Troubleshooting Slow Webpage Downloads

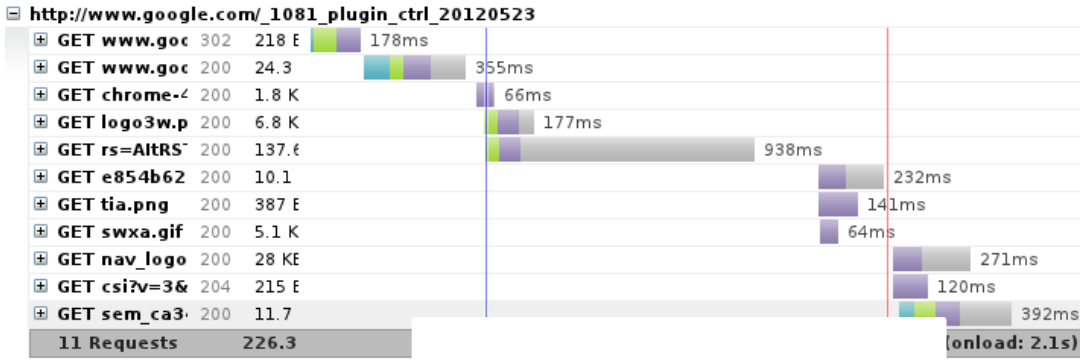
Heng Cui, **Ernst Biersack**
Eurecom,
Sophia Antipolis, France

Motivation

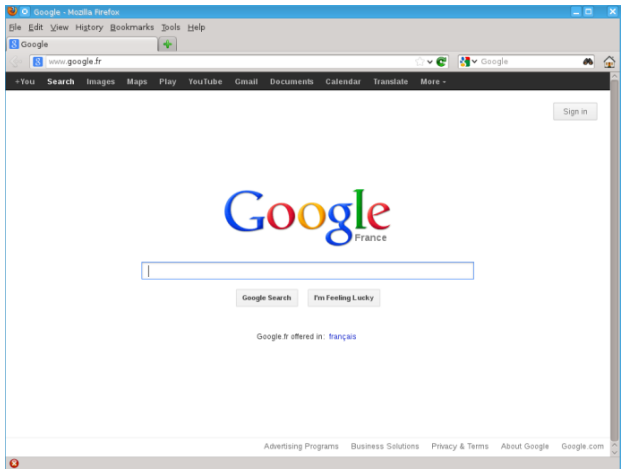
- Web browsing performance is important:
  - for end-user
    - ***10 seconds*** page load time is the limit for keeping the user's attention on one web page [1].
  - for service vendor
    - From [2]: for Microsoft's Bing, *2 second* delay load time:
      - 4% reduction in clicks
      - 4% loss in revenue
      - …
- ISP hotline:
  - People complain about poor performance **to their ISP .**

[1] ITU-T Recommendation G.1030 (2005), Estimating end-to-end performance in IP networks for data applications.
[2] Gomez White Paper. Why Web Performance Matters: Is Your Site Driving Customers Away?

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.

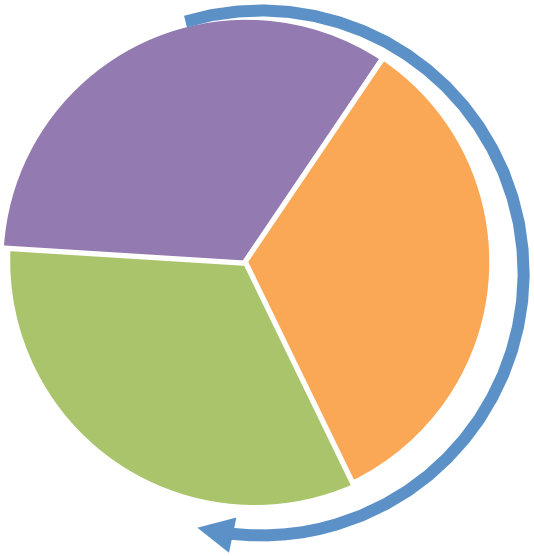
- From the end-user perspective → **Quality of Experience (QoE)**

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

Quality of Experience

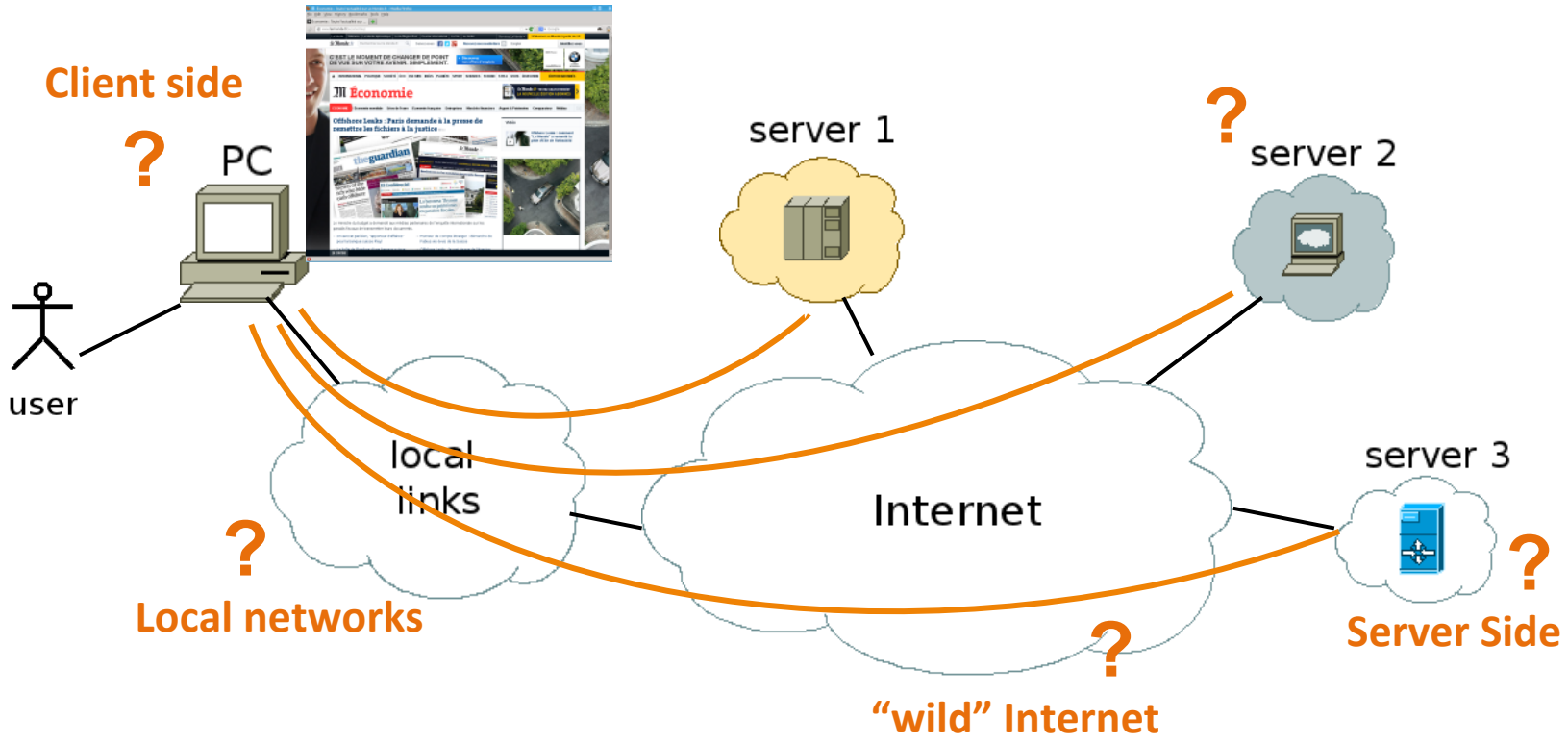
- When we say **QoE** for web browsing**:**
  - we **use**:
    - Performance metric that can be measured in an objective manner. (**Page Load Time**)
    - Subjective feedback as a complement (through a button).
  - we do **not** focus on:
    - Designing subject metrics to measure QoE
    - Mean Opinion Score (MOS)



Troubleshooting Web Browsing

EURECOM
Sophia Antipolis

End-to-End Path for Web Browsing

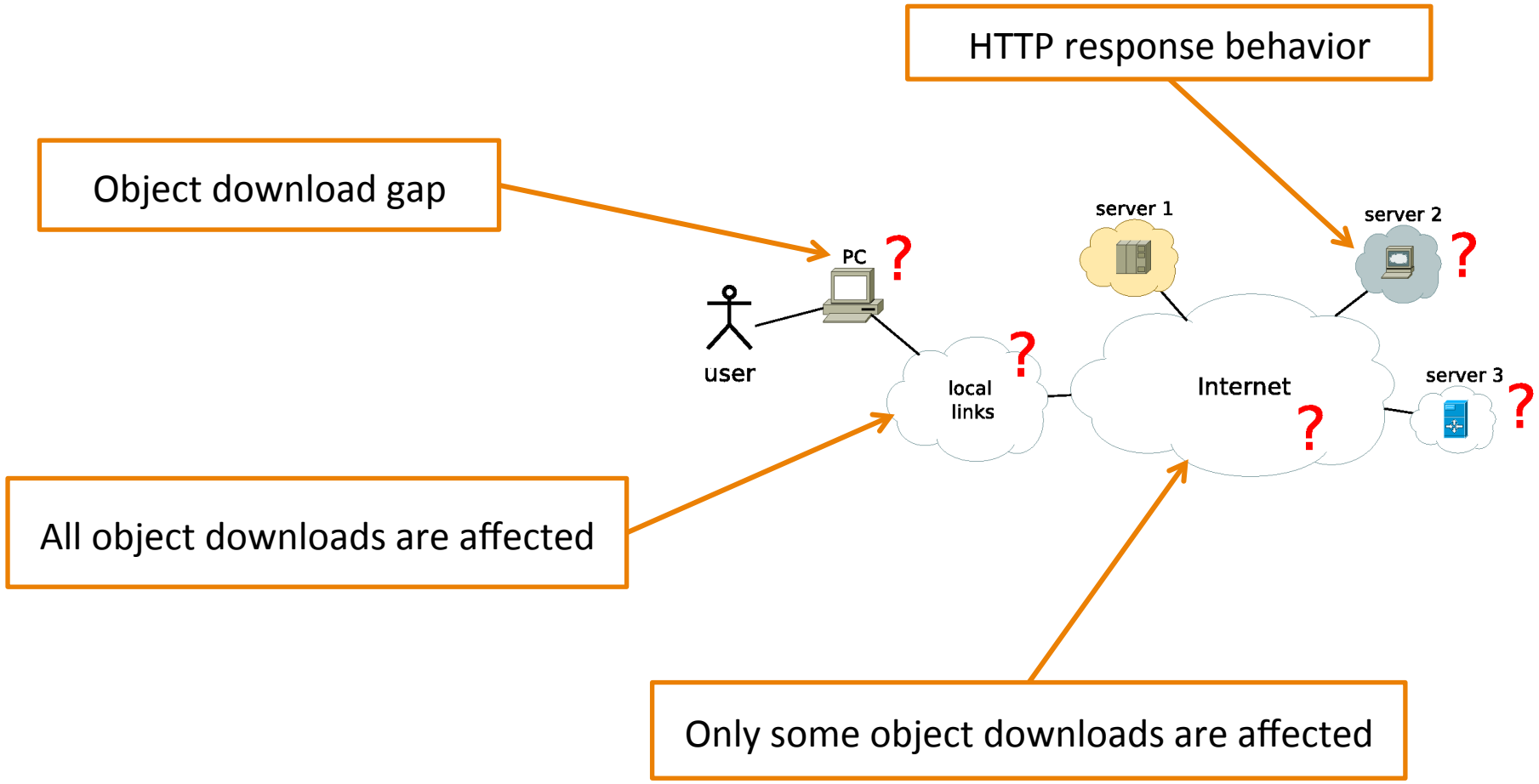


A slow-down of any component along the path does matter.

Q: Which component bears the major responsibility?

Intuitions

- How to discover different limitations:

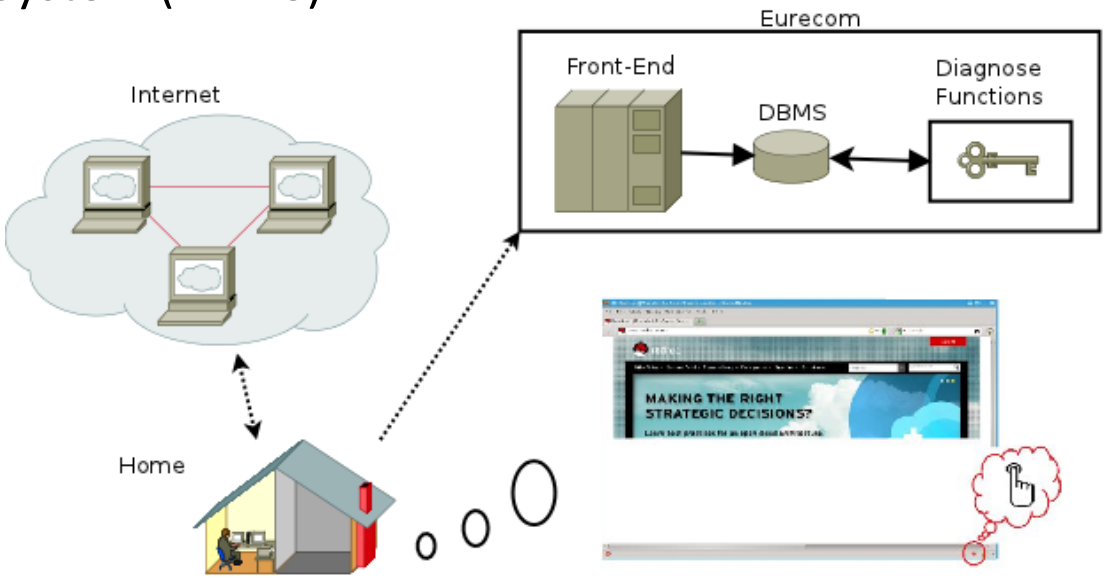


“FireLog” Architecture

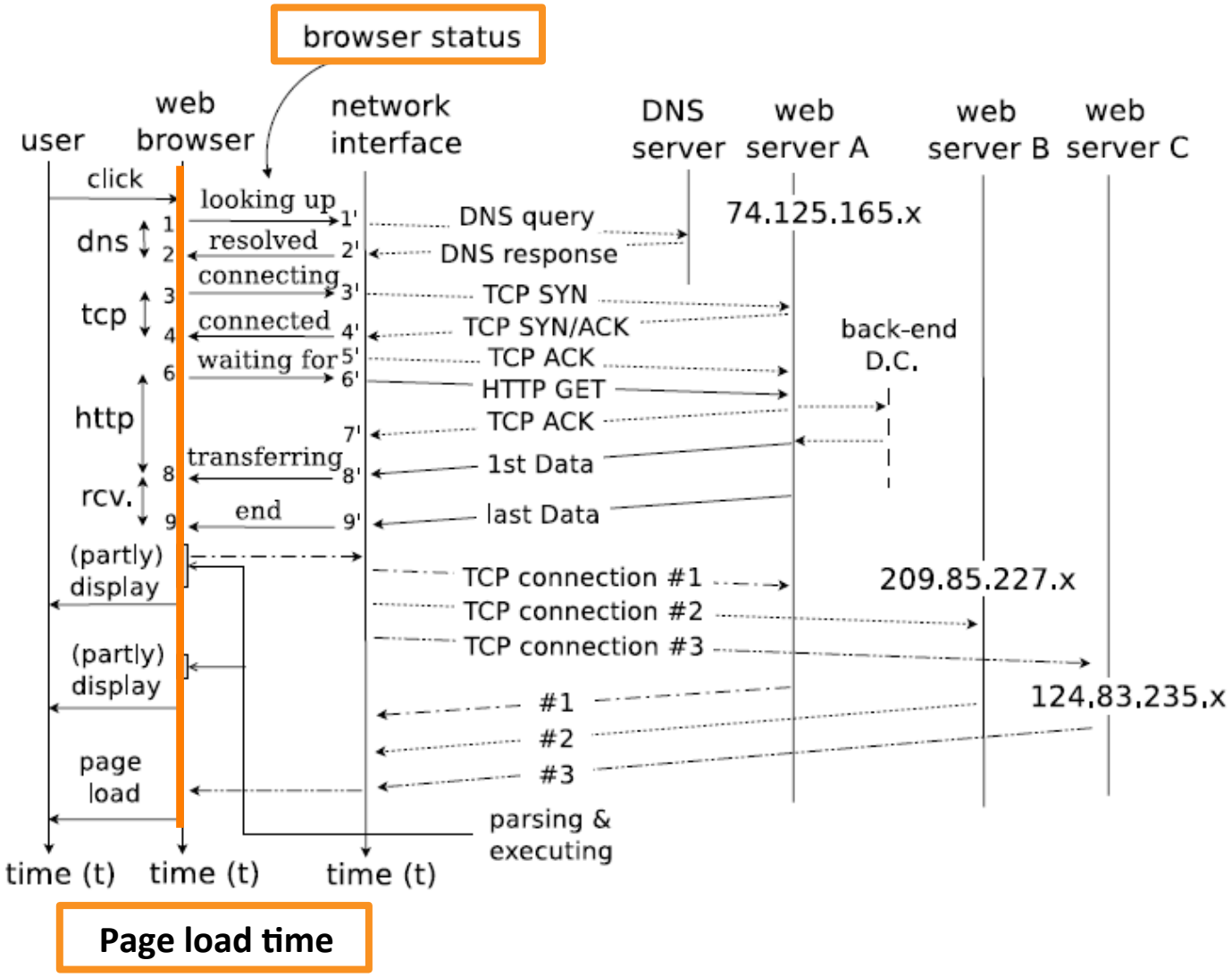
Develop “FireLog”:

- Plug-in at Web browser (e.g. Firefox)
  - No low level packet capture
  - All measurement **<u>at one place</u>**

- Server Repository for post-processing
  - Database Management System (DBMS)
  - Diagnosis engine



Client: Metrics Measured at Browser



Server: Collected Raw Data

Collect events per object → DB

firelog :

File Edit View Bookmarks Settings Help

Column	Type
client_ip_by_server	inet
method	character varying
content_load_ts	timestamp without time zone
load_ts	timestamp without time zone
session_start	timestamp without time zone
session_url	character varying
if_complete_cache	integer
localaddress	inet
localport	integer
remoteaddress	inet
remoteport	integer
response_code	integer
http_request_bytes	integer
http_header_bytes	integer
http_body_bytes	integer
http_cache_bytes	integer
dns_start	timestamp without time zone
dns_end	timestamp without time zone
dns_time	bigint
syn_start	timestamp without time zone
syn_end	timestamp without time zone
tcp_cnxting	bigint
send_ts	timestamp without time zone
send	bigint

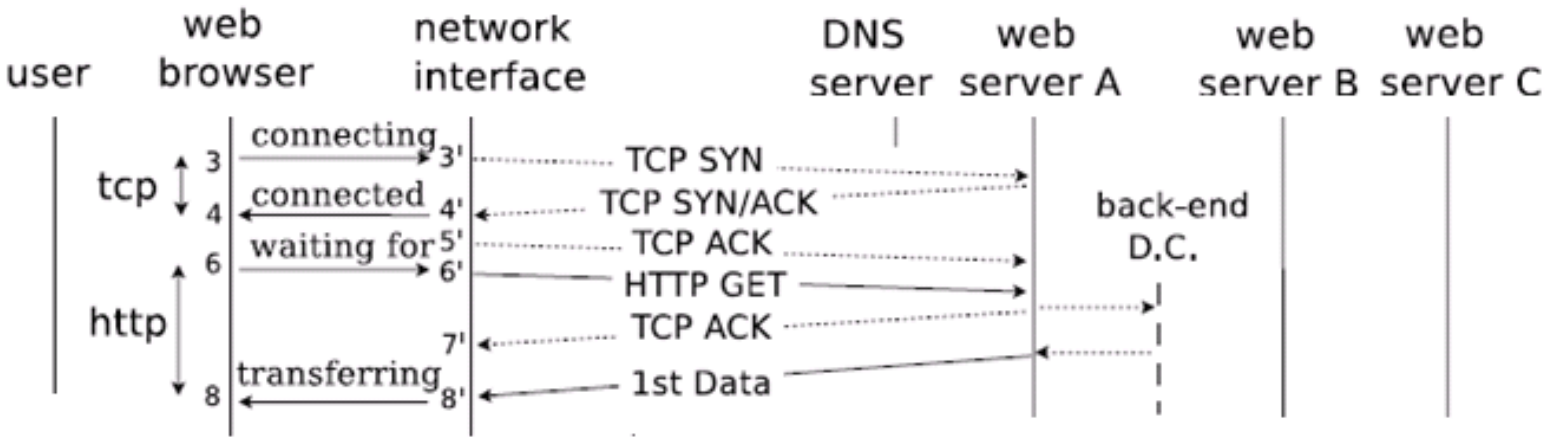
Server: Transform Metrics to Limitation Scores

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.

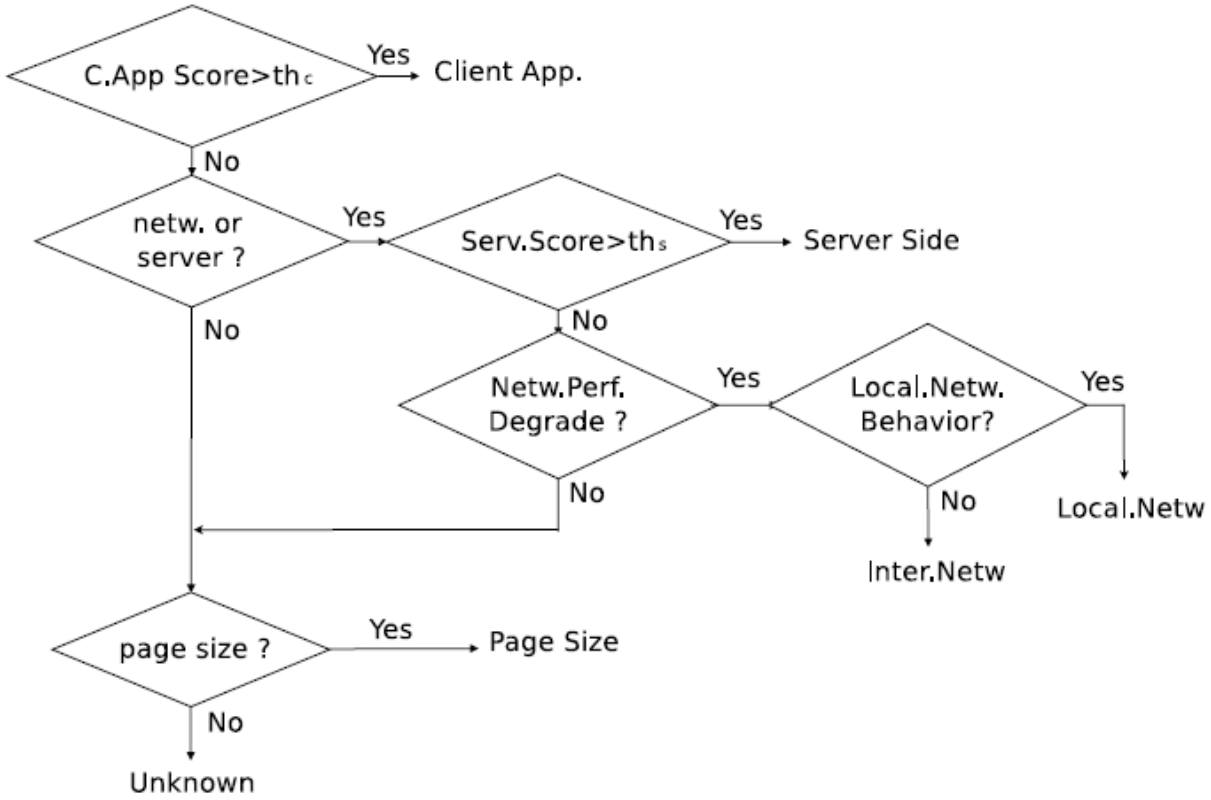


Server: Diagnosis Engine

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.

PLT > 10 sec. [1]

		Totally Browsed			Web Pages with "High Load Time"		
user	duration	*#page*	*#domain*	*#object*	*#page*	*#domain*	*#object*
$A(FR)$	5 month	3,451	579	501,102	808	247	142,939
$B(FR)$	3 month	1,788	263	87,898	281	114	24,406
$C(CN)$	2 month	3,766	535	317,700	466	183	63,619

Limitation Causes for "High Load Time" Pages

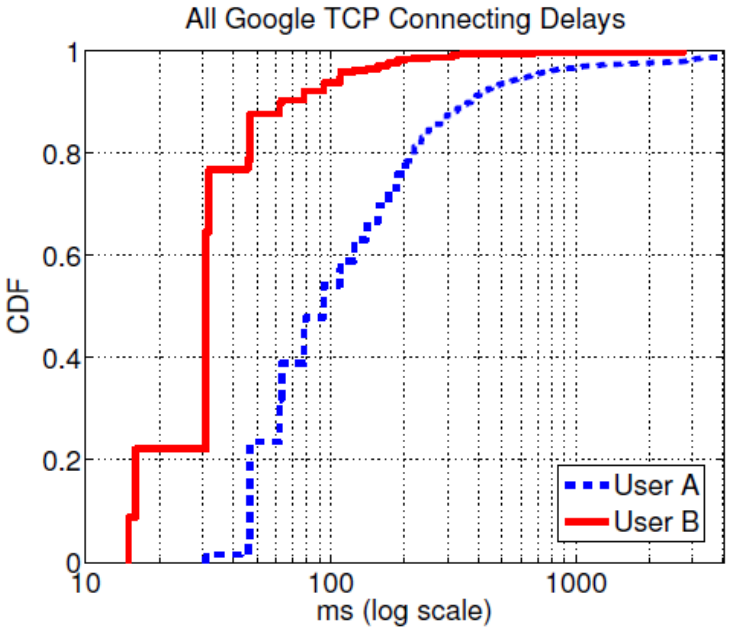
User	Main cause				
	Client	Server	Local access	Internet	others
A	**21%**	4%	**29%**	**32%**	14%
B	**28%**	**39%**	9%	10%	14%
C	**21%**	**44%**	9%	6%	20%

[1] ITU-T Recommendation G.1030 (2005), Estimating end-to-end performance in IP networks for data applications

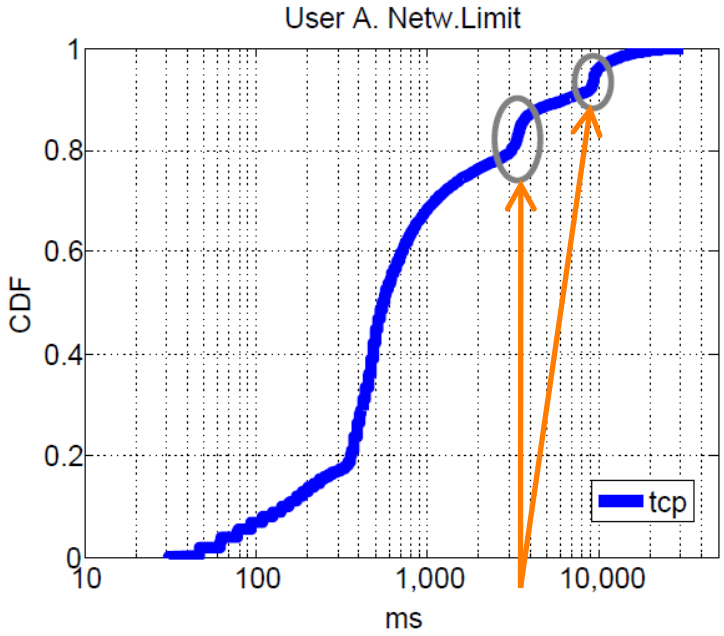
Limitation by Networks

User	Main cause				
	Client	Server	Local access	Internet	others
A	**21%**	4%	**29%**	**32%**	14%
B	**28%**	**39%**	9%	10%	14%
C	**21%**	**44%**	9%	6%	20%

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



Network is the problem (e.g. loss)



3 or 9 seconds. Typical RTO.

Conclusion

We present a diagnosis system, FireLog:
- client side measurement
- server side repository

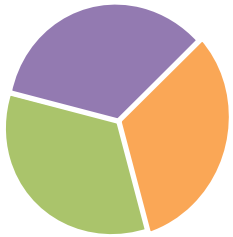
Diagnosis model for troubleshooting Web browsing.
- carefully design the model
- lab experiments for thresholds

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 enclients
    - 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?

EURECOM
Sophia Antipolis