Promise of Push

HTTP/2 Web Performance

@ColinBendell
Director, CTO Office

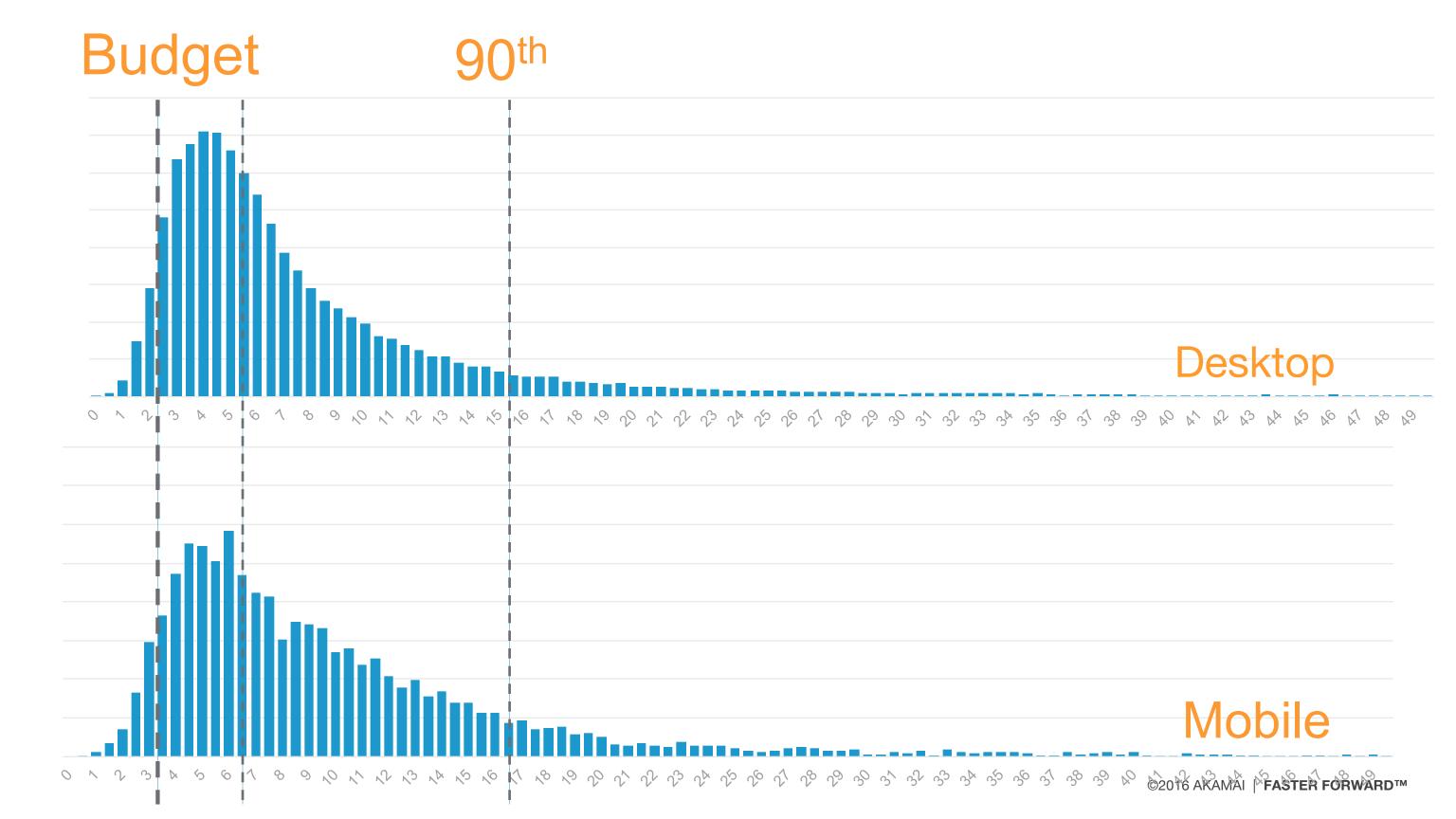
Hi. I'm Colin

- Dad,
- Coder,
- Runner,
- (part crazy)



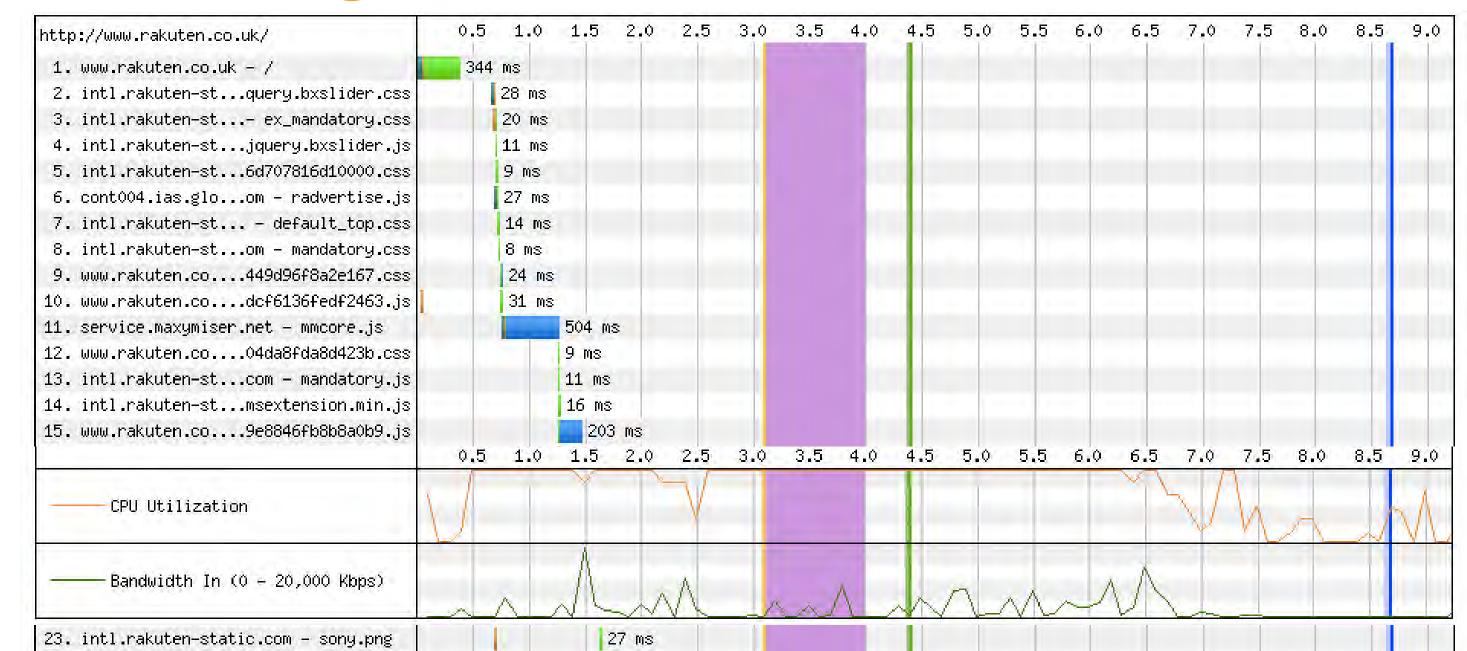


Chapter 1: What is (h2) PUSH?



WebPageTest

Looking for culprits



#PerfMatters

Perf 101

▼ Wait

CachingProtocol optRoute opt.Sharding

Use time

machine

▼ Bits

gzip/br
Resp. img
jp2 /webp
webfonts

Reqs

webpack / concatspriting / inlininglazyload

Render

Critical CSSjavascriptAMP

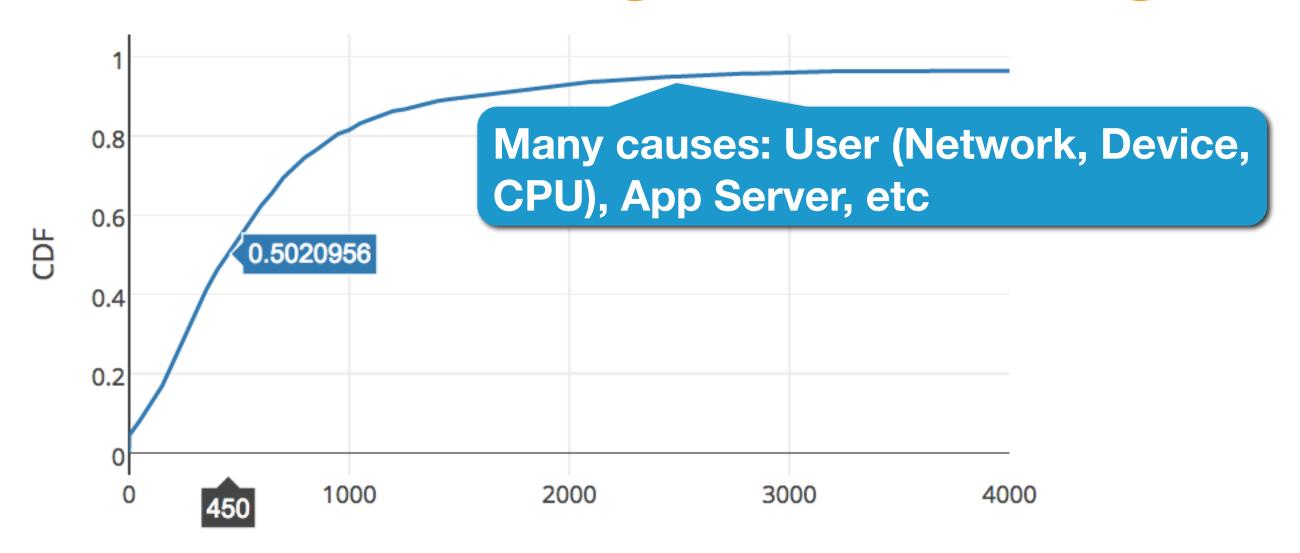
Chroma
Sub-Sampling





Time-To-First-Byte

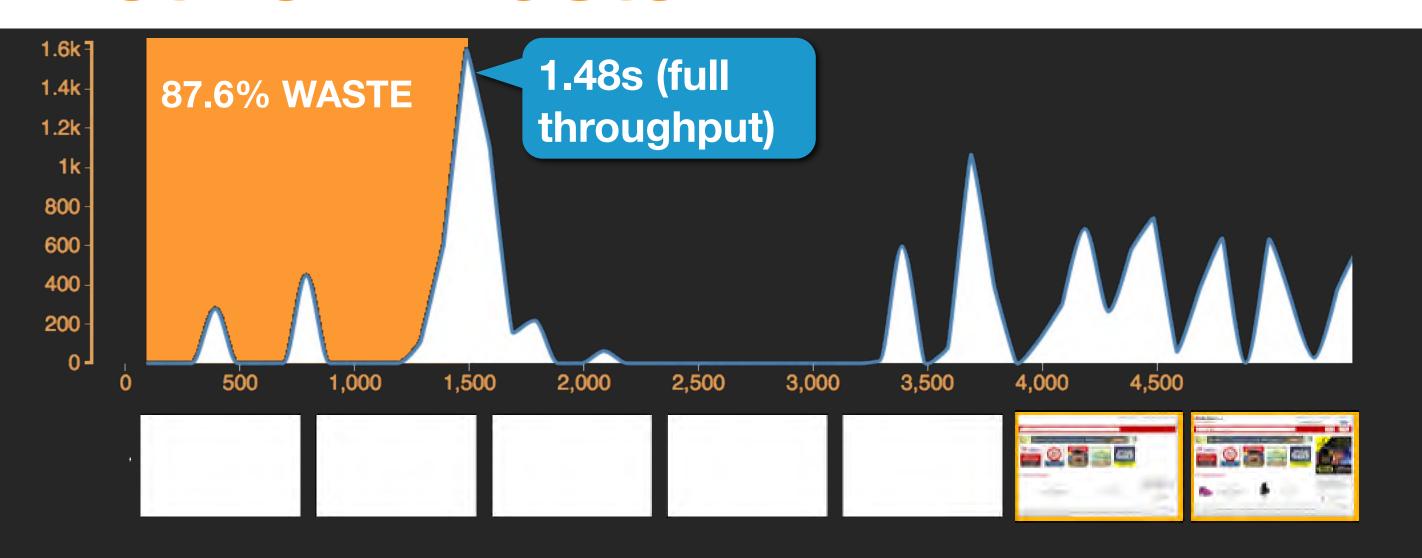
25% of Perf. Budget is waiting



Time To First Byte of base page (ms)

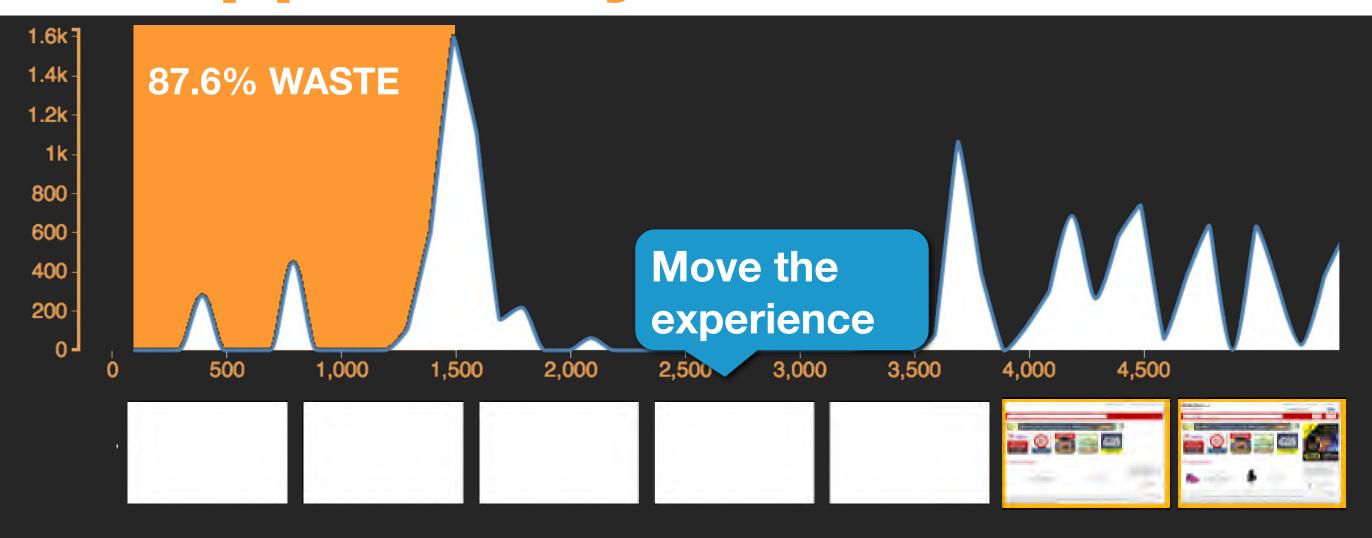
Introducing

Network Waste



Network Waste

An opportunity to PUSH content



PUSH_PROMISE

When you get it Right:







Don't we already have PUSH?

How to PUSH

Polling / Long Polling

```
GET /messages HTTP/1.1
Host: queue.example.com
Authorization: ...
```

```
HTTP/1.1 200 OK
Content-Length: 100
{ "from": "adbram@montague.net",
  "to": "sampson@capulet.com",
  "message: "Do you bite your
             thumb at us, sir?
```

One-to-One with

WebSockets

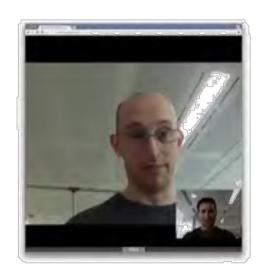
Not cacheable; meant for special apps, not webpages

```
var ws = new WebSocket('wss://example.com/socket');
ws.onmessage = function(msg) {
  if(msg.data instanceof Blob) {
    processBlob(msg.data);
  } else {
    processText(msg.data);
function sendMessage(msgText) {
  ws.send(msgText);
```

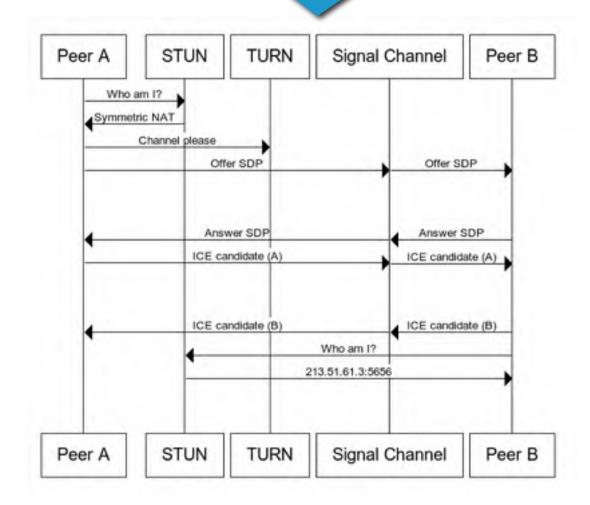
Peer-to-Peer with

WebRTC

Great for video chat, not webpages







How to PUSH

Server-Side-Events

```
GET /messages HTTP/1.1
Host: api.example.com
Accept: text/event-stream
Authorization: ...
Last-Event-ID: 41
```

```
HTTP/1.1 200 OK
Connection: keep-alive
Content-Type: text/event-stream
Transfer-Encoding: chunked
id: 42
event: receiveMessage
data: { "from": "adbram@montague.net",
  "to": "sampson@capulet.com",
  "message: "Do you bite your
             thumb at us, sir?}
id: 43
event: stageAction
data: { "from": "adbram@montague.net",
  "to": "sampson@capulet.com",
  "message: "Do you bite your
             thumb at us, sir?}
                                          RWARD™
```

How to PUSH

Server-Side-Events

Close – but realtime & text only (not resources)

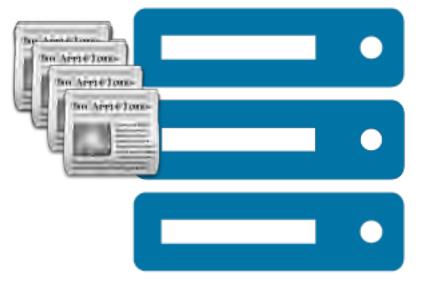
```
HTTP/1.1 200 OK
Connection: keep-alive
Content-Type: text/event-stream
Transfer-Encoding: chunked
id: 42
event: recieveMessage
data: { "from": "adbram@montague.net",
  "to": "sampson@capulet.com",
  "message: "Do you bite your
             thumb at us, sir?}
id: 43
event: stageAction
data: { "from": "adbram@montague.net",
  "to": "sampson@capulet.com",
  "message: "Do you bite your
             thumb at us, sir?}
```

```
<script type="text/javascript">
var evtSource = new EventSource("/messages");
evtSource.addEventListener("receiveMessage",
  function(e) {
    var newElement = document.createElement("p");
    var obj = JSON.parse(e.data);
    newElement.innerHTML = "Message: " + obj.message;
    eventList.appendChild(newElement);
  }, false);
</script>
```

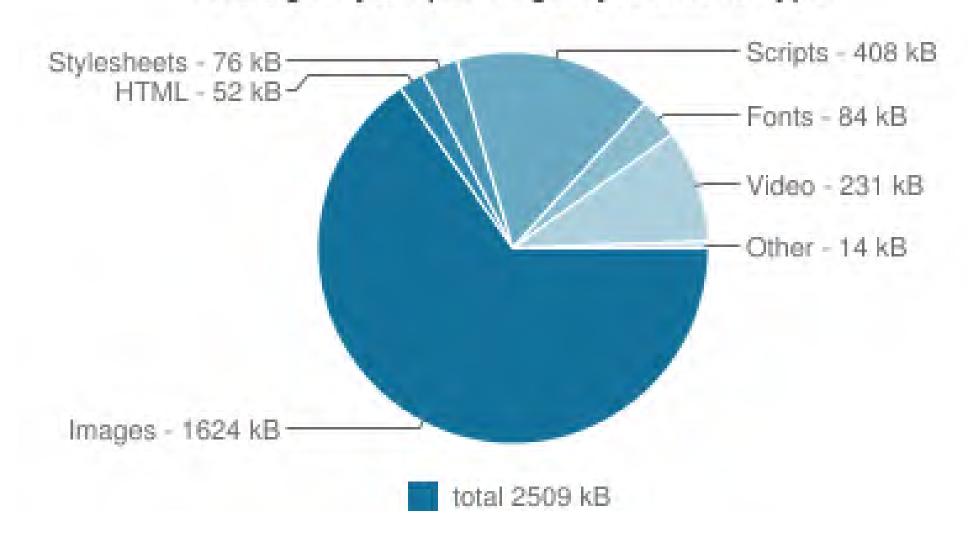


TCP Slow Start



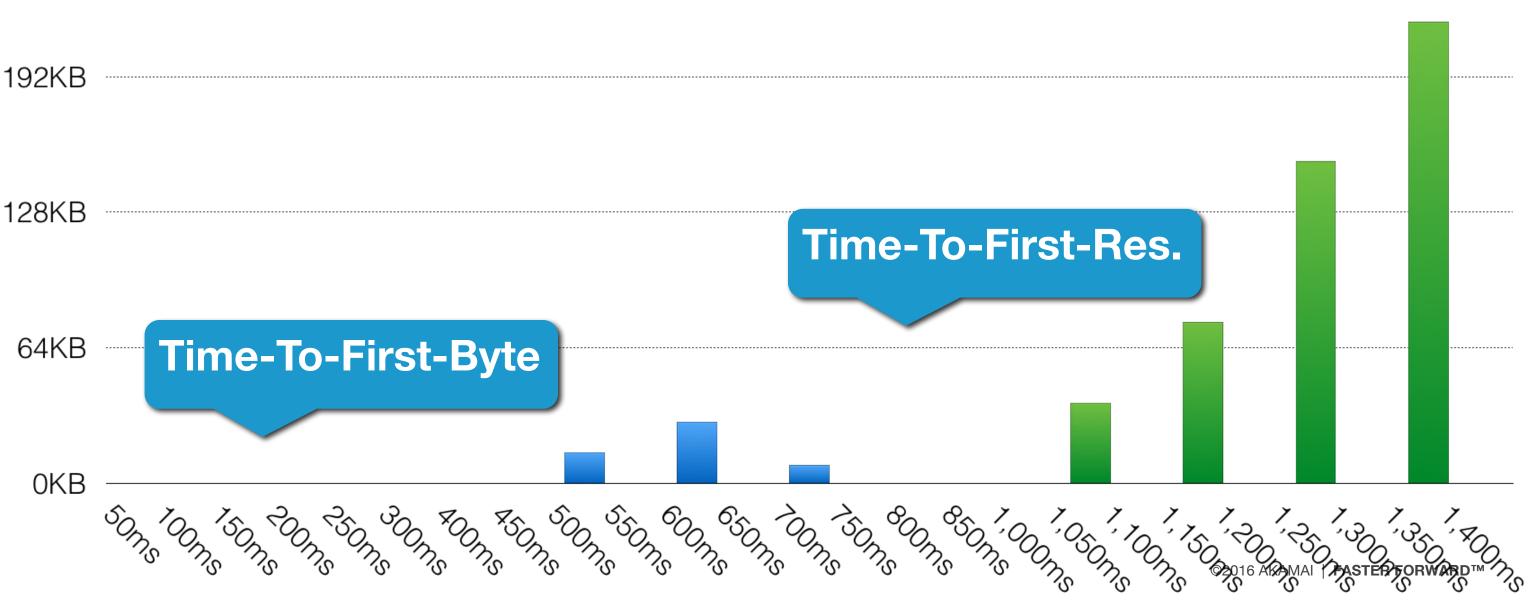


Average Bytes per Page by Content Type



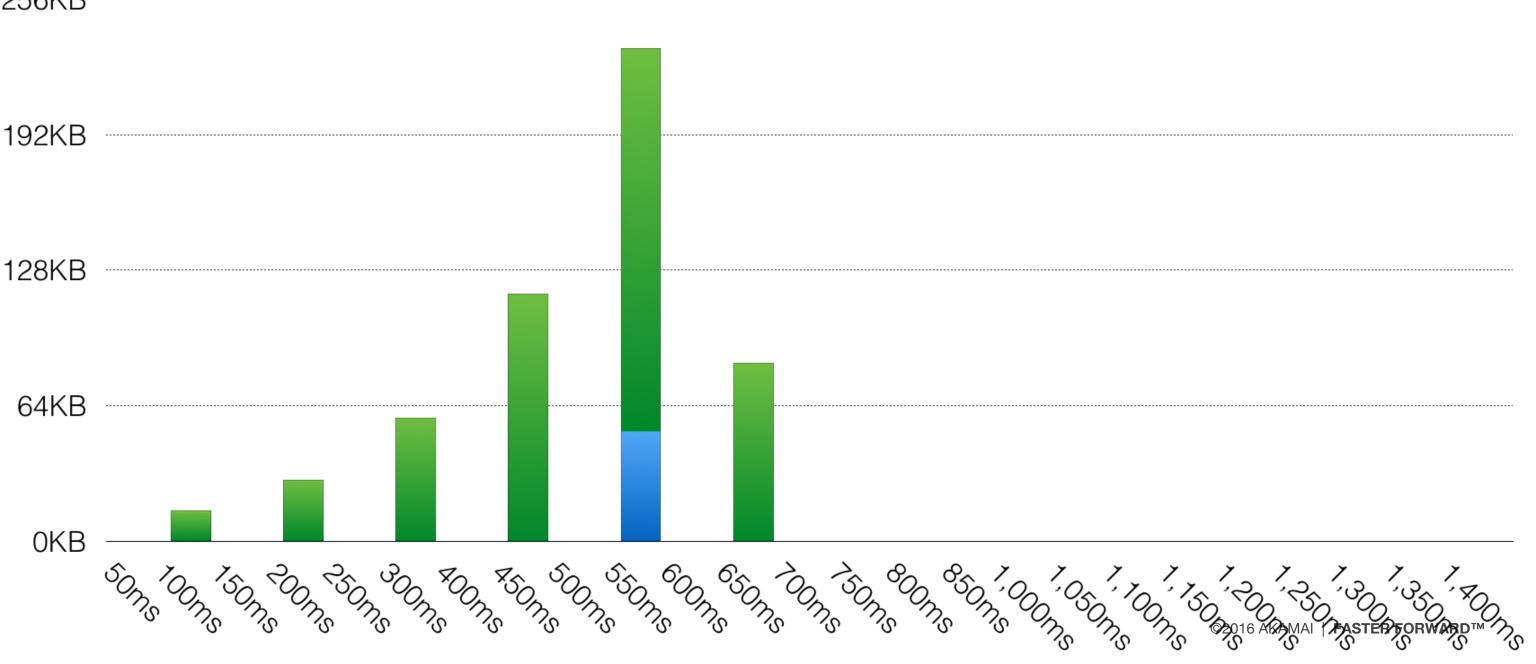
Client Side Receive Window

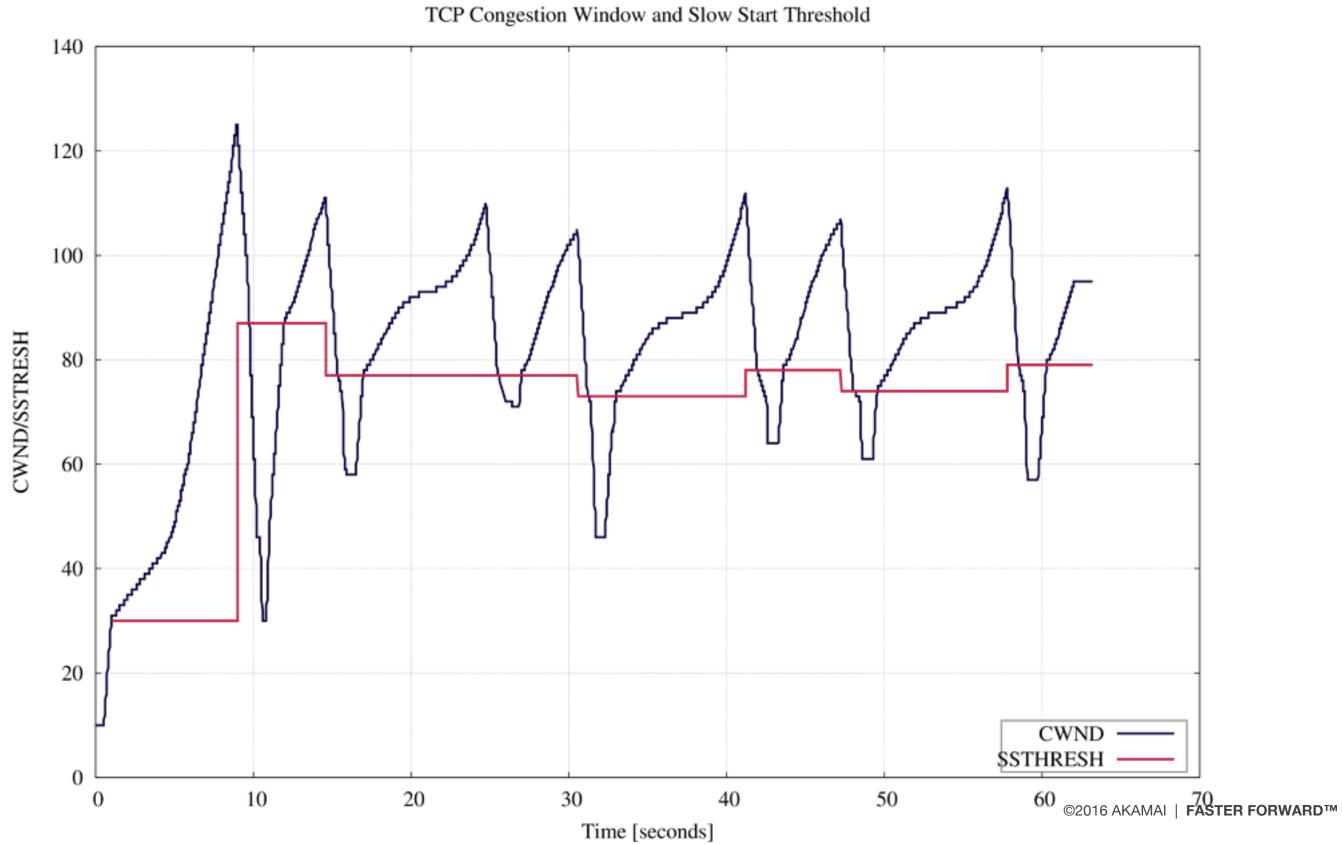
20mbps@100ms RTT



HTTP/2 PUSH:

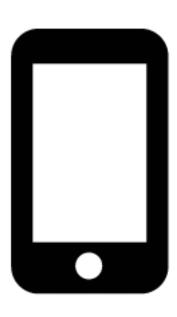
Warm the TCP Connection





HTTP/2 PUSH_PROMISE

Frame Basics



SETTINGS

ENABLE_PUSH: 0

HEADERS

:method: GET

. . .

RST_STREAM

PUSH_PROMISE

:url: /push.css

user-agent: chrome



:status: 200

• •



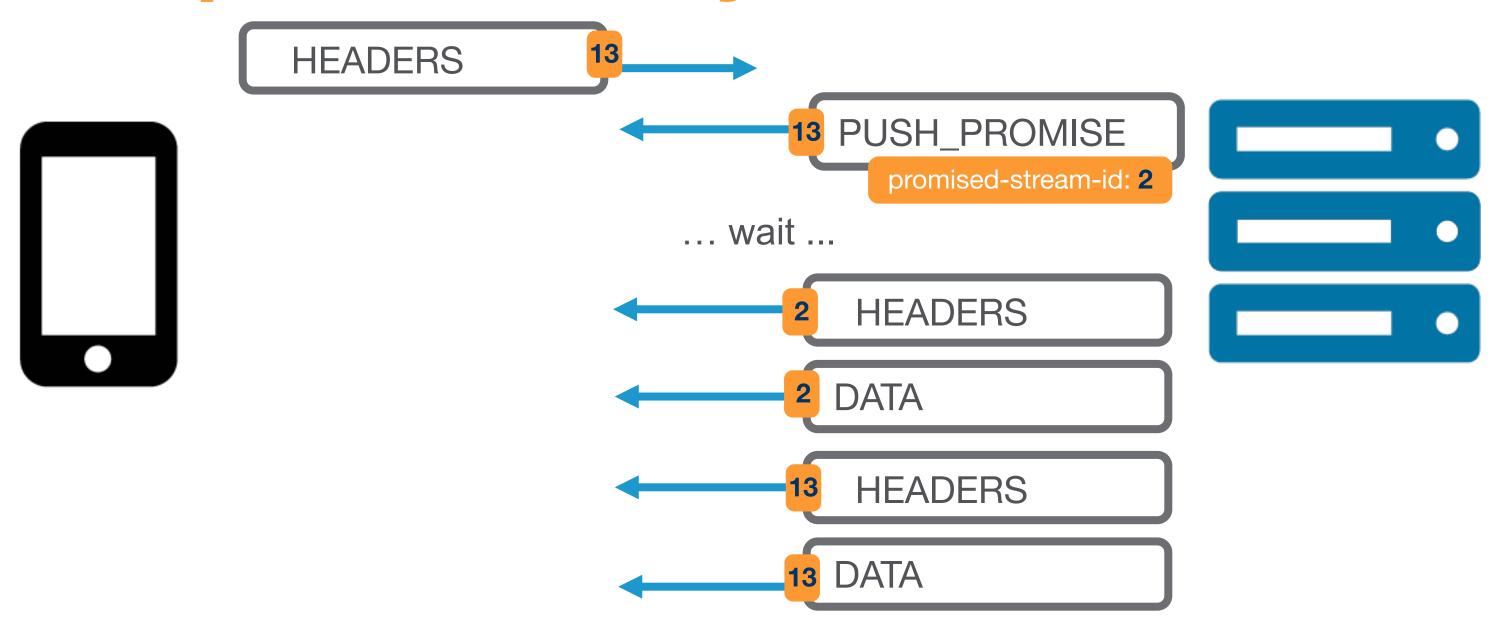
DATA

a:hover {color: blue}

. . .

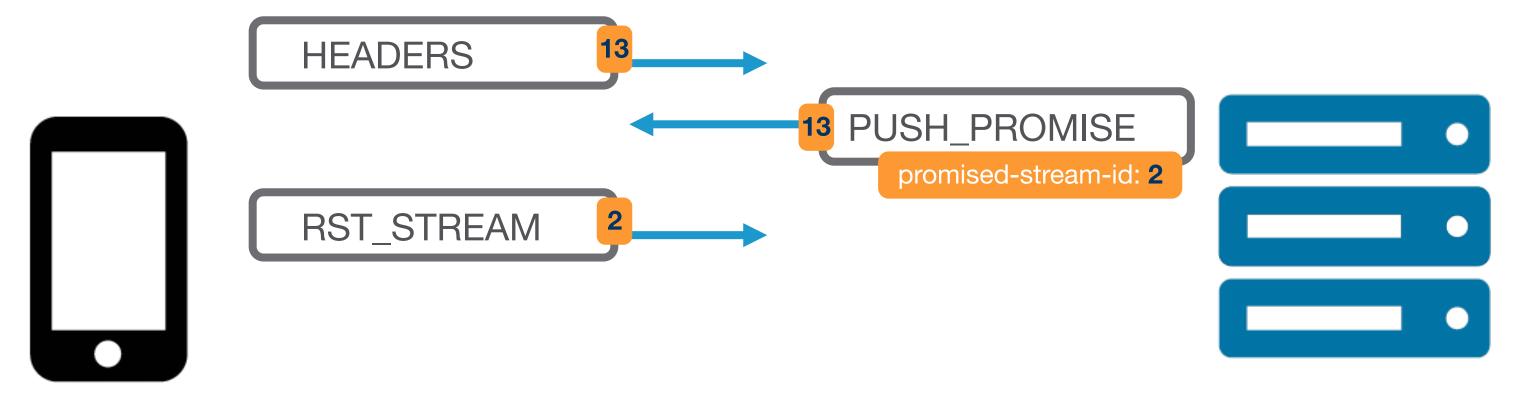
HTTP/2 PUSH_PROMISE

Expected Lifecycle



HTTP/2 PUSH_PROMISE

Expected Lifecycle





HTTP/2

PUSH_PROMISE

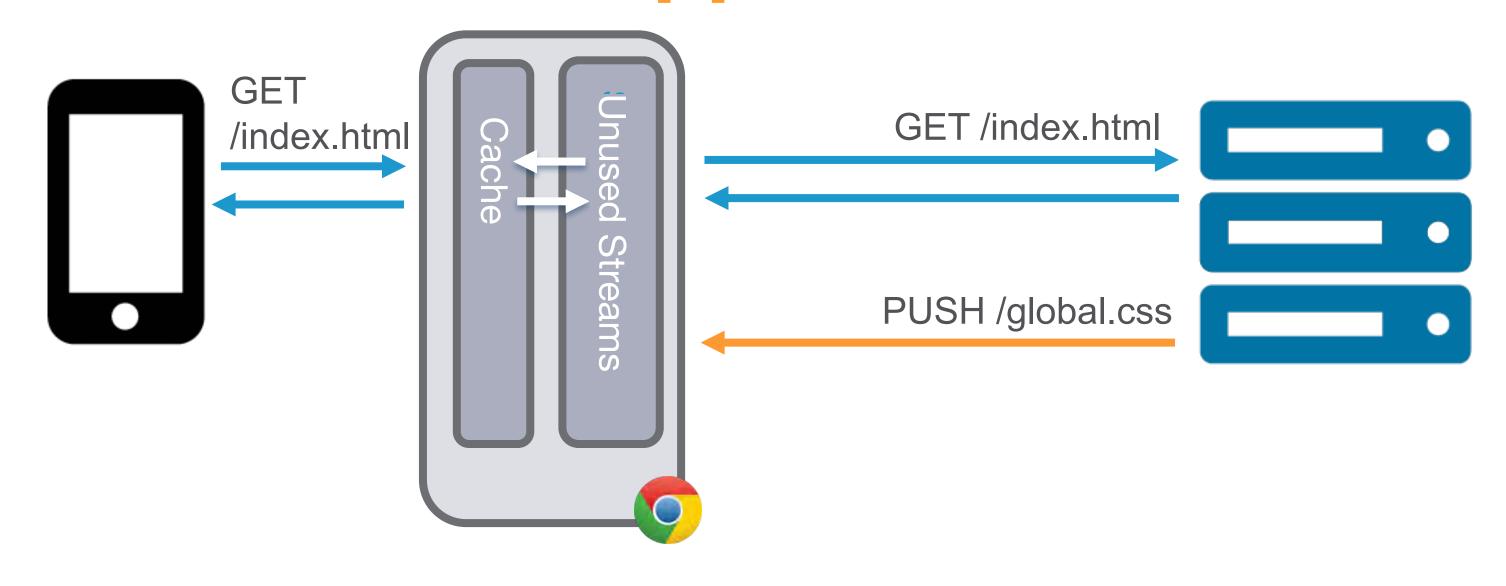
Push not accepted

```
0.230] send RST_STREAM frame <length=5, flags=0x07, stream_id=4>
```

Chapter 2: Can i Push?

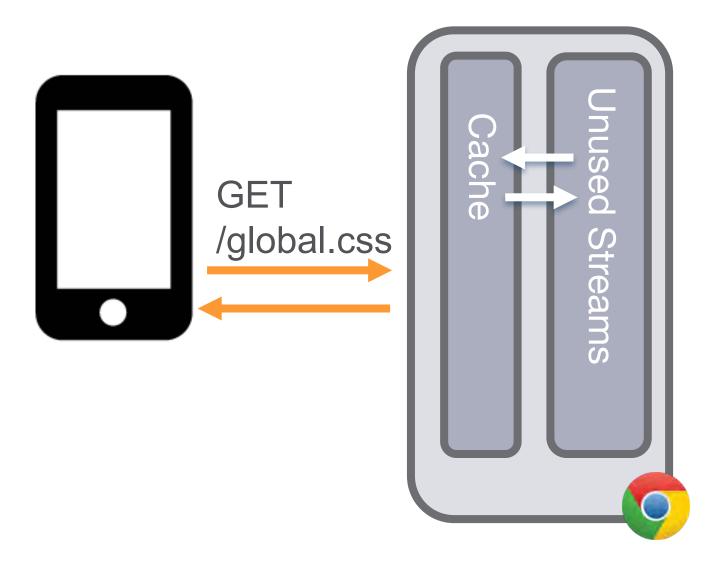
HTTP/2 PUSH

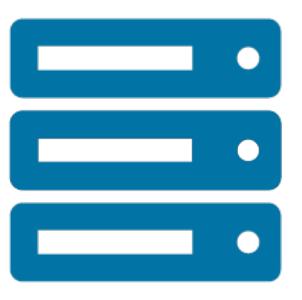
Browsers & Apps

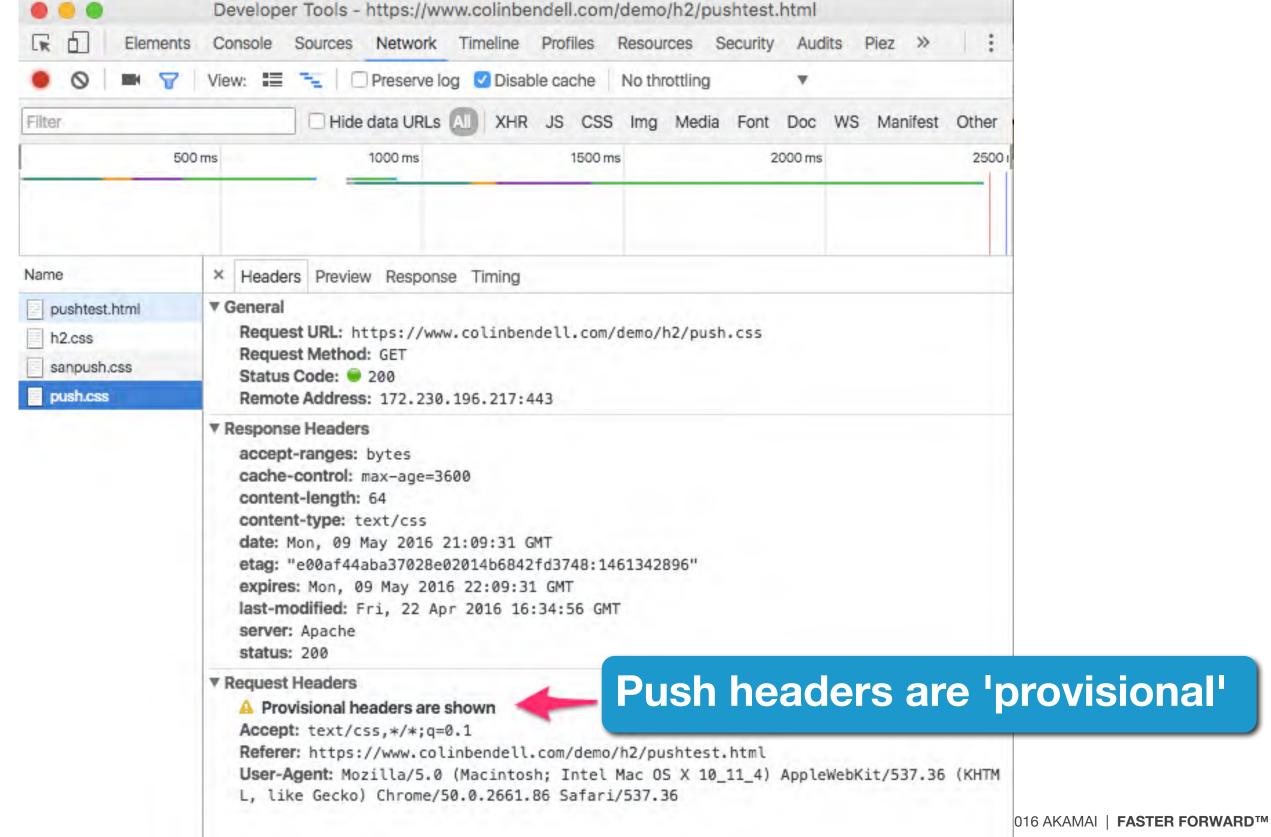


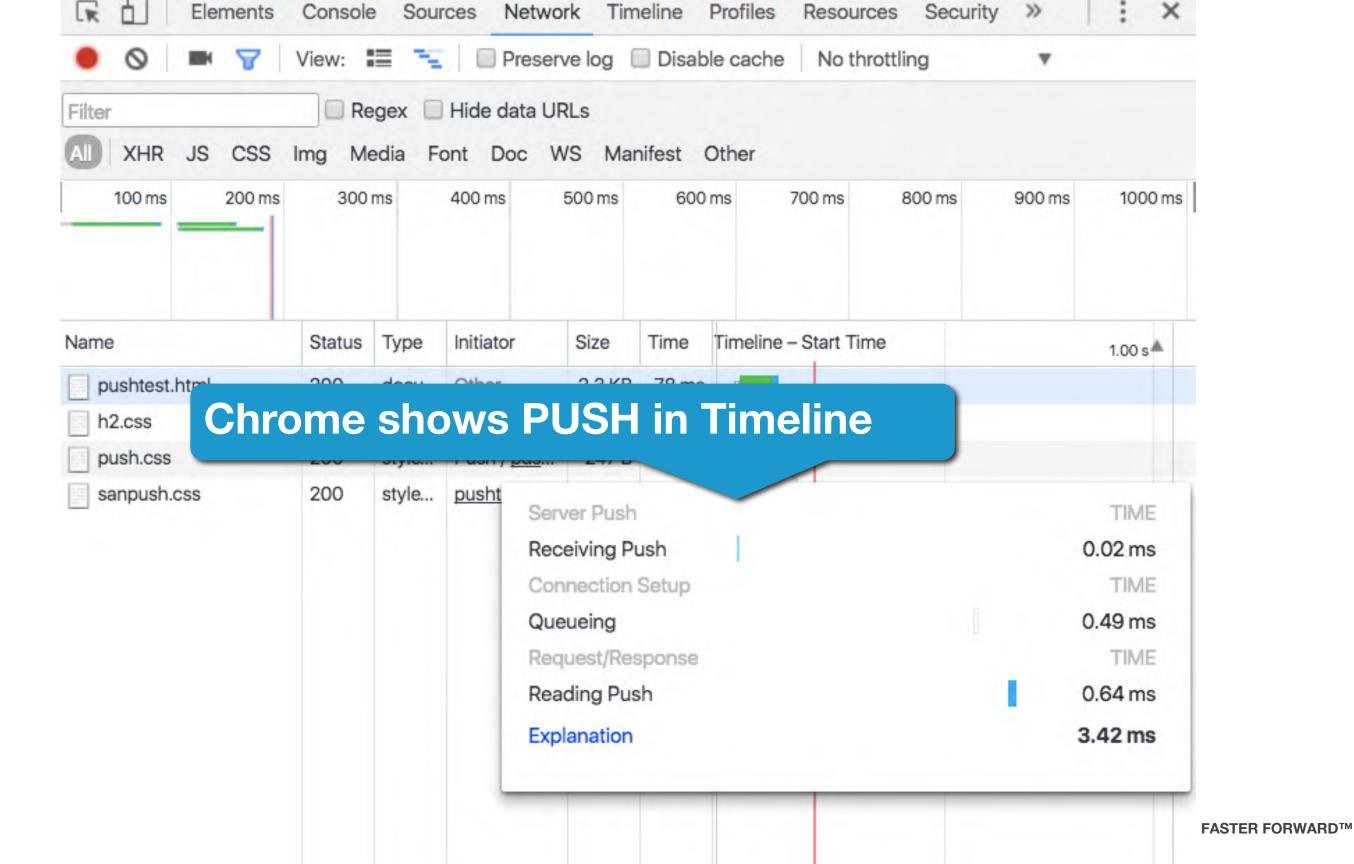
HTTP/2 PUSH in a Secret Cache / Queue

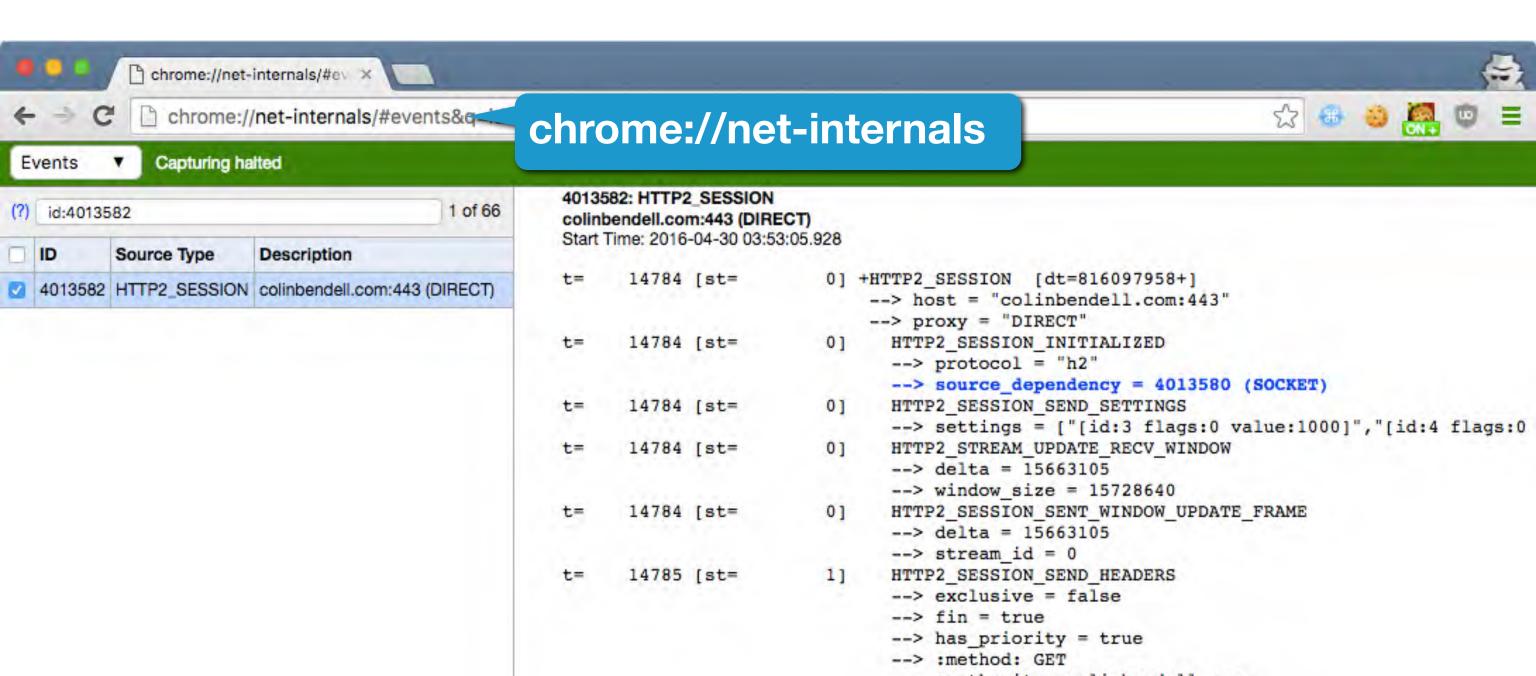
Browsers & Apps











HTTP/2 (Chrome)

Adopted Push Stream

Adopted pushes are logged (not a frame)

chrome://net-internals to find unclaimed pushes

Host	Proxy	ID	Protocol Negotiated	Active streams	Unclaimed pushed	Max	Initiated	Pushed	Pushed and claimed	Abandoned	Received frames	Secu
1,bendell.ca:443	direct://	4052647	h2	0	0	100	1	0	0	0	2	true
www.colinbendell.com:443	direct://	4052633	h2	1	1	100	2	3	0	0	18	true

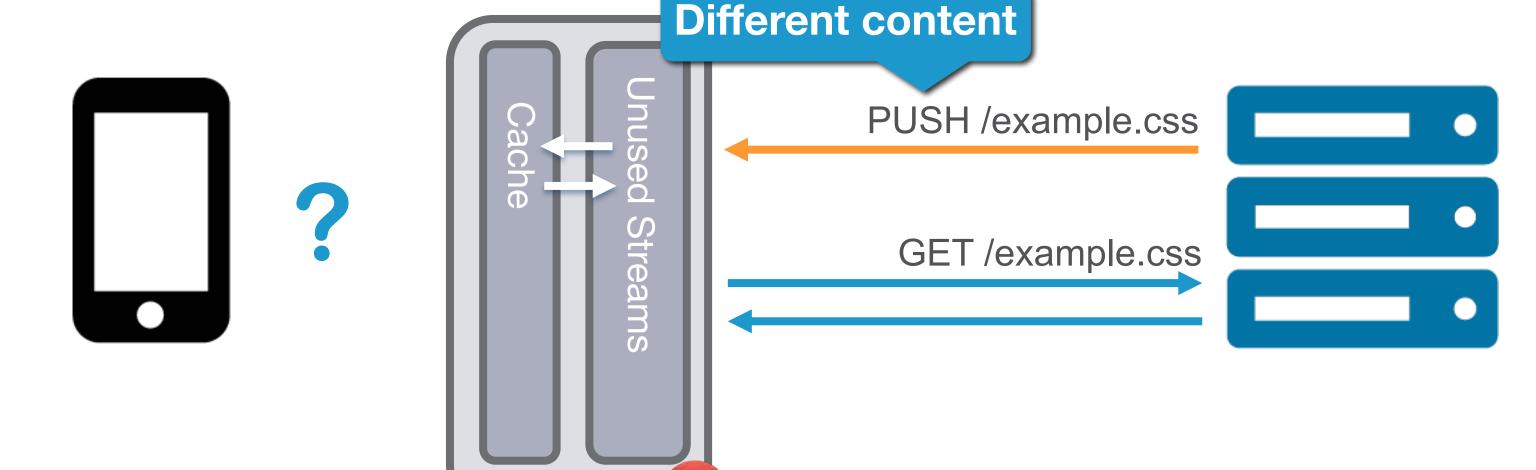
^{*} PUSH is kept in memory for 5 minutes and then get deleted if unclaimed

Does the browser behave as I expect?

Announcing: canipush.com



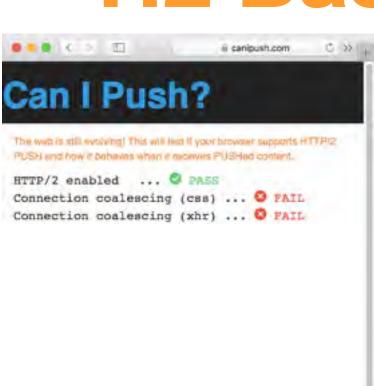
Browser Test Harness

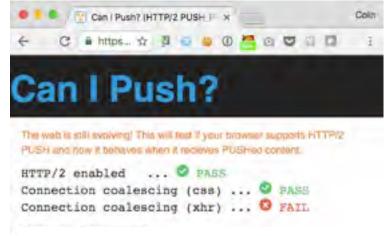


H2 Basics: Coalescing

https://canipush.com/?test=1 https://canipush.com/?test=2

H2 Basics: Coalescing











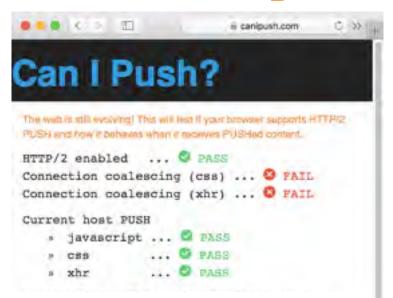




Simple Push

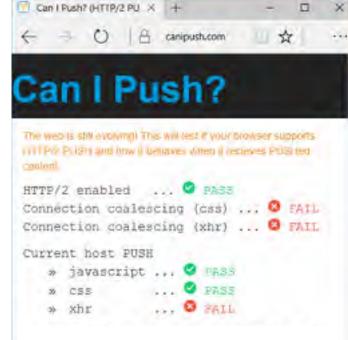
https://canipush.com/?test=3

Simple Push











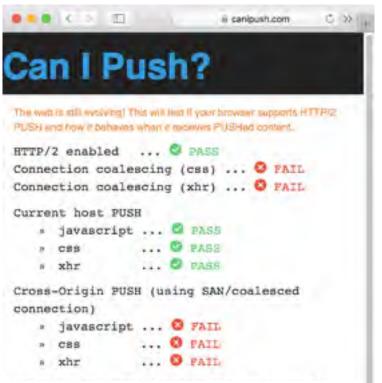


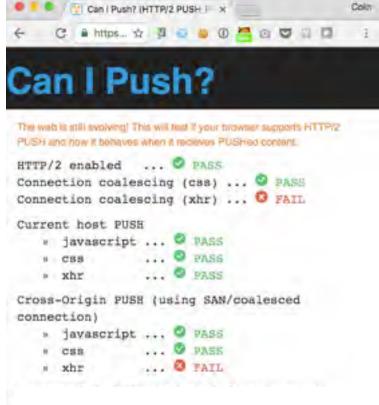


Sharded Resource (Same Cert)

https://canipush.com/?test=4

Sharded Resource (Same Cert)













Should | Push?

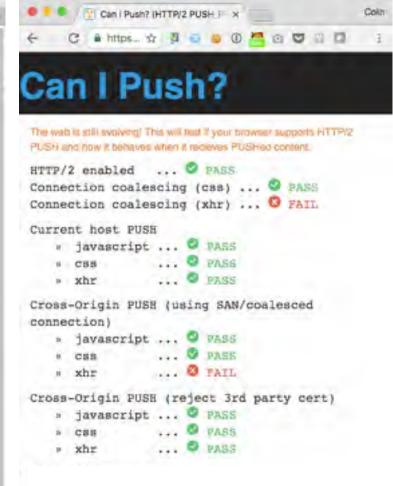


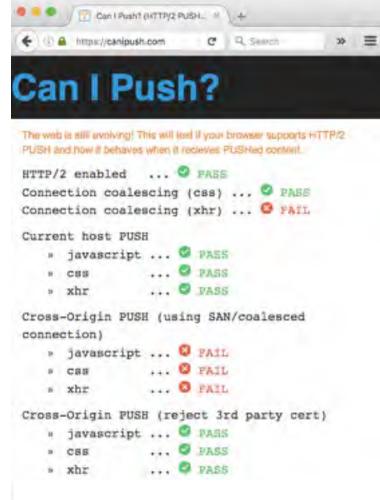
Sharded Resource (Diff Cert)

https://canipush.com/?test=5

Shard w/ diff cert (should reject)









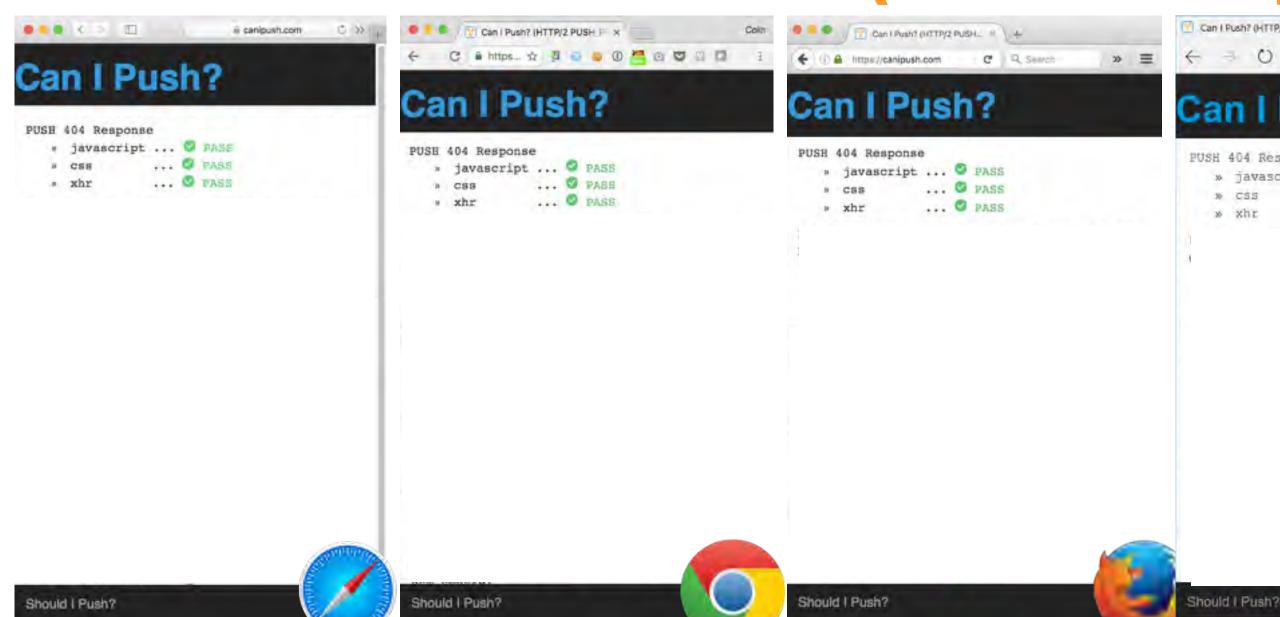




Error pages (eg: 404)

https://canipush.com/?test=6

Shard w/ diff cert (should reject)

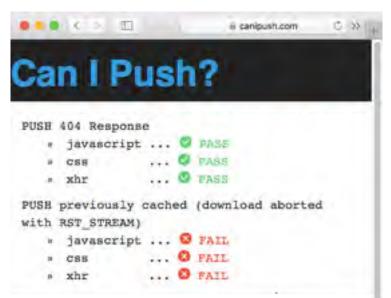


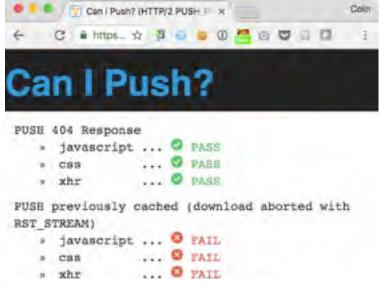


Previously Cached

https://canipush.com/?test=7

Previously Cached















Browser Support

Implementation notes

- Cross Origin XHR are different
- PUSHed content must be pulled (no programatic alerting – use SSE for that)
- FF uses init flow control of 64kb
- Browsers don't (yet) sync caches

What is coming?

Cache Digests for HTTP/2 draft-ietf-httpbis-cache-digest-01

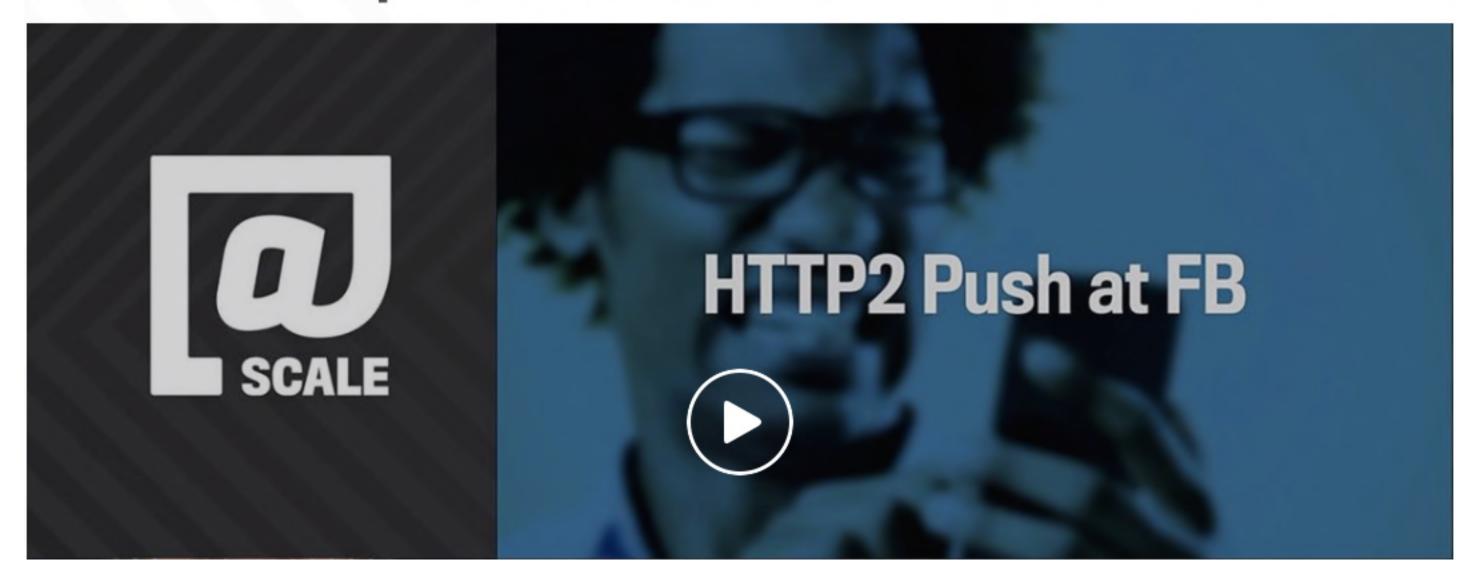
Abstract

This specification defines a HTTP/2 frame type to allow clients to inform the server of their cache's contents. Servers can then use this to inform their choices of what to push to clients.

Note to Readers

Chapter 3: Should i Push?

HTTP2 server push: Lower latencies around the world









Contrary to what the hive mind may be saying, h2 push for the general case will make your site measurably slower not faster.

RETWEETS

LIKES

9











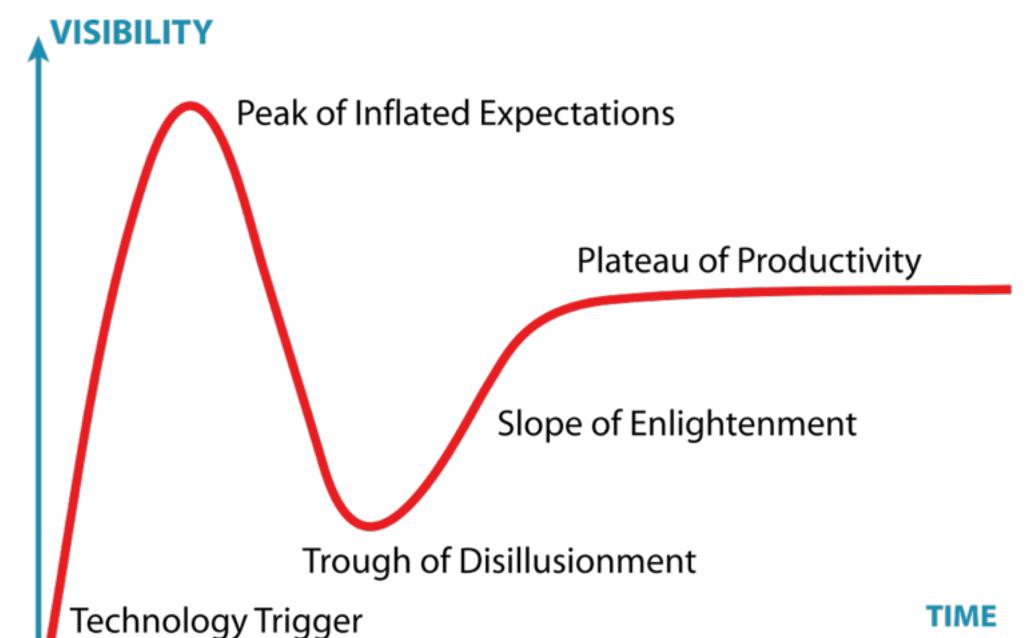


12:17 PM - 26 Oct 2016



Gartner's Hype-Cycle

Where is HTTP/2 PUSH?

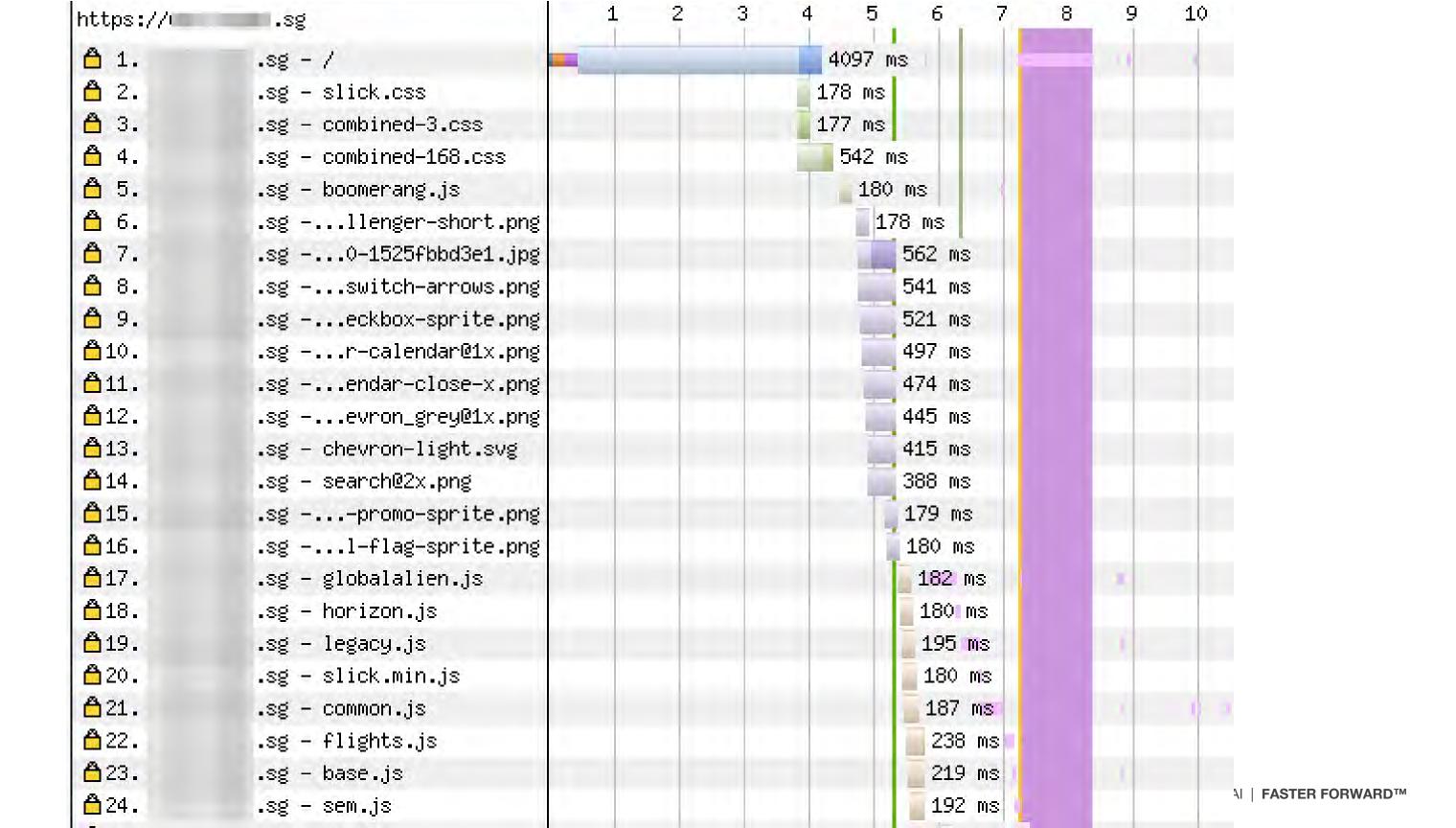


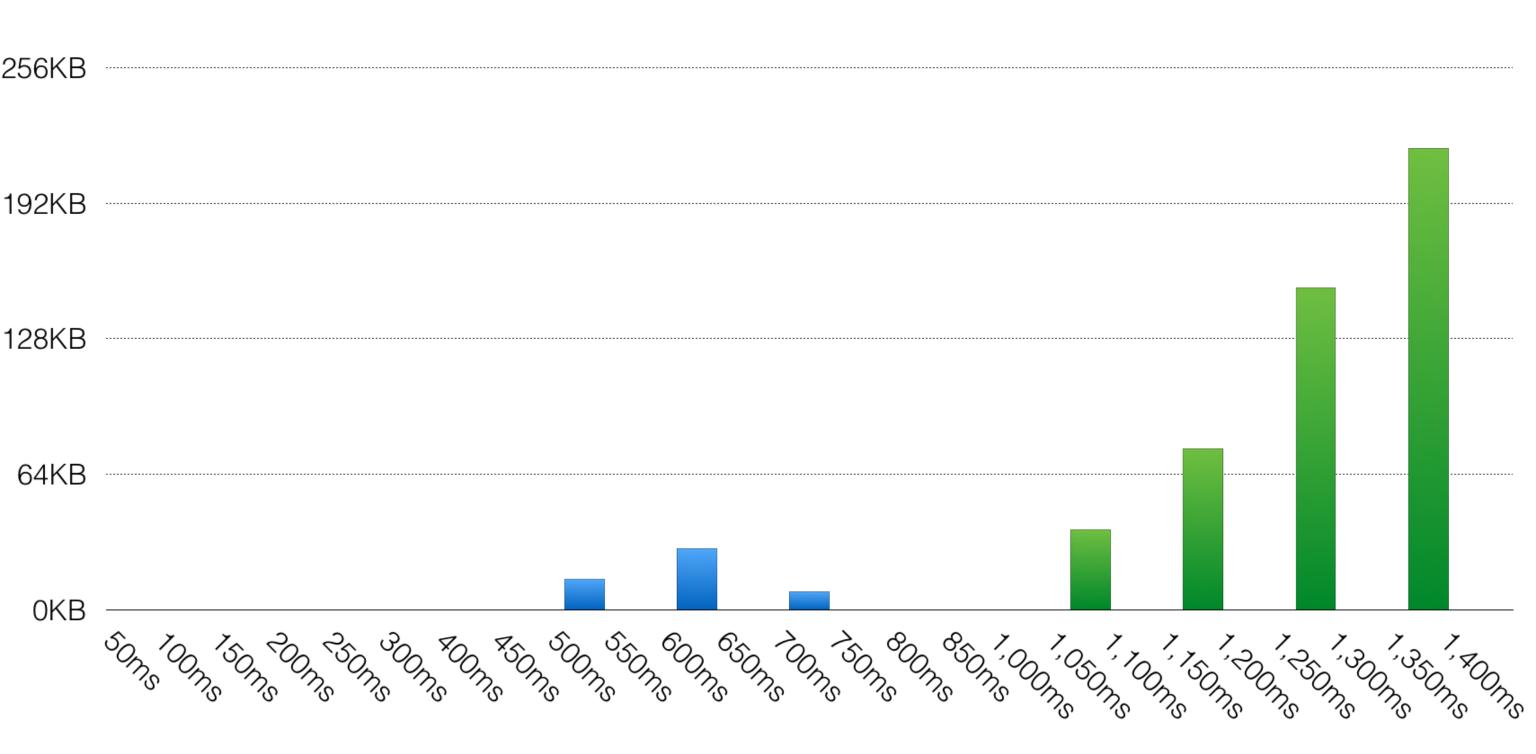
How do you push?

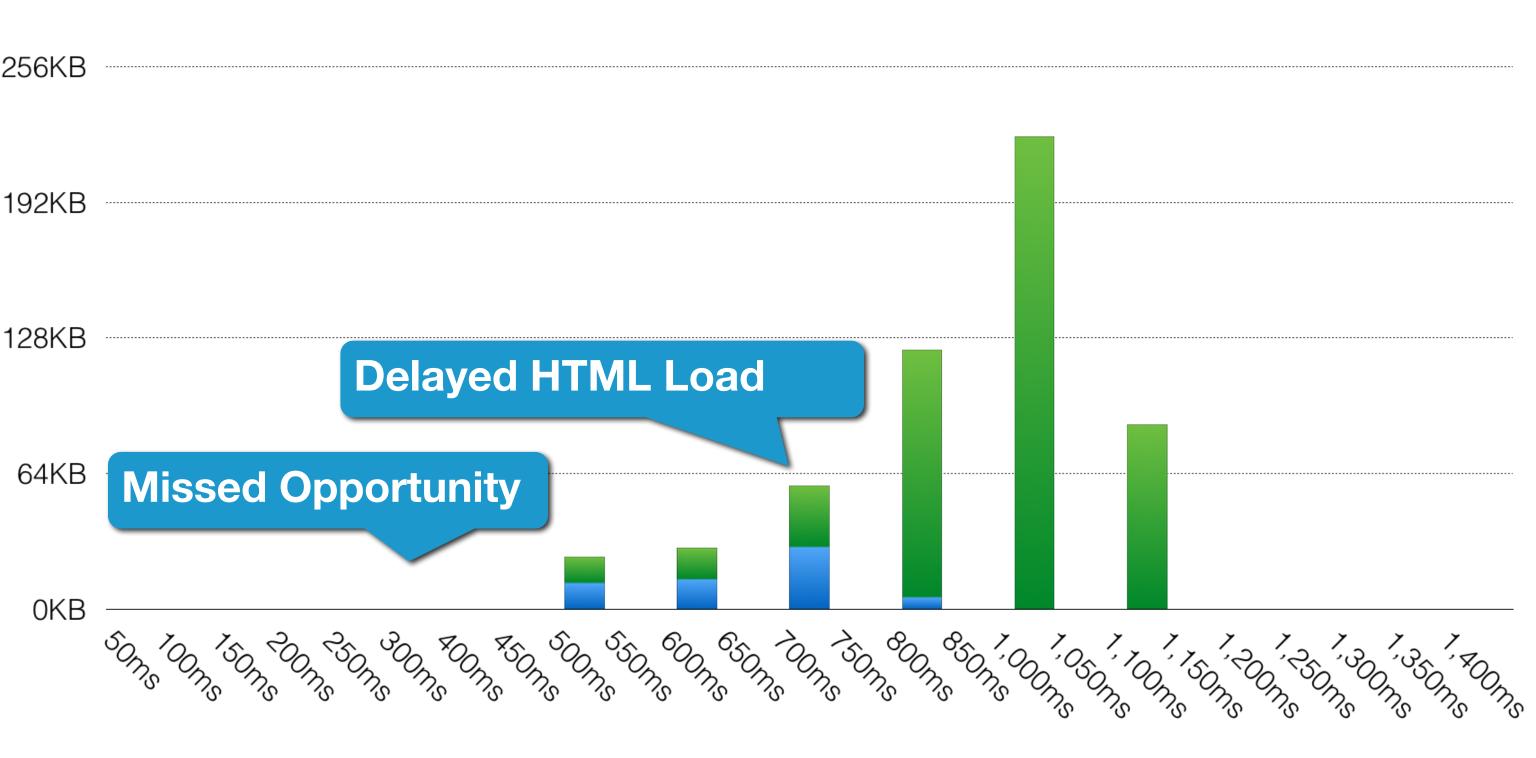
Hint (late) with Headers

Link: <URL>; rel=preload; as=style

```
HTTP/2 200 OK
Content-Type: text/html
Link: /jquery.js; rel=preload; as=script
Link: /global.css; rel=preload; as=style; nopush
```

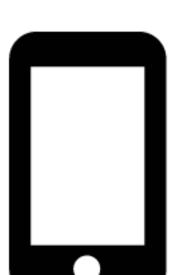






HTTP/2 PUSH_PROMISE

Why can't you send Link early?



HEADERS

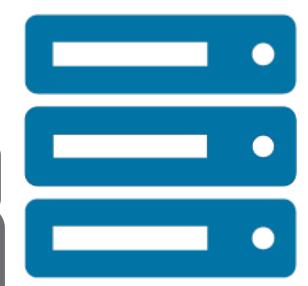
:method: GET

...



:status: 200

Link: ... rel=preload



DATA

a:hover {color: blue}

. . .

Proposal

```
[Docs] [txt pdf] [Tracker] [Email] [Nits]
```

Versions: 00

Network Working Group

Internet-Draft

Intended status: Informational

Expires: May 4, 2017

K. Oku

DeNA Co., Ltd.

October 31, 2016

An HTTP Status Code for Indicating Hints draft-kazuho-early-hints-status-code-00

Abstract

This memo introduces an informational status code for HTTP that can be used for indicating hints to help a client start making preparations for processing the final response.

Status of This Memo

TER FORWARD™

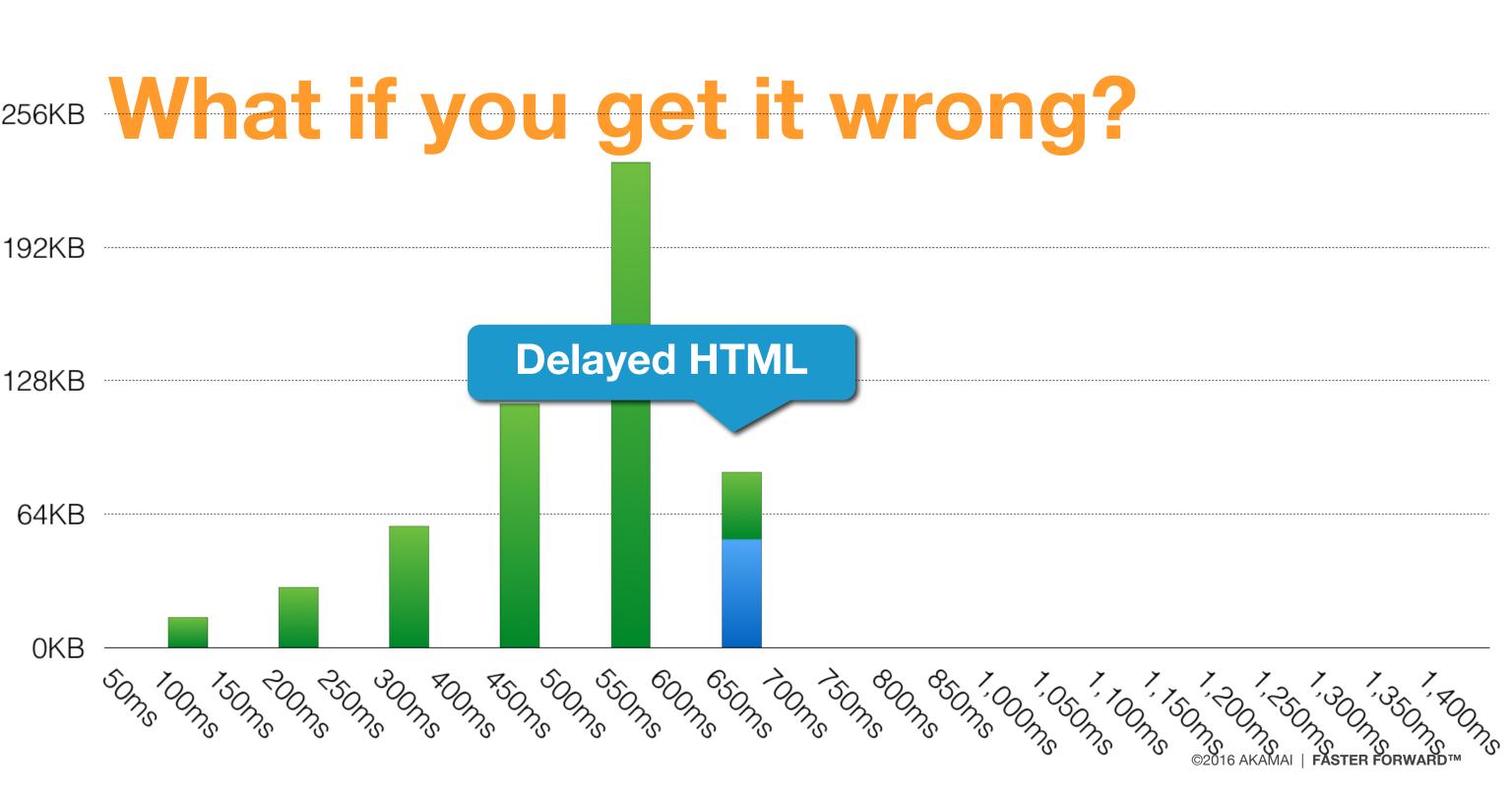
RFC 7540:

The server SHOULD send PUSH_PROMISE (Section 6.6) frames prior to sending any frames that reference the promised responses. This avoids a race where clients issue requests prior to receiving any PUSH_PROMISE frames.

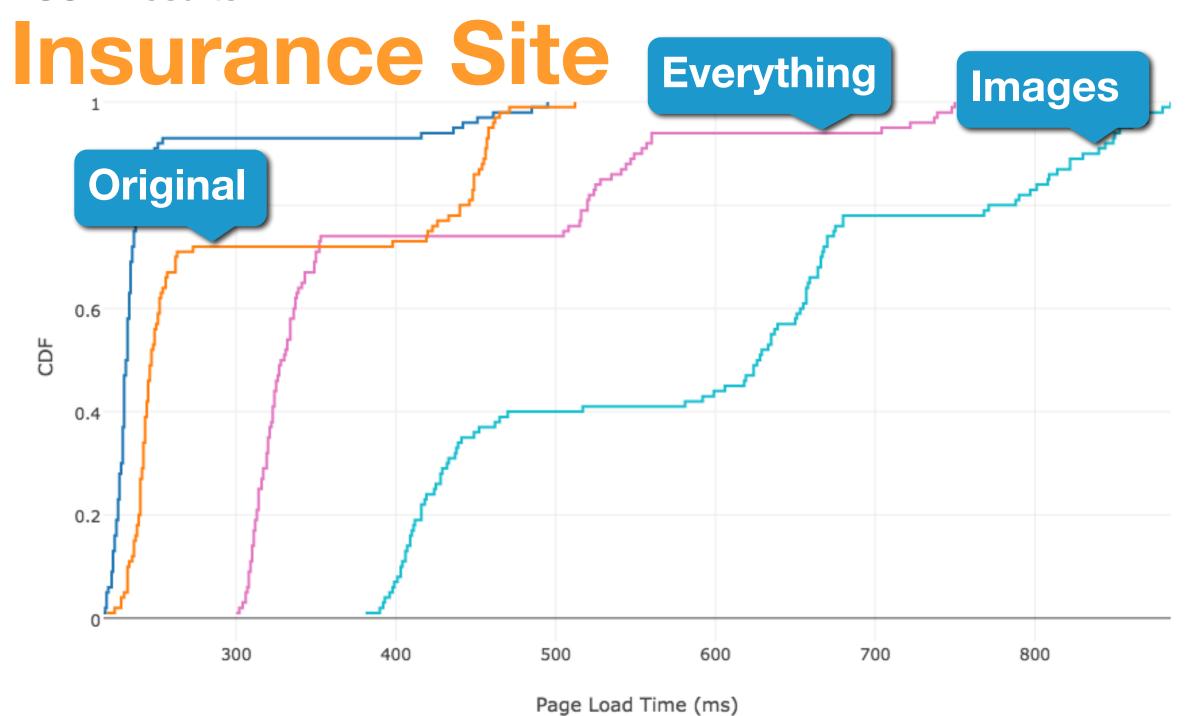
How do you push?

Push while waiting

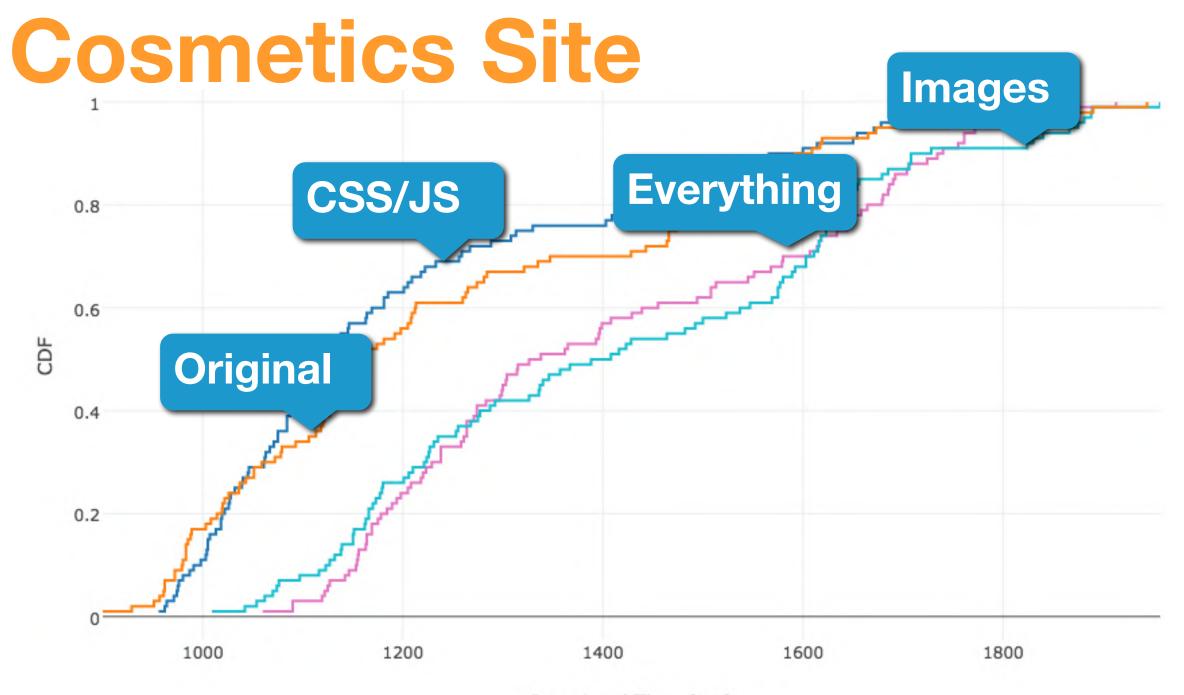
```
public class MyServlet extends HttpServlet
  public void doGet(HttpServletRequest request, HttpServletResponse response)
   String uri = request.getRequestURI();
    if ("/index.html".equals(uri))
      String resourceToPush = "/js/jquery.js";
      RequestDispatcher dispatcher =
              request.getRequestDispatcher(resourceToPush);
      dispatcher.push(request);
```



PUSH Results



PUSH Results



Automatic PUSH_PROMISE

But when you get it Right:







https://goo.gl/M98bWr

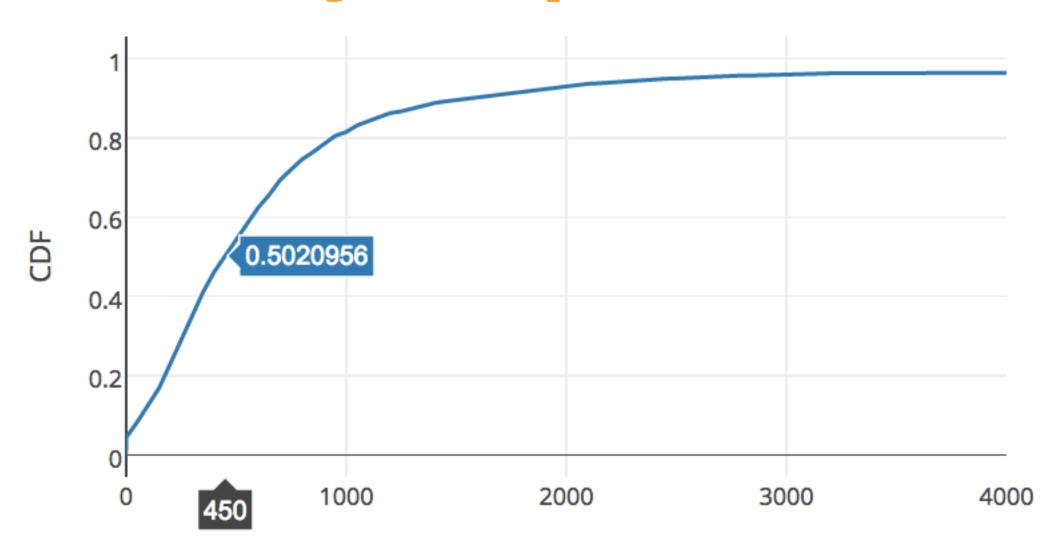
Google' Recommendations

- 1. Push just enough to fill idle network time, and no more
- 2. Push resources in right order.
- 3. Use special strategies to track the client-side cache
- 4. Use the right cookies when pushing resources that vary by cookie
- 5. Use server push to fill the initial cwnd and use preload to reveal the remaining critical or hidden resources.



Time-To-First-Byte

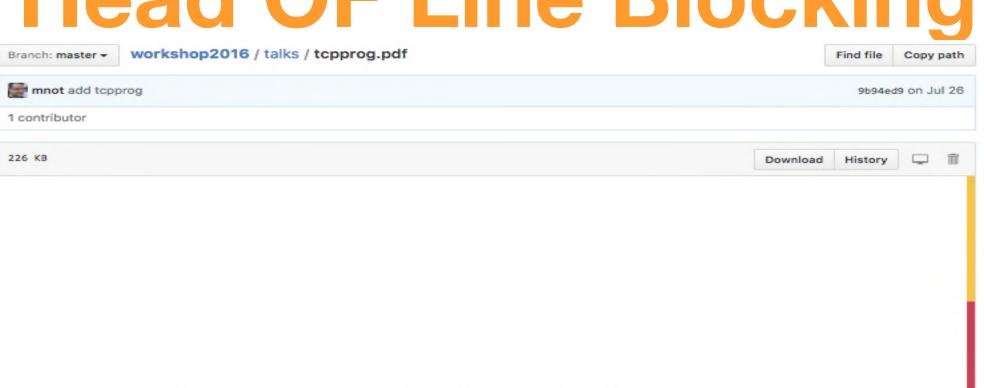
TTFB delays impact TCP HOL



Time To First Byte of base page (ms)

Minimally send to TCP Send Buffer

TCP Head OF Line Blocking



Programming TCP for responsiveness



DeNA Co., Ltd. Kazuho Oku

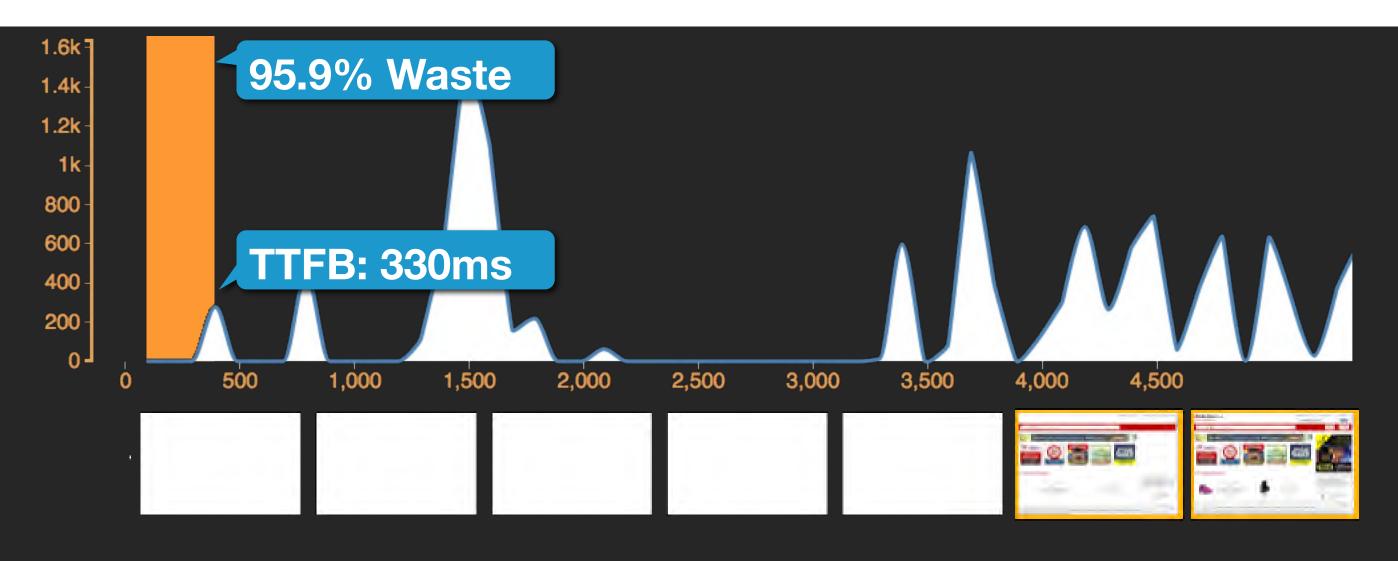


AMAI | **FASTER FORWARD™**

Announcing: shouldipush.com

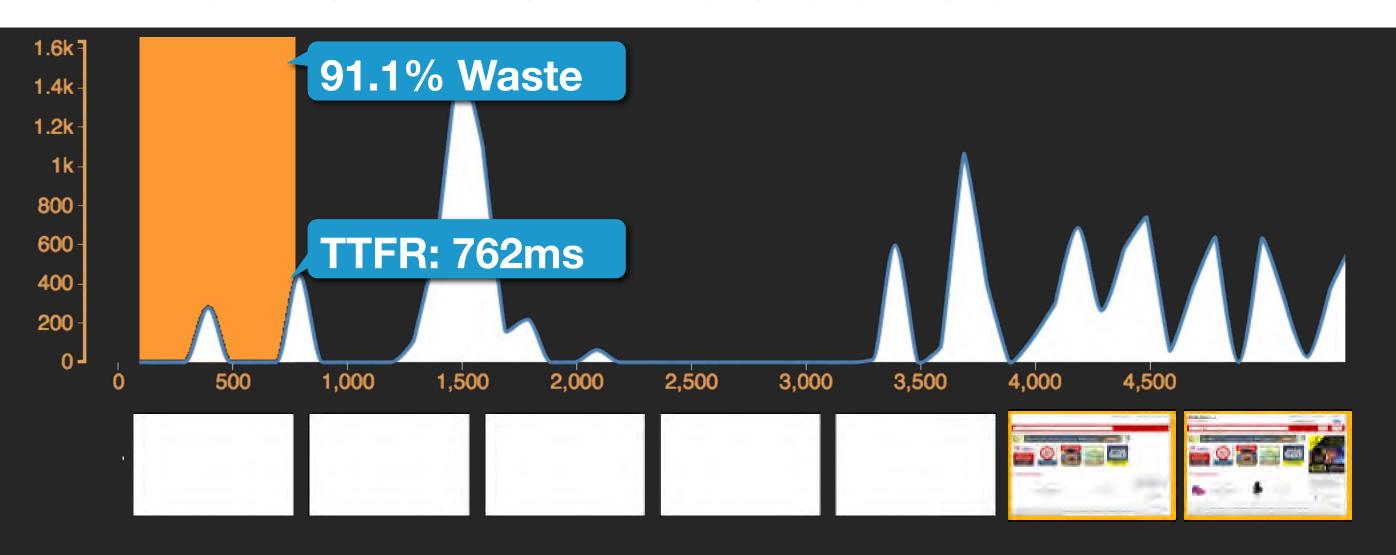
www.rakuten.co.uk

TTFB: Network Waste

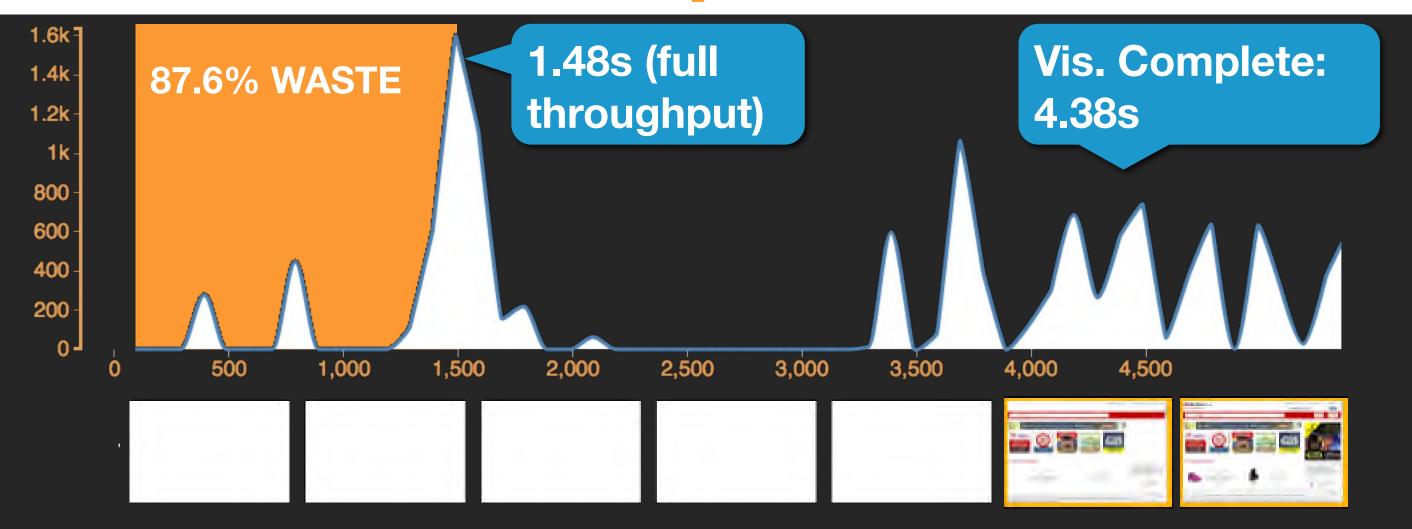


www.rakuten.co.uk

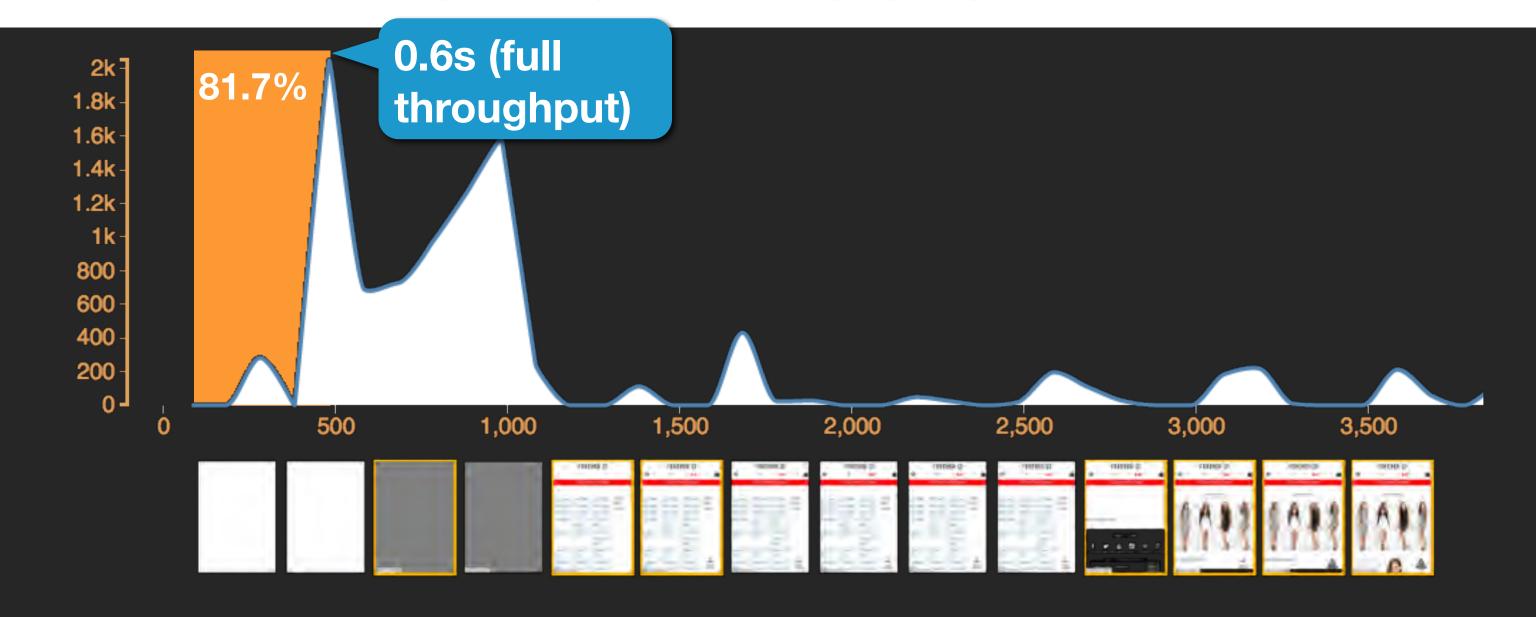
Time-to-First-Resource



33.7% Vis. Complete Time



WPT: Network Waste



Browser Support

Implementation notes

- Measures opportunity to use the network
- Synthetic indicator metric NOT silver bullet
- Uses WPT results

Chapter 4: Epilogue

PUSH

Questions & Predictions

- First request: ClientHints, Session Cookies
- PUSH to surrogate but not client
- Cache Management / Invalidation
- Next page resources
- DDOS countermeasures using PUSH



HTTP/2 PUSH

Take Action

Optimize for the Browser

- Cache Everything
- Remove 3rd Party Content
- Federated development: move APIs to the 1st party domain
- Plan for coalescing (use TLS SANs)

Experiment

- shouldipush.com
- Test TTFB globally

Thanks! @ColinBendell