Congestion Control for High-Bandwidth-Delay-Product Networks: XCP vs. HighSpeed TCP and QuickStart

Sally Floyd

September 11, 2002

ICIR Wednesday Lunch

Outline:

- Description of the problem.
- Description of HighSpeed TCP and QuickStart.
- Description of XCP.
- Evaluation of differences (open discussion).

The Problem:
Congestion Control for High-Bandwidth-Delay-Product Networks

- Sustaining high congestion windows:

A Standard TCP connection with:

– 1500-byte packets;
– a 100 ms round-trip time;
– a steady-state throughput of 10 Gbps;

would require:

– an average congestion window of 83,333 segments;
– and at most one drop (or mark) every 5,000,000,000 packets
  (or equivalently, at most one drop every 1 2/3 hours).

This is not realistic.

The Problem, Part II:

- High throughput without parallel TCP connections?
- Starting up with high congestion windows?
- Making prompt use of newly-available bandwidth?
- Better ways of achieving high thoughput with low delay?

The Solution Space:

- At one end of the spectrum:

Simplier, more incremental, and more-easily-deployable changes to the current protocols:

  - HighSpeed TCP (TCP with modified parameters);
  - QuickStart (an IP option to allow high initial congestion windows.)

- At the other end of the spectrum:

More powerful changes with a new transport protocol, and more explicit feedback from the routers:

  - XCP

- And other proposals along the simplicity/deployability/power spectrums.

What is HighSpeed TCP:

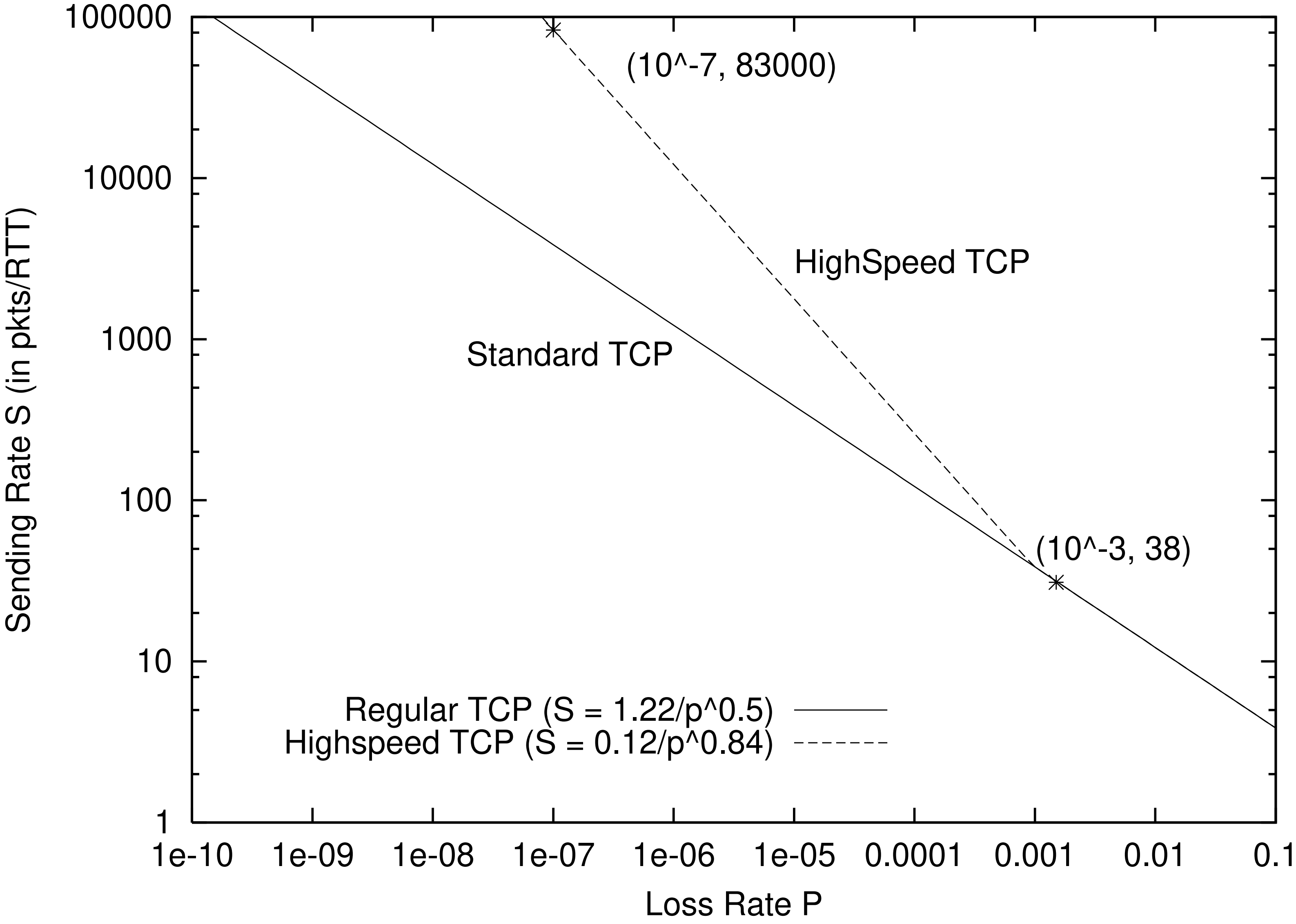
- Just like Standard TCP when cwnd is low.

- More aggressive than Standard TCP when cwnd is high.
  - Uses a modified TCP response function.

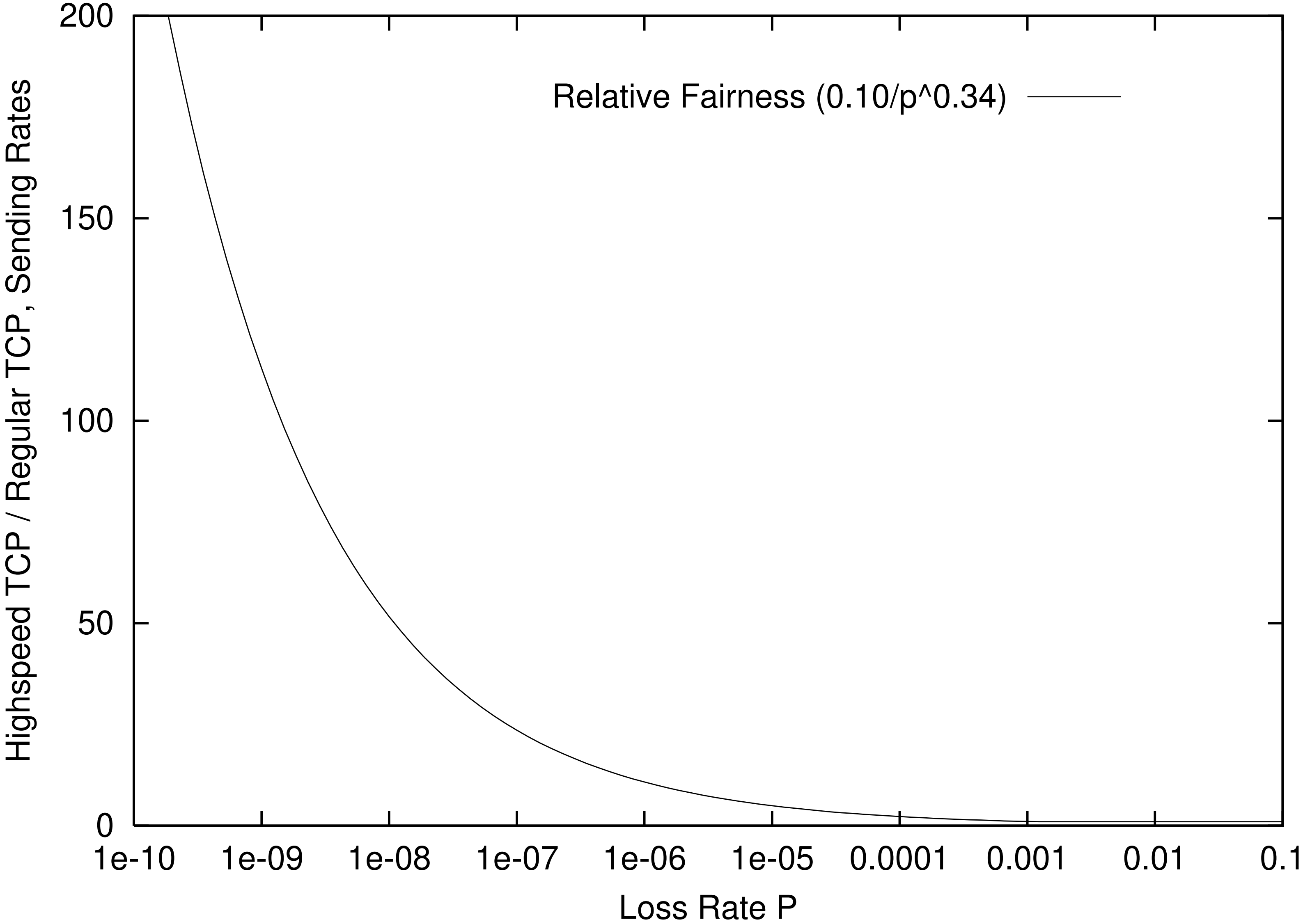
- HighSpeed TCP can be thought of as behaving as an aggregate of N TCP connections at higher congestion windows.

Joint work with Sylvia Ratnasamy and Scott Shenker, additional contributions from Evandro de Souza, Deb Agarwal, Tom Dunigan.

HighSpeed TCP: the modified response function.



HighSpeed TCP: Relative fairness.



HighSpeed TCP: The Gory Details:

w	a(w)	b(w)
38	1	0.50
118	2	0.44
221	3	0.41
347	4	0.38
495	5	0.37
663	6	0.35
851	7	0.34
1058	8	0.33
1284	9	0.32
1529	10	0.31
1793	11	0.30
2076	12	0.29
2378	13	0.28

HighSpeed TCP requires Limited Slow-Start:

- Limited Slow-Start for TCP with Large Congestion Windows:
  - Agent/TCP set max_ssthresh_ N
  - During the initial slow-start, increase the congestion window by at most N packets in one RTT.

```
For each arriving ACK in slow-start:
  If (cwnd <= max_ssthresh)
     cwnd += MSS;
  else
     K = 2 * cwnd/max_ssthresh ;
     cwnd += MSS/K ;
```

What is QuickStart?

- In an IP option in the SYN packet, the sender's desired sending rate:
  - Routers on the path decrement a TTL counter,
  - and decrease the allowed initial sending rate, if necessary.

- The receiver sends feedback to the sender in the SYN/ACK packet:
  - The 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 with AIMD using normal methods.

From an initial proposal by Amit Jain

The Quick-Start Request Option for IPv4

```
     0          1          2          3
+----------+----------+----------+----------+
| Option   | Length=4 |  QS TTL  | Initial  |
|          |          |          | Rate     |
+----------+----------+----------+----------+
```

- Explicit feedback from all of the routers along the path would be required.
- This option will only be approved by routers that are significantly under-utilized.
- No per-flow state is kept at the router.

Questions:

- Would the benefits of Quick-Start be worth the added complexity?
  – SYN and SYN/ACK packets would not take the fast path in routers.

- Is there a compelling need to add some form of congestion-related feedback from routers such as this (in addition to ECN)?

- Is there a compelling need for more fine-grained or more frequent feedback than Quick-Start?

- If so, are there other mechanisms that would be preferable to Quick-Start?

What is XCP?

- Congestion Control for High Bandwidth-Delay Product Networks
  - by Dina Katabi, Mark Handley, and Charlie Rohrs.

- XCP (eXplicit Control Protocol) has the goals of stability,
fair bandwidth allocation, high utilization, small standing queue size,
and near-zero packet drops.

- Specific goals:
  - Minimizing oscillations.
  - High delay-bandwidth-product connections.
  - Minimizing the transfer delay of short flows.
  - Fairness between flows with different RTTs.

- No per-flow-state is maintained in routers.

XCP: the End Nodes

- The packet header contains:
  - current cwnd,
  - rtt estimate,
  - feedback
    (Initialized to the desired increase in bytes in the cwnd, per ACK.)

- Routers modify the feedback field.

- At the sender, for each ACK:

cwnd $< -$ max (cwnd + feedback, packet size)

XCP: the Routers

- Routers deal with efficiency and fairness separately.

- The efficiency controller computes the desired change in the number of arriving bytes in a control interval (i.e., an average RTT), based on the spare bandwidth and persistent queue.

- The fairness controller uses AIMD to allocate the increase or decrease to individual packets.

- This requires a few additions and three multiplications per packet.

- Policing agents can be used at the edge of the network for security.

The Relative Evaluation?

- HighSpeed TCP should be done whether or not XCP is done also.

- What about QuickStart?
Would QuickStart delay the deployment of something stronger like XCP?
Is something stronger like XCP needed?

- If HighSpeed TCP and QuickStart were deployed, how pressing would it be to deploy something stronger?

Architectural sub-themes favoring incremental deployment:

- A goal of incremental deployment in the current Internet.

- Steps must go in the fundamantally correct, long-term direction, not be short-term hacks.

- Robustness in heterogeneous environments valued over efficiency of performance in well-defined environments.

- A preference for simple mechanisms, but a skepticism towards simple traffic and topology models.

- Learning from actual deployment is an invaluable step.

- The Internet will continue to be decentralized and fast-changing.

Architectural sub-themes favoring bolder actions?

- We can't make changes all that often, so why don't we just get it right?

- High-delay-bandwidth environments shouldn't require a delay-bandwidth-product of buffering (or even a fraction of a delay-bandwidth-product of buffering) at the routers.

References:

- HighSpeed TCP:

http://www.icir.org/floyd/hstcp.html

- QuickStart:

http://www.icir.org/floyd/papers/draft-amit-quick-start-01.txt

- XCP:

http://www.acm.org/sigcomm/sigcomm2002/papers/xcp.html