Lecture 16: Multi-path TCP

CSE 222A: Computer Communication Networks
Alex C. Snoeren

Thanks: Costin Raiciu
Lecture 16 Overview

- TCP review
- MPTCP Overview
TCP Connection Setup

HTTP server listening on port 80
TCP Connection Setup

SYN
DPORT 80 SPORT 12572
TCP Connection Setup

SYN/ACK
DPORT 80 SPORT 12572
TCP Connection Setup

ACK
DPORT 80 SPORT 12572
TCP Data Transmission

Browser: Send 3KB of data

HTTP Server: Read Request
TCP Data Transmission

Browser: Send 3KB of data

HTTP Server: Read Request
TCP Data Transmission

Browser: Send 3KB of data

HTTP Server: Read Request

Data: 1000-2000 → Data: 0-1000
TCP Data Transmission

Data: 2001-3000
Data: 1001-2000
Data: 1-1000

Browser: Send 3KB of data
HTTP Server: Read Request

CSE 222A – Lecture 16: Multipath TCP
TCP: Lost Packets

Browser: Send 3KB of data

HTTP Server: Read Request
TCP: Reordering

Browser: Send 3KB of data

HTTP Server: Read Request
TCP: Sequence #s and ACKs


Browser: Send 3KB of data  HTTP Server: Read Request
TCP: Sequence #s and ACKs

Browser: Send 3KB of data

HTTP Server: Read Request
TCP: Sequence #s and ACKs

Browser: Send 3KB of data

HTTP Server: Read Request
TCP: Sequence #s and ACKs

Browser: Send 3KB of data

HTTP Server: Read Request

ACK 1001  ACK 2001  ACK 3001
Multipath TCP

- MPTCP is a drop in replacement for TCP
  - Works with unmodified applications
  - Over the existing network
Drop-in TCP Replacement

The sender stripes packets across paths.

user space
Socket API

MPTCP

IP

addr

The receiver puts the packets in the correct order.

MPTCP

addr

addr_1

addr_2
MPTCP Operation

SYN
MP_CAPABLE X

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

SYN/ACK
MP_CAPABLE Y
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

SYN
JOIN Y
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

SYN/ACK
JOIN X

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

options

SEQ 1000
DSEQ 10000
DATA
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
1000
DSEQ
10000
options

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
5000
DSEQ
11000
options

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
DSEQ
options

…
1000
10000

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
DSEQ
options

…
5000
11000

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
1000
DSEQ
10000
options

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
5000
DSEQ
11000
options
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

ACK
2000

CSE 222A – Lecture 16: Multipath TCP
MPTCP Operation

STATE 1
CWND
Snd.SEQNO
Rcv.SEQNO

STATE 2
CWND
Snd.SEQNO
Rcv.SEQNO

DATA
SEQ
2000
DSEQ
11000
...
options
Each path runs its own congestion control, to detect and respond to the congestion it sees.

But link the congestion control parameters, so as to move traffic away from the more congested paths.
Design goal 1:  
Be fair to regular TCP

- To be fair, Multipath TCP should take as much capacity as TCP at a bottleneck link, no matter how many paths it is using.
Design goal 2:
MPTCP should use efficient paths

● Each flow has a choice of a 1-hop and a 2-hop path.
● How should split its traffic?
Design goal 2:
MPTCP should use efficient paths

- If each flow split its traffic 1:1 ...
Design goal 2:
MPTCP should use efficient paths

- If each flow split its traffic 2:1 ...
Design goal 2: 
**MPTCP should use efficient paths**

- If each flow split its traffic 4:1 ...
Design goal 2: MPTCP should use efficient paths

- If each flow split its traffic $\infty:1$ ...
Theoretical solution (Kelly+Voice 2005; Han, Towsley et al. 2006) MPTCP should send all its traffic on its least-congested paths.

Theorem. This will lead to the most efficient allocation possible, given a network topology and a set of available paths.
Design goal 3: Perform as well as TCP

- Design Goal 2 says to send all your traffic on the least congested path, in this case 3G. But this has high RTT, hence it will give low throughput.

Goal 3a. A Multipath TCP user should get at least as much throughput as a single-path TCP would on the best of the available paths.

Goal 3b. A Multipath TCP flow should take no more capacity on any link than a single-path TCP would.
Design goals

- Goal 1. Be fair to TCP at bottleneck links redundant
- Goal 2. Use efficient paths ...
- Goal 3. as much as we can, while being fair to TCP
- Goal 4. Adapt quickly when congestion changes
- Goal 5. Don’t oscillate
TCP congestion control

- Maintain a congestion window $w$.

- Increase $w$ for each ACK, by $1/w$

- Decrease $w$ for each drop, by $w/2$
MPTCP congestion control

- Maintain a congestion window $w_r$, one window for each path, where $r \in R$ ranges over the set of available paths.

- Increase $w_r$ for each ACK on path $r$, by
  
  $\min_{S \subseteq R : r \in S} \frac{\max_{s \in S} \frac{w_s}{\text{RTT}_s^2}}{\left( \sum_{s \in S} \frac{w_s}{\text{RTT}_s} \right)^2}$

- Decrease $w_r$ for each drop on path $r$, by $w_r/2$
MPTCP congestion control

- Maintain a congestion window $w_r$, one window for each path, where $r \in R$ ranges over the set of available paths.

- Increase $w_r$ for each ACK on path $r$, by

  $\min_{S \subseteq R : r \in S} \max_{s \in S} \frac{w_s}{RTT_s^2} \left( \sum_{s \in S} \frac{w_s}{RTT_s} \right)^2$

- Decrease $w_r$ for each drop on path $r$, by $w_r/2$
MPTCP congestion control

- Maintain a congestion window $w_r$, one window for each path, where $r \in R$ ranges over the set of available paths.
- Increase $w_r$ for each ACK on path $r$, by
  $$\min_{S \subseteq R : r \in S} \frac{\max_{s \in S} w_s / RTT_s^2}{\left( \sum_{s \in S} w_s / RTT_s \right)^2}$$
- Decrease $w_r$ for each drop on path $r$, by $w_r/2$
Discussion

- How much of the Internet can be pooled?
- What are the implications for network operators?
- How should we fit multipath congestion control to CompoundTCP or CubicTCP?
- Is it worth using multipath for small flows?
For Next Class…

- Read and review Jigsaw paper
- Keep going on projects!
  - Checkpoint 2 only 1 week away