Demystify the Impact of E2E Latency on Mobile QoE

Dr. Jie Hui (jie.hui@t-mobile.com), Principal Engineer
QoE Lab, T-Mobile U.S.A
Jun 25, 2013

Acknowledgements to T-Mobile team and contributors: Kevin Lau, Himesh Bagley, Jeffery Smith, Pablo Tapia, Jason Chang, Aaron Drake, Antoine Tran
Introduction of Me

15 years experience in mobile networking R&D, passionate about **Quality of User Experience (QoE)** and **E2E Cross-Layer Optimizations**

- QoS over UMTS, ZTE, 3G R&D
- QoS over Wi-Fi, NCSU, ECE
- QoS over Wi-Max, Intel Labs, Wireless Network Lab
- QoE over Cellular Network, T-Mobile, QoE Lab
Agenda

• Mobile trends and challenges
• Latency mysteries
• Call for actions
Mobile Trend and Challenges
Aggressive Mobile Data Growth

Mobile Traffic as % of Global Internet Traffic = Growing 1.5x per Year & Likely to Maintain Trajectory or Accelerate

“2013 Internet Trends”, KCPB

Mobile voice traffic stays steady
Mobile data traffic grows exponentially

“The State of the Internet”, Q4 2012, Akamai
### Faster Mobile Network Speed

- Access technology faster – GSM, EDGE, UMTS, HSPA, HSPA+, LTE
- Mobile connection speed faster

#### Average Speed

<table>
<thead>
<tr>
<th></th>
<th>2011 Q4</th>
<th>2012 Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-1</td>
<td>1.2</td>
<td>3.4</td>
</tr>
<tr>
<td>US-2</td>
<td>1.5</td>
<td>4.0</td>
</tr>
<tr>
<td>US-3</td>
<td>1.3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

#### Peak Speed

<table>
<thead>
<tr>
<th></th>
<th>2011 Q4</th>
<th>2012 Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-1</td>
<td>2.0</td>
<td>6.5</td>
</tr>
<tr>
<td>US-2</td>
<td>2.8</td>
<td>7.0</td>
</tr>
<tr>
<td>US-3</td>
<td>1.9</td>
<td>5.2</td>
</tr>
</tbody>
</table>

---

“The State of the Internet”, Q4 2011/2012, Akamai
## Mobile Device Processing Faster

### Snapdragon platform examples

<table>
<thead>
<tr>
<th></th>
<th>Snapdragon 800</th>
<th>Snapdragon 600</th>
<th>Snapdragon 400</th>
<th>Snapdragon 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Up to 2.3 GHz Quad Krait 400 CPU</td>
<td>Up to 1.9 GHz Quad Krait 300 CPU</td>
<td>Up to 1.7 GHz Dual Krait 300 CPU</td>
<td>Up to 1.4 GHz Quad Cortex A5 CPU</td>
</tr>
<tr>
<td>GPU</td>
<td>Adreno 330 GPU</td>
<td>Adreno 320 GPU</td>
<td>Adreno 305 GPU</td>
<td>Adreno 203 GPU</td>
</tr>
<tr>
<td>DSP</td>
<td>Hexagon, QDSP65V5A, 600MHz</td>
<td>Hexagon, QDSP6V4, 500MHz</td>
<td>Hexagon, QDSP6V4, 500MHz</td>
<td>QDSP5, 384MHz</td>
</tr>
<tr>
<td>Video</td>
<td>4k x 2k UHD video capture/playback</td>
<td>1080p HD video</td>
<td>1080p HD video</td>
<td>720p HD Video (30/15 fps)</td>
</tr>
<tr>
<td>Modem</td>
<td>3G/4G World/multimode LTE on select processors</td>
<td>No modem</td>
<td>3G/4G World/multimode LTE on select processors</td>
<td>3G CDMA/UMTS/GSM on select processors</td>
</tr>
</tbody>
</table>

Is The Mobile Experience Faster?

People expects mobile experience as fast as desktop

**Usability Engineering 101**

<table>
<thead>
<tr>
<th>Delay</th>
<th>User reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 100 ms</td>
<td>Instant</td>
</tr>
<tr>
<td>100 - 300 ms</td>
<td><em>Feels sluggish</em></td>
</tr>
<tr>
<td>300 - 1000 ms</td>
<td>Machine is working...</td>
</tr>
<tr>
<td>1 s+</td>
<td>Mental context switch</td>
</tr>
<tr>
<td>10 s+</td>
<td>I'll come back later...</td>
</tr>
</tbody>
</table>

Mobile web 30% faster  
Desktop web 5.7% faster

“Building faster mobile websites”, Ilya Grigorik, Google  
“Is the web getting faster?”, Google Analytics Blog
Great, But There Are Still Challenges

"Is the web getting faster?", Google Analytics Blog
Latency Mysteries
Mobile Browsing QoE Investigations

http://www.google.com

<table>
<thead>
<tr>
<th>Radio State Transition Time</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2.0</th>
<th>2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <a href="http://www.google.com">www.google.com</a> - /</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>256 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. <a href="http://www.google.com">www.google.com</a> - _j/f1340409.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145 ms</td>
<td>82 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. <a href="http://www.google.com">www.google.com</a> - clismo_m.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>267 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. <a href="http://www.google.com">www.google.com</a> - logo.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>123 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. <a href="http://www.google.com">www.google.com</a> - throw-46.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>258 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. <a href="http://www.google.com">www.google.com</a> ...choochoo-mete-1.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>170 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. <a href="http://www.google.com">www.google.com</a> ...choochoo-radio-2.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. <a href="http://www.google.com">www.google.com</a> ...test/new001.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. <a href="http://www.google.com">www.google.com</a> - sigpke_s.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. <a href="http://www.google.com">www.google.com</a> ...new_logo.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. <a href="http://www.google.com">www.google.com</a> ...fanico.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. <a href="http://www.google.com">www.google.com</a> ...49fada844677676.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. <a href="http://www.google.com">www.google.com</a> ...297f/hobi/hobi559.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. <a href="http://www.google.com">www.google.com</a> ...fax.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. <a href="http://www.google.com">www.google.com</a> ...box.png</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. <a href="http://www.google.com">www.google.com</a> - css</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mystery 1 – Long DNS Lookup Time

- Aggregated average DNS lookup (network side) = 222 ms
- Curious Georges looked into distribution and outliers (10s+) from device side
Root Cause: Cross-layer Analysis

Data Stall #1: Radio network controller corrupts downlink packets due to key out-of-synch after radio state transition (Cell PCH $\rightarrow$ Cell FACH)

Label: $\uparrow$ DNS query  $\downarrow$ DNS response  $\downarrow$ Corrupted DNS response
Improvements After Fix

When DNS lookup is initiated from Cell_PCH state

DNS Lookup Time (s)
Before vs. After DataStall #1 Fix

Saving avg. 3s [*] during erroneous scenario

3.7s → 0.7s

[*] Note: results provide by device testing of 200+ samples.
Mystery 2 – Long TCP Connect Time

Delay up to 11s

2nd RTT: TCP connect time

TCP SYN
TCP SYN/ACK
TCP SYN Retran
TCP SYN/ACK Retran
TCP ACK
TCP ACK Retran

TCP Connect Time (s)
Before vs. After DataStall #2 Fix

Slight improvement → Data Stall #3

Root Cause: Radio network controller scheduler unnecessarily buffer DL packets up to 11s due to a bug in credit based scheduling algorithm between NodeB and RNC → causes UL RLC retransmissions and RESETs.

[*] Note: results provide by device testing of 2000+ samples.
Mystery 3 – TCP Connect Time Still Long

Root Cause: Radio network controller UL scheduler drops UL packets during channel switching from DCH to FACH, same problem happens during HS→R99, DCH→FACH

[*] Note: results provide by device testing of 2000+ samples.
Mystery 4 – Long HTTP Response Time

Keystroke HTTP response time

Google Search Keystroke Response Time Correlation with Radio States

High latency during low power state

82% spent over the air
Cross-Layer Analysis After Search Traffic Change

Keystroke HTTP requests and responses mostly happen in high power state (Cell_DCH) on dedicated channel due to large packet size (~1200B) therefore experience high data rate and low latency.

Google Search Keystroke Response Time (s) Before vs. After Traffic Tuning

Before: 1.6s  
After: 0.4s
Misperception

Demystify

It is an elastic pipe!

- Shared channel
- Power limitation
- Traffic dependent – You have the power to tune your traffic to be radio aware!
Perhaps The Biggest Mystery Of All…
Call for Actions
Inspirations - Fast Web Evolution and Open Innovations

http://www.evolutionoftheweb.com
More Open Innovations & Inter-disciplinary R&D

CS
(application, web development)

EE
(telecommunication-engineering)

Application Layer

Link Layer

Cross Layer

CS

EE
Area 1:
Transient State Performance and Resilience

Problem:
- **Automatically** characterize and instrument transient state mobile application performance
- Packet loss and retransmissions observed during transient states [*]
- Cross-layer analysis and optimizations still not fully explored!

Area 2: Interactive Traffic Modeling and Scheduling

Problem:
- 3GPP radio scheduling designed mostly with Full Buffer Traffic model
- Higher layer scheduling (RNC, SGW) is not considered

Evaluation assumption: NodeB or eNB queues always full of packets

“Traffic models impact on OFDMA scheduling design”, P Ameigeiras et al. EURASIP Journal on Wireless Communications and Networking 2012
Quality of User Experience (QoE) Lab

Unique approach

– Measure end user experiences from the device perspective
– Seek success by improving not only averages but also the spread and long tail
– Leverage key industry and academic partners for joint R&D and Real User Measurement (RUM)
– Analyze cross functionally to achieve cross layer intelligence and automation
Thank You!
Questions?
Jie.hui@T-Mobile.com