart:
  - Knowing behavior in the current Internet doesn't help.
  - Evaluation is about understanding possible behavior in the future Internet.
  - This is hard.

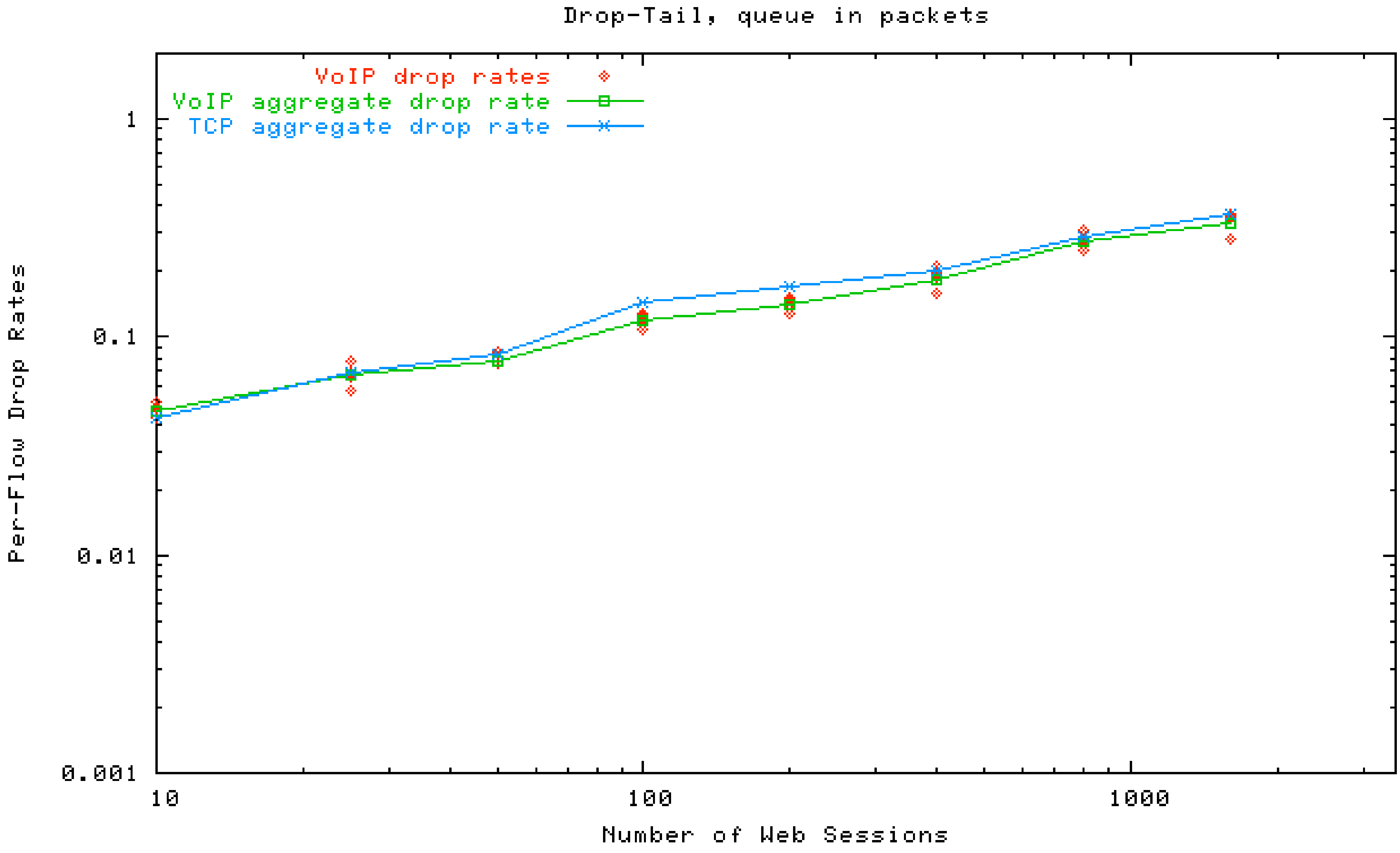
Example: congestion control for VoIP

- TFRC (TCP-Friendly Rate Control):
  - The same average sending rate, in packets per RTT, as a TCP flow with the same loss event rate.
  - More slowly-responding than TCP -
    - Doesn't halve the sending rate in response to a single loss.
  - The mechanism:
    - The receiver calculates the loss event rate.
    - The sender calculates the allowed sending rate for that loss event rate.

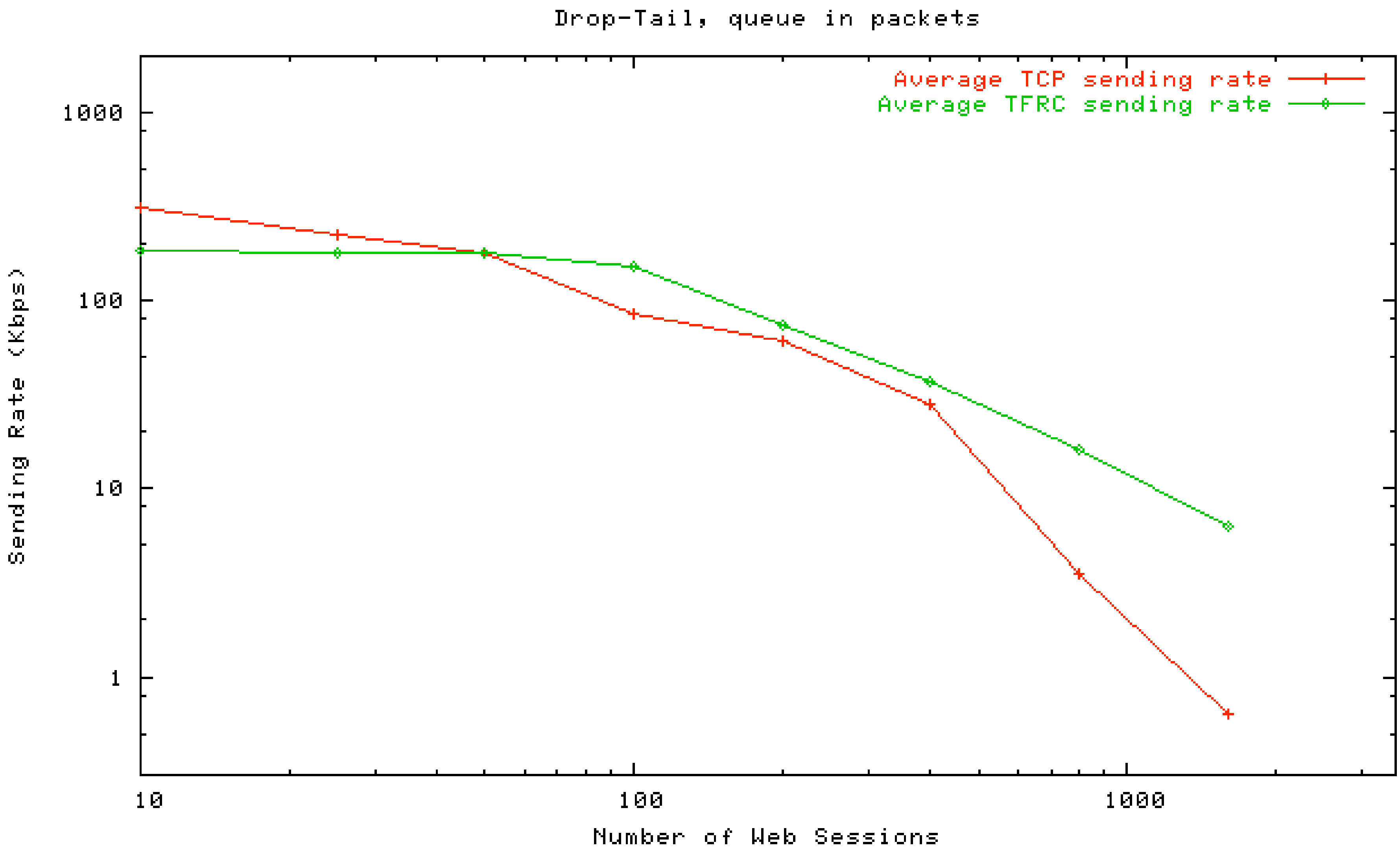
VoIP TFRC:

- A variant of TFRC for flows with small packets:
  - Sending at most 100 packets per second.
- The goal:
  - The same sending rate in bytes per second as TCP flows with large packets and the same packet drop rate.
- The problem:
  - Works fine when flows with small packets receive a similar packet drop rate as flows with large packets…

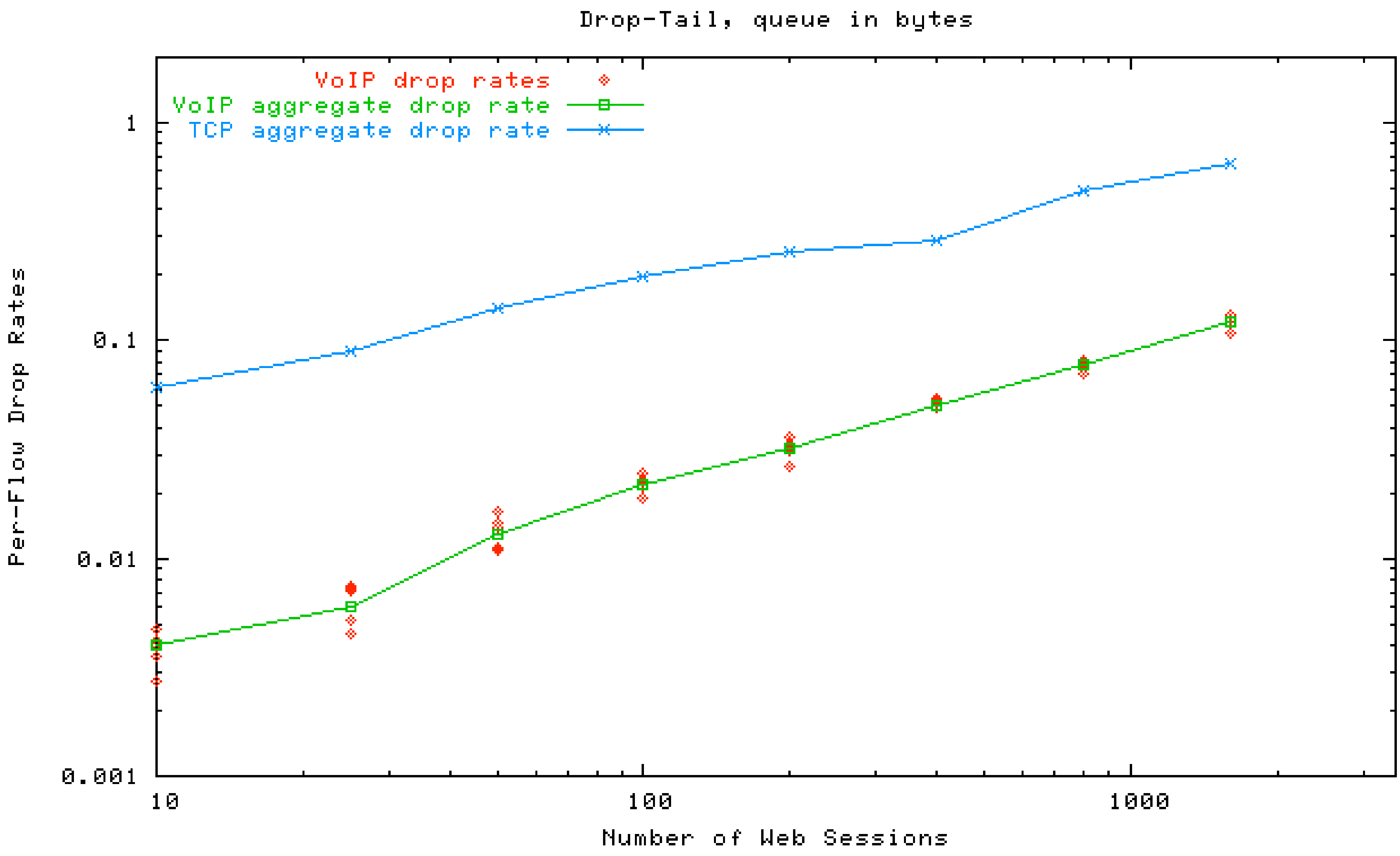
VoIP TFRC, Queue in Packets:



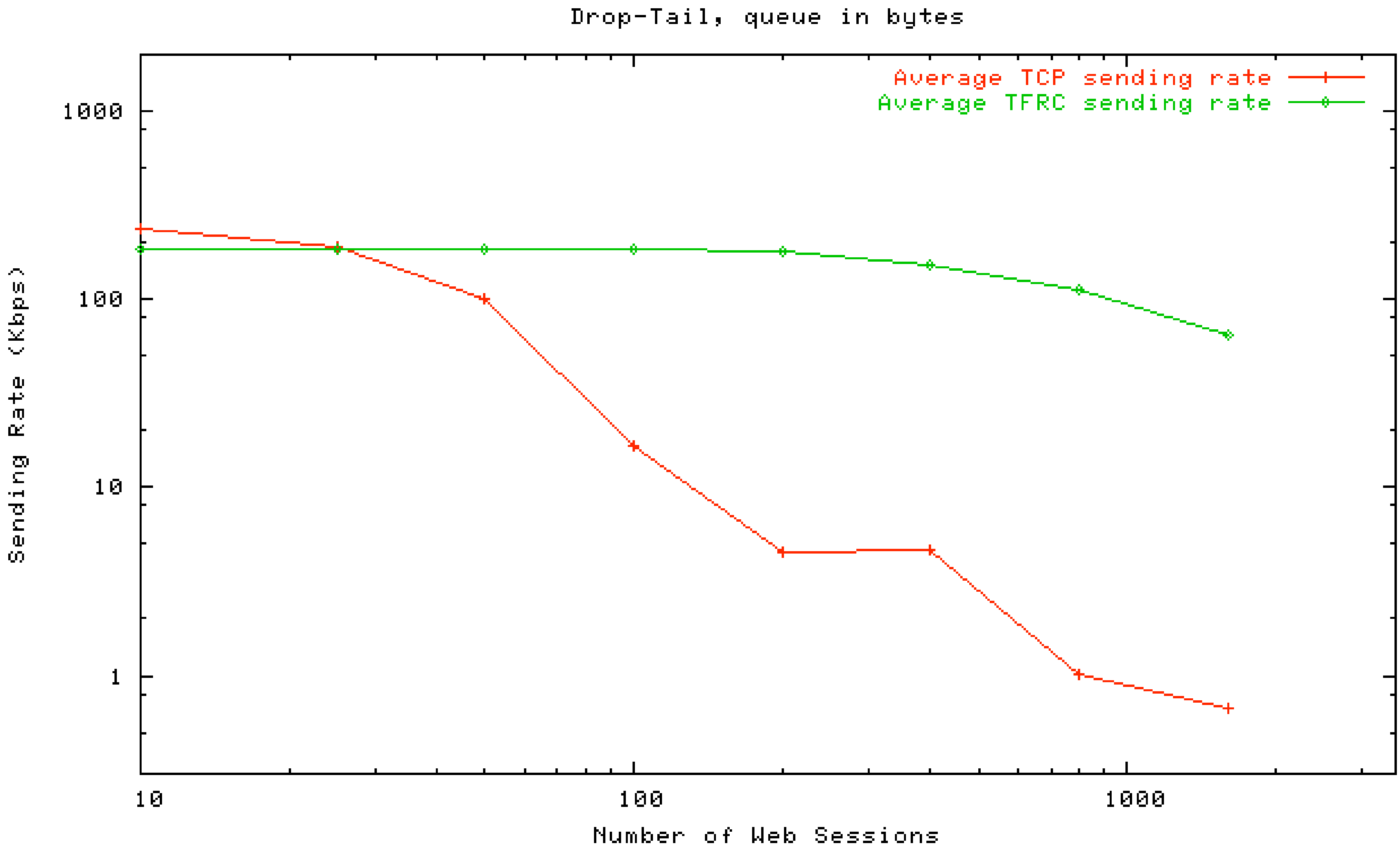
VoIP TFRC, Queue in Packets:



VoIP TFRC, Queue in Bytes:



VoIP TFRC, Queue in Bytes:



VoIP TFRC:

- What are queues like in congested routers in the Internet today and tomorrow?
  - Queue in packets, bytes, or K-byte buffers?
    - Cisco routers have pools of fixed-size buffers, e.g., of 1500B, 600B, and 80B. What is the effect on packet-dropping?
  - DropTail or AQM (Active Queue Management)?
  - If AQM:
    - Are all packets dropped with the same probability (e.g.,Cisco)?
    - All bytes dropped with the same probability?
- Should transport be changed to accommodate small-packet flows, or should routers be changed?

Conclusions: Questions

- How do our models affect our results?
- How do our models affect the relevance of our results to the current or future Internet?
- What kinds of tools do we need to improve our understanding of models?

References:

- Metrics for the Evaluation of Congestion Control Mechanisms, S. Floyd, internet-draft draft-floyd-transport-models-00.txt, 2005.
- Internet Research Needs Better Models, Sally Floyd and Eddie Kohler, Hotnets-I, 2002.
- On Traffic Phase Effects in Packet-Switched Gateways, S. Floyd and V. Jacobson, Internetworking: Research and Experience, 1992.
- HighSpeed TCP for Large Congestion Windows, S. Floyd, RFC 3649, 2003.

References, continued:

- **Evaluating Quick-Start for TCP**. Pasi Sarolahti, Mark Allman, and Sally Floyd. February 2005.
- TCP Friendly Rate Control (TFRC) for Voice: VoIP Variant, Sally Floyd, internet-draft draft-ietf-dccp-tfrc-voip-02.txt, work in progress, July 2005.