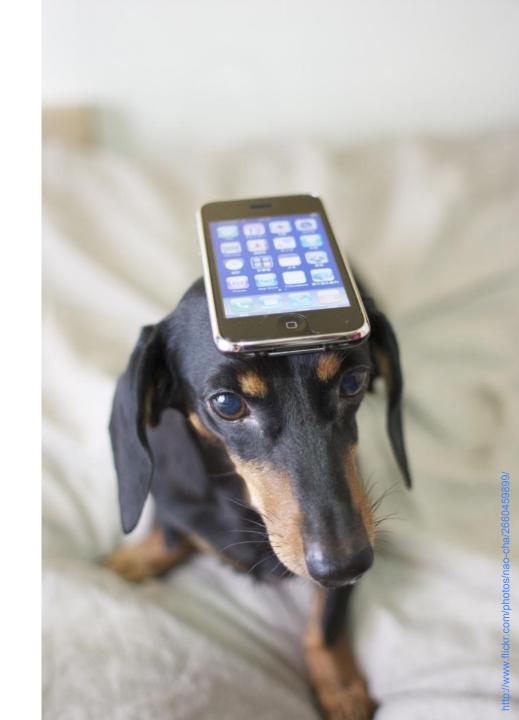


# Why Mobile Performance is Hard

Matt Welsh mdw@google.com

Google, Inc.



#### In a nutshell:

Despite 20 years of research and engineering, mobile performance still sucks

Networking advancements (e.g., LTE) help but do not solve the problem

Need to start thinking end-to-end, not just about latency and bandwidth

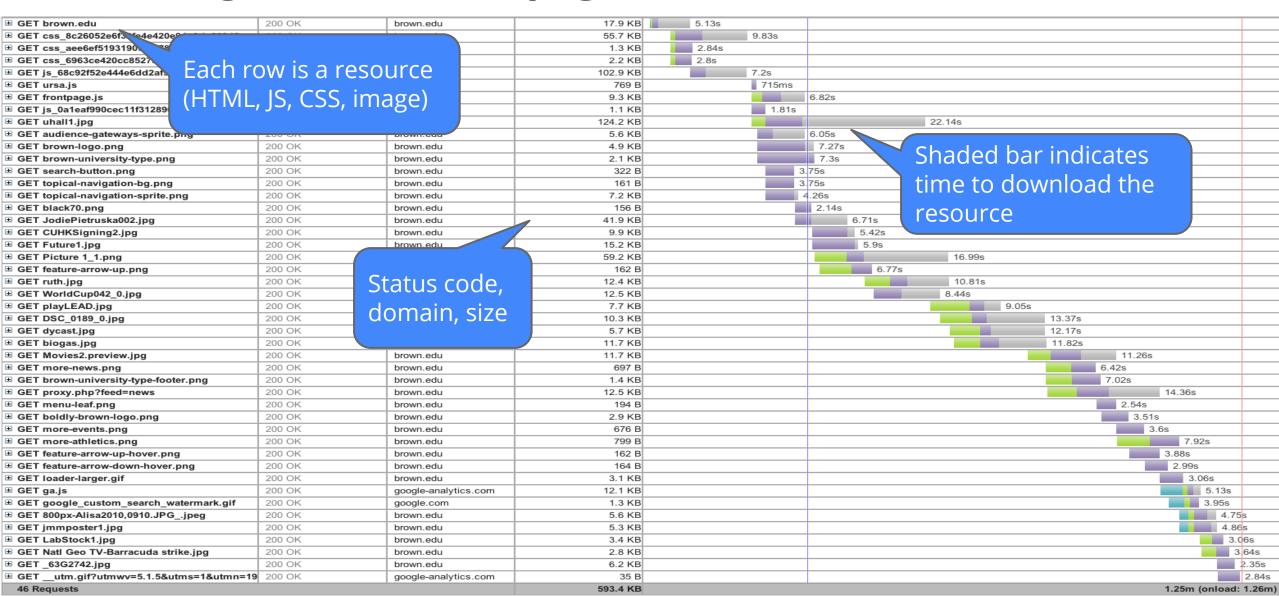
A tour through some of the big, hard problems we still face





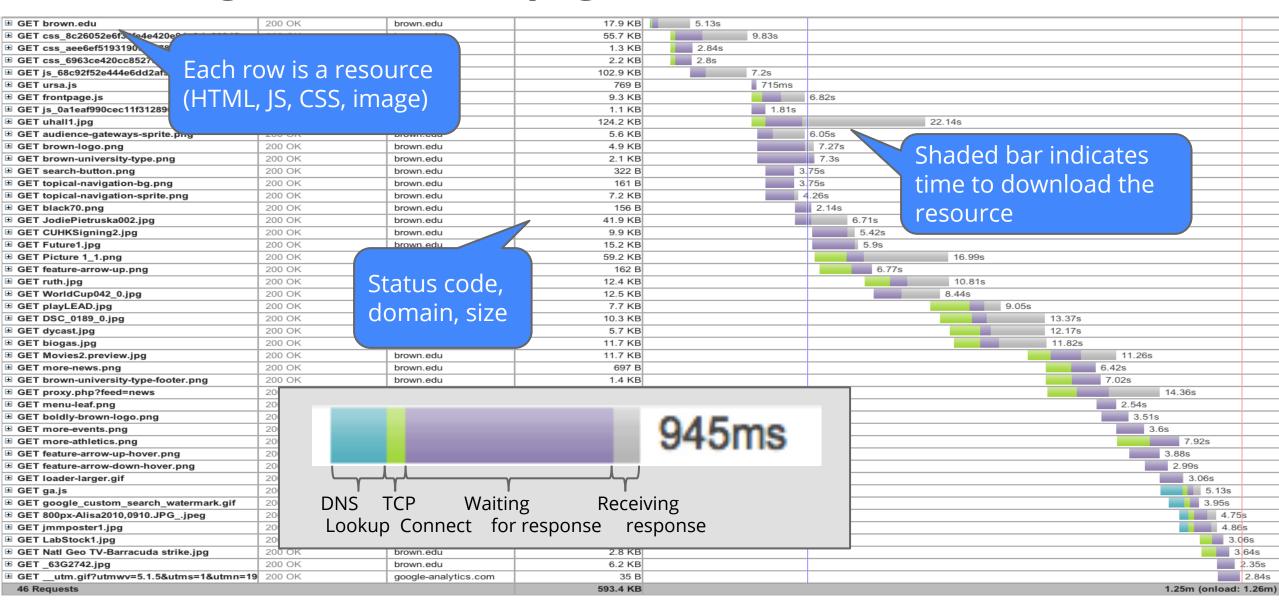
# Case study: Mobile web page load times

## Visualizing a mobile web page load (brown.edu)



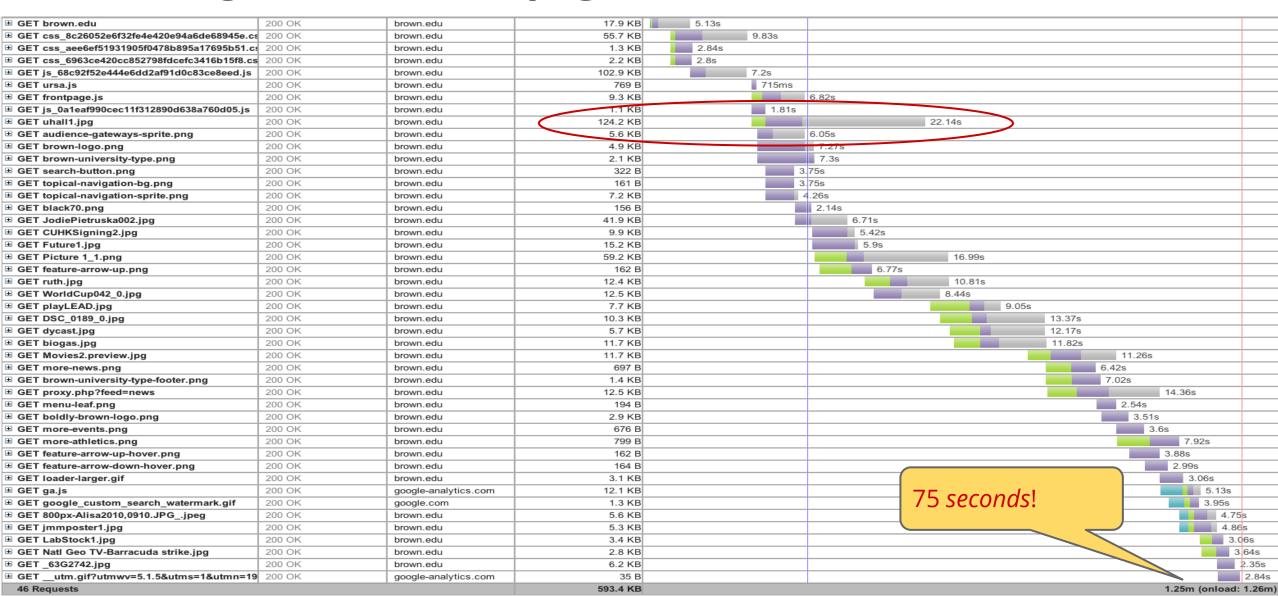


## Visualizing a mobile web page load (brown.edu)





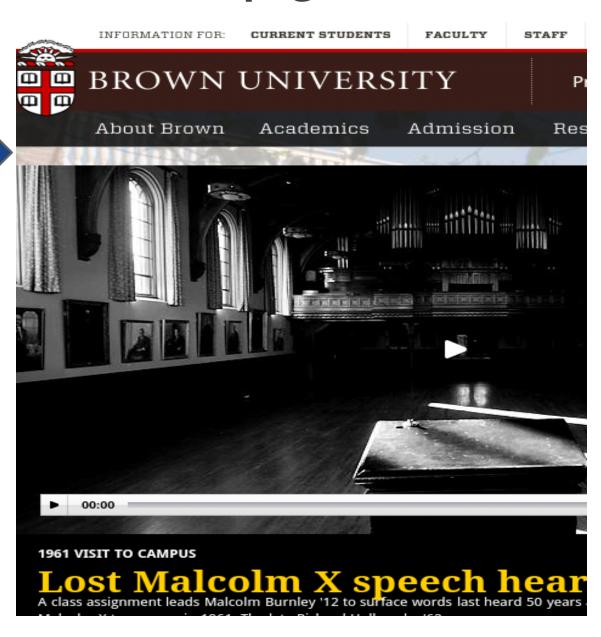
# Visualizing a mobile web page load (brown.edu)





#### brown.edu's mobile home page

124 KB, 1800 x 800 background image that is completely obscured



#### The web was not designed for mobile:

High RTTs

Mismatch between HTTP, TCP, and 3G/4G technologies

Poorly tuned browsers running on slow CPUs with limited memory

Badly designed pages that fail to account for limited device capabilities

Lack of good tools to measure mobile web performance





# You can't fix what you can't measure

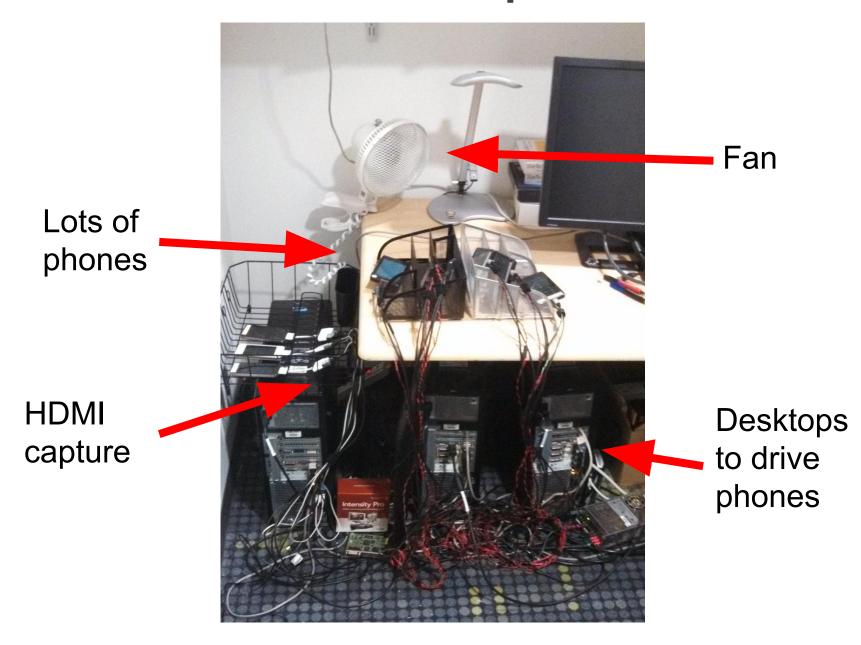








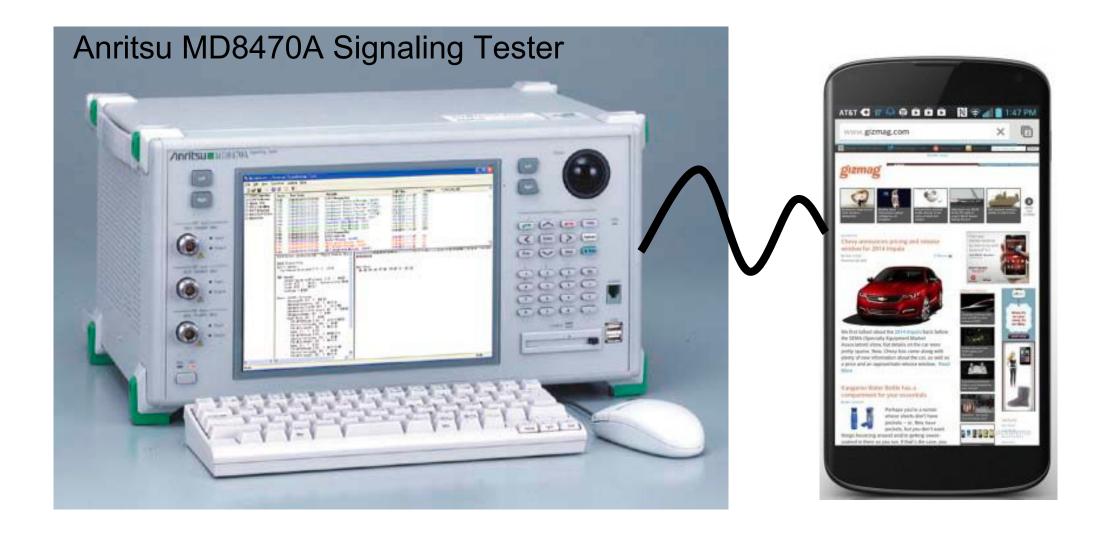




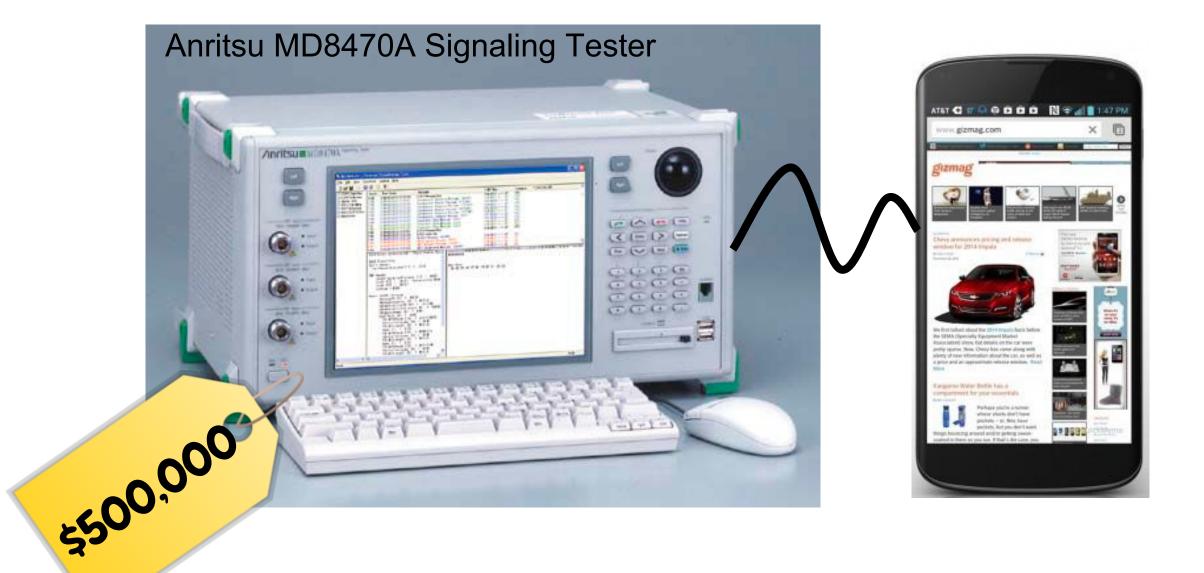


















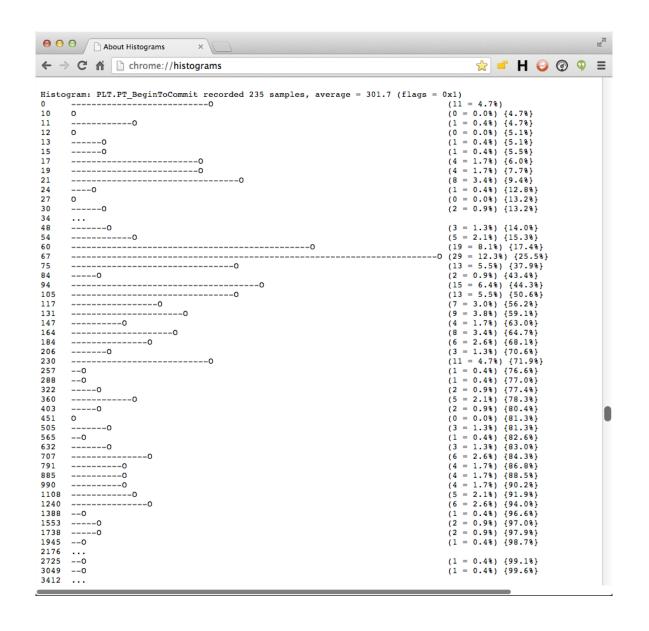
# Mobile measurement at scale: Chrome User Metrics

#### **Chrome User Metrics**

Chrome collects anonymous data on hundreds of performance metrics from users who opt in:

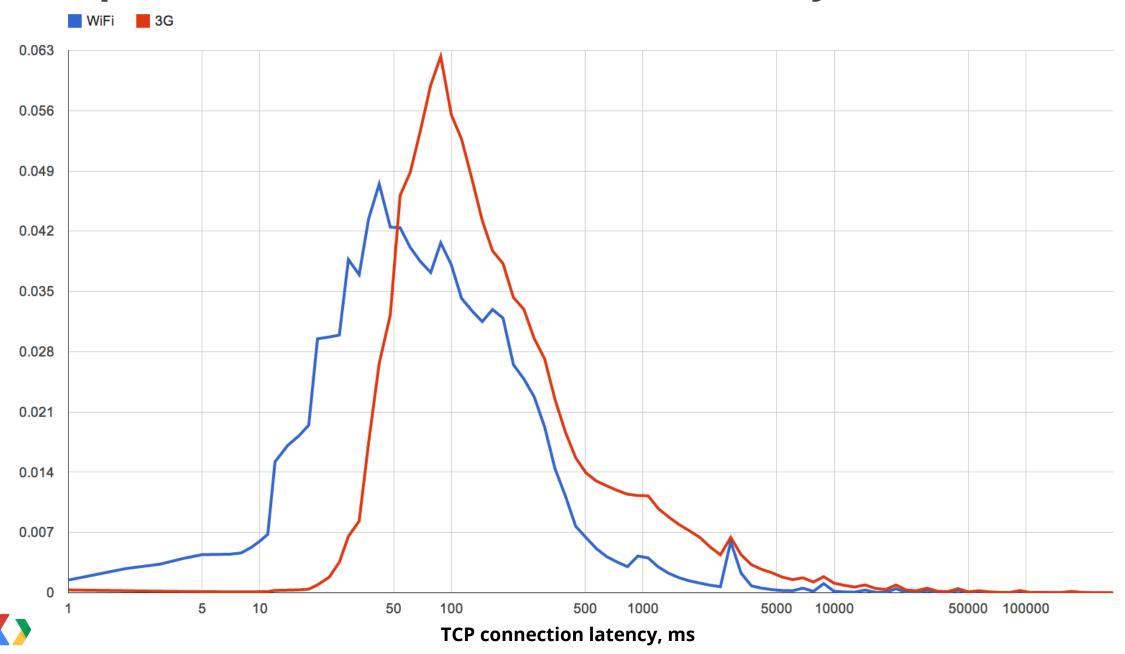
 page load times, DNS lookups, TCP connections, cache and graphics performance, SSL, you name it...

check out chrome://histograms to see your own data

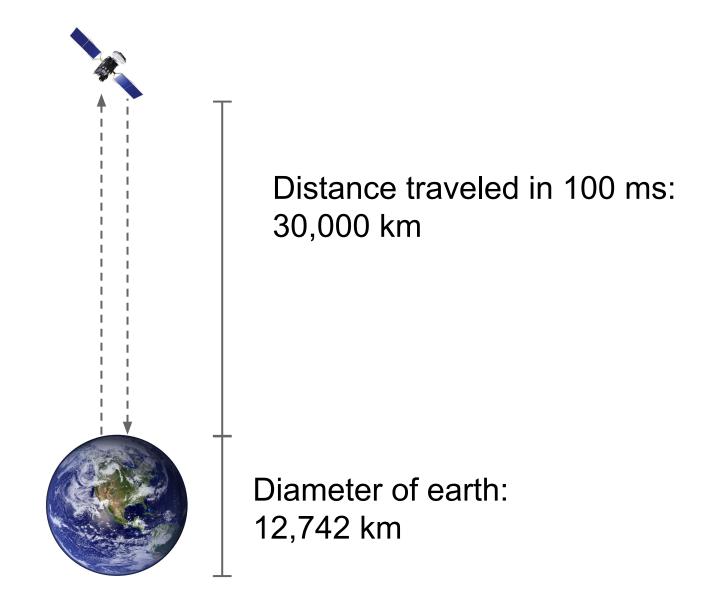




#### Global performance trends, as measured by Chrome

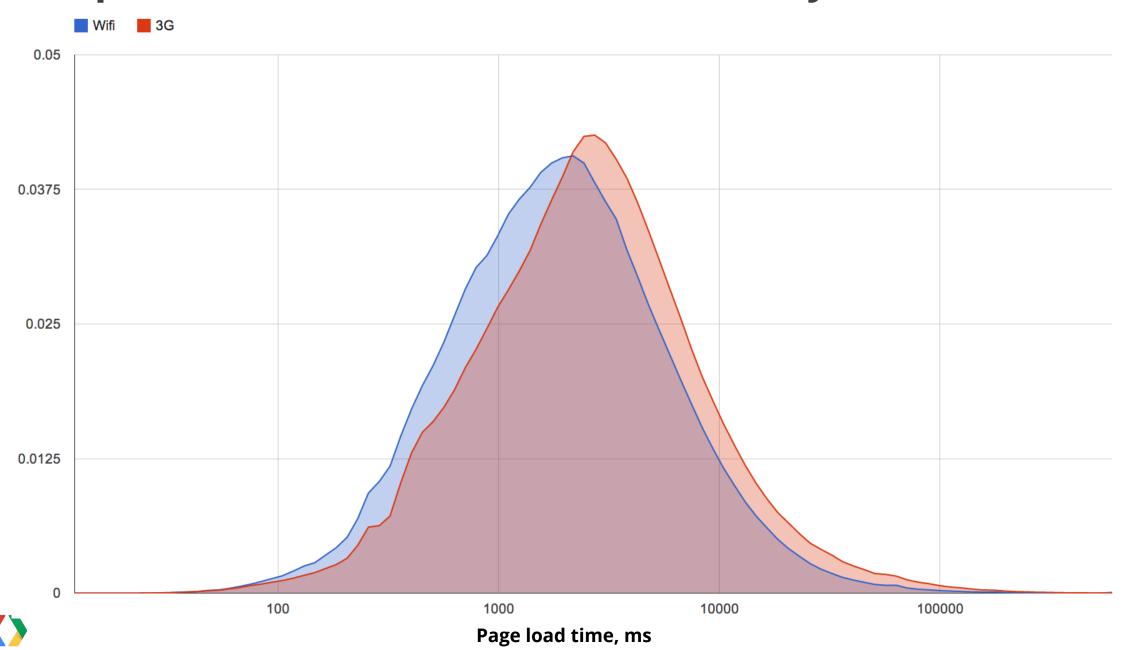


#### What does 200 ms RTT look like?

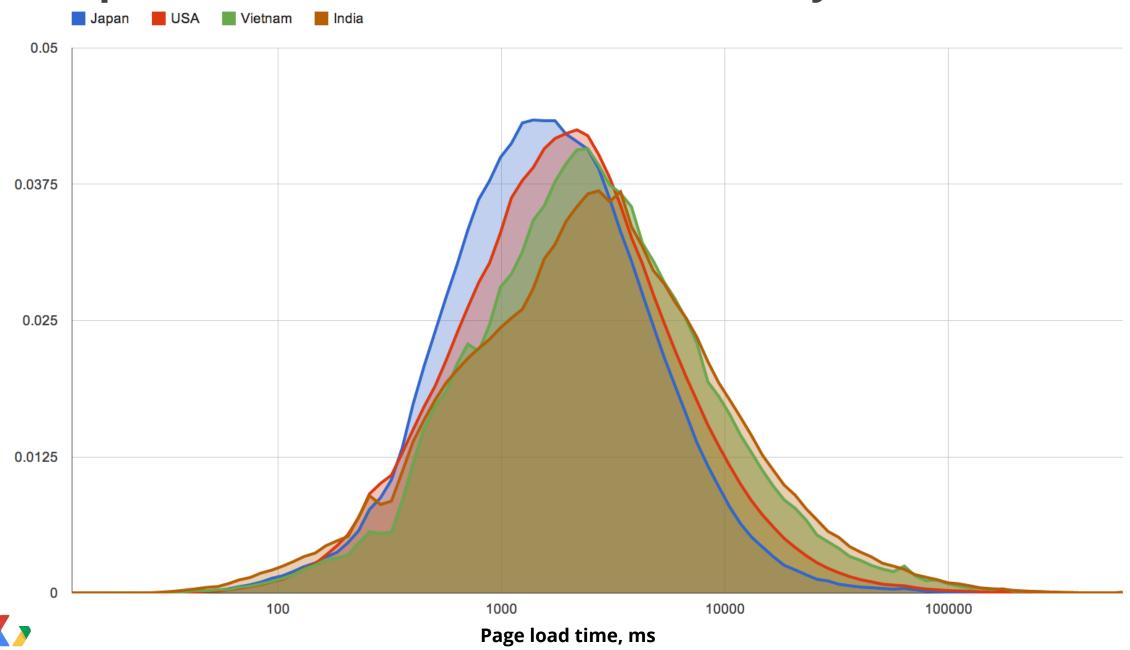




# Global performance trends, as measured by Chrome



#### Global performance trends, as measured by Chrome





# Delivering optimizations at scale

#### **SPDY - a replacement for HTTP**

Already built into Chrome, Firefox, Opera, Kindle Fire. You probably use it every day.

#### **Key features:**

Use a single TCP connection and multiplex multiple requests/responses

Out-of-order responses: no more head-of-line blocking

Header compression: Eliminates ~85% of headers

Server-initiated resource push

Baseline for the HTTP/2.0 standard being developed by IETF.



#### **QUIC - Quick UDP Internet Connections**

New transport protocol based on **UDP** 

Multiple sessions multiplexed over a single UDP "stream"

Packet pacing and FEC to reduce data loss

Customizable congestion control

Zero-RTT connection establishment (send hello + request back to back)

More: http://blog.chromium.org/2013/06/experimenting-with-quic.html



#### Chrome network stack optimizations

#### DNS preresolution:

Keep track of top 10 domains you visit, DNS resolve them at startup

#### TCP preconnect:

Pre-resolve and pre-connect to most common domains on a page

#### Omnibox preloading:

Resolve DNS and TCP connections as you type into the omnibox

#### Resource prefetching:

Use k rel="prefetch"> and k rel="subresource"> hints in pages

#### Page prerendering:

Render entire page in the background as you type in the omnibox

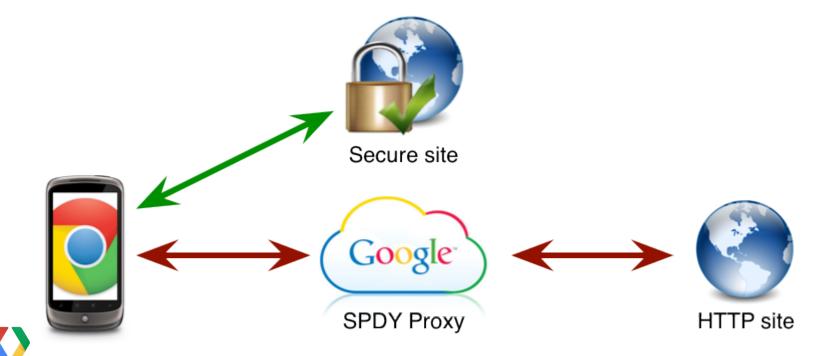


#### **Chrome Mobile data compression**

Goal: Optimize page load times, while saving users money

Can't fix every website on the planet, so provide a free proxy service to optimize sites instead.

Proxy only HTTP (not HTTPS or Incognito tabs)





#### **Optimizations**

Compress images to WebP format

Reduce image resolution and quality (slightly)

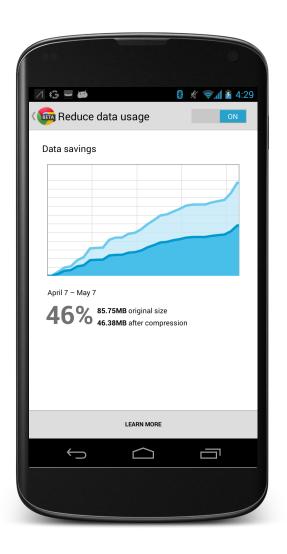
Minify JavaScript and CSS

Gzip everything

Deliver all page content via SPDY

Cache web content in Google's datacenters

These techniques result in 50% compression on average





### Some of the challenges we face...

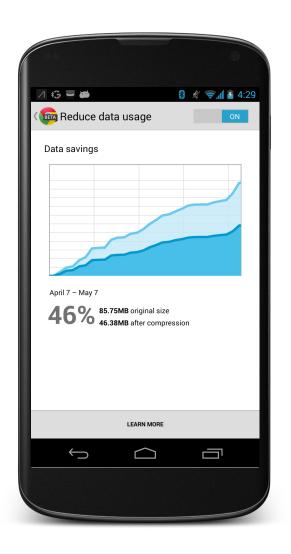
Proxying the whole web is not easy. (Or cheap.)

Not all websites do proper geolocation using the X-Forwarded-For header (French news site redirects you to English version)

Stripping JS comments broke rendering on a site (Using JS comments as a fingerprint of site content)

Using SPDY converts headers to lowercase (One site relied on case-sensitive header match to work)

Yelp initially (erroneously) blocked our traffic (Looks like clickspam from Google)







# What the mobile research community should work on

#### Understanding how apps interact with the network

Why does a YouTube playback stall over a fast LTE connection?

Why does doing a Google search take so long on cellular?

How does Ajax calls on a web page interact with 3G RRC behavior?

How do handoffs impact app performance?

Not enough to just look at the network: Need to look cross layer.



## Why doesn't my phone last all day?

Lots of good work on energy management.

Some of this work assumes you can modify every app -- ain't gonna happen.

Tons of inefficiency between the apps, OS, and network stack.

Screen and network are the main sources of energy drain. (Not CPU.)

Solutions have to work in the real world: Not just in the lab. (Not OK to black out half the screen while watching a video...)



#### Where is my data and who has access to it?

Cell phones are basically tracking devices.

e.g., Google aggregates anonymous GPS data from many users to build real-time traffic maps.

But many apps are more nefarious: Stealing personal information, phone numbers, contacts, photos, location, ...

Mobile researchers: Please stop spying on users for the purpose of doing research studies.

(Read up on human subjects and IRB requirements at your university.)





#### What should the web look like 10 years from now?

#### The web is primarily a mobile platform.

But web pages are primarily designed for desktop users.

Not enough overlap between the web dev and mobile research communities.

Web devs get excited about things like using sockets from JavaScript.

Given the tremendous capabilities of mobile devices, what should the web really look like, if you designed it for 10 years from now?

Are we still going to be loading web "pages" in "tabs" and using "bookmarks" to access "URLs"?



# Thank You!



