Summer Intern Showcase 2016
Hello, I am Utkarsh Goel

- Division: PDG
- Department: Foundry Web Experience
- Title: Research Engineer
- Hiring Manager: Moritz Steiner
- School Name: Montana State University
- Graduation: 2017/Computer Science
- Location: San Francisco, CA

How I spent my summer at Akamai:

- HTTP/2 Performance in Cellular Networks.
- Are 3rd Parties Slowing Down the Web? [Ongoing]

Memorable moment from your summer experience:

- Got two papers accepted to ACM MobiCom 2016 [Top Scientific Conference]
Web Performance

• HTTP/1.x
  • Creates **Head-of-Line** blocking.
  • Can not download objects **in parallel**.
  • Slows down webpages.

• HTTP/2
  • Request-response **multiplexing**.
  • **Eliminates** Head-of-line blocking.
  • Designed to speedup webpages.

• Previous investigations on HTTP/2 performance
  • Some show **improvements**.
  • Others show **degradation**.
Technical Problem

• In the event of loss
  • Single TCP connection hurts performance of all HTTP/2 streams.
  • Congestion window reduces to half

• For 6 TCP connections in HTTP/1
  • Cumulative congestion window is still larger than HTTP/2

• Cellular networks experience more loss than wired last mile networks
  • About 32% connections overall

• Unclear as to how HTTP/2 performs in cellular networks, especially under loss?
  • Need to understand if we should adopt HTTP/2 for mobile Web.
  • Today, we don’t optimize TCP specifically for HTTP/2
Challenges

• Need to correlate HTTP/2 performance with loss

  • Option 1: **Use RUM**
    • RUM does not record TCP loss.

  • Option 2: **Use RUM and TCP logs together**
    • Map RUM and TCP records.
    • **Extremely low overlap** between TCP logs and RUM records for HTTP/2 page loads.
      • Only 1 RUM overlap in 30 million TCP logs.

• **Could not utilize RUM and TCP logs**
Our Approach – Data Collection

• **Captured TCP traces** from inside a T-Mobile AANP region
  • Serves traffic to (only) cellular clients.
  • About 50K connections
  • Observe connection characteristics at every 70ms interval
    • Latency
    • Retransmission rate (loss rate)
    • Throughput
    • Time gap between loss events

• **Insights**
  • Connections experience loss **multiple times** during their lifetimes.
  • Losses are often **clustered**
    • **Multiple packets are lost** during a loss event
Our Approach – Experimental Setup

• Connection classification
  
  • Based on **how often loss occurs**
    • Or, time gap between loss events
    • **Good connection**: Time gap > 1 second
    • **Median connection**: Time gap about 500 ms
    • **Bad connection**: Time gap < 250 ms
  
  • Simulation on AWS instances
    • **Replayed** TCP connections obtained from a cellular network
    • Measure HTTP/2 webpage load time under different loss conditions.
Overview of Results

• Spinning globe page - Multiple very small objects
  • Server sends many objects in parallel during TCP slow start.
  • HTTP/2 reduces page load time under all network conditions, compared to HTTP/1.

• A page with few large objects
  • HTTP/2 increases page load time
    • ICW is one-sixth
    • Congestion window of single HTTP/2 connection reduces at every loss event
  • HTTP/1.1 reduces page load time
    • Cumulative ICW for 6 HTTP/1 connections remains larger
    • Delivers more data in lesser round trips.

• Real webpages – 2MB, 6MB, 11MB
  • HTTP/2 reduces page load time for 2MB, but increases for 6MB and 11MB pages.
Rethinking Sharding for HTTP/2

A real 8 MB webpage – designed using HTTP Archive data

Page Load Time (Seconds)

- h1 (CWND 10, 6 Connections)
- h2 (CWND 10, 1 Connection)
- h2 (CWND 10, Sharded)
- h2 (CWND 60, 1 Connection)

Network Condition
- Good
- Median
- Poor
Some Take Aways

• Traditional **network simulators do not emulate** cellular networks accurately
  • Inject loss randomly

• We offer a **novel tool** to simulate cellular networks that emulates
  • clustered loss events
  • And does NOT inject loss randomly

• **Isolating large downloads** on separate connections improves mobile Web performance.

• Potential recommendations to FEO to **not disable sharding** for HTTP/2 enabled webpages.

• https://gsd.akamai.com/jira/browse/FOUNDRY-104
Questions?

Utkarsh Goel
ugoel@akamai.com

Thank you
Background Slides
Webpage with few large objects

Page Load Time (Seconds)

- h1 (CWND 10, 6 Connections)
- h2 (CWND 10, 1 Connection)
- h2 (CWND 10, 2 Connections)
- h2 (CWND 10, 3 Connections)
- h2 (CWND 10, 6 Connections)
- h2 (CWND 10, 10 Connections)
- h2 (CWND 60, 1 Connection)

Network Condition

- Good
- Median
- Poor
The Spinning Globe page

![Box plots showing page load time for different network conditions and HTTP versions.](image)

- **HTTP/1.1 (6 Connections)**
- **HTTP/2 (1 Connection)**
- **HTTP/2 (2 Connections)**
- **HTTP/2 (3 Connections)**
Sharding

CDF of Requests

Page Load Time (Seconds)

h2

Sharding Type A

Sharding Type B

Sharding Type C

Montana State University

Akamai

Faster Forward