

Standardizing New Congestion Control Algorithms

Sally Floyd

Workshop on High-speed TCP
Microsoft
February 5-6, 2007

Slides: <http://www.icir.org/floyd/talks.html>

What is the problem?

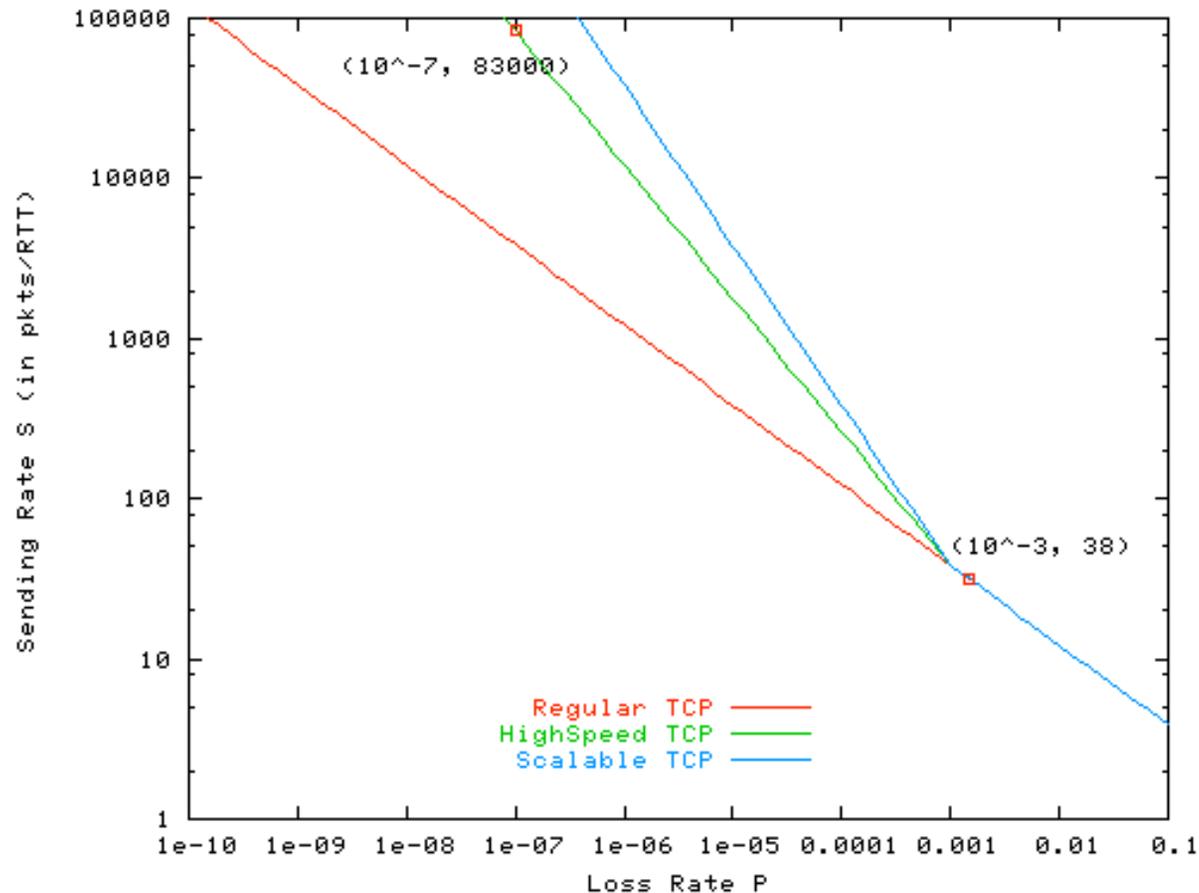
- **The problems:**
 - There are many proposed congestion control mechanisms.
 - Some TCP implementations use congestion control that has not been through IETF process.
 - E.g., Linux and BIC TCP.
 - Windows Server “Longhorn” and Compound TCP?
 - ...
- Goals of “**Specifying New Congestion Control Algorithms**”, draft-floyd-tsvwg-cc-alt-00.txt, Sally Floyd and Mark Allman:
 - Encourage new congestion control mechanisms to go through IETF review. (High-speed networking, robustness in wireless environments, alternate semantics for ECN, and more.)
 - Give guidelines to authors and reviewers for considering congestion control mechanisms for Experimental status.

Experimental status:

- Experimental RFCs for congestion control would indicate, in the abstract, the status:
 - Safe to deploy in the global Internet, or not?
 - Environments where the protocol is not recommended?
- Examples:
 - RFC 3649, HighSpeed TCP (2003):
 - safe to deploy in the Internet.
 - RFC 4782, Quick-Start (2007):
 - for controlled environments.

RFC 3649, HighSpeed TCP

- HighSpeed TCP is a minimal change sufficient to allow TCP to use high-bandwidth paths.



RFC 3649, HighSpeed TCP, from the abstract:

- “The proposals in this document are experimental. While **they may be deployed in the current Internet**, they do not represent a consensus that this is the best method for high-speed congestion control. In particular, we note that alternative experimental proposals are likely to be forthcoming, and it is not well understood how the proposals in this document will interact with such alternative proposals.”

RFC 4782, Quick-Start:

- QuickStart with TCP, for setting the initial window:
- In an IP option in the TCP SYN packet,
 - the **sender's desired sending rate**:
 - Routers on the path decrement a TTL counter,
 - and decrease the allowed sending rate, if necessary.
- **The TCP receiver sends feedback to the sender** in the SYN/ACK packet:
 - The TCP sender knows if all routers on the path participated.
 - The sender has an RTT measurement.
 - The sender can set the initial congestion window.
 - The TCP sender continues using normal congestion control.

RFC 4782, Quick-Start, from the abstract:

- “This document describes many **paths where Quick-Start Requests would not be approved**. These paths include all paths containing routers, IP tunnels, MPLS paths, and the like that do not support Quick-Start. These paths also include paths with routers or middleboxes that drop packets containing IP options. Quick-Start Requests could be difficult to approve over paths that include multi-access layer-two networks. This document also describes **environments where the Quick-Start process could fail with false positives**, with the sender incorrectly assuming that the Quick-Start Request had been approved by all of the routers along the path.”
- “**As a result of these concerns**, and as a result of the difficulties and seeming absence of motivation for routers such as core routers to deploy Quick-Start, **Quick-Start is being proposed as a mechanism that could be of use in controlled environments**, and not as a mechanism that would be intended or appropriate for ubiquitous deployment in the global Internet.”

Guidelines from draft-floyd-tsvwg-cc-alt:

- Fairness to TCP, SCTP, and DCCP.
- Using spare capacity?
- Difficult environments.
- Investigating a range of environments.
- Protection against congestion collapse.
- Fairness within the proposed mechanism.
- Performance with misbehaving nodes and attackers.
- Response to sudden or transient events.
- Incremental deployment?
- To add:
 - Robust with different queue management mechanisms.
 - Robust with current Internet infrastructure (including middleboxes)?

Fairness to TCP:

- “In environments where standard congestion control is able to make reasonable use of the available bandwidth the proposed change should not significantly change this state.”
- “For instance, in a situation where each of N flows uses $1/N$ the network capacity, a new congestion control scheme should not significantly deviate from this state. For instance, a flow using an alternate congestion controller that took half the capacity and left each of the remaining N flows with $1/2N$ of the capacity would be suspect.”

Using spare capacity:

- “Alternate congestion control algorithms may take up spare capacity in the network, but **may not steal significant amounts of capacity** from flows using currently standardized congestion control.”

Difficult Environments.

- “An assessment of proposed algorithms in difficult environments such as paths containing **wireless links** and paths with **reverse-path congestion**. In addition, proposed algorithms should be evaluated in situations where the bottleneck has **high and low levels of statistical multiplexing**.”

Investigating a Range of Environments.

- “Proposed alternate congestion controllers should be assessed in a range of environments. For instance, proposals should be investigated across a range of bandwidths and round-trip times.”
- “A particularly important aspect of evaluating a proposal for standardization is in **understanding where the algorithm breaks down**. Therefore, particular attention should be paid to extending the investigation into areas where the proposal does not perform well.”

Protection Against Congestion Collapse:

- “The alternate congestion control mechanism should either stop sending when the packet drop rate exceeds some threshold [RFC3714], or should include some notion of "full backoff".”
- “For "full backoff", at some point the algorithm would reduce the sending rate to one packet per round-trip time and then exponentially backoff the time between single packet transmissions if congestion persists.”

Fairness within the Alternate Congestion Control Algorithm.

- “In environments with multiple competing flows using the alternate congestion control algorithm, the proposal should **explore how bandwidth is shared** among the competing flows.”

Performance with Misbehaving Nodes and Outside Attackers.

- “The proposal should explore how the alternate congestion control mechanism performs with **misbehaving senders, receivers, or routers**. In addition, the proposal should explore how the alternate congestion control mechanism performs with **outside attackers**. This can be particularly important for congestion control mechanisms that involve explicit feedback from routers along the path.”

Responses to Sudden or Transient Events

- “The proposal should consider how the alternate congestion control mechanism would perform in the presence of transient events such as **sudden congestion**, a **routing change**, or a **mobility event**.”

Incremental Deployment:

- “The proposal should discuss whether the alternate congestion control mechanism allows for **incremental deployment in the targeted environment**. If the alternate congestion control mechanism is intended only for specific environments, the proposal should consider how this intention is to be carried out.”

Bullets to add to the draft:

- Robust with **different queue management mechanisms**.
- Robust with **current Internet infrastructure** (including middleboxes)?
- Suggestion from Jitu:
 - Add **minimum requirements** necessary for widespread deployment in the Internet.

Informational RFCs: a possible first path.

- For congestion control mechanisms that are not yet ready to be considered for Experimental, an **Informational RFC** would be useful describing the algorithms, and giving pointers to what is known so far about performance.

Transport Modeling Research Group (TMRG)

- **Metrics for the Evaluation of Congestion Control Mechanisms.**
 - S. Floyd, internet-draft draft-irtf-tmrg-metrics-06, work in progress, December 2006.
- **Tools for the Evaluation of Simulation and Testbed Scenarios.**
 - S. Floyd and E. Kohler, internet-draft draft-irtf-tmrg-tools-03, work in progress, December 2006.
- **An NS2 TCP Evaluation Tool Suite.**
 - Yong Xia and Gang Wang, internet-draft draft-irtf-tmrg-ns2-tool-00, February 2007, not yet submitted.

Extra Viewgraphs

Metrics for the Evaluation of Congestion Control Mechanisms

- Throughput, delay, and packet drop rates.
- Response to sudden changes or to transient events; Minimizing oscillations in throughput or in delay.
- Fairness and convergence times.
- Robustness for challenging environments.
- Robustness to failures and to misbehaving users.
- Deployability.
- Security.
- Metrics for specific types of transport.

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?

Tools for Evaluating Scenarios in Simulations, Experiments, and Analysis: Characterizing Aggregate Traffic on a Link

- Distribution of per-packet **round-trip times**:
 - Measurements: Jiang and Dovrolis.
- Distribution of per-packet **sequence numbers**:
 - Measurements: distribution of connection sizes.
- Distribution of **packet sizes**.
- Ratio between forward and **reverse path traffic**.
- Distribution of per-packet **peak flow rates**.
 - Measurements: Sarvotham et al.
- Distribution of **transport protocols**.
- Typical **bandwidth and packet drop rates** for congested links.

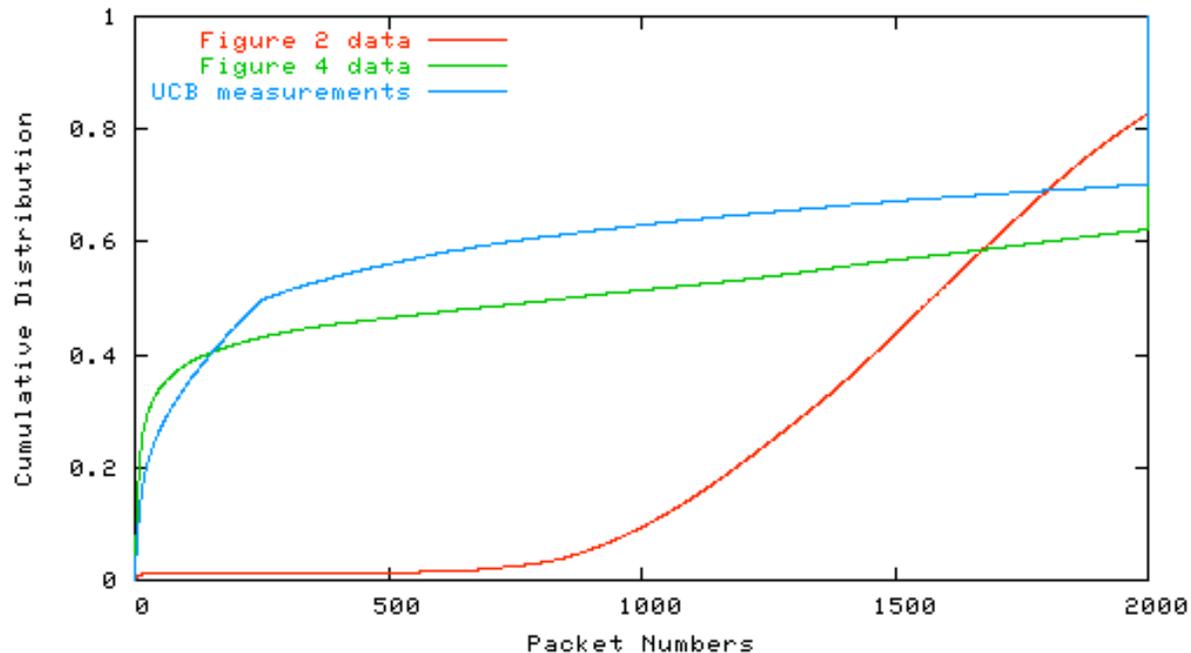
Tools for Evaluating Scenarios in Simulations, Experiments, and Analysis: Characterizing Paths

- **Synchronization Ratio.**
 - Determined by queue management (Drop-Tail or RED), level of statistical multiplexing, traffic mix, etc.
- **Drop or mark rates as a function of packet size.**
 - Determined by queue structure
 - Affects congestion control for small-packet flows.
- **Drop rates as a function of burst size.**
- **Drop rates as a function of sending rate.**
 - E.g., determined by the level of statistical multiplexing.

The Effect of Background Traffic on Congestion Control Dynamics:

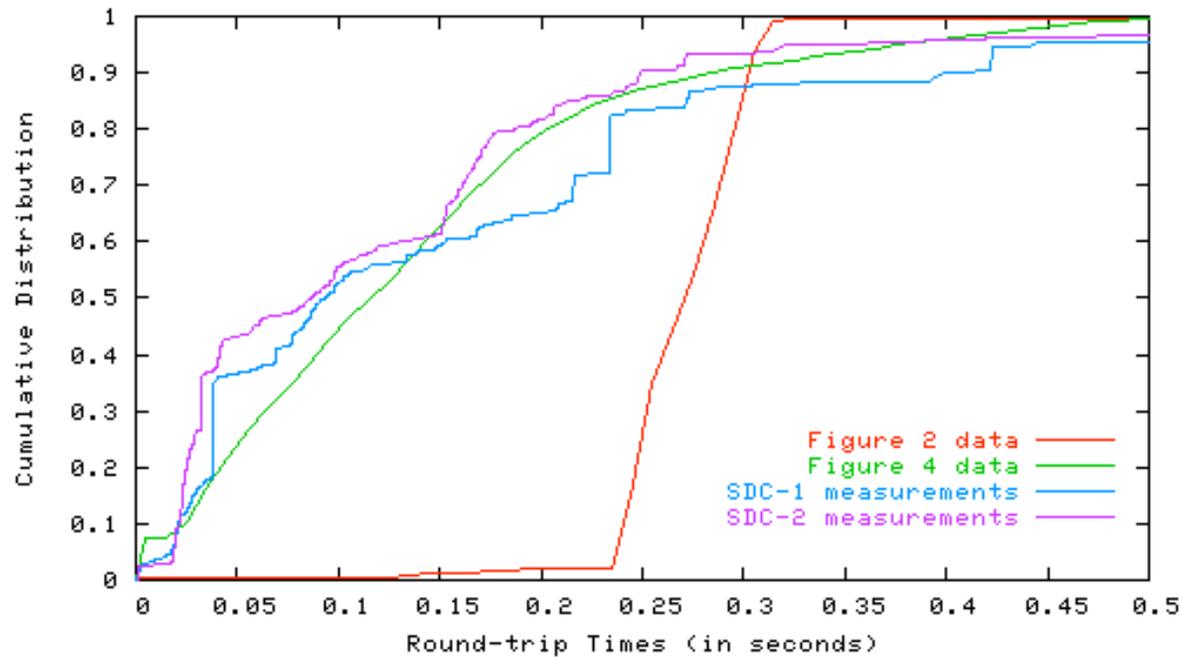
- **A Step toward Realistic Performance Evaluation of High-Speed TCP Variants**, S. Ha, Y. Kim, L. Le, I. Rhee and L. Xu, PFLDnet2006.
- **The Effect of Reverse Traffic on the Performance of New TCP Congestion Control Algorithms for Gigabit Networks**, S. Mascolo and F. Vacirca, PFLDnet2006.
- ...
- **Observations on the Dynamics of a Congestion Control Algorithm: the Effects of Two-Way Traffic**, L. Zhang, S. Shenker, and D. Clark, SIGCOMM 1991.

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 the end-to-end path: drop rates as a function of packet size

- Relevant for:
 - evaluating congestion control for VoIP and other small-packet flows.
 - E.g., **TFRC for Voice: the VoIP Variant**, draft-ietf-dccp-tfrc-voip-02.txt,
- Measurements:
 - compare drop rates for large-packet TCP, small-packet TCP, and small-packet UDP on the same path.
- There is a **wide diversity in the real world**:
 - Drop-Tail queues in packets, bytes, and in between.
 - RED in byte mode (Linux) and in packet mode (Cisco).
 - Routers with per-flow scheduling:
 - with units in Bps or in packets per second?