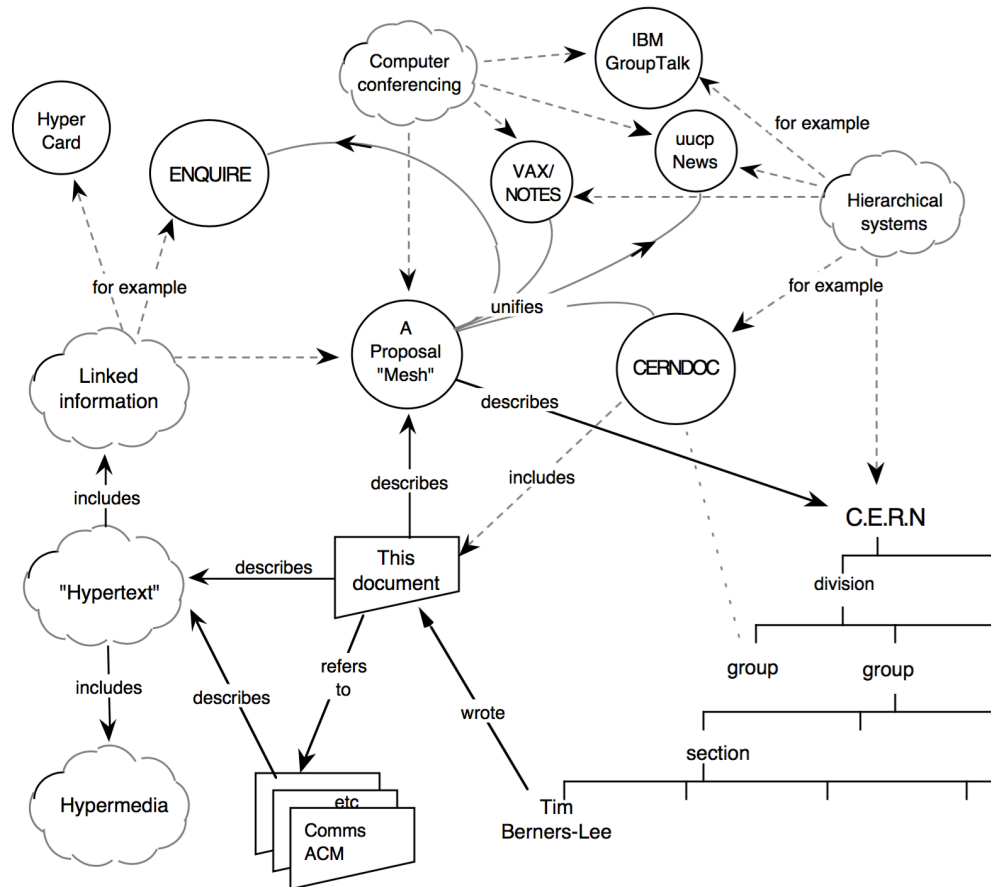


# Web

Lecture 23, Computer Networks (198:552)

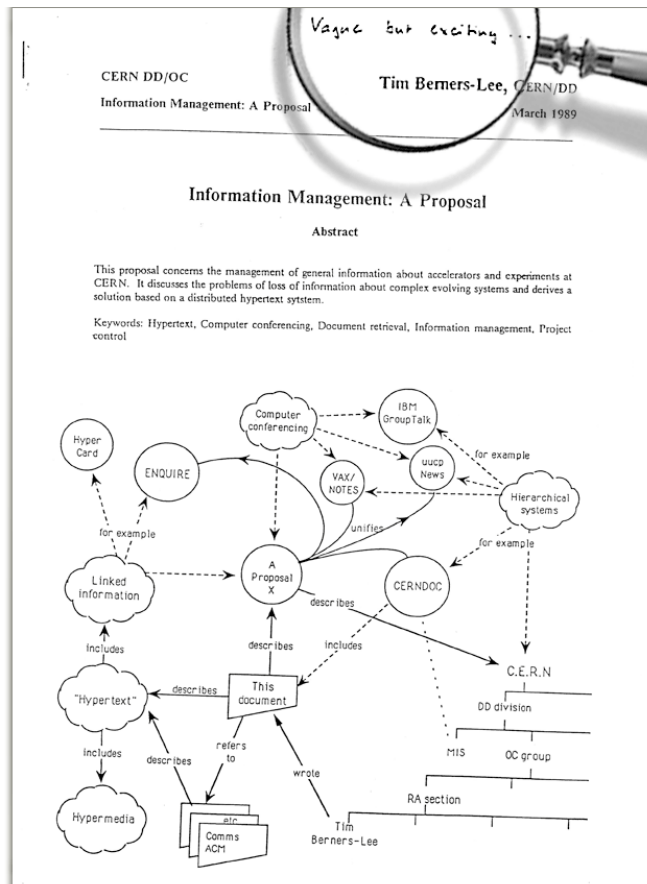
# Information management: A proposal



Sir Tim Berners-Lee

<https://cds.cern.ch/record/369245/files/dd-89-001.pdf>

# Information management: A proposal



“Vague, but exciting”

# Prescient ideas in the original proposal

- A distributed system linking content together
- Non-linear text navigation (“hypertext”)
- Access from different kinds of systems
- Access to different kinds of data and media
- The possibility of automated search and analysis
- Highly dynamic content and live links

# Structure of a web page: Rendered

A white rectangular search input field with a thin gray border and a subtle drop shadow. On the right side of the field is a small, colorful microphone icon, indicating voice search functionality.

Google Search

I'm Feeling Lucky

# Structure of a web page: Source code

`decoration:underline}.hRvfyE #fsett a:hover{text-decoration:underline}.hRvfyE a:hover{text-decoration:underline}.fbar p{display:inline}.fbar a,#fsettl{text-decoration:none;white-space:nowrap}.fbar{margin-left:-27px}.Fx4vi{padding-left:27px;margin:0 !important;In26Ec{padding:0 !important;margin:0 !important}#fbarcnt{display:block}.smiUbb img{margin-right:4px}.smiUbb a,.M6hT6 #swml a{text-decoration:none}.fmulti{text-align:center}.fmulti #fsr{display:block;float:none}.fmulti #fuser{display:block;float:none}.fmuser{line-height:25px}#fsr{float:right;white-space:nowrap}#fsl{white-space:nowrap}#fset{background:#fff;border:1px solid #999;bottom:30px;padding:10px 0;position:absolute;box-shadow:0 4px rgba(0,0,0,2);box-shadow:0 2px 4px rgba(0,0,0,2);text-align:left;z-index:104}#fsett a{display:block;line-height:44px;padding:0 20px;text-decoration:none;white-space:nowrap}.p2Kmcc{padding:8px 16px;margin-right:10px}.B4GxFc{margin-left:166px}.fbar{background:#f2f2f2;border-top:1px solid #e4e4e4;line-height:40px;min-width:980px}.yBWvie{margin-left:166px}.fbar p,.fbar a,#fsettl,#fsett a{color:#5f6368}.fbar a: hover,#fsett a: hover{color:#333}.fbar{font-size:small}#fuser{float:right}.smiUbb{margin-left:166px;line-height:15px;color:#5f6368};</style><style>#fsl{margin-left:30px}#fsr{margin-right:30px}.fmulti #fsl{margin-left:0}.fmulti #fsr{margin-right:0}</style><div class="EvHmz hRvfyE" id="fbar"><div class="fbar"><div class="b2hzT"> <style>.b0KoTc{color:rgba(0,0,0,.54);padding-left:27px;.b0KoTc .UnWzvd{margin-left:13px}.Q8LRLC{font-size:15px;.b0KoTc{margin-left:30px;text-align:left}</style> <div class="b0KoTc"> </div></div><span id="fsr">a href="https://www.google.com/intl/en_us/policies/privacy/?fg=1" ping="/url?sa=t&amp;rct=j&amp;source=webhp&amp;url=https://www.google.com/intl/en_us/policies/privacy/%3Ffg%3D1&amp;ved=0ahUKEwI74_rVjLeAhVhkeAKHYVjCKQ08awCCAq">Privacy</a><a class="Fx4vi" href="https://www.google.com/intl/en_us/policies/terms/?fg=1" ping="/url?sa=t&amp;rct=j&amp;source=webhp&amp;url=https://www.google.com/intl/en_us/policies/terms/%3Ffg%3D1&amp;ved=0ahUKEwi74_rVjLeAhVhkeAKHYVjCKQ08qwCCAQ">Terms</a><span style="display:inline-block;position:relative"><a class="Fx4vi" href="https://www.google.com/preferences?hl=en" id="fsettl" aria-controls="fsett" aria-expanded="false" aria-haspopup="true" role="button" onclick="foot set" ping="/url?`

- HTML: text content and (basic) formatting
  - `<a href="...">link text</a>`
- JavaScript: content rendering
  - `document.getElementById("x")`
- CSS: styling of an entire page
  - `link { color: red }`
- Media: images, videos, etc.

# Structure of a web page: DOM

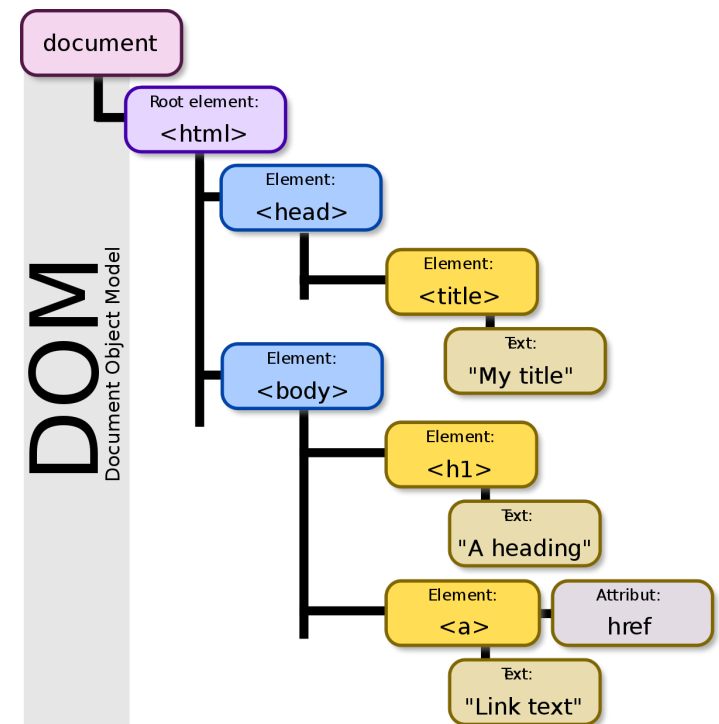
## Document object model

Hierarchical representation of elements

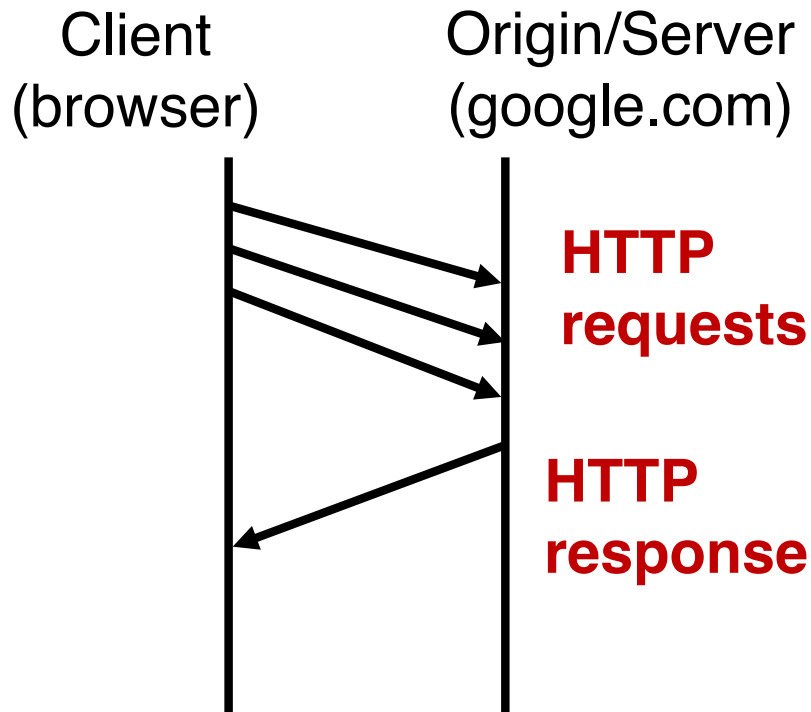
*Lexical* dependencies

Interface for processing, not rendering

Common language for content  
*specification* (HTML) and *formatting* (JS,  
CSS)



# Structure of a web page: The network



HttpDetect (EffeTech HTTP Sniffer)

File View Sniffer Help

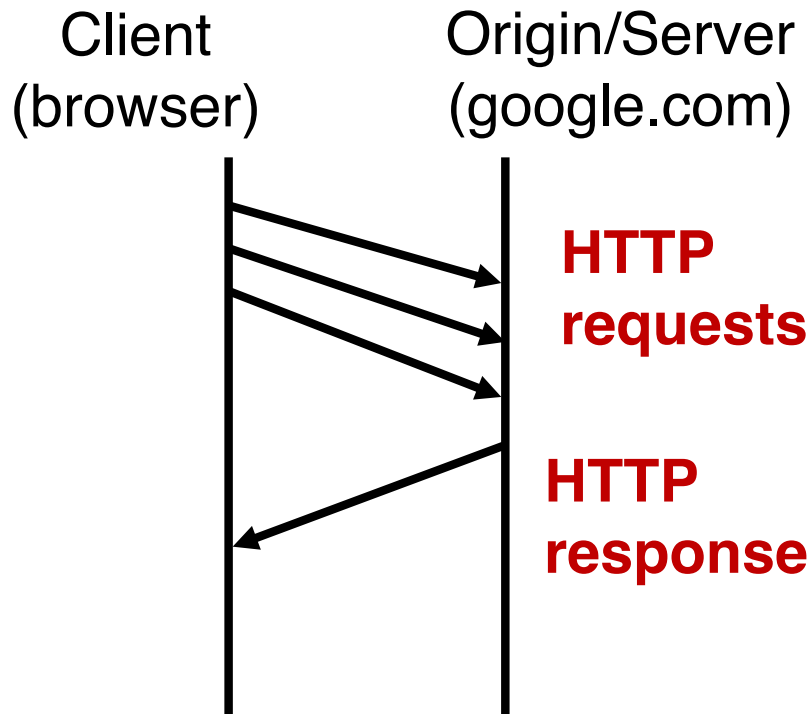
No.	Time	Client (IP:PORT)	Server (IP:PORT)	URL	File...	Status
69	Ju...	192.168.1.3 :3057	www.effetechn.com :80	/	35660	FIN, 200
70	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/style.css		FIN, 304
71	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/logo_main.jpg		FIN, 304
72	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/chinese_edition.gif		FIN, 304
73	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/space.gif		FIN, 304
74	Ju...	192.168.1.3 :3059	www.effetechn.com :80	/images/arrow_small.gif		FIN, 304
75	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/award_tucows_4ratel...		FIN, 304
76	Ju...	192.168.1.3 :3060	www.etherdetect.com...	/images/logo_ms.gif	628	FIN, 200
77	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_ibm.gif	1217	FIN, 200
78	Ju...	192.168.1.3 :3059	www.effetechn.com :80	/images/award_FileHungry_5s...		FIN, 304
79	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/award_softwareseeke...		FIN, 304
80	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_mit.gif	259	FIN, 200
81	Ju...	192.168.1.3 :3060	www.etherdetect.com...	/images/logo_ms.gif		Requested
82	Ju...	192.168.1.3 :3059	www.effetechn.com :80	/images/award_webaward2002e...		FIN, 304
83	Ju...	192.168.1.3 :3058	www.effetechn.com :80	/images/ed_small.gif	24269	FIN, 200
84	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_cornell.gif	2027	FIN, 200
85	Ju...	192.168.1.3 :3059	www.effetechn.com :80	/images/flag_detail.gif	1026	FIN, 200
86	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_reuters.gif	1822	FIN, 200
87	Ju...	192.168.1.3 :3059	www.effetechn.com :80	/images/flag_demo.gif	1013	FIN, 200
88	Ju...	192.168.1.3 :3059	www.etherdetect.com...	/images/flag_demo.gif	1013	FIN, 200

HTTP Request Header	HTTP Response Header
GET /images/logo_ibm.gif HTTP/1.1	HTTP/1.1 200 OK
Accept: */*	Date: Sat, 07 Jun 2003 13:32:07 GMT
Referer: http://www.effetechn.com/	Server: Apache/1.3.27
Accept-Language: zh-cn	Last-Modified: Mon, 14 Apr 2003 14:11:33 GMT
Accept-Encoding: gzip, deflate	ETag: "bdae-4c1-3e9ac195"
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)	Accept-Ranges: bytes
Host: www.etherdetect.com	Content-Length: 1217
Connection: Keep-Alive	Keep-Alive: timeout=5, max=100

Ready Buffer: 3% URLs: 95 Packets: 393



# Structure of a web page: The network



Requests based on DOM or JS (xmlHttpRequest/AJAX)

HttpDetect (EffeTech HTTP Sniffer)

No.	Time	Client (IP:PORT)	Server (IP:PORT)	URL	File...	Status
69	Ju...	192.168.1.3 :3057	www.effetech.com :80	/	35660	FIN, 200
70	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/style.css		FIN, 304
71	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/logo_main.jpg		FIN, 304
72	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/chinese_edition.gif		FIN, 304
73	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/space.gif		FIN, 304
74	Ju...	192.168.1.3 :3059	www.effetech.com :80	/images/arrow_small.gif		FIN, 304
75	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/award_tucows_4ratel...		FIN, 304
76	Ju...	192.168.1.3 :3060	www.etherdetect.com...	/images/logo_ms.gif	628	FIN, 200
77	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_ibm.gif	1217	FIN, 200
78	Ju...	192.168.1.3 :3059	www.effetech.com :80	/images/award_FileHungry_5s...		FIN, 304
79	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/award_softwareseeke...		FIN, 304
80	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_mit.gif	259	FIN, 200
81	Ju...	192.168.1.3 :3060	www.etherdetect.com...	/images/logo_ms.gif		Requested
82	Ju...	192.168.1.3 :3059	www.effetech.com :80	/images/award_webaward2002e...		FIN, 304
83	Ju...	192.168.1.3 :3058	www.effetech.com :80	/images/ed_small.gif	24269	FIN, 200
84	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_cornell.gif	2027	FIN, 200
85	Ju...	192.168.1.3 :3059	www.etherdetect.com :80	/images/flag_detail.gif	1026	FIN, 200
86	Ju...	192.168.1.3 :3061	www.etherdetect.com...	/images/logo_reuters.gif	1822	FIN, 200
87	Ju...	192.168.1.3 :3059	www.etherdetect.com :80	/images/flag_demo.gif	1013	FIN, 200
88	Ju...	192.168.1.3 :3059	www.etherdetect.com :80	/images/flag_demo.gif	1013	FIN, 200

HTTP Request Header	HTTP Response Header
GET /images/logo_ibm.gif HTTP/1.1	HTTP/1.1 200 OK
Accept: */*	Date: Sat, 07 Jun 2003 13:32:07 GMT
Referer: http://www.effetech.com/	Server: Apache/1.3.27
Accept-Language: zh-cn	Last-Modified: Mon, 14 Apr 2003 14:11:33 GMT
Accept-Encoding: gzip, deflate	ETag: "bdae-4c1-3e9ac195"
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)	Accept-Ranges: bytes
Host: www.etherdetect.com	Content-Length: 1217
Connection: Keep-Alive	Keep-Alive: timeout=5, max=100

Ready Buffer: 3% URLs: 95 Packets: 393

# Polaris: Faster Page Loads using Fine-Grained Dependency Tracking

Ravi Netravali et al., NSDI '16

(ACK: Material by Ravi Netravali)

# Web Performance

- Users demand fast page loads
- Slow page loads lead to lost revenue and low search rank

## Research: Site Speed Is Hurting Your Everyone's Revenue

IAN LURIE // MAY 9 2014

Site speed, site speed, site speed. Everyone around me is sick of hearing me say that because I've pushed it on every client Portent's had since, oh, 2008.

**Google** Webmaster Central Blog

Official news on crawling and indexing sites for the Google index

Using site speed in web search ranking

Friday, April 09, 2010

Webmaster Level: All

You may have heard that here at Google we're obsessed with speed, in [our products](#) and [on the web](#). As part of that effort, today we're including a new signal in our search ranking algorithms: site speed. Site speed reflects how quickly a website responds to web requests.

## How One Second Could Cost Amazon \$1.6 Billion In Sales

Research on U.S. Net habits suggests that if this sentence takes longer than a second to load, many citizens will have clicked elsewhere already. If you've got the patience (or are European) read on for more shocking data on not dawdling.

## It's Official: Google Now Counts Site Speed As A Ranking Factor

Matt McGee on April 9, 2010 at 2:00 pm

Google has [kept a promise](#) it made last year: Site speed is now a ranking factor in Google's algorithm, and is already in place for U.S. searchers. But Google also cautions web site owners not to sacrifice relevance in the name of faster web pages, and even says this new ranking factor will impact very few queries. More on that below, but first the background on today's announcement from Google Fellow Amit Singhal and Matt Cutts, head of Google's web spam team.

### Why Page Speed Matters

The first warning that site speed was on Google's radar came last November, when [Cutts said](#) there

## Google Rank Website On Loading Time of the Page

By: Harsh Agrawal | In: SEO | Last Updated: 18/03/2015

Back in 2010, the [Google webmaster](#) team indicated that they will start ranking websites on their page loading time. Websites which take ages to load slows down the user experience and they are considering this factor seriously. Apart from other parameters like meta descriptions, Google will also consider Page load time as one of the reasons for your website search engine ranking.

## How Website Speed Actually Impacts Search Ranking

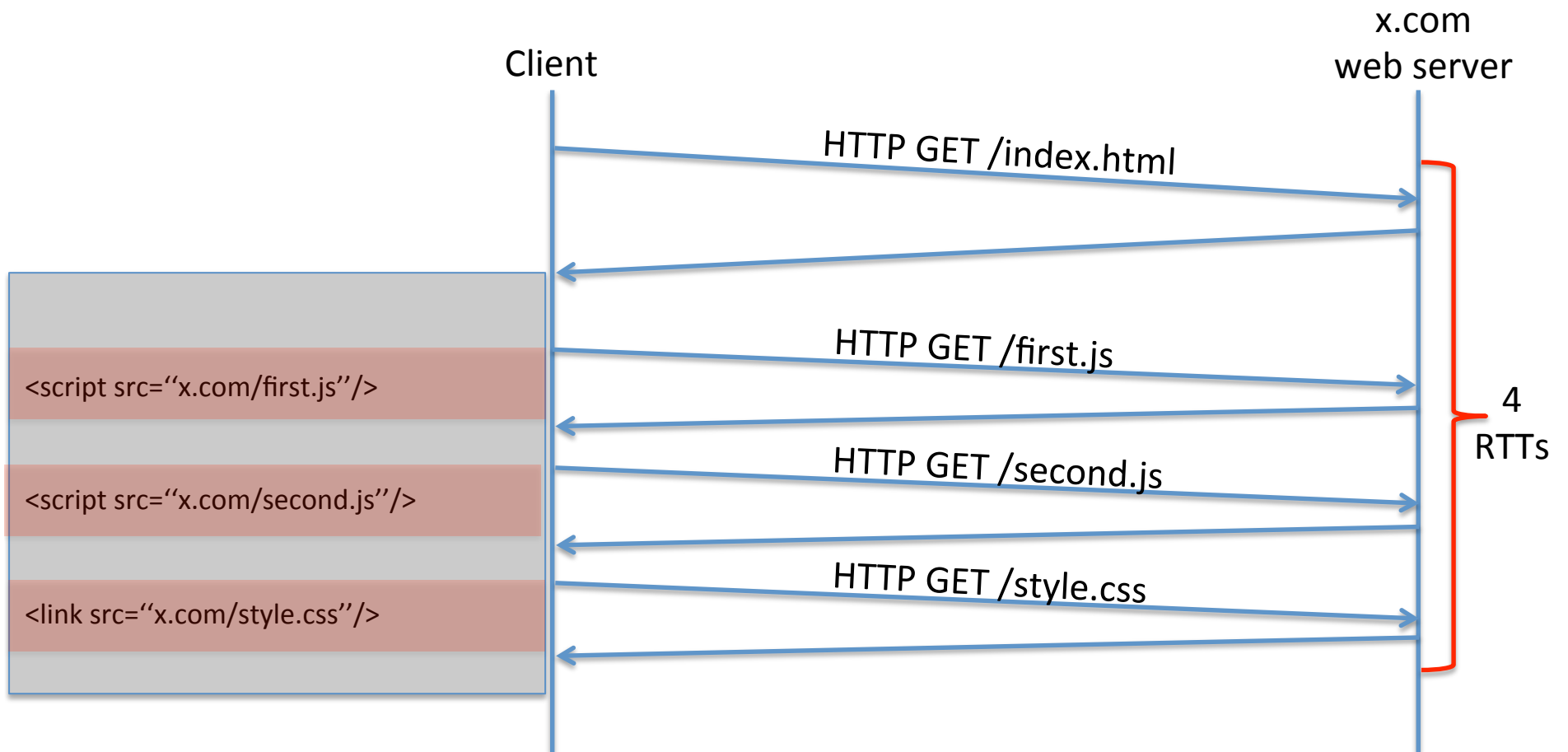
[On-page SEO](#)

The author's views are entirely his or her own (excluding the unlikely event of hypnosis) and may not always reflect the views of Moz.



Google uses a multitude of factors to determine how to rank search engine results. Factors are either related to the content of a webpage itself (the text, its URL, the ti

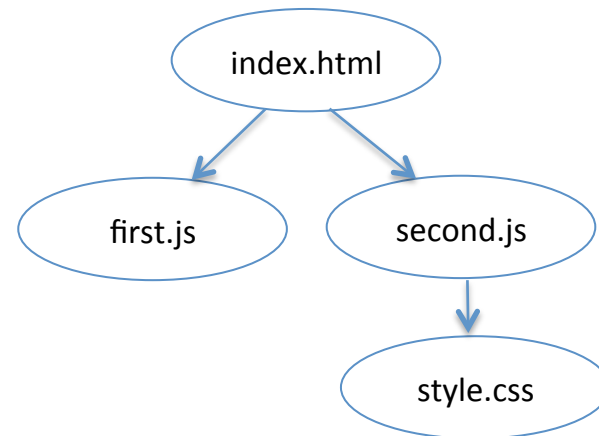
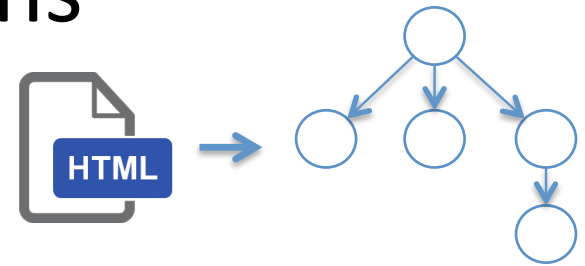
# Page Load



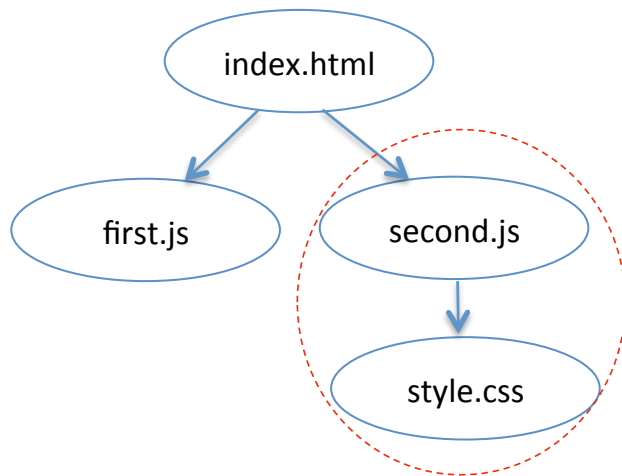
# Dependency Graphs

Model page loads as directed acyclic graphs

- Page load time = time to completely resolve dependency graph

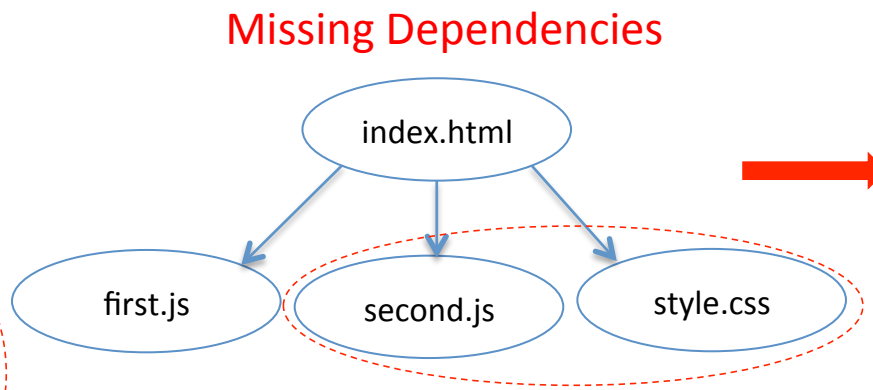


# Dependency Graphs



**True Dependency Graph**

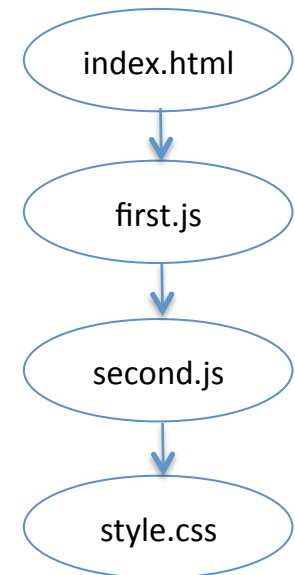
3 RTTs



**Lexical HTML Dependencies**

```
<script src="x.com/first.js"/>
<script src="x.com/second.js"/>
<link src="x.com/style.css"/>
```

**Conservative Assumptions**

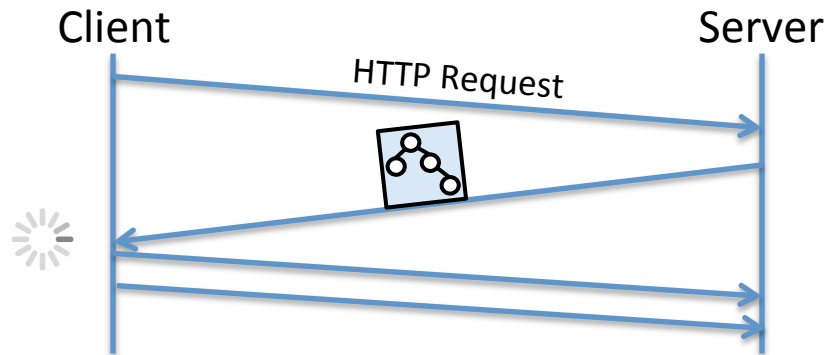


**HTML Tag Order**

4 RTTs

# Outline

- Scout: tracks fine-grained dependencies between page's objects
  - Traditional dependency graphs **miss 30%** of edges
- Polaris: dynamic client-side scheduler written in JavaScript
  - Uses fine-grained dependencies to reduce page load times



- **34% faster (1.3 seconds)** on 12 Mbits/s link with 100 ms RTT

# Scout

- Scout tracks many different dependencies across a page's state

## 3 Types of Dependencies

### Write/Read

**first.js**

```
x = 6;
```

**second.js**

```
y = x + 5;
```

### Read/Write

**first.js**

```
x = [1,3,5];
```

**second.js**

```
y = x.length;
```

**third.js**

```
x.push(7);
```

### Write/Write

**first.js**

```
alert("first message");
```

**second.js**

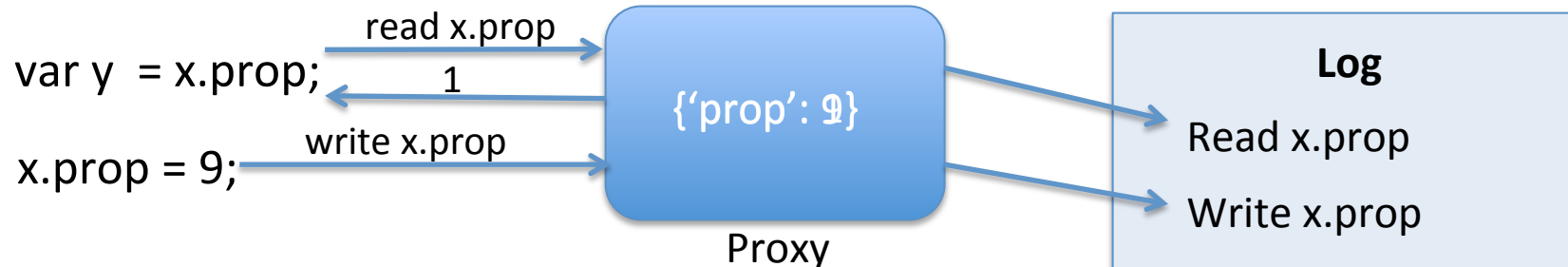
```
alert("second message");
```



# Tracking Dependencies

- JS proxy objects

```
var x = {'prop': 1}; new Proxy({'prop': 1}, log_handlers);
```

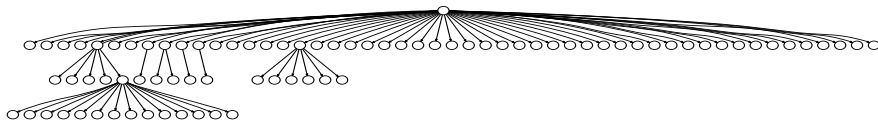


- Many others described in paper
  - Global variables
  - Recursive proxying (e.g., x.y.z)
  - DOM (e.g., document.getElementById("foo"))

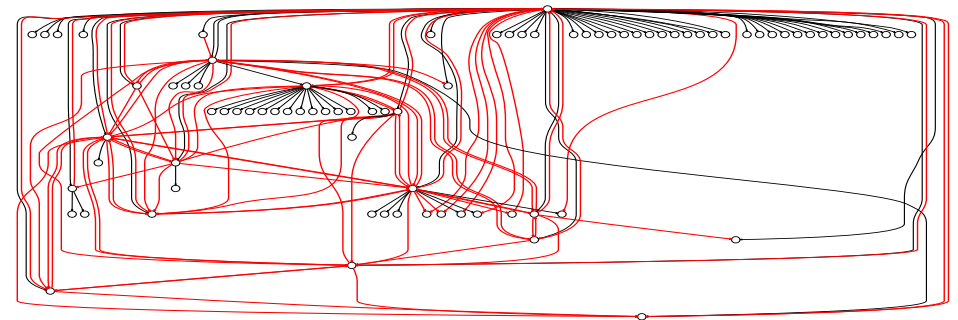
# What we found

- Traditional graphs miss 30% of edges
- 80.8% of pages have altered critical paths

weather.com



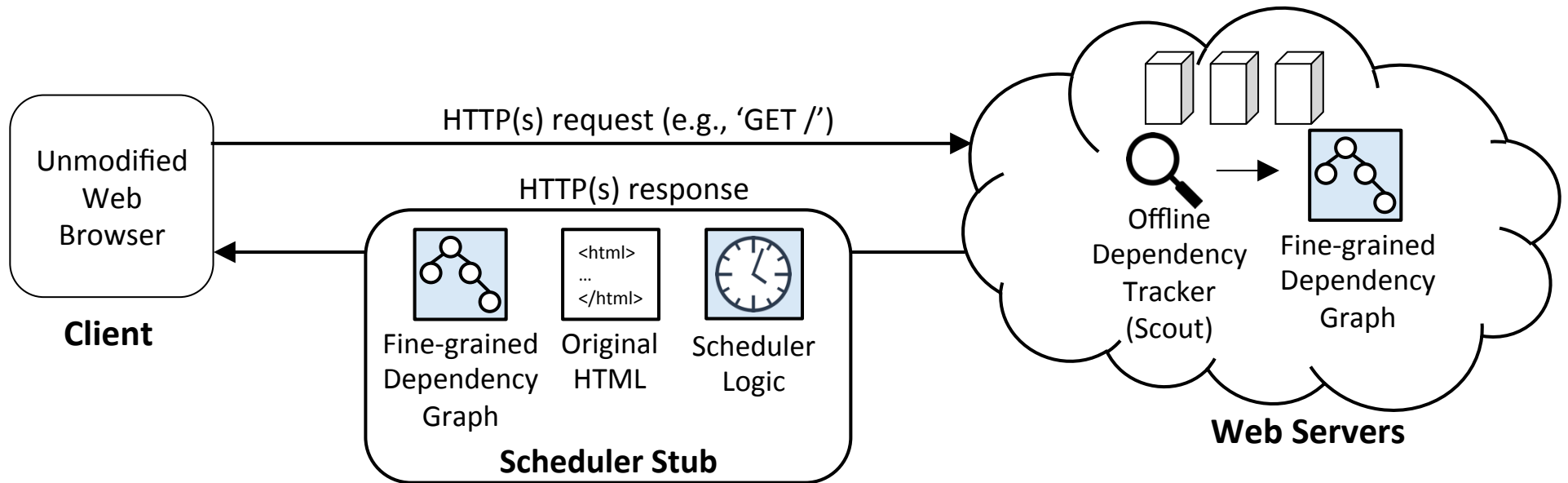
Traditional Dependency Graph



Complete Dependency Graph

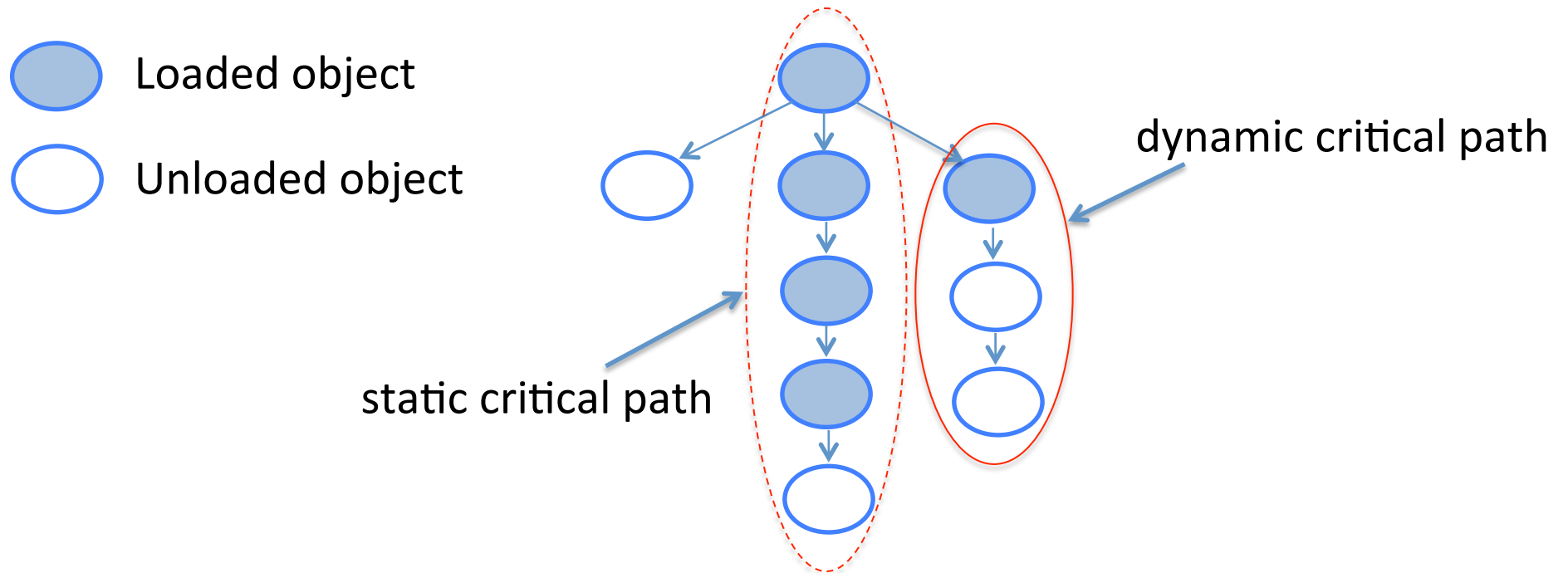
**Adding all edges → removes conservative assumptions → faster page loads!**

# Polaris

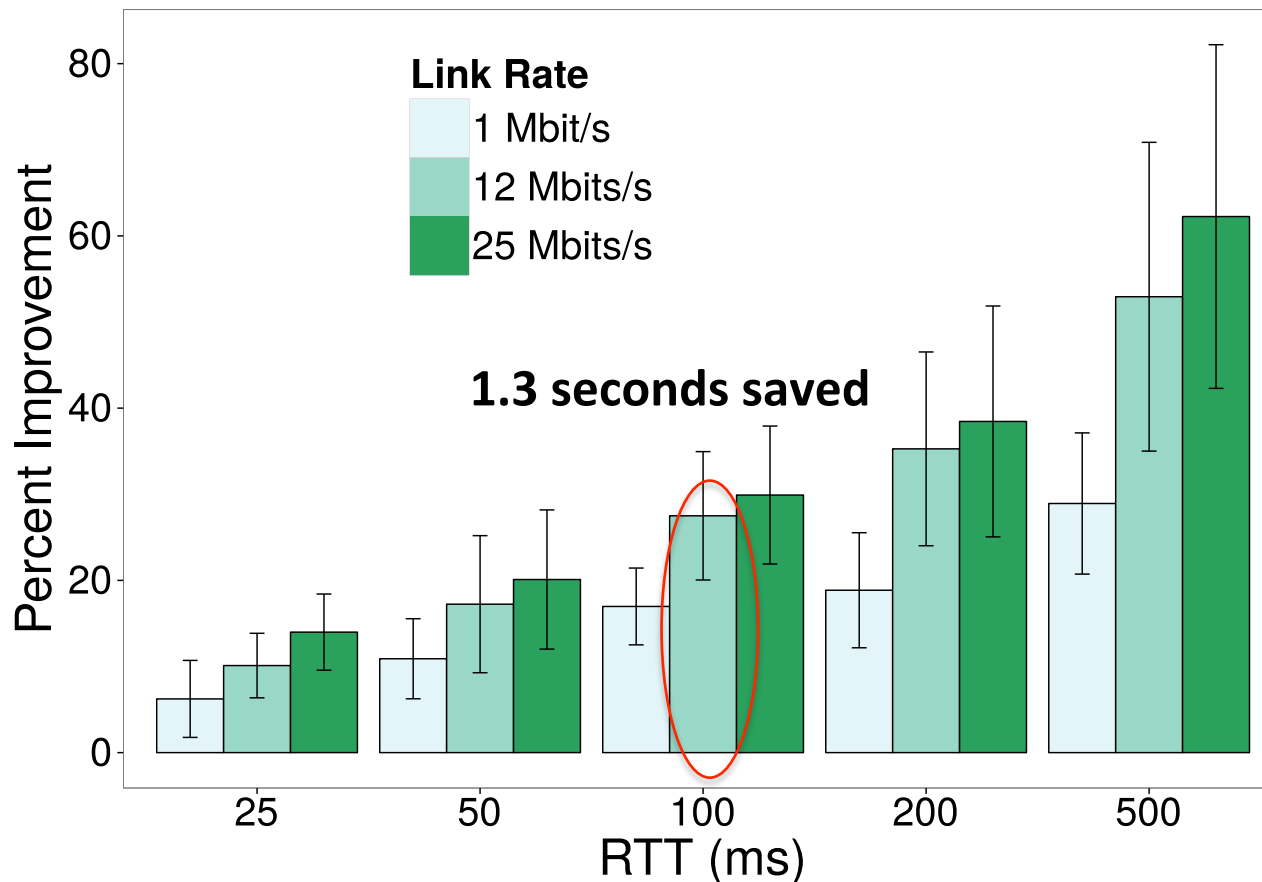


# Request Scheduling with Polaris

Always fetch objects on the **dynamic critical path**

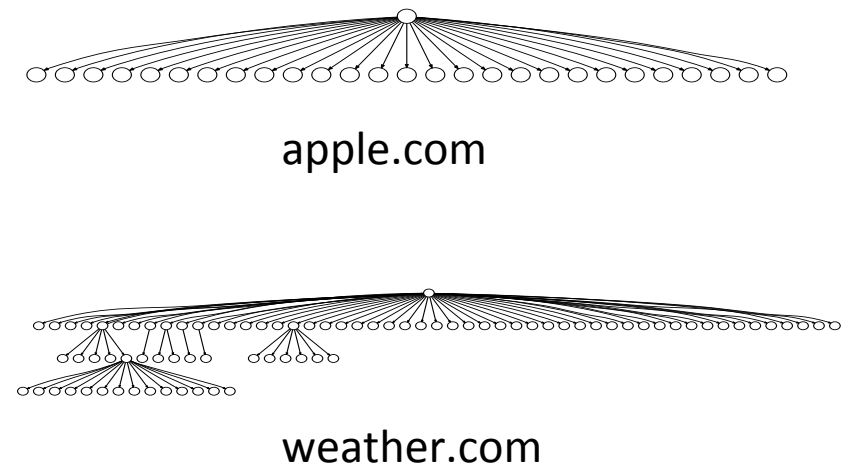
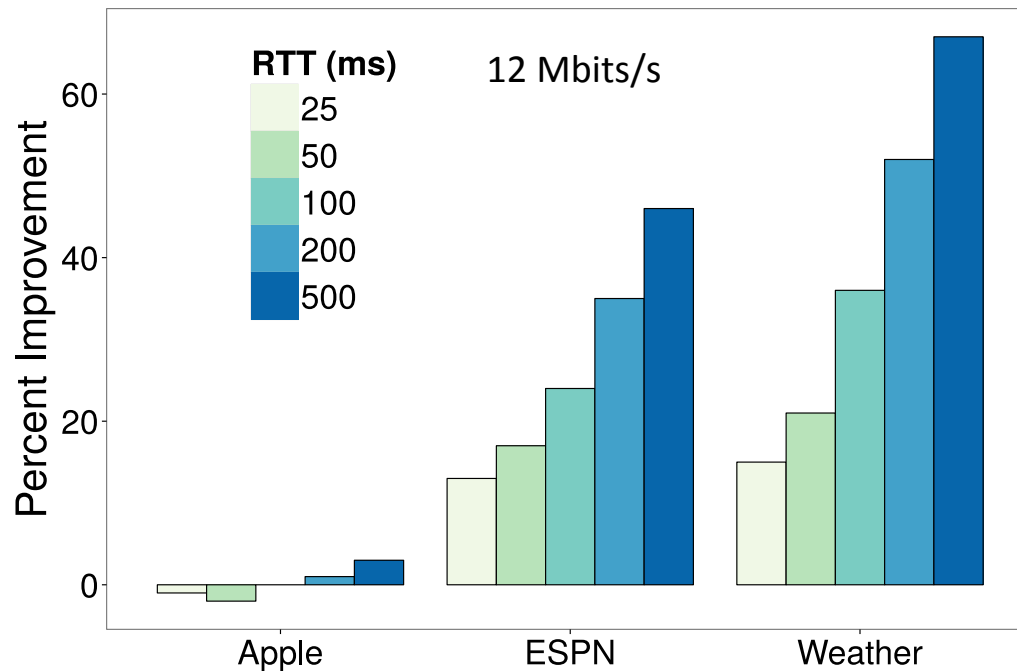


# Evaluating Polaris



- Gains increase with increasing RTT
- Gains increase with increasing link rate
- Baseline is Firefox (which does speculative execution)
- Large error bars: page structure matters too!

# Impact of Dependency Graph Structure



- Apple: scheduling doesn't matter (all requests have same priority)
- Weather: short and long chains so scheduling matters

# Conclusion

- Browsers today are constrained by uncertainty!
  - Conservative assumptions lead to higher page load times
  - Klotski (NSDI '15), WProf (NSDI '13), Browser Developer Tools
- Scout: tracks fine-grained dependencies between page's objects
  - Prior dependency graphs **miss 30%** of edges
- Polaris: dynamic client-side scheduler written in JavaScript
  - Uses Scout's fine-grained dependencies to reduce page load times
  - **34% faster (1.3 seconds)** on 12 Mbits/s link with 100 ms RTT

<http://web.mit.edu/polaris>

polarisweb@mit.edu