Troubleshooting Slow Webpage Downloads

Heng Cui, Ernst Biersack
Eurecom,
Sophia Antipolis, France
• Web browsing performance is important:
  
  • for end-user
    
      • **10 seconds** page load time is the limit for keeping the user's attention on one web page [1].

  • for service vendor
    
      • From [2]: for Microsoft’s Bing, 2 second delay load time:
        • 4% reduction in clicks
        • 4% loss in revenue
        • ...

  • ISP hotline:
    
      • People complain about poor performance **to their ISP**.

---

What is a Web Page Browsing?

We see

This page is very simple. Larger pages become more complicated.
Our Goal

Find poor web browsing experience and explain it in a reasonable way.

- From the end-user perspective → Quality of Experience (QoE)
Why is Qualify of Service Not Enough?

• QoE ≠ QoS
  • RTT 100ms → 200ms?
  • throughput 200 KB/s → 100KB/s?
  • bandwidth 4 Mbs → 10Mbs?

• Tools that do NOT QoE
  • SAM knows
  • Netalyzr
• When we say **QoE** for web browsing:
  • we **use**:
    • Performance metric that can be measured in an objective manner. *(Page Load Time)*
    • Subjective feedback as a complement (through a button).

• we do **not** focus on:
  • Designing subject metrics to measure QoE
  • Mean Opinion Score (MOS)
Troubleshooting Web Browsing
A slow-down of any component along the path does matter.

Q: Which component bears the major responsibility?
- How to discover different limitations:

  - Object download gap
  - All object downloads are affected
  - Only some object downloads are affected

HTTP response behavior

Troubleshoot Web Browsing
Develop “FireLog”:
• Plug-in at Web browser (e.g. Firefox)
  • No low level packet capture
  • All measurement at one place

• Server Repository for post-processing
  • Database Management System (DBMS)
  • Diagnosis engine
Server: Collected Raw Data

Collect events per object → DB

<table>
<thead>
<tr>
<th>Column</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>client_ip_by_server</td>
<td>inet</td>
</tr>
<tr>
<td>method</td>
<td>character varying</td>
</tr>
<tr>
<td>content_load_ts</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>load_ts</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>session_start</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>session_url</td>
<td>character varying</td>
</tr>
<tr>
<td>if_complete_cache</td>
<td>integer</td>
</tr>
<tr>
<td>localaddress</td>
<td>inet</td>
</tr>
<tr>
<td>localport</td>
<td>integer</td>
</tr>
<tr>
<td>remoteaddress</td>
<td>inet</td>
</tr>
<tr>
<td>remoteport</td>
<td>integer</td>
</tr>
<tr>
<td>response_code</td>
<td>integer</td>
</tr>
<tr>
<td>http_request_bytes</td>
<td>integer</td>
</tr>
<tr>
<td>http_header_bytes</td>
<td>integer</td>
</tr>
<tr>
<td>http_body_bytes</td>
<td>integer</td>
</tr>
<tr>
<td>http_cache_bytes</td>
<td>integer</td>
</tr>
<tr>
<td>dns_start</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>dns_end</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>dns_time</td>
<td>bigint</td>
</tr>
<tr>
<td>syn_start</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>syn_end</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>tcp_cnxting</td>
<td>bigint</td>
</tr>
<tr>
<td>send_ts</td>
<td>timestamp without time zone</td>
</tr>
<tr>
<td>send</td>
<td>bigint</td>
</tr>
</tbody>
</table>
Compute limitation scores:
Q: Is the server the performance bottleneck?
A: Need to compute Server Score.
- e.g. Serv.Score= HTTP/TCP
- TCP: measure network Round Trip Time
- HTTP: measure server side processing
  - High value indicates that server is bottleneck.
Compute relevant scores and compare against thresholds.

**Diagnosis Scheme:**
- Decision Tree
- Threshold-based

Validate thresholds by lab experiments.
Home User Browsing in the “Wild”

Three user deployments of several months.

<table>
<thead>
<tr>
<th>User</th>
<th>duration</th>
<th>Totally Browsed</th>
<th>Web Pages with “High Load Time”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#page</td>
<td>#domain</td>
</tr>
<tr>
<td>A(FR)</td>
<td>5 month</td>
<td>3,451</td>
<td>579</td>
</tr>
<tr>
<td>B(FR)</td>
<td>3 month</td>
<td>1,788</td>
<td>263</td>
</tr>
<tr>
<td>C(CN)</td>
<td>2 month</td>
<td>3,766</td>
<td>535</td>
</tr>
</tbody>
</table>

Limitation Causes for “High Load Time” Pages

<table>
<thead>
<tr>
<th>User</th>
<th>Main cause</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Client</td>
</tr>
<tr>
<td>A</td>
<td>21%</td>
</tr>
<tr>
<td>B</td>
<td>28%</td>
</tr>
<tr>
<td>C</td>
<td>21%</td>
</tr>
</tbody>
</table>


PLT > 10 sec. [1]

Troubleshoot Web Browsing
Limitation by Networks

<table>
<thead>
<tr>
<th>User</th>
<th>Main cause</th>
<th>Client</th>
<th>Server</th>
<th>Local access</th>
<th>Internet</th>
<th>others</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>21%</td>
<td>4%</td>
<td>29%</td>
<td>32%</td>
<td>14%</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>28%</td>
<td>39%</td>
<td>9%</td>
<td>10%</td>
<td>14%</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>21%</td>
<td>44%</td>
<td>9%</td>
<td>6%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Use “Google” as a reference to detect local access problem

Network is the problem (e.g. loss)

Troubleshoot Web Browsing

3 or 9 seconds. Typical RTO.
We present a diagnosis system, FireLog:
- client side measurement
- server side repository

Diagnosis model for troubleshooting Web browsing.
- carefully design the model
- lab experiments for thresholds

Long-term wild deployments show that:
(i) It is not sufficient that only rely on the network causes (by RTT) to explain poor performance;
(ii) Client side factors MUST be considered;
(iii) Server side MUST BE considered

Today’s QoS measurements ignore (ii) and (iii).
Future Work

- **Build more intelligent diagnosis system**
  - Use external information
    - WIFI signal strength and interference
    - Measurements of other encllients
    - Additional active measurements
  - What-if not a single cause, but multiple causes.
  - How to improve diagnosis using information available from external network probes.
    - mPlane: a new European FP7 project

- **User Privacy**
  - How to better protect user privacy?
    - Do also diagnosis in the end client

- **More experiments.**
  - Port to browsers other than Firefox
  - Port to mobile phones
Thanks
Questions?