Update on
Router Mechanisms to Support End-to-End Congestion Control

Sally Floyd
Lawrence Berkeley National Laboratory
http://www-nrg.ee.lbl.gov/floyd/

Internet End-to-End Research Group
July 8-9, 1997

Joint work with Kevin Fall

URL ftp://ftp.ee.lbl.gov/papers/collapse.ps

Scenarios for congestion collapse:

Congestion collapse occurs when the network is increasingly busy, but little useful work is getting done.

Problem: Paths clogged with unnecessarily-retransmitted packets. [Nagle 84]
Fix: Modern TCP retransmit timer and congestion control algorithms [J88].

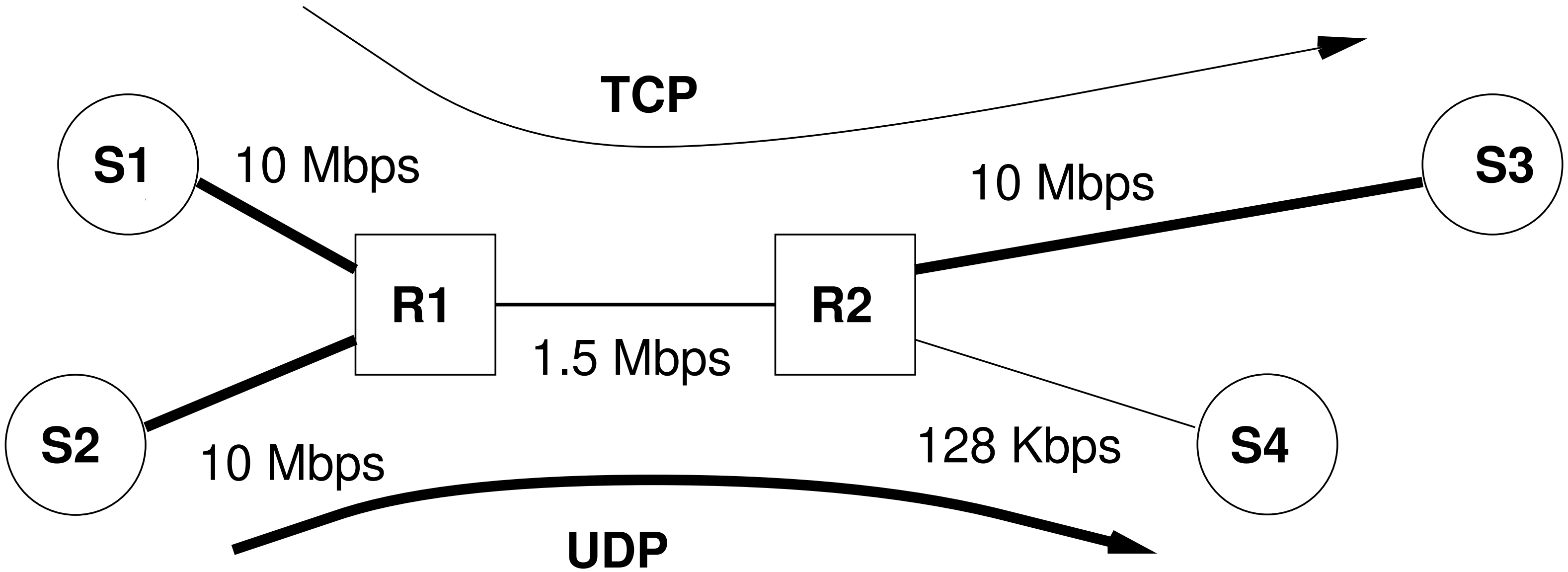
Problem: Paths clogged with fragments of packets invalidated because another fragment (or cell) has been discarded along the path. [KM, 1987]
Fix: MTU discovery [KMM 1988], Early Packet Discard in ATM networks [RF95].

Scenarios for congestion collapse, continued:

Problem: Paths clogged with packets that are discarded before they reach the receiver.

Fix: End-to-end congestion control.

Disincentives at the routers for flows that do not use either end-to-end congestion control or explicit bandwidth allocation mechanisms.



Scenarios for congestion collapse, continued:

Problem: Networks where a small fraction of the bytes actually belong to application-level data because of high levels of control traffic or small data packets with large packet headers.

This scenario is only related to congestion if the congestion itself results in increasing levels of control traffic (e.g., from reliable multicast, routing?) or increasingly large ratios of header information to application data in packets.

Fix: Not a present danger. Design to avoid increasing levels of control traffic as levels of congestion increase.

Two orthogonal issues:

1. Identifying high-bandwidth flows (e.g., from random sampling).

2. Determining which high-bandwidth flows to regulate.

- The TCP-friendly test.

- The test for responsiveness.

- The very-high-bandwidth test.

The TCP-friendly test.

A flow is accepted as TCP-friendly if the arrival rate is no more than
1.45 B/(R $\sqrt{p}$),
for B the upper bound on the packet size in bytes,
R the lower bound on the roundtrip time in seconds, and
p the packet drop rate at that queue over that time interval.

The time interval should be large relative to a typical roundtrip time.

The test for responsiveness.

A flow is judged to be unresponsive if the packet drop rate increases by a factor of four, but the flow's arrival rate does not decreased to below 90% of its previous value.

Basis: For a TCP flow with persistent demand, if the long-term packet drop rate of the connection increases by a factor of x, then the arrival rate from the source should decrease by a factor of at least $\sqrt{x}$.

The very-high-bandwidth test.

A flow is identified as very-high-bandwidth, and needing to be regulated, if its estimated arrival rate is greater than the maximum arrival rate of a TCP flow with 512-byte packets, a 50 msec roundtrip time, and the measured packet drop rate, AND it is using a “disproportionate share of the link bandwidth”.

A flow is using a “disproportionate share of the link bandwidth” if its fraction of the best-effort arrival rate is more than $\log(3n)/n$, for n the number of flows with recent packet drops. This function is close to one for n equal to 2, and grows slowly as a multiple of 1/n.