

Measuring the Web: Part II

Website Complexity

Prof. Georgios Smaragdakis, Ph.D.

The Web is the Media to reach Billions of People



HTTP as the Narrow Waist of the Future Internet

Lucian Popa
U.C. Berkeley / ICSI

Ali Ghodsi
U.C. Berkeley

Ion Stoica
U.C. Berkeley

ABSTRACT

Over the past decade a variety of network architectures have been proposed to address IP's limitations in terms of flexible forwarding, security, and data distribution. Meanwhile, fueled by the explosive growth of video traffic and HTTP infrastructure (*e.g.*, CDNs, web caches), HTTP has become the

In this paper, we take this trend to its logical conclusion and consider the scenario where HTTP becomes the de facto "narrow waist" of the Internet—that is, the vast majority of traffic runs over HTTP instead of directly over IP, and HTTP itself might run on top of network layers other than IP.

Given such scenario, we argue that we should start eval-

Welcome to Your Timeline — Preview

Right now, only you can see your timeline. This gives you a chance to:

1. Review what's on your timeline now, and add or hide whatever you want.
2. Click Publish Now or wait until **your timeline goes live on March 16**. Learn more.

Start Tour **Publish Now**

Now
February
2012
2011
2010
2009
2008
2007
2003
2002
Born

Embedded objects

Georgios Smaragdakis

Update Info Activity Log

Works at T-Labs
Studied Computer Science at Boston University
Lives in Berlin, Germany

Friends 541 Photos 210 Map 12 Likes 67

Status Photo Place Life Event

What's on your mind?

Georgios Smaragdakis Yesterday

Friends See All

Vijay Erramilli 91 mutual friends
Nagia Chanaki 7 mutual friends
Nikolaos Laoutaris 71 mutual friends
Dimitris Skoulaxinos 25 mutual friends

Advertisements
>120 Billion Market

sponsored

% Aktion: iPhone GRATIS!
handyservice.de

Deutschlands größtes Mobilfunkversandhaus. Top-Händler, günstige Tarife und Gratis Prämien oder Abzahlungen sicher und sicher!

KATIE MELUA - Berlin
ticketmaster.de

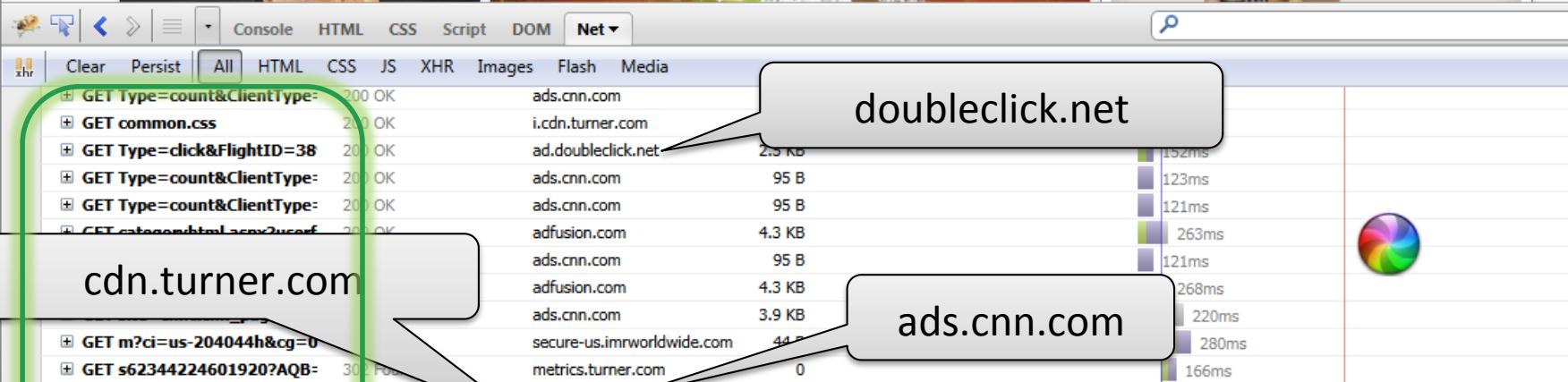
Am 9. Juli live auf dem Classic Open Air Gendarmenmarkt. Katie Melua auf Secret Symphony Tour 2012. www.ticketmaster.de

Dein Auslandsjahr
schueleraustausch.net

Hol dir jetzt kostenlos die Kataloge aller Organisationen und vergleiche die Angebote mit unserem Vergleichstool



updated 9:02 p.m. EDT, Sat June 11, 2011

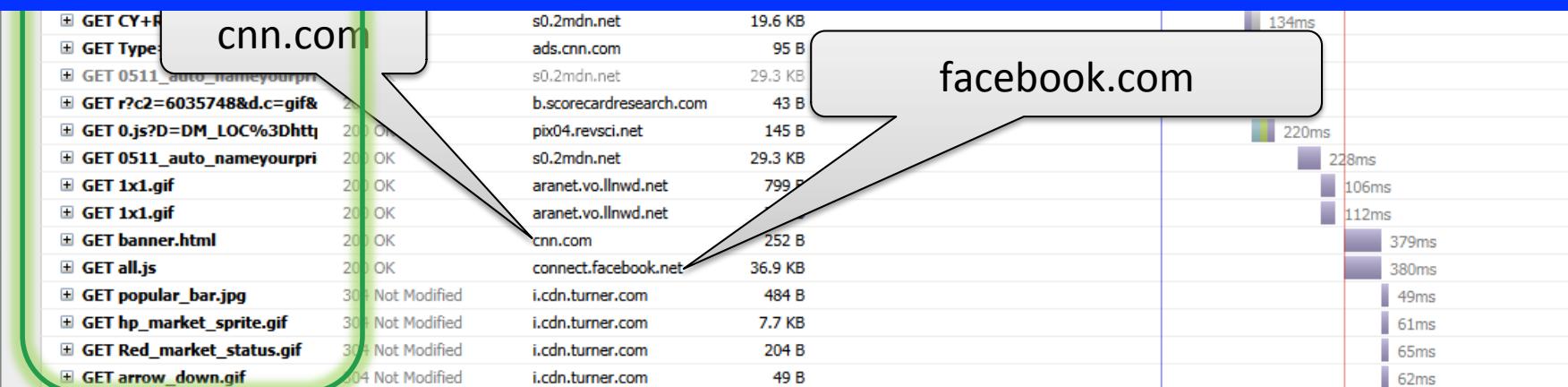


cdn.turner.com

doubleclick.net

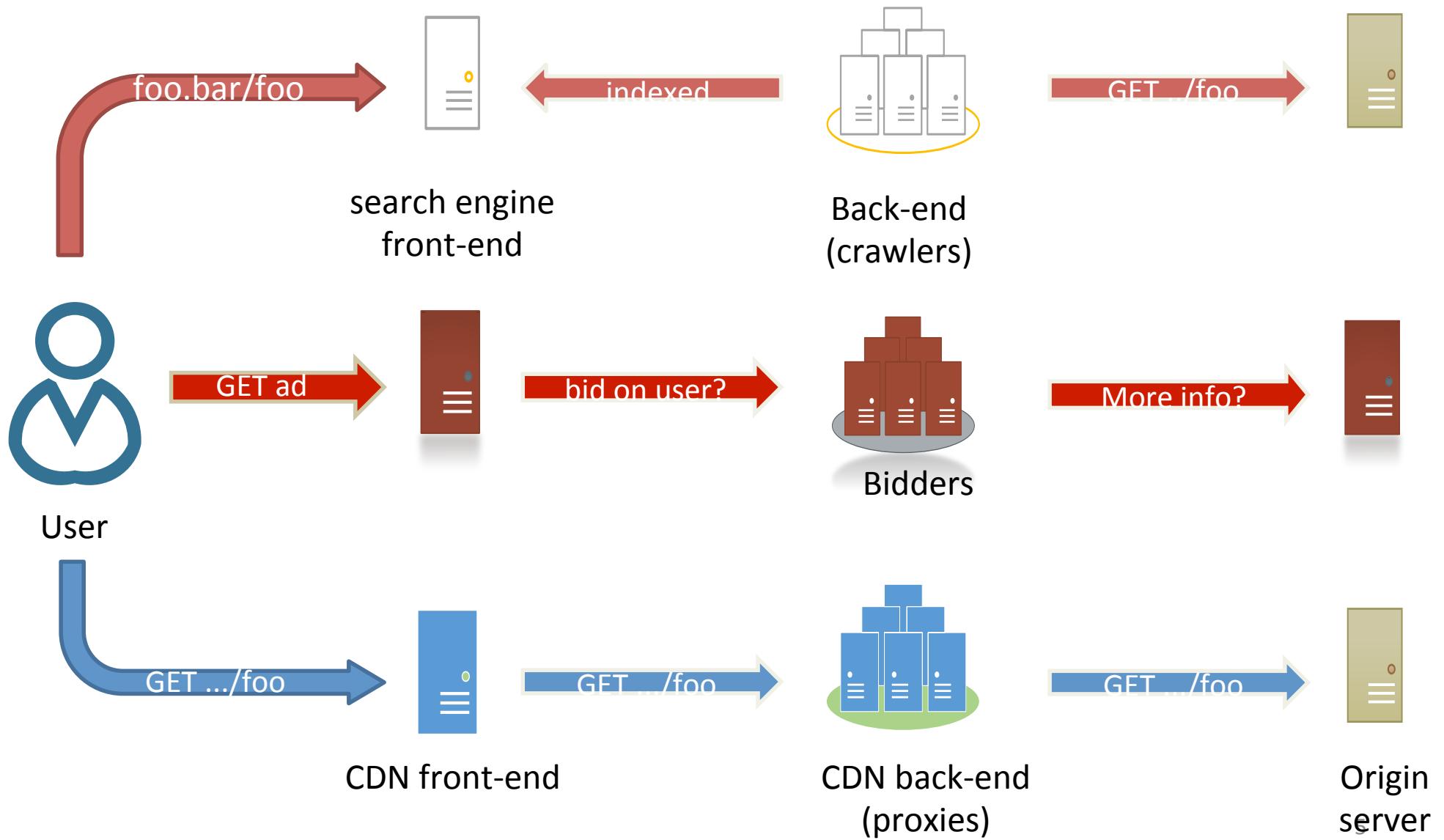
ads.cnn.com

Websites today are very complex!
Diverse content from many servers and third party services

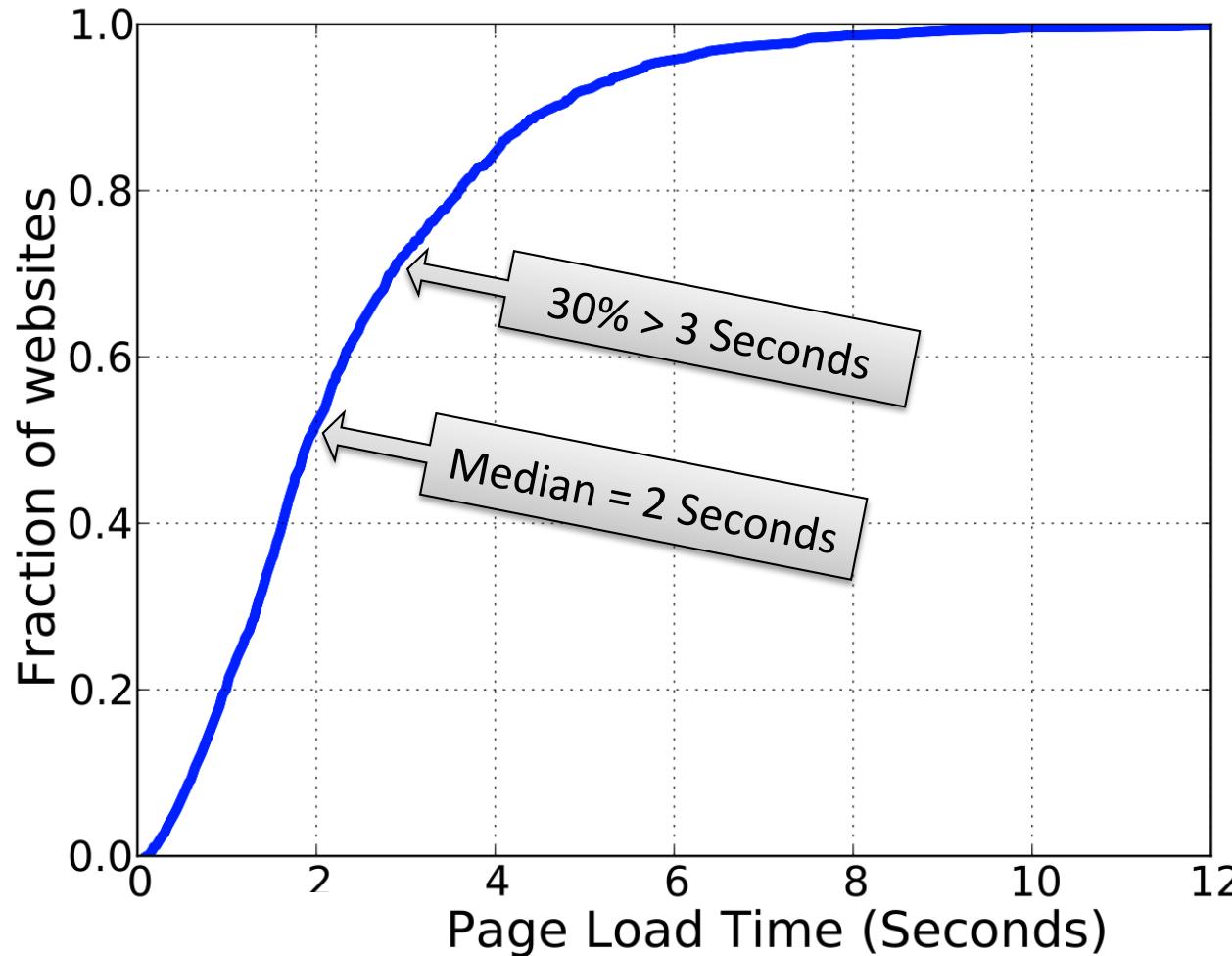


facebook.com

Behind the Curtain



Users see slow loading websites!



How long a person will wait for page to load before navigating away

8 seconds



5 seconds



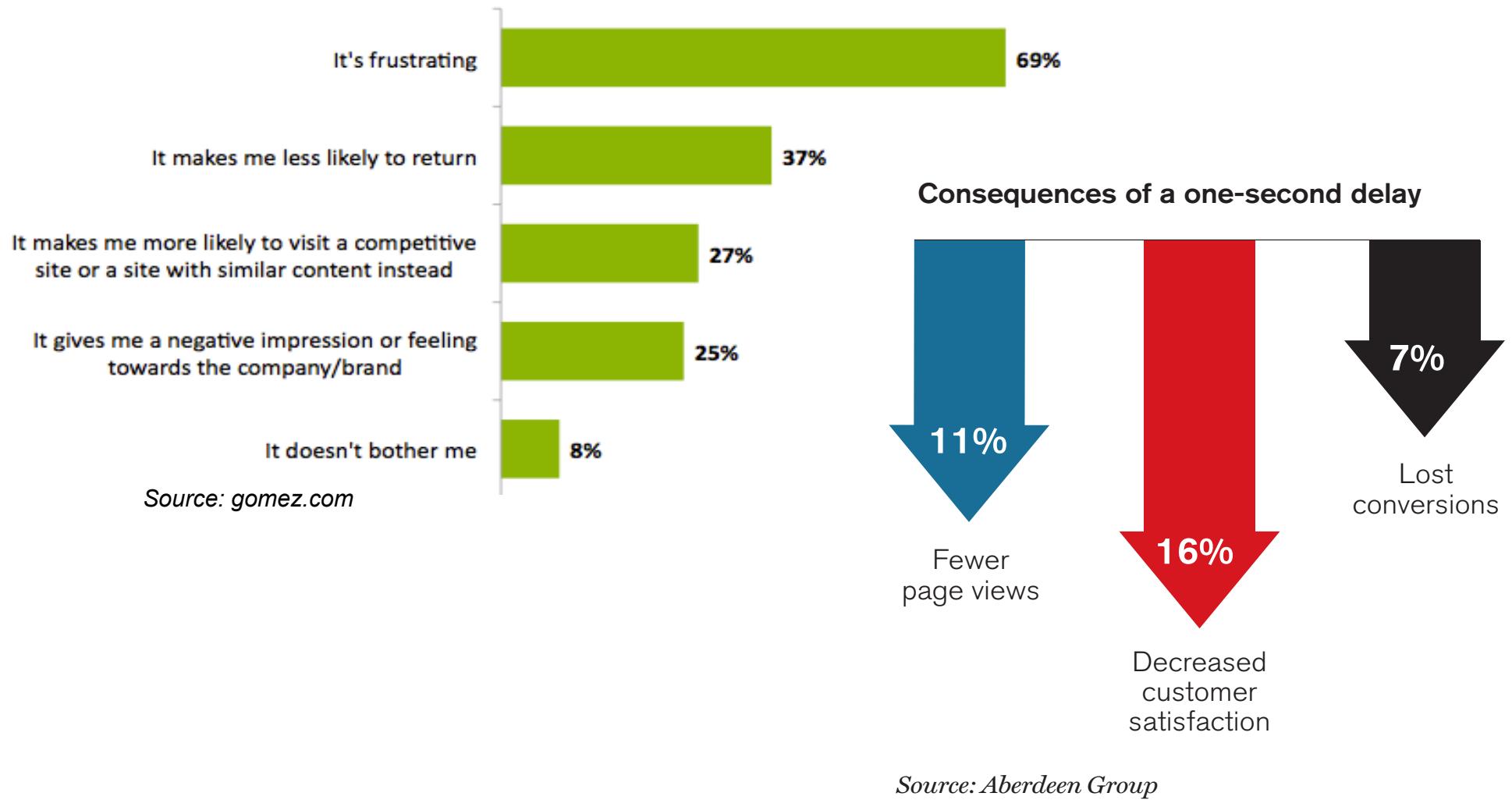
3 seconds



Source: Akamai

67% of users encounter “slow” sites once a week (gomez.com)

Why does load time matter?



Why does load time matter?

Research Findings



Industry leaders have demonstrated the impact of web performance on business value:

RESEARCH TEAM	BETTER WEB PERFORMANCE RESULTS IN
Google	2 X increased productivity ¹
Shopzilla	12% increase in revenue; 25% increase in page views for every 5 seconds of gain ²
AOL	2 X more page views for fastest customers ³
Microsoft Bing	5% more revenue per user for every 2 seconds of improvement ⁴
Amazon	1% more revenue for every 100 milliseconds of improvement ⁵
Yahoo	9% more traffic for every 400 milliseconds of improvement ⁶

References:

- ¹<http://assets.en.oreilly.com/1/event/29/Keynote%20Presentation%202.pdf>
- ²<http://en.oreilly.com/velocity2009/public/schedule/detail/7709>
- ³<http://en.oreilly.com/velocity2009/public/schedule/detail/7579>

- ⁴<http://en.oreilly.com/velocity2009/public/schedule/detail/8523>
- ⁵<http://sites.google.com/site/glinden/Home/StanfordDataMining.2006-11-28.ppt>
- ⁶<http://www.slideshare.net/stoyan/yslow-20-presentation>

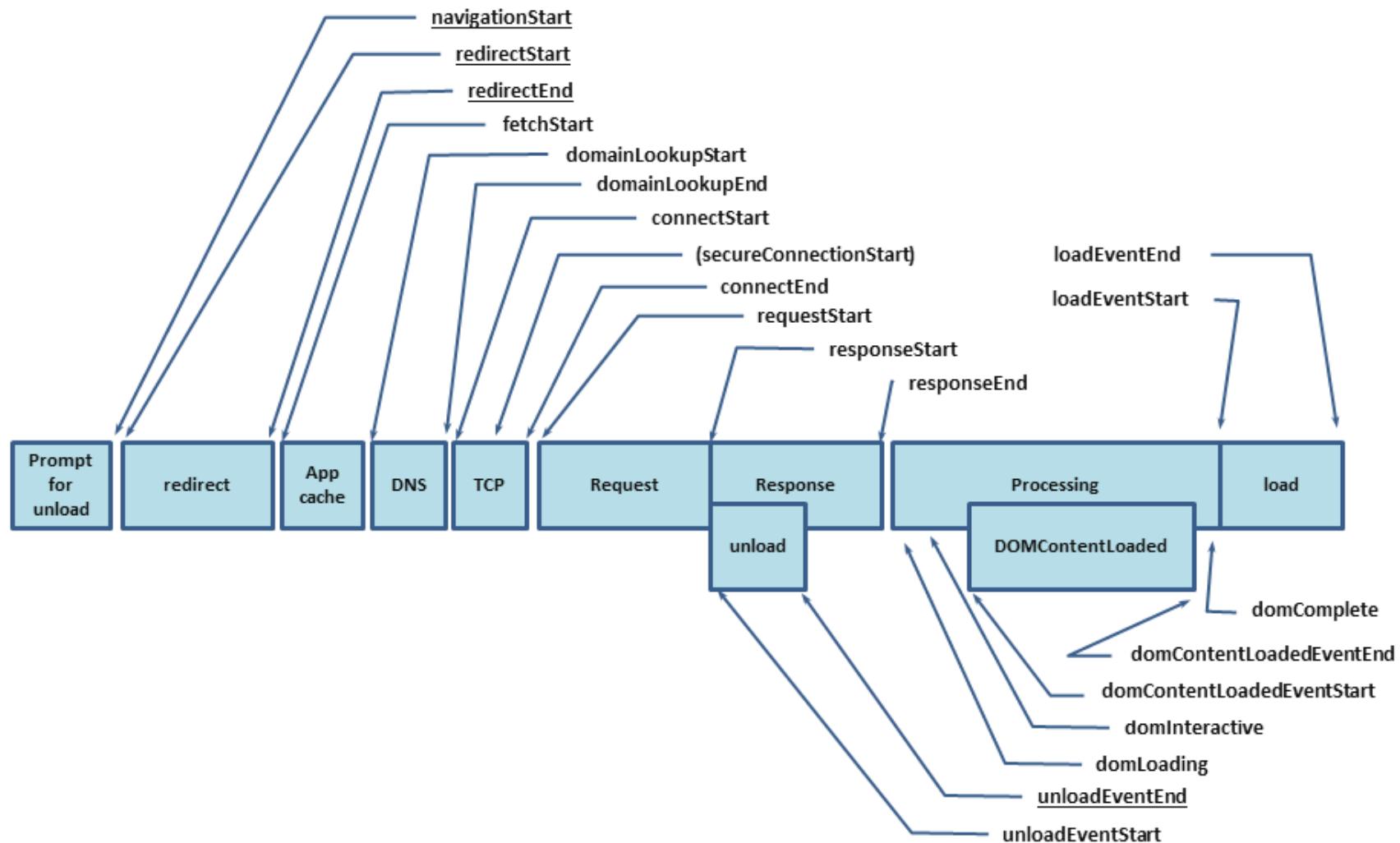
Survey: <http://www.strangeloopnetworks.com/assets/Uploads/SO-Datasheet.pdf>

Implications for:
Website owners
End users
Browser developers
Customization

Overview

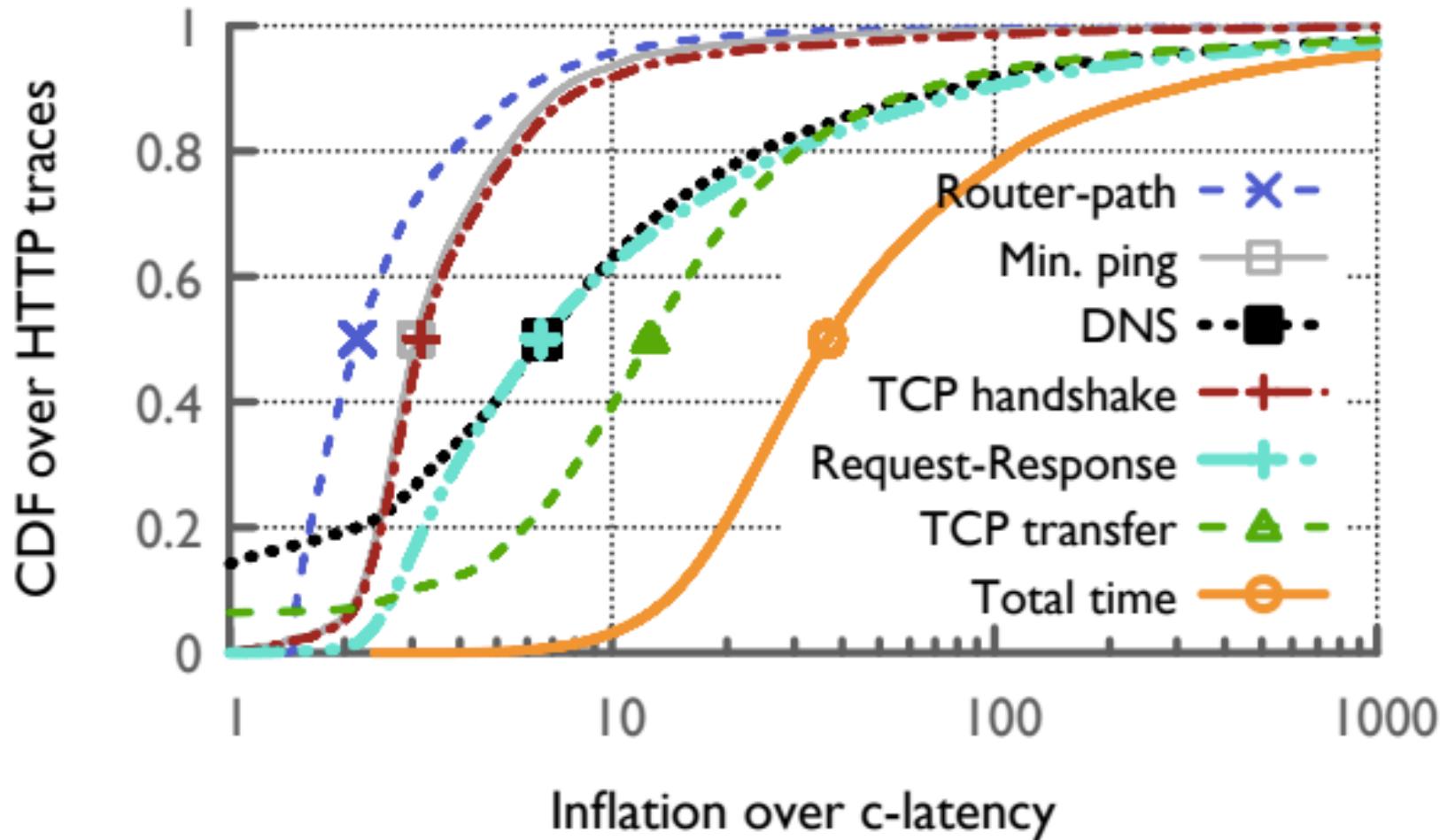
- Comprehensive study of *website complexity*
 - Analysis of sites across *rank* and *category*
 - Content and Service level metrics
- Key metrics that impact performance

Navigation Time



<https://dvcs.w3.org/hg/webperf/raw-file/tip/specs/NavigationTiming/Overview.html>

Bottleneck is a Moving Target



Source: "Why Internet is so Slow?!" PAM 2017

Useful Tools

- **Firefox Firebug**



You can download it in your firefox browser

<https://getfirebug.com/>

- **Chrome Developer tool**

View -> Developer -> Develop Tools



(and PageSpeed tool by Google)

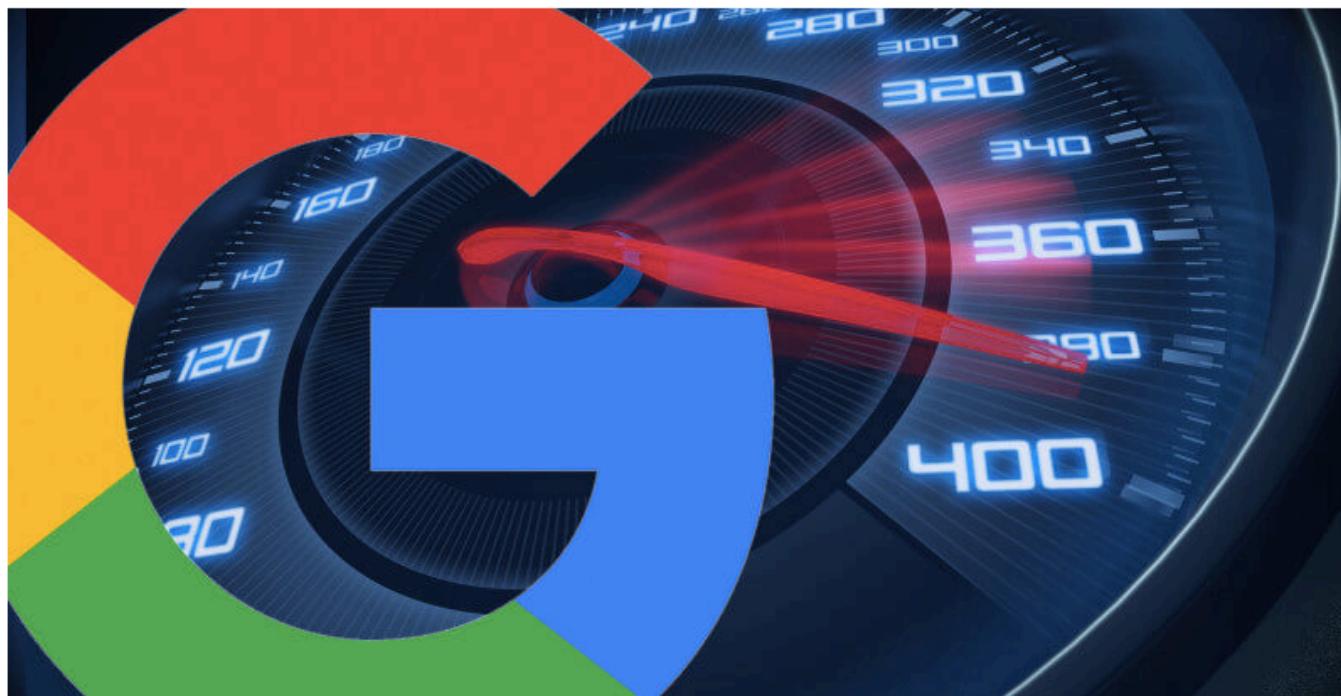


Page Download is now a Metric in PageRank!

The Google Speed Update: Page speed will become a ranking factor in mobile search

Starting in July 2018, Google will finally use mobile page speed as a ranking in their mobile search results.

Barry Schwartz on January 17, 2018 at 12:30 pm



Google Page Speed Example

https://developers.google.com/speed/pagespeed/insights/?url=www.tu-berlin.de

PageSpeed Insights HOME DOCS

http://www.tu-berlin.de/ ANALYZE

MOBILE DESKTOP

85

https://www.tu-berlin.de/menue/home/

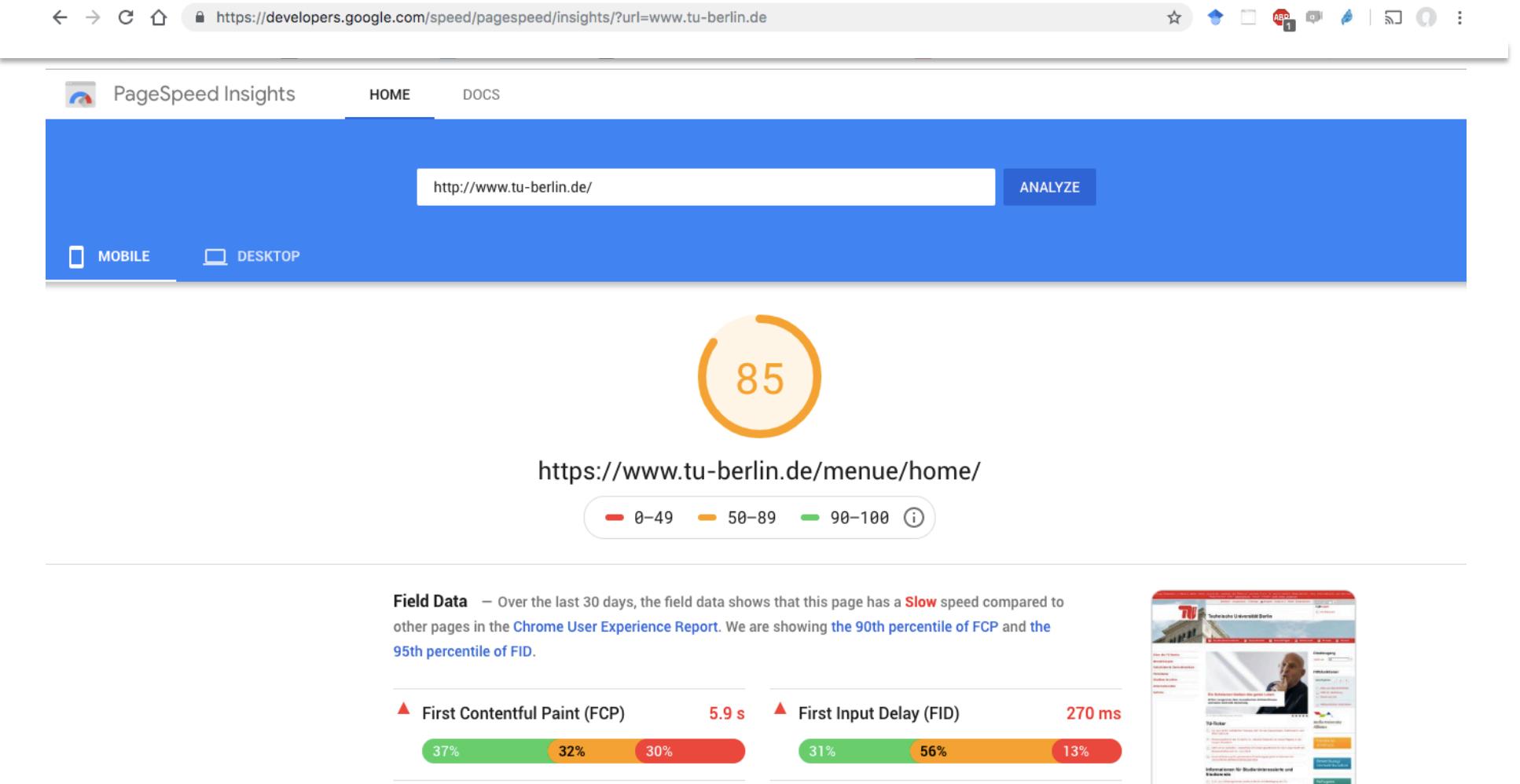
0-49 50-89 90-100

Field Data — Over the last 30 days, the field data shows that this page has a **Slow** speed compared to other pages in the [Chrome User Experience Report](#). We are showing the [90th percentile of FCP](#) and the [95th percentile of FID](#).

▲ First Contentful Paint (FCP) 5.9 s ▲ First Input Delay (FID) 270 ms

37% 32% 30%

31% 56% 13%



Chrome Developer Tool Example

The screenshot shows the Chrome Developer Tools Network tab for the URL <https://www.tu-berlin.de/menue/home/>. The Network tab is active, displaying a timeline of requests and a detailed table of resources.

Network Timeline: The timeline shows the progression of network requests from 500 ms to 2500 ms. Multiple parallel requests are visible, with colors indicating different types or initiators.

Resource Table: The table lists the following resources:

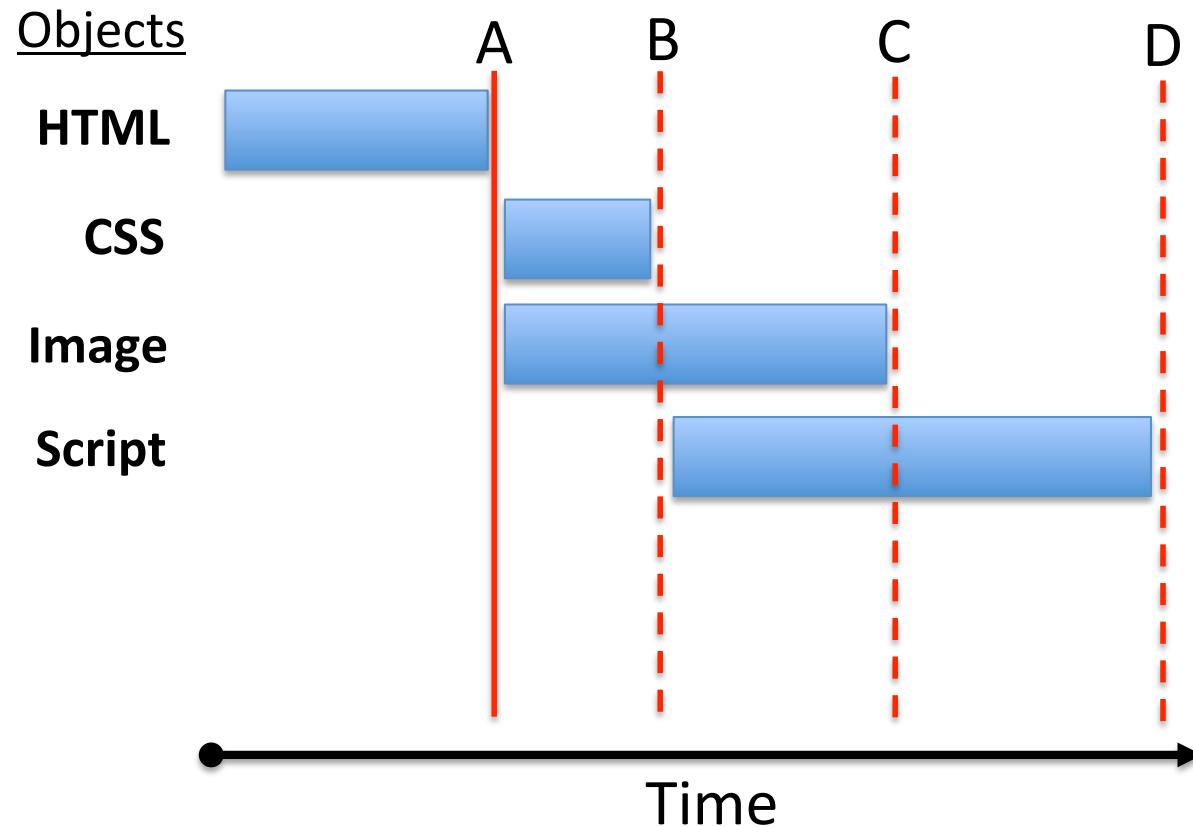
Name	Status	Type	Initiator	Size	Waterfall
icon-link-empfehlen.gif	200	gif	(index)	456 B	
piwik.js	200	script	(index):342	65.6 KB	
dotted-bg.gif	200	gif	(index)	556 B	
icon_link_up.gif	200	gif	(index)	900 B	
print.css?1459765087	200	stylesheet	(index)	2.3 KB	
icon-prev.gif	200	gif	jquery-1.11.3.mi...	695 B	
icon-next.gif	200	gif	jquery-1.11.3.mi...	696 B	
piwik.php?action_name=TU%20Berlin%...	200	gif	piwik.js:44	456 B	
favicon.ico	200	vnd.micr...	Other	914 B	

Total Statistics: 60 requests | 870 KB transferred | 1.1 MB resources | Finish: 2.14 s | DOMContentLoaded: 1.85 s | Load: 2.10 s

Measurement Setup

- Selecting websites
 - 1,700 websites from Quantcast top 20k (by Alexa)
 - Primary focus on landing (home) page
 - Annotated with Alexa Categories
- Tools
 - Firefox + Firebug
 - No Local Caching
- Approach
 - 4 vantage points (3 Amazon EC2, 1 UCR)
 - Every 60 second one page loaded
 - ~30 measurements per site per vantage point over 9 weeks

Example Site Download



Website Complexity Study (ca 2011)

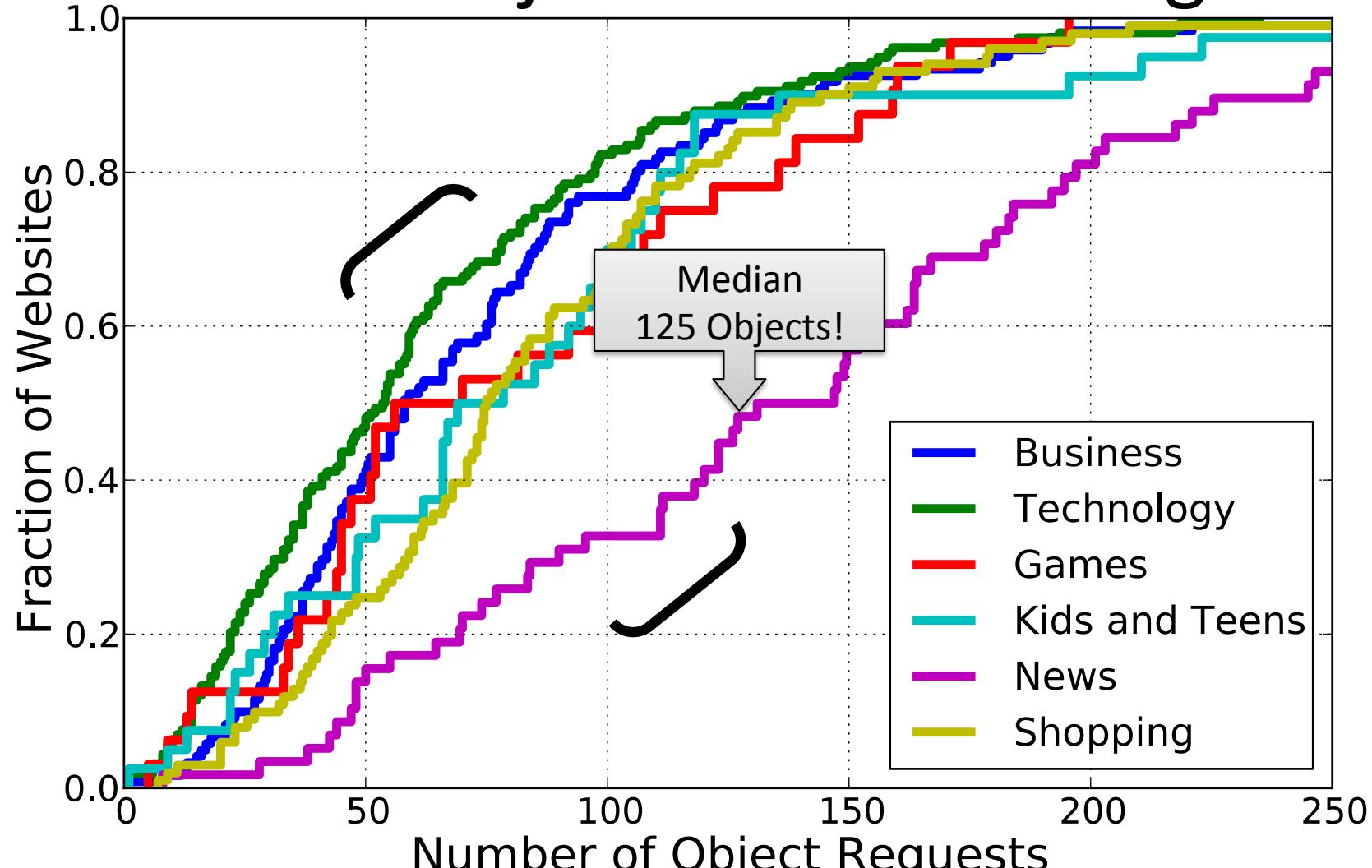
- Introduction
- Measurement Setup

- **Complexity**
 - Content-level
 - Service-level



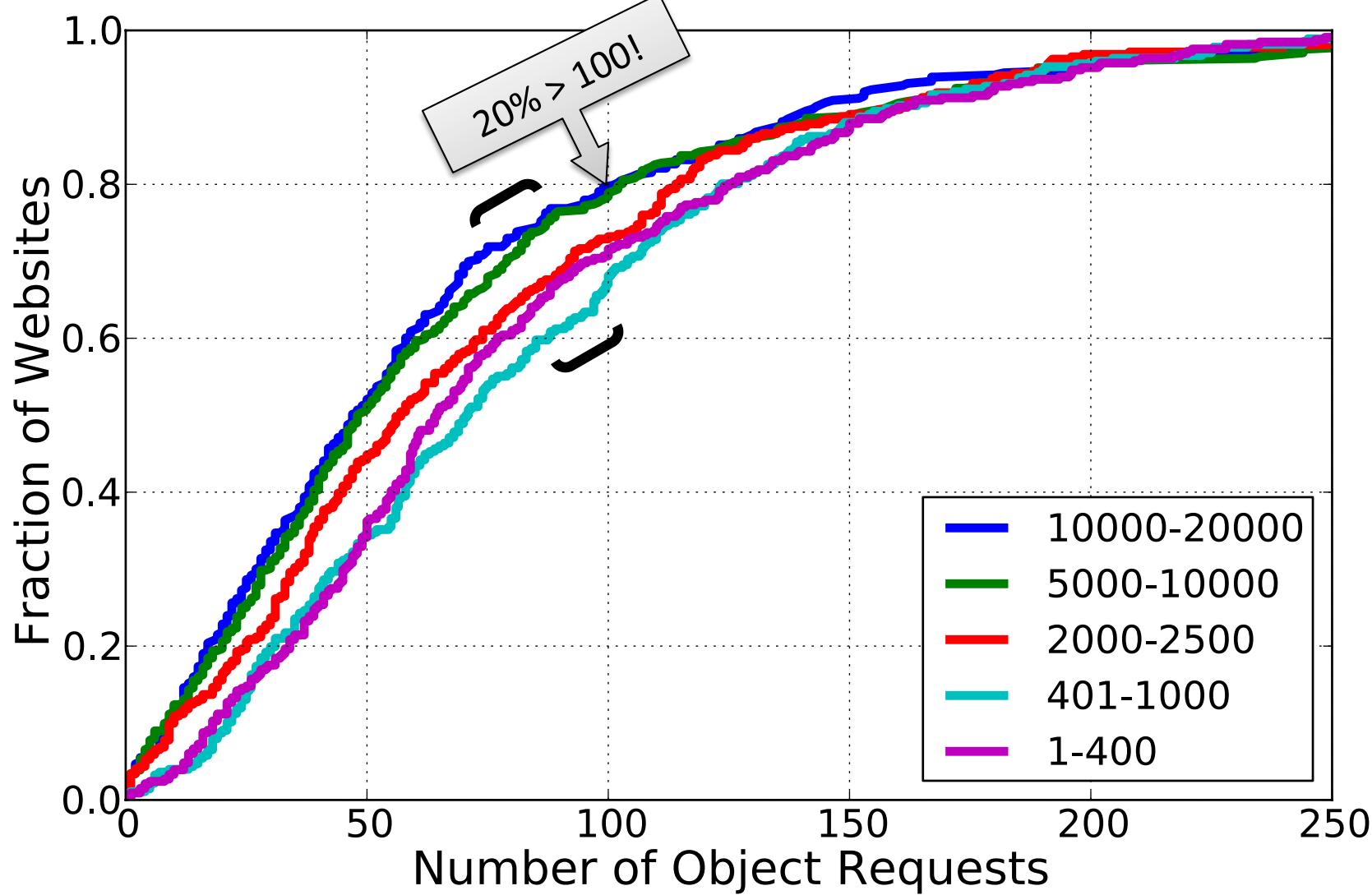
- Performance Implications

Number of Objects: Across Categories



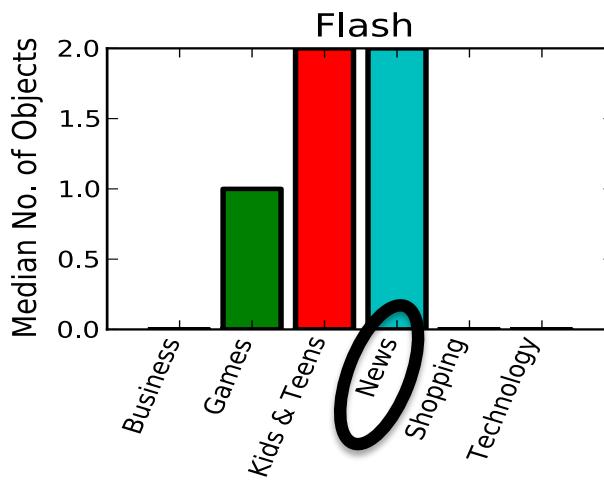
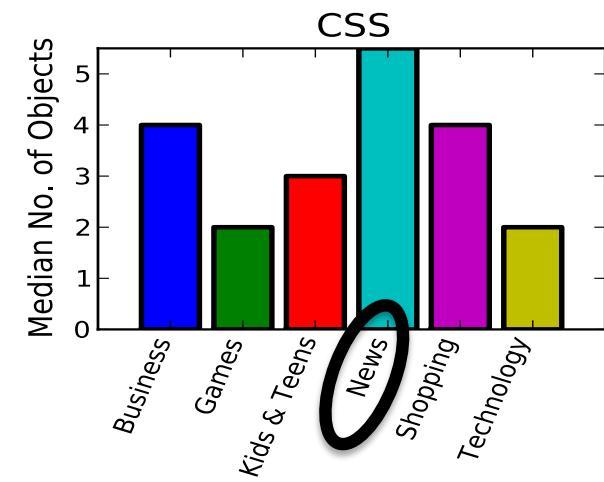
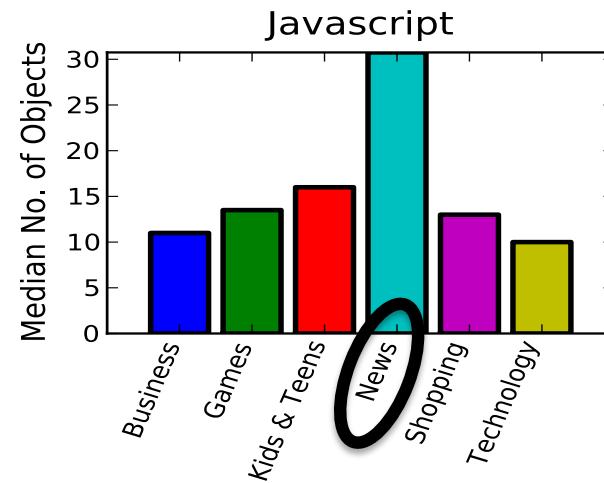
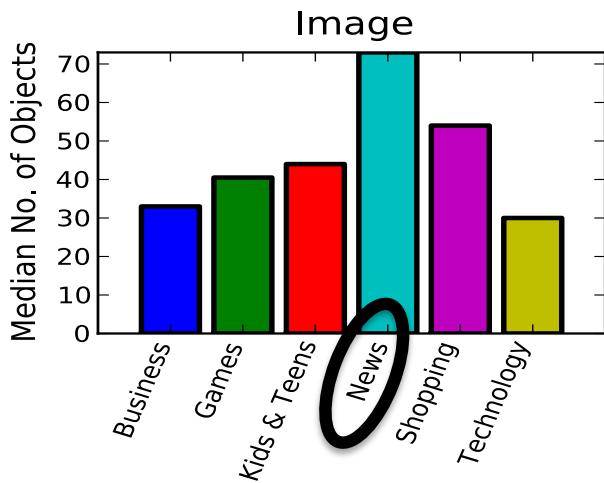
Median Site = 57 Objects!

Number of Objects: Across Ranks



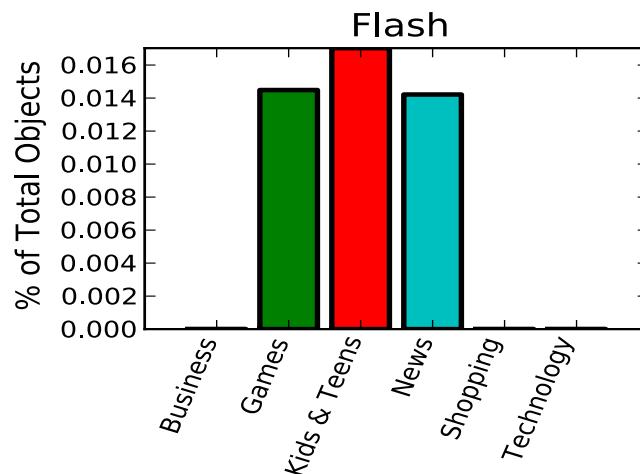
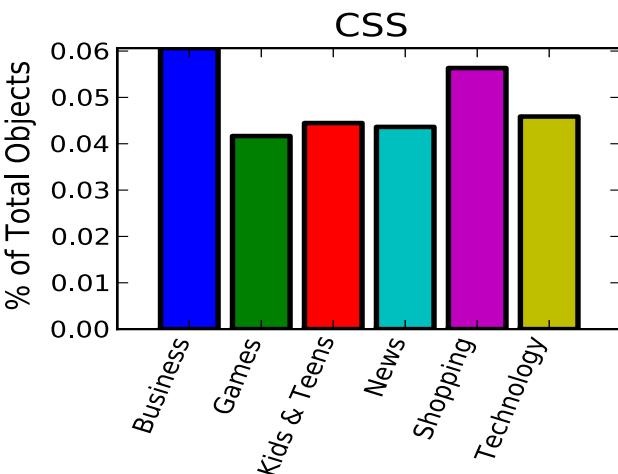
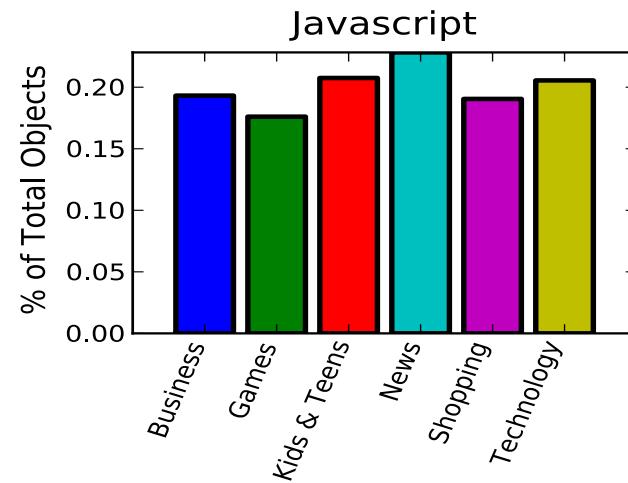
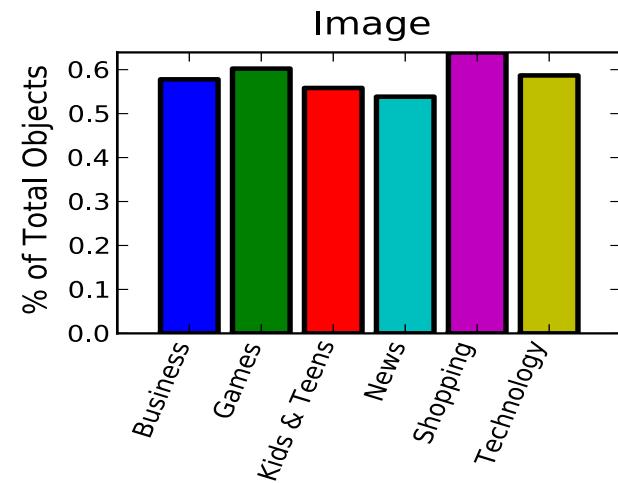
Not as much difference across rank ranges

Types of Content



Median site: 33 Images, 10 JavaScript, 3 CSS, 0 Flash

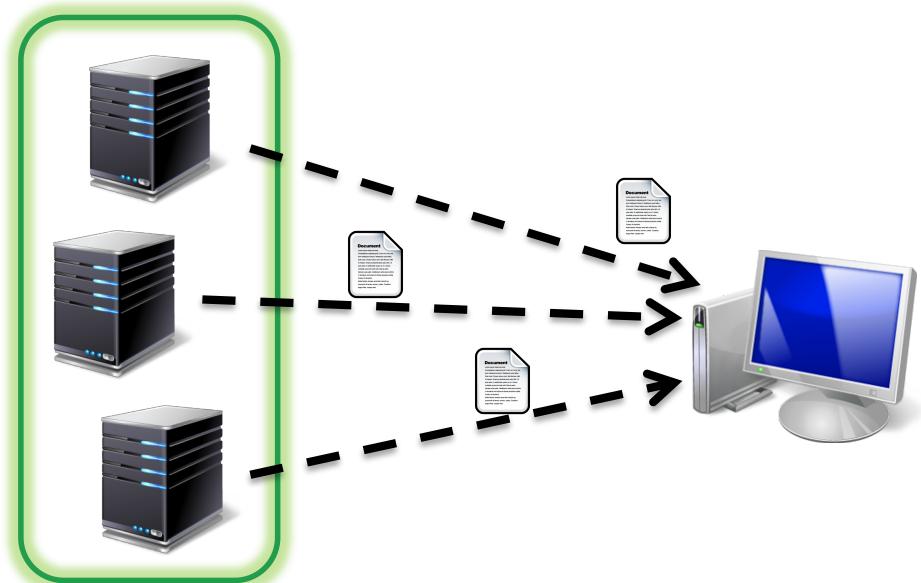
Normalized types of content



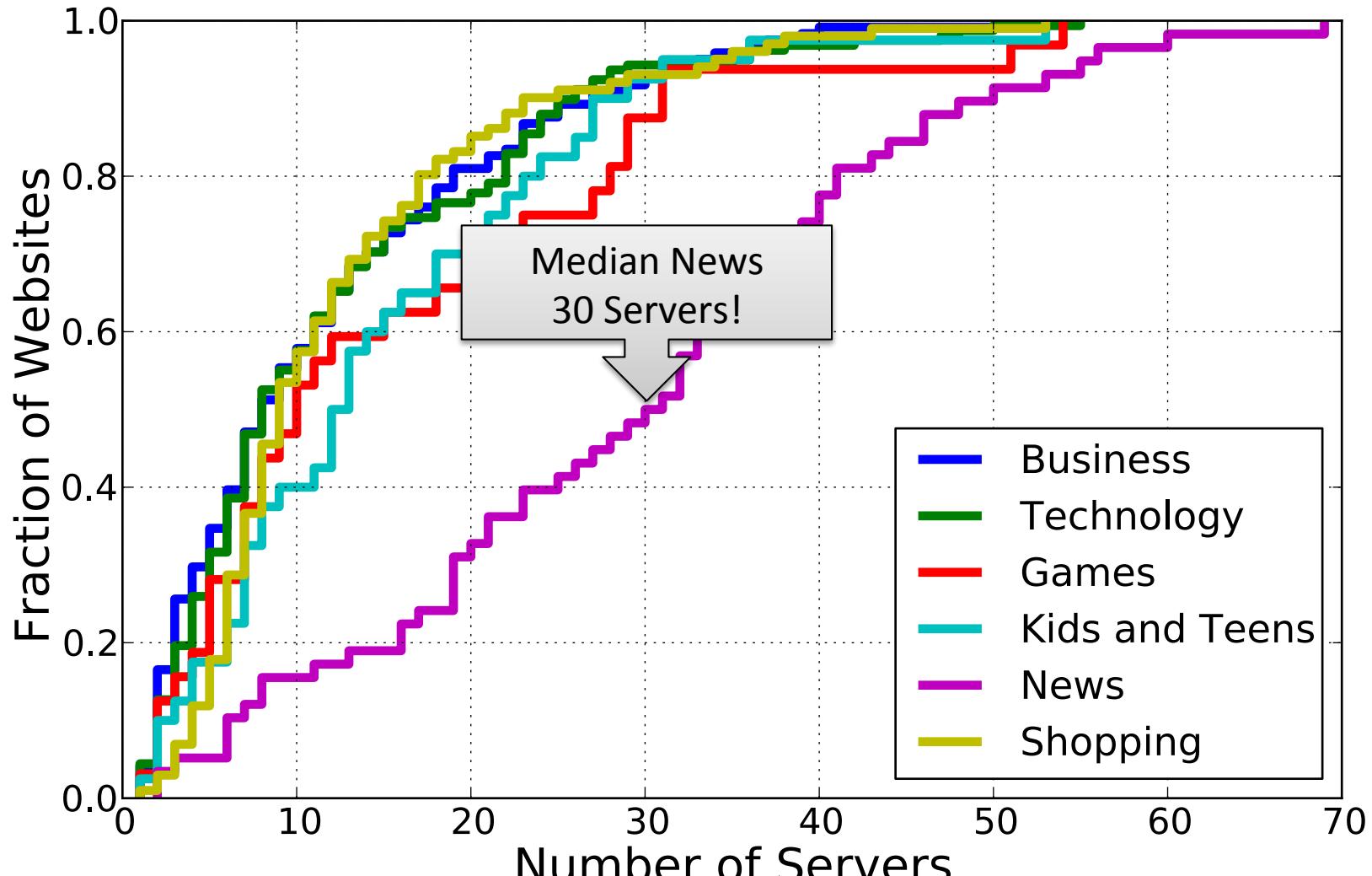
Type % Mostly Homogeneous; Flash Skewed

Website Complexity Study (ca 2011)

- Introduction
- Measurement Setup
- **Complexity**
 - Content-level
 - **Service-level (3rd Party servers)**
- Performance Implications

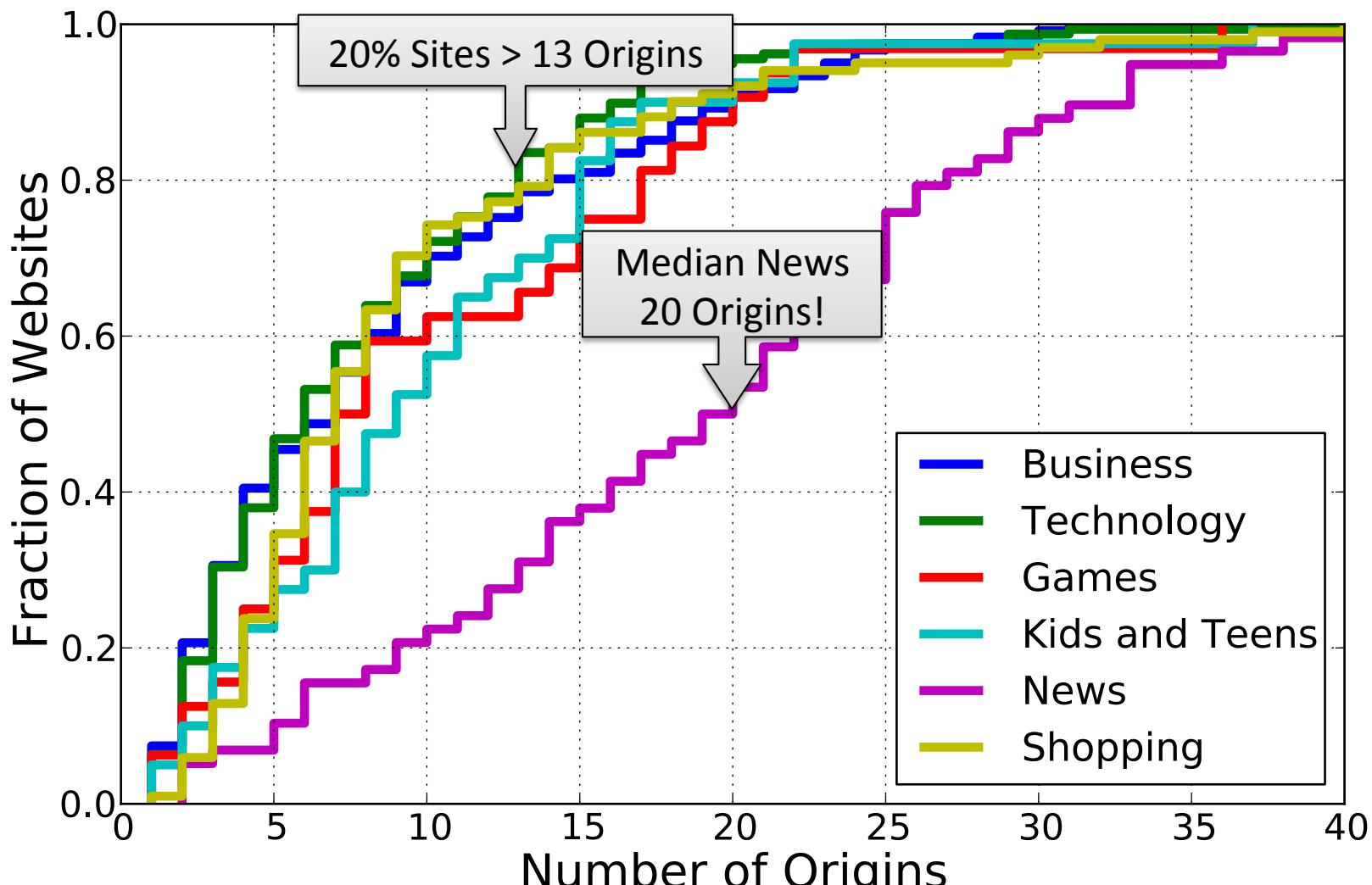


Number of Servers



Median site requires contacting 8 servers

Number of Origins



Median site requires contacting 6 origins

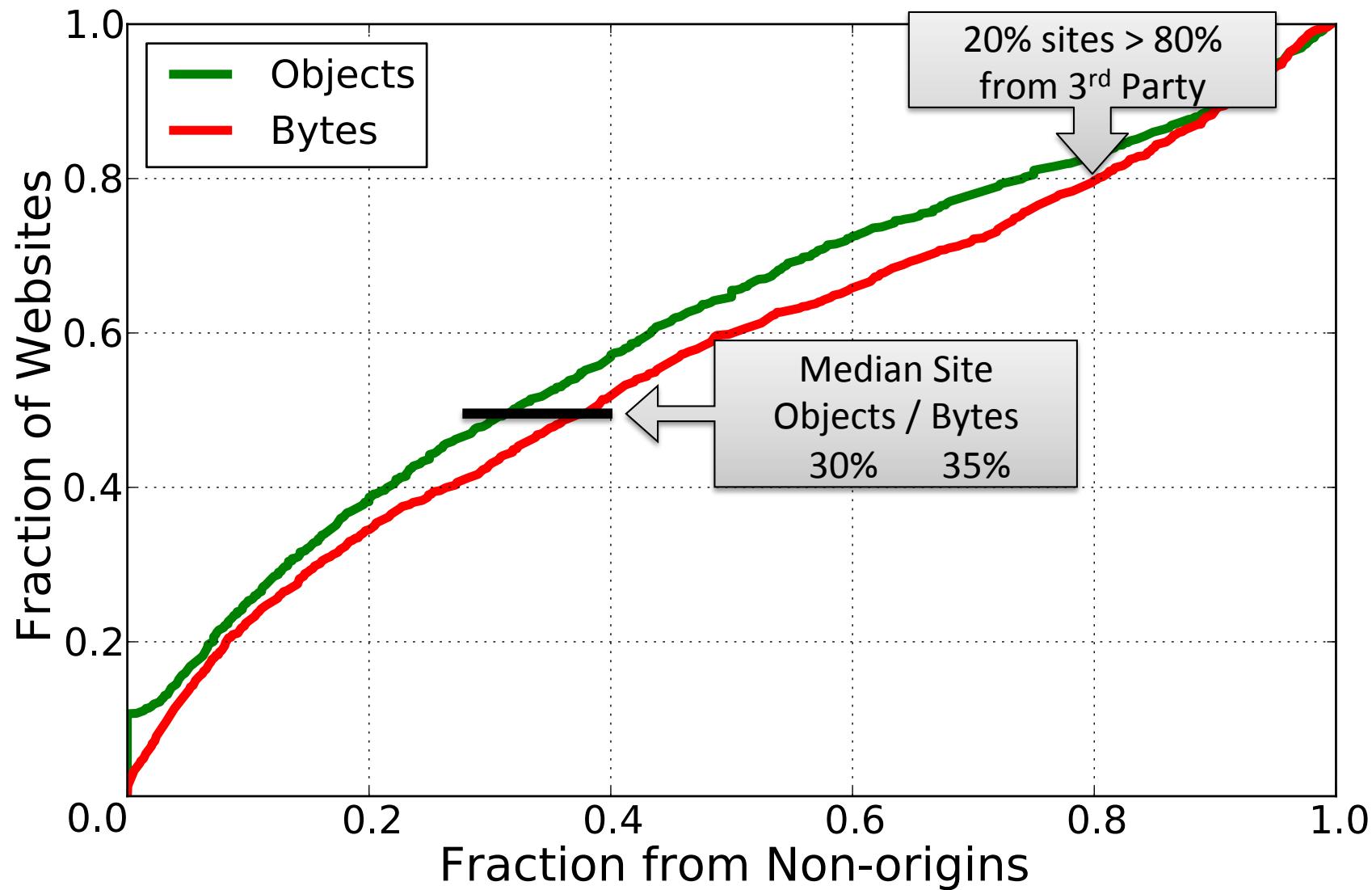
Popular non-origin providers

Name	% of sites
google-analytics	58
doubleclick	45
quantserve	30
scorecardresearch	27
2mdn	24
googleadservices	18
facebook	17
yieldmanager	16

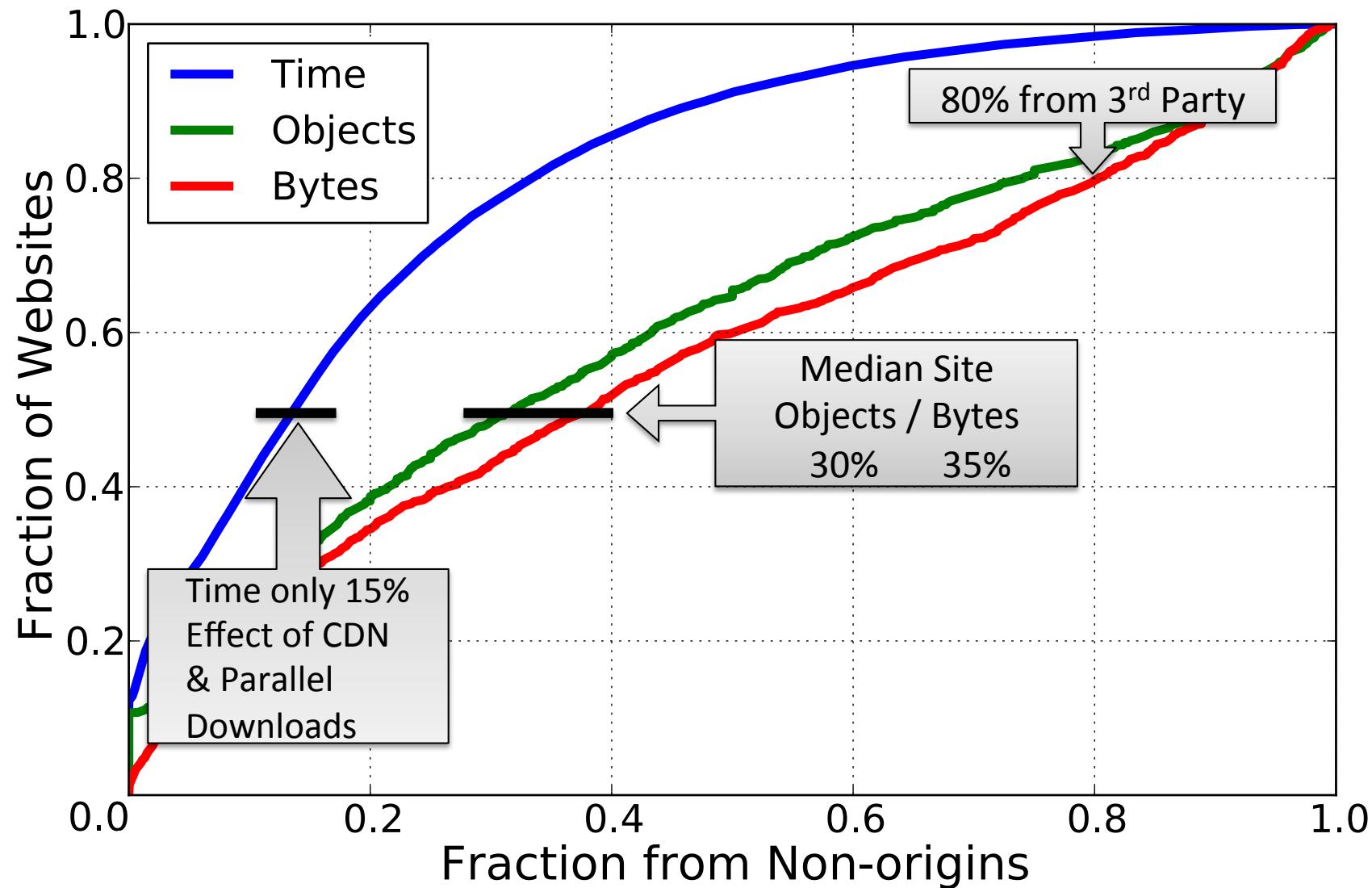
Most common services: Analytics & Advertising

Most common objects: Image (small!) and Javascript

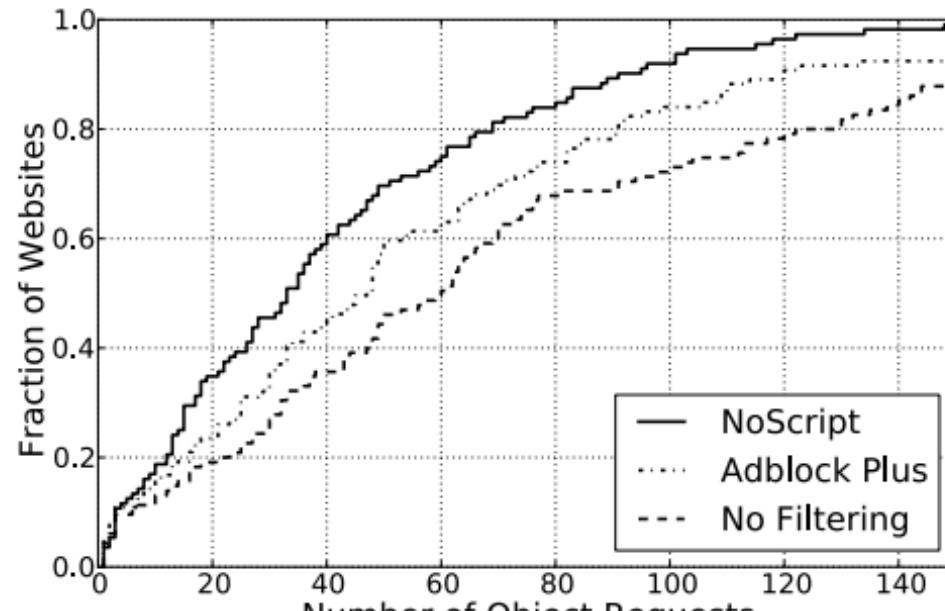
Contribution of non-origin services



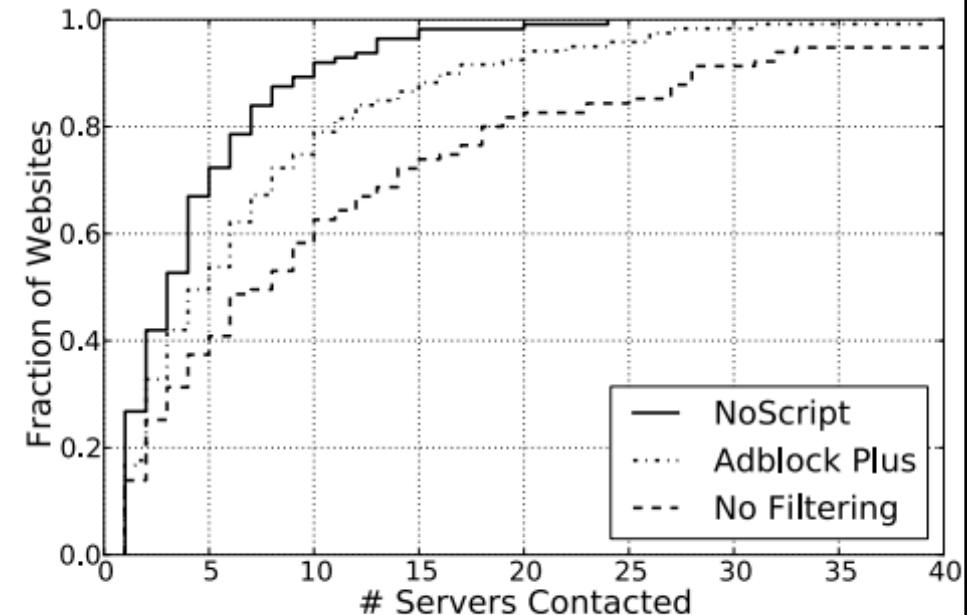
Contribution of non-origin services



Effect of Ad Blocking

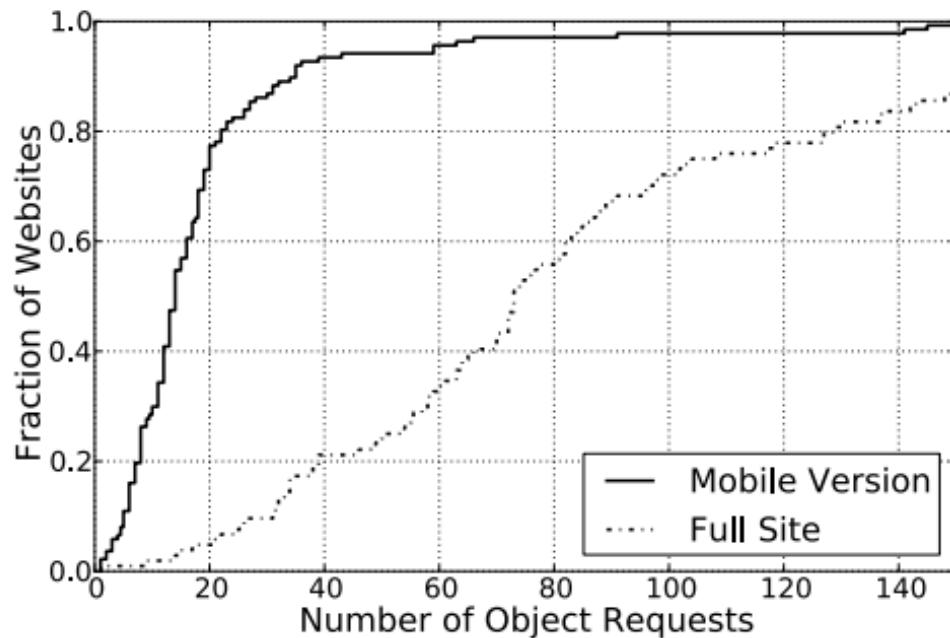


(a) No. of objects

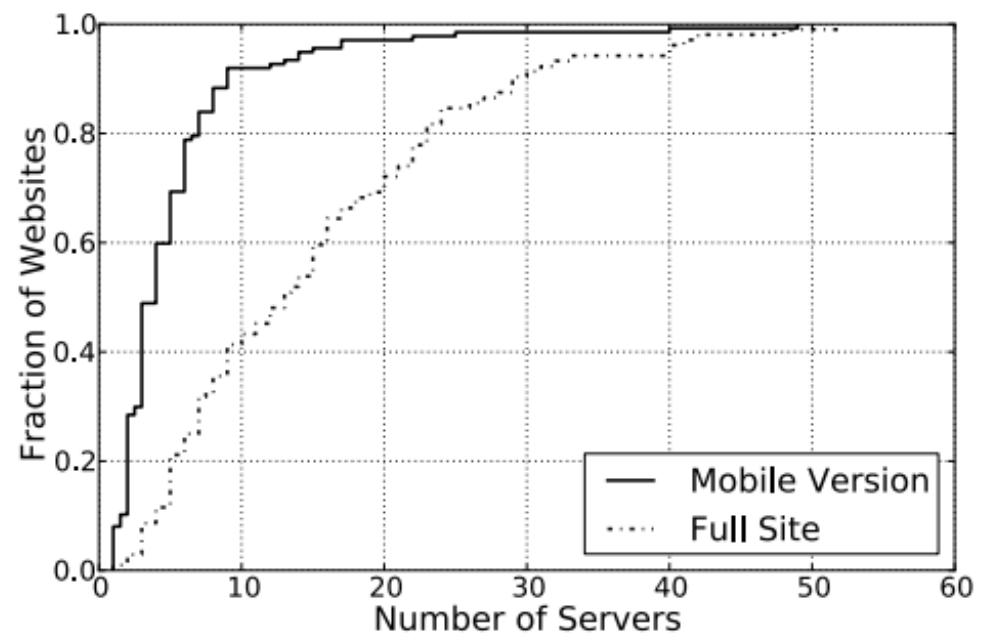


(b) Total no. of servers contacted

Optimizations for Mobiles



(a) No. of objects



(b) No. of servers

Discussion: Many other variables

- Client-side plugins
 - NoScript reduces #objects by half!
- Mobile-specific customizations
 - Mobile version reduces #objects to a quarter
- Landing vs. non-landing pages
 - Non-landing seem less complex

Readings

“Understanding Website Complexity:
Measurements, Metrics, and Implications”.
Michael Butkiewicz, Harsha V. Madhyastha, and
Vyas Sekar. IMC 2011

Acknowledgement

Material presented in these slides is borrowed from presentations
by:

Michael Butkiewicz, Harsha Madhyastha (UC Riverside),
Vyas Sekar (CMU), and Enric Pujol (TU Berlin)