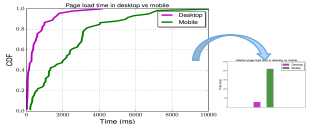
Significantly improving Mobile Web Performance

Javad Nejati, Yi Cao, Conor Kelton, Aruna Balasubramanian (In collaboration with WINGS, PACE, and Computer Vision Lab)

Stony Brook University

Project #1: Web Performance Bottlenecks

- **Problem**: Mobile browsing is an order-of-magnitude slower compare to desktop browsing



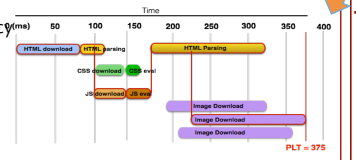
- **Research Question:** Are the bottlenecks in mobile browsing similar to desktop browsing? Can the same kinds of optimizations be applied?

WProf-M: Understanding Mobile browser bottlenecks

- The page load process involves both computation and network activities that are inter-dependent.
- The relations between these activities form a dependency graph.

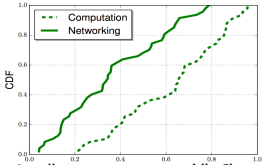
WProf-M: Uncovers the dependency graph and bottlenecks in mobile browsers.

The bottleneck (or critical path) is shown in red in the figure



Key Result: Computation is the bottleneck in mobile browsers. On desktops, network is the bottleneck

Figure shows the CDF of the fraction of compute and network activities on the critical path



Loading 100 pages on mobile Chrome over a 4G connection

WProf-M published at WWW 2016

Ongoing work

- Design WProf-X, a visualization tool that will allow Web page developers to easily detect bottlenecks and perform what-if analysis
- Study the page load bottlenecks when loading a page in developing regions

Project Webpage: wprofx.cs.stonybrook.edu

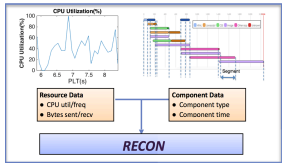
Project #2: Modeling

- **Problem**: Energy consumption of mobile Web performance is critical, but not well studied
- Optimizations that improve page load time (PLT) may effect energy differently because PLT depends on the critical path, while energy depends on all page load activities

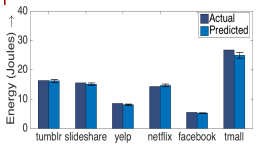
RECON: Modeling Energy Consumption of Web Pages

Idea: App Semantics + Resource Monitoring

- Extract low-level page load semantics from WProf-M
- Keep track of coarse-grained resource consumption
- Combine the two for accurate energy modeling



Evaluation: We used RECON to predict energy consumption for 80 Web pages. Mean prediction error less than 10%.



Explanatory power: RECON can explain why energy consumption changes

Example: After inlining, energy increases but PLT decreases. Why?

Because HTML parsing is power hungry, and inlining increases parsing energy

netflix.com before inlining *netflix.com* after inlining

RECON published at Sigmetrics 2017

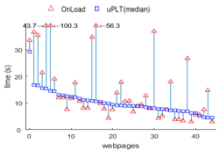
TCP Throughput Modeling (Ongoing)

- TCP Throughput model unchanged since 1998
- Existing models make assumptions (for eg., that the congestion window and loss rate are independent) that is no longer true
- Our goal is to design a new TCP throughput model and use it to predict the throughput of HTTP/2 versus HTTP

Project #3: Improving Web QoE

Question: Are traditional Page Load Time (PLT) metrics such as OnLoad measuring user Quality of Experience (QoE)

Our Approach: Define user-perceived PLT (uPLT) as the time when users perceive the page to be loaded



User study across 50 users and 45 Web pages show that uPLT and OnLoad poorly correlate

WebGaze: Improving QoE using Gaze

Insight: Loading objects on the page that are "interesting" to the user first can improve user QoE.

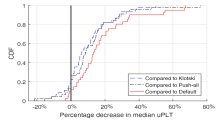
WebGaze idea: Identify ``interesting'' objects on the Web page by leveraging gaze information across large number of users.

Gaze user study: 50 users, 45 Web pages, using commodity tracker

Key Takeway: Gaze track across users form a pattern. Regions of high collective fixation (fraction of users who fixate on a region) and low collective fixation exist

WEbGaze: Server pushes objects with high collective fixation to load first

Evaluation: User study with 300 users show QoE improvements



WebGaze published at NSDI 2017

Ongoing work

- Define and measure uPLT for mobile browsing
- Design a new metric that can be measured systematically (without user studies) that correlates well with uPLT

Project Webpage: gaze.cs.stonybrook.edu

We gratefully acknowledge funding from the National Science Foundation and Google