eyeorg

A Platform for Crowdsourcing Web Quality of Experience Measurements

Matteo Varvello
Telefónica Research

Jeremy Blackburn
Telefónica Research

David Naylor
Carnegie Mellon University

Dina Papagiannaki
Google Inc.

Web **quality of experience** matters a lot

amazon

1 second slowdown

▼$1.6 Billion in sales per year

Google

0.4 second slowdown

▼8 Million searches per day

A lot of people are working to improve **page load time (PLT)**

RESEARCH

Polaris [NSDI '16]
Shandian [NSDI '16]
Klotski [NSDI '15]

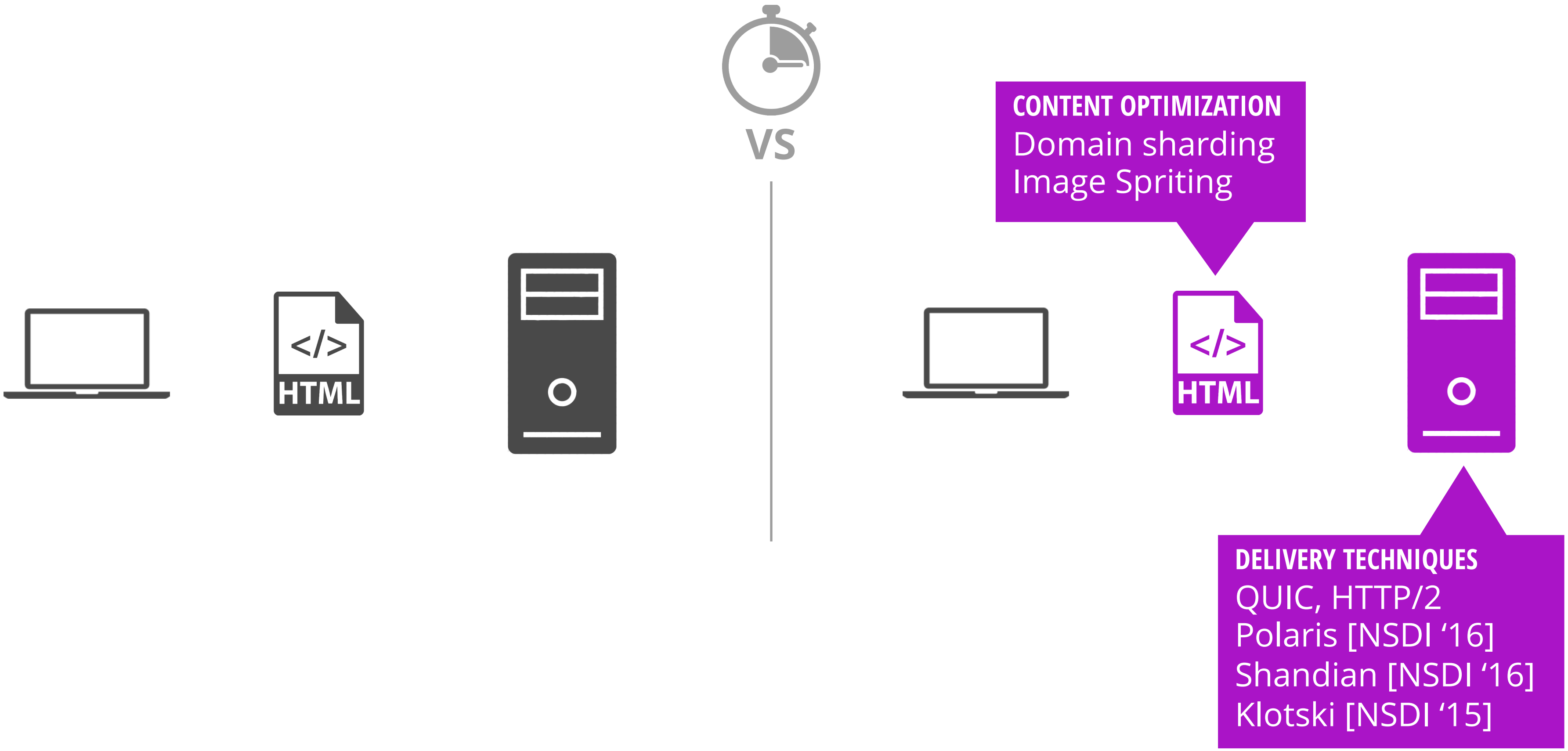
STANDARDS

QUIC [Google]
SPDY [Google]
HTTP/2 [IETF]

CDNs

Akamai
Level 3
CloudFlare
Limelight
CacheFly
MaxCDN
Instart Logic
Speedera
EdgeCast
Aryaka
Incapsula
Aryaka
...

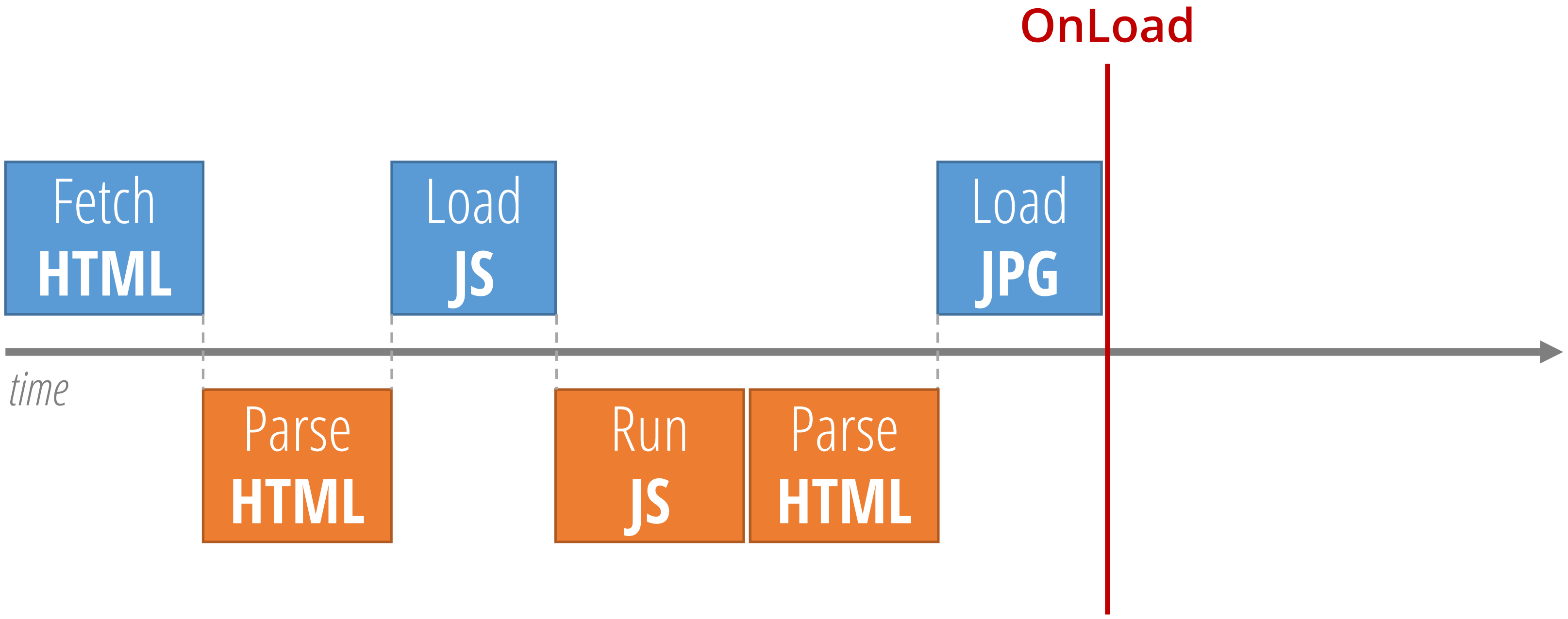
Measuring PLT is important for evaluating new technologies



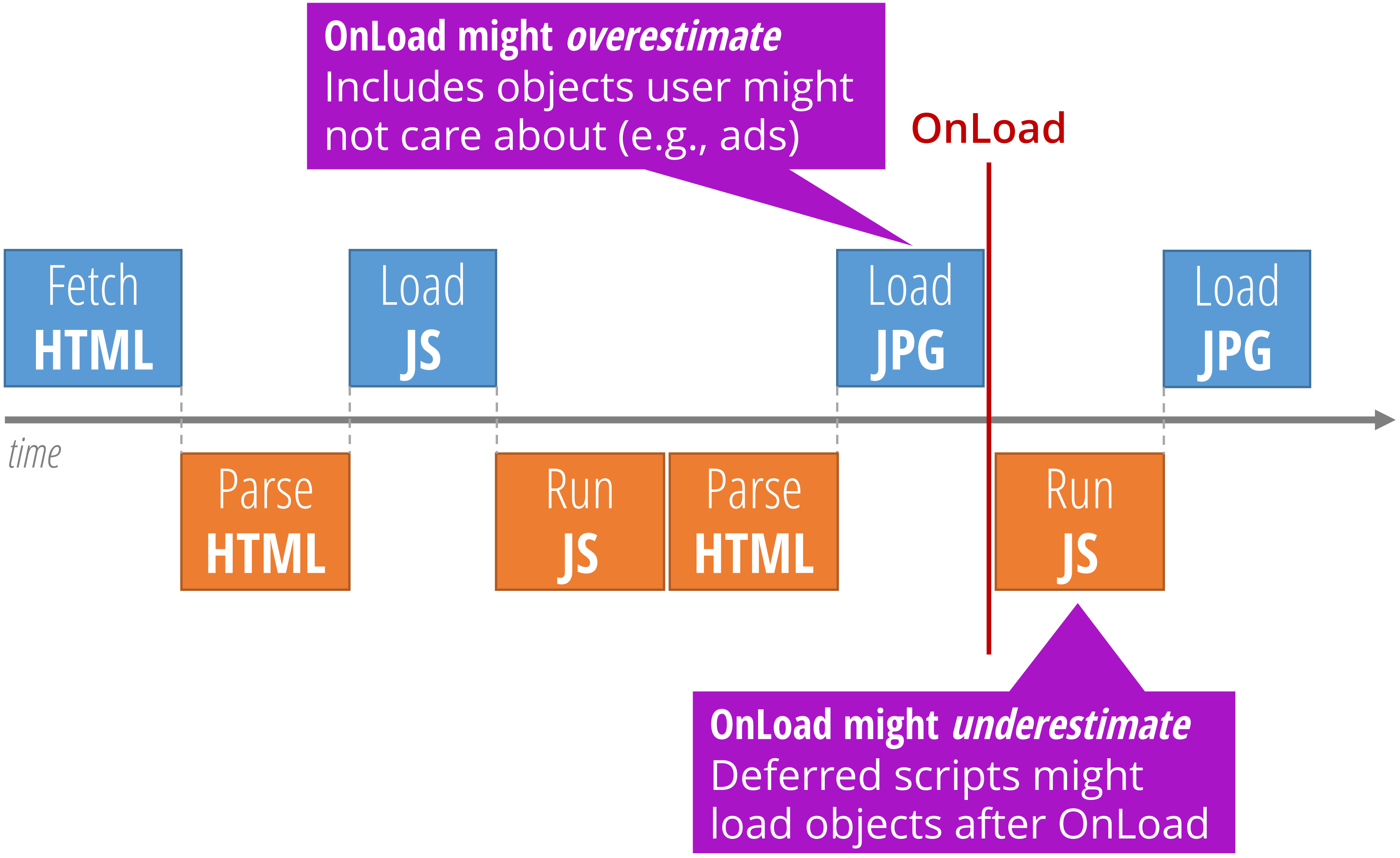
CONTENT OPTIMIZATION
Domain sharding
Image Spriting

DELIVERY TECHNIQUES
QUIC, HTTP/2
Polaris [NSDI '16]
Shandian [NSDI '16]
Klotski [NSDI '15]

PLT is usually measured with *OnLoad*



OnLoad might not reflect *user-perceived* PLT



OnLoad might *overestimate*
Includes objects user might not care about (e.g., ads)

OnLoad

Fetch **HTML**

Load **JS**

Load **JPG**

Load **JPG**

time

Parse **HTML**

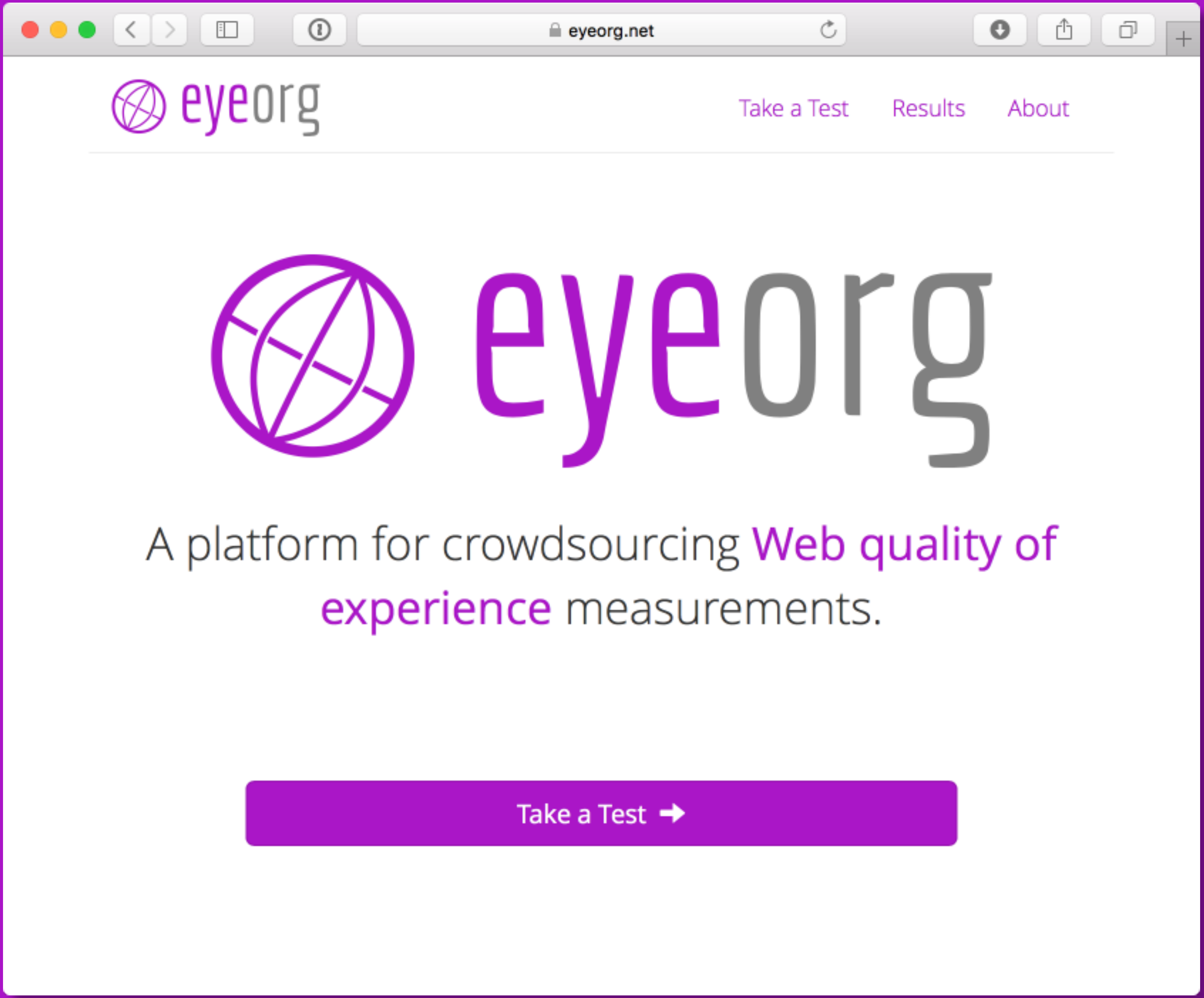
Run **JS**

Parse **HTML**

Run **JS**

OnLoad might *underestimate*
Deferred scripts might load objects after OnLoad

How do we measure *User-Perceived* Page Load Time?



A platform for crowdsourcing Web quality of experience measurements.

Challenges

1. **Consistent experience**
   Participants have different software and network conditions

2. **Quantitative responses**
   It's hard to express when a page "seems loaded"

3. **Trustworthy results**
   Crowd workers are not always reliable

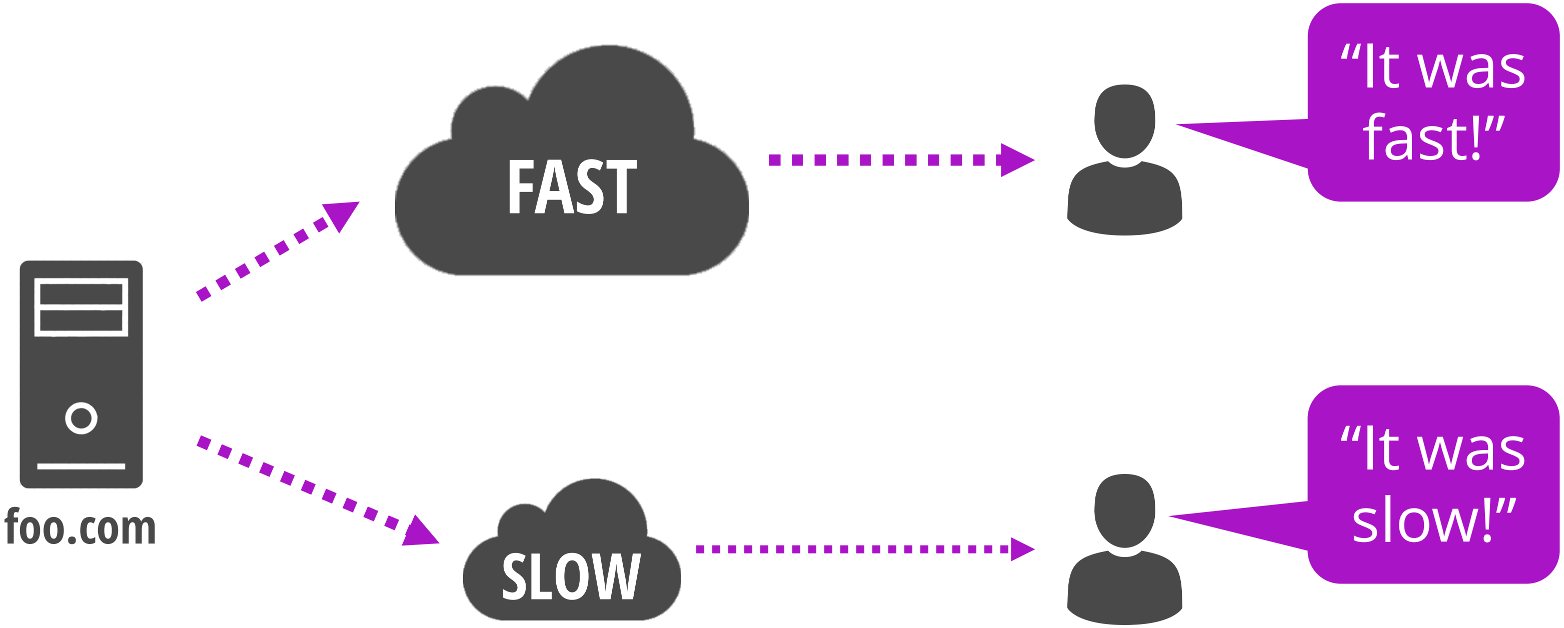
Challenges

1. **Consistent experience**
   Participants have different software and network conditions

2. **Quantitative responses**
   It's hard to express when a page "seems loaded"

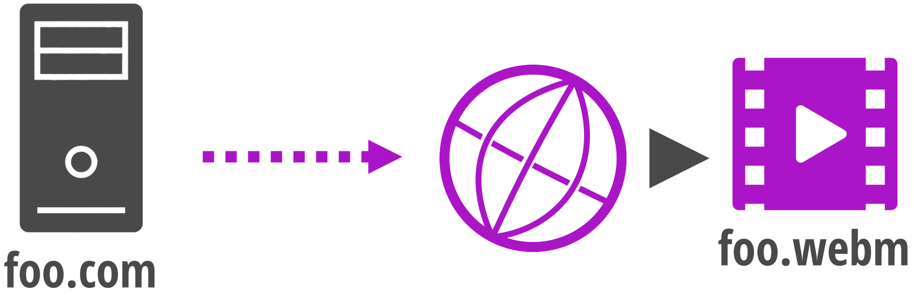
3. **Trustworthy results**
   Crowd workers are not always reliable

Participants’ network connections impact their responses

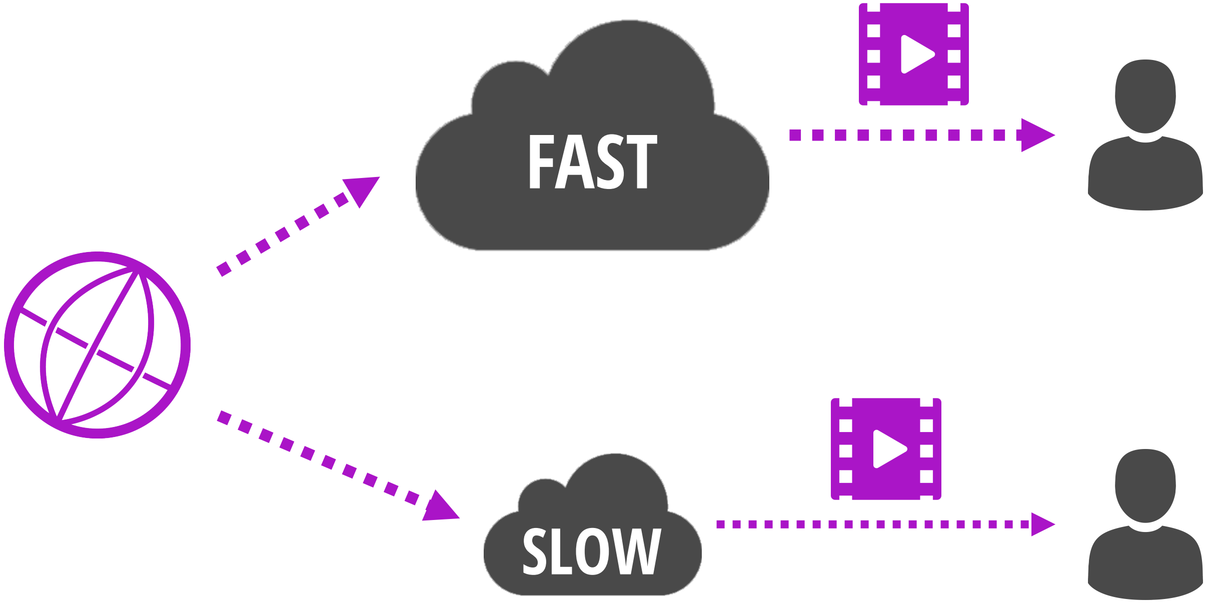


Videos of pages loading look the same to everyone

Capture videos in advance



Serve videos, not sites, during tests



Challenges

1 **Consistent experience**
Participants have different software and network conditions

2 **Quantitative responses**
It's hard to express when a page "seems loaded"

3 **Trustworthy results**
Crowd workers are not always reliable

Challenges

1. **Consistent experience**
   Participants have different software and network conditions

2. **Quantitative responses**
   It’s hard to express when a page “seems loaded”

3. **Trustworthy results**
   Crowd workers are not always reliable

We designed two types of test

Timeline

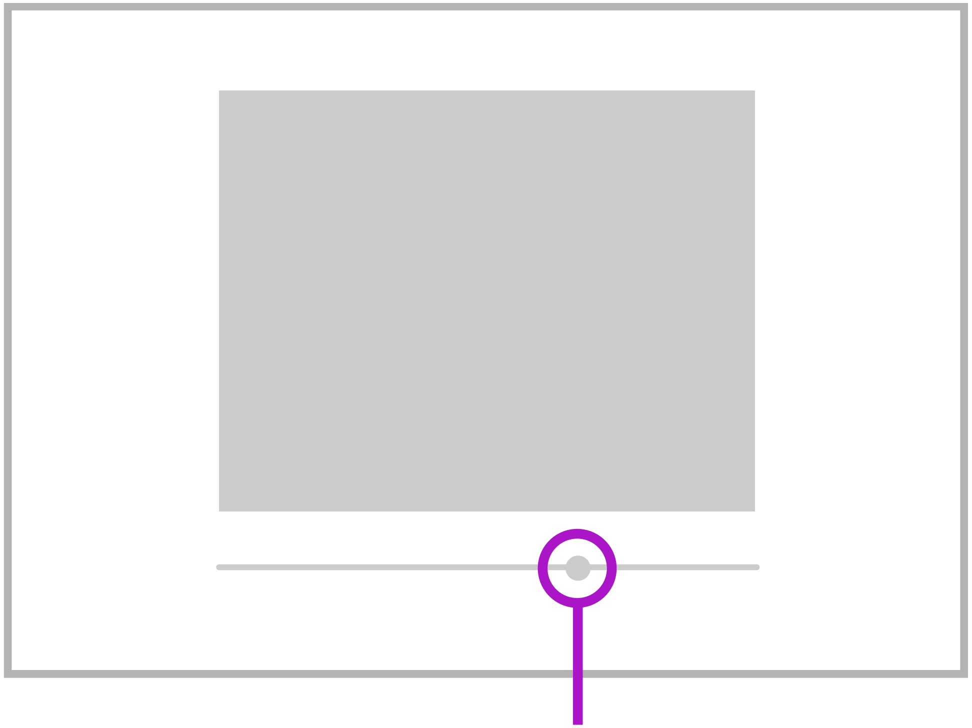
When does the page look “ready to use”?

A/B

Which version loaded faster?

Timeline

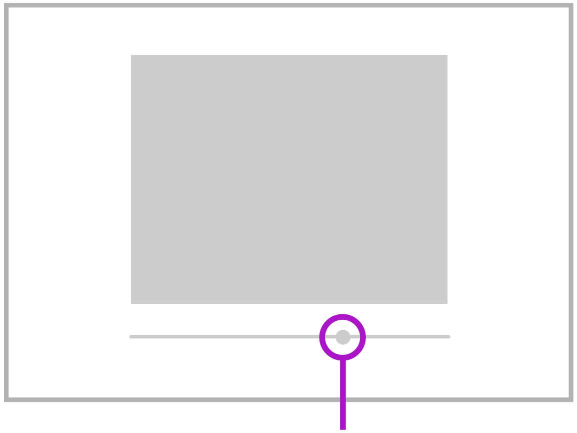
When does the page look “ready to use”?



Drag the slider to scrub through the video until the page appears “ready to use.”

Timeline

When does the page look "ready to use"?



Drag the slider to scrub through the video until the page appears "ready to use."

"Scrub bar"

Rather than standard HTML5 video controls

Preload the video

To avoid "is the page in the video still loading, or is the video itself still loading?"

Frame rewind

*When user submits, offer the **earliest similar frame** to correct for overshooting*

We designed two types of test

Timeline

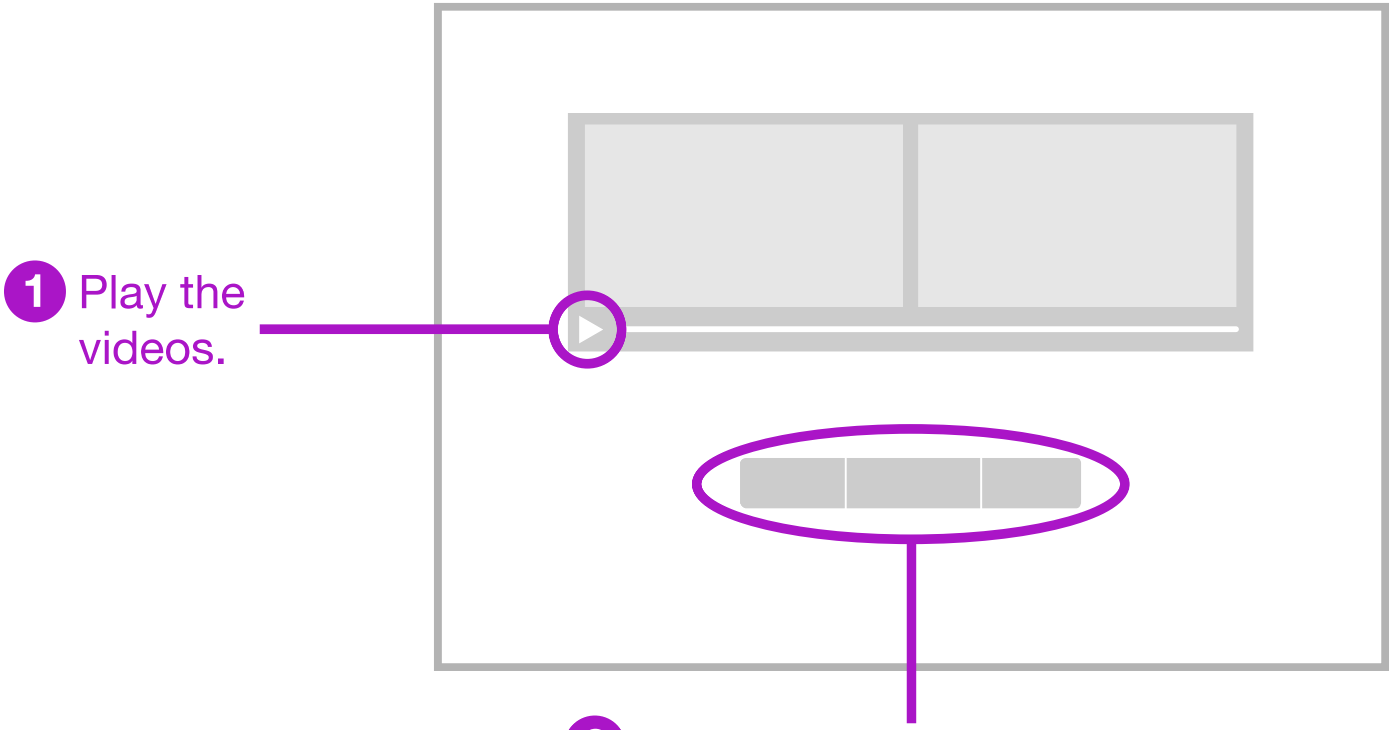
When does the page look “ready to use”?

A/B

Which version loaded faster?

A/B

Which version loaded faster?

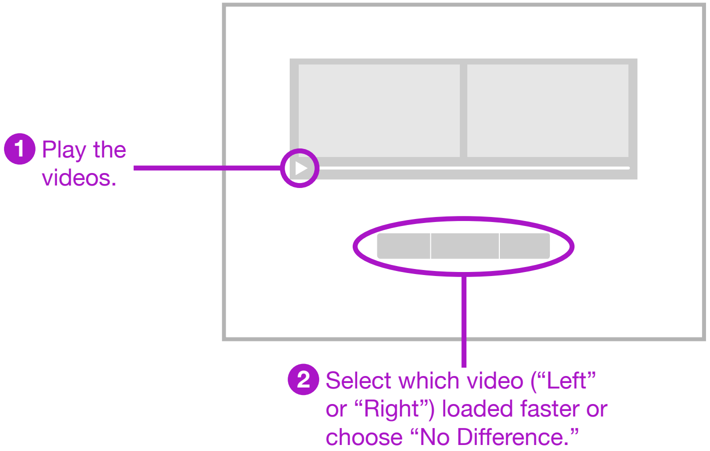


1 Play the videos.

2 Select which video ("Left" or "Right") loaded faster or choose "No Difference."

A/B

Which version loaded faster?



1 Play the videos.

2 Select which video ("Left" or "Right") loaded faster or choose "No Difference."

Head-to-head comparison

No need to decide precise PLT; simpler to just choose winner

Single video

So A and B never get out of sync

Random order

A is not always left, B is not always right

We designed two types of test

Timeline

When does the page look “ready to use”?

A/B

Which version loaded faster?

Challenges

1 **Consistent experience**

Participants have different software and network conditions

2 **Quantitative responses**

It’s hard to express when a page “seems loaded”

3 **Trustworthy results**

Crowd workers are not always reliable

Challenges

1 Consistent experience
Participants have different software and network conditions

2 Quantitative responses
It's hard to express when a page "seems loaded"

3 Trustworthy results
Crowd workers are not always reliable

Eyeorg filters responses using techniques from HCI literature

Evaluation Campaign

100
crowdsourced workers

100
trusted participants as ground truth

20
sites from Alexa top 1M

Filtering techniques:

1. Control questions
2. Engagement
3. Soft rules
4. Wisdom of the Crowd

Challenges

1 **Consistent experience**
Participants have different software and network conditions

2 **Quantitative responses**
It's hard to express when a page "seems loaded"

3 **Trustworthy results**
Crowd workers are not always reliable

Challenges

1. **Consistent experience**
   Participants have different software and network conditions

2. **Quantitative responses**
   It’s hard to express when a page “seems loaded”

3. **Trustworthy results**
   Crowd workers are not always reliable

We ran three measurement campaigns on eyeorg

1. **PLT metrics**
   How well do existing metrics capture user-perceived PLT?

2. **HTTP/1.1 vs. HTTP/2**
   Do users perceive a PLT difference between the two?

3. **Ad Blockers**
   Do users perceive a PLT difference between popular ad blockers?

We ran three measurement campaigns on eyeorg

1 **PLT metrics**

How well do existing metrics capture user-perceived PLT?

2 HTTP/1.1 vs. HTTP/2

Do users perceive a PLT difference between the two?

See Paper

3 Ad Blockers

Do users perceive a PLT difference between popular ad blockers?

We use **timeline tests** to compare **PLT metrics**

PLT Metric Campaign

1000
crowdsourced workers

100
sites from Alexa top 1M

$120
total cost

1.5 days
to collect responses

For each site, measure PLT 5 ways:

1| OnLoad
(from HAR)

2| First Visual Change (FVC)
3| Last Visual Change (LVC)
4| SpeedIndex
(from video)

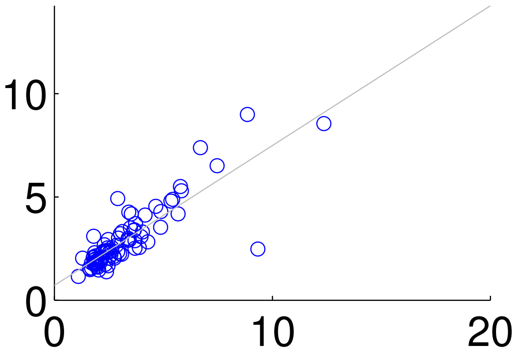
5| User-Perceived PLT
(from eyeorg)

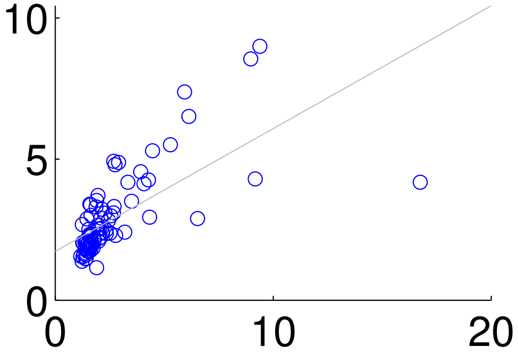
OnLoad and **First Visual Change** correlate best with UPLT

OnLoad

Correlation:

0.85





SpeedIndex

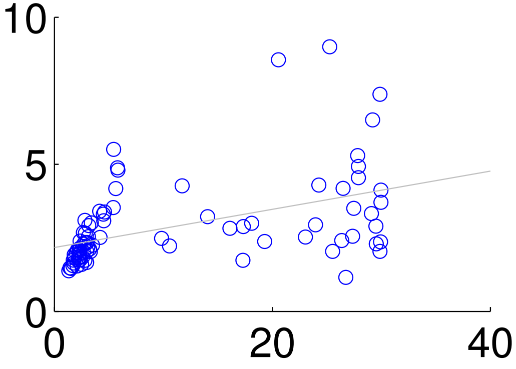
Correlation:

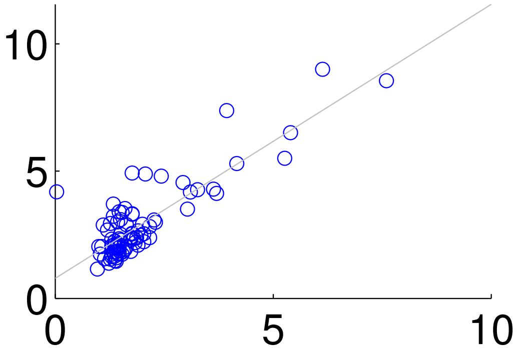
0.68

LastVisualChange

Correlation:

0.47



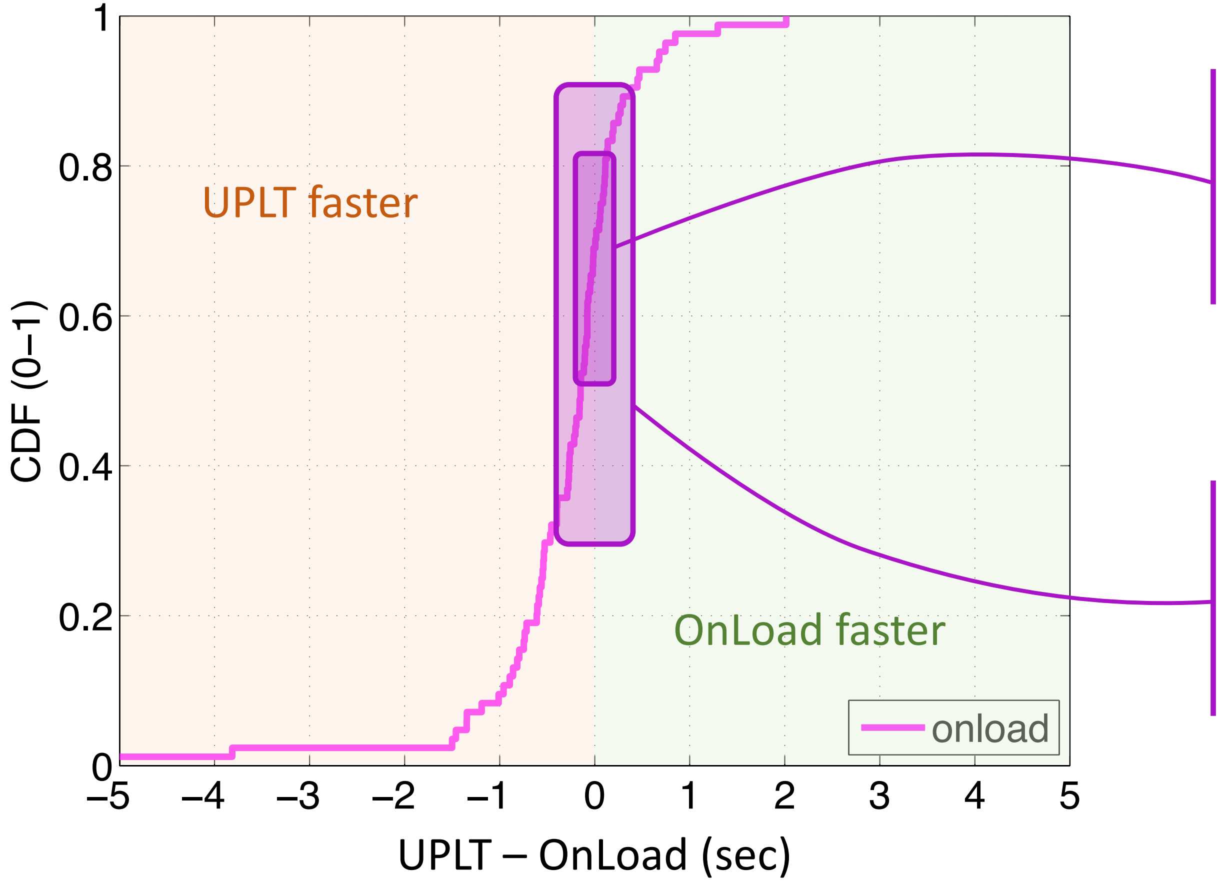


FirstVisualChange

Correlation:

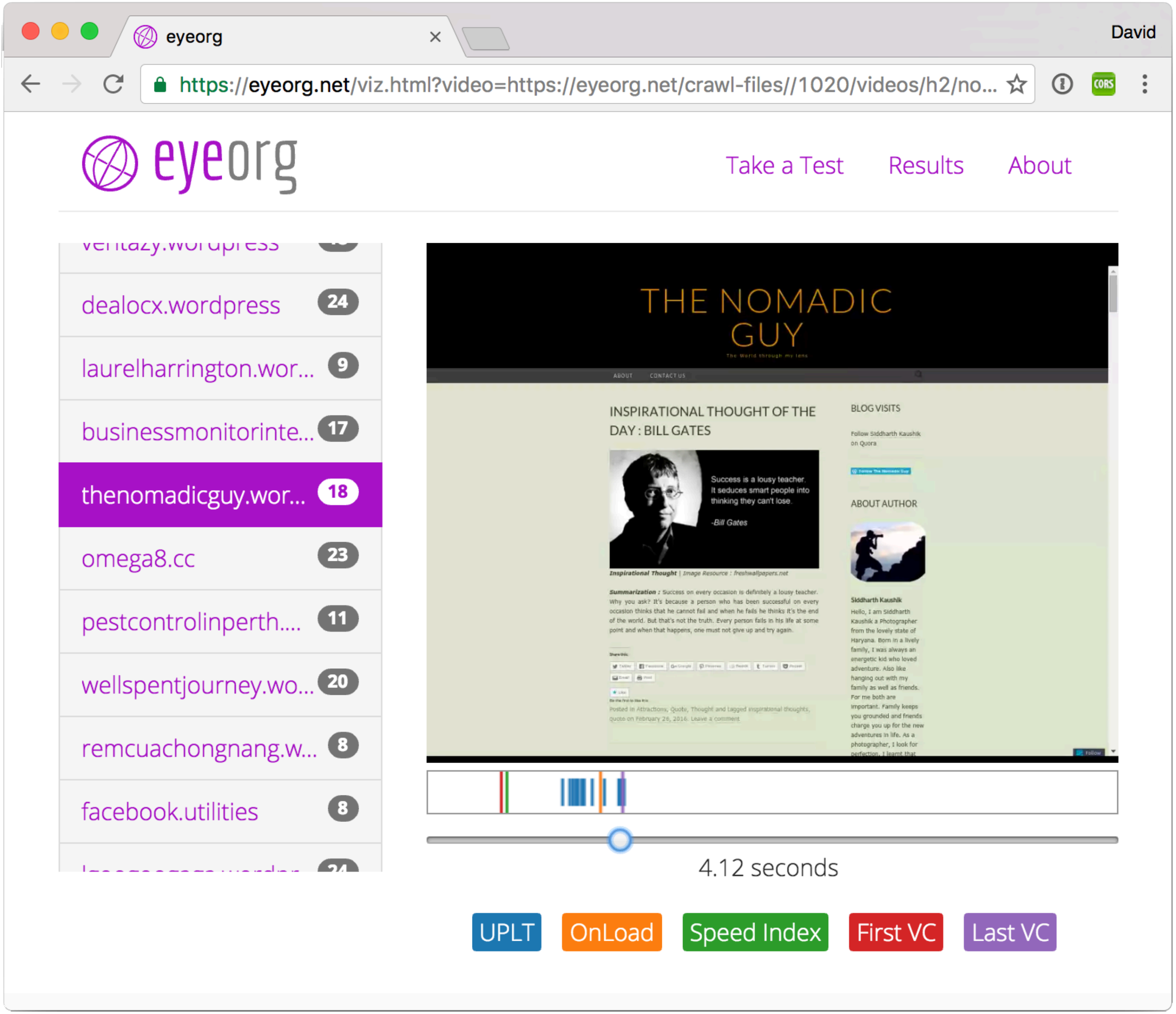
0.84

OnLoad is usually within 1 second of UPLT



For 30% of sites, onload within 100 ms of UPLT

For 60% of sites, onload within 200 ms of UPLT



eyeorg
David
https://eyeorg.net/viz.html?video=https://eyeorg.net/crawl-files//1020/videos/h2/no...
eyeorg
Take a Test
Results
About

veritazy.wordpress	
dealocx.wordpress	24
laurelharrington.wor...	9
businessmonitorinte...	17
thenomadicguy.wor...	18
omega8.cc	23
pestcontrolinperth....	11
wellspentjourney.wo...	20
remcuachongnang.w...	8
facebook.utilities	8

THE NOMADIC GUY

The World through my lens

ABOUT CONTACT US

INSPIRATIONAL THOUGHT OF THE DAY : BILL GATES

Success is a lousy teacher. It seduces smart people into thinking they can't lose.

-Bill Gates

Inspirational Thought | *Image Resource : freshwallpapers.net*

Summarization : Success on every occasion is definitely a lousy teacher. Why you ask? It's because a person who has been successful on every occasion thinks that he cannot fail and when he fails he thinks it's the end of the world. But that's not the truth. Every person fails in his life at some point and when that happens, one must not give up and try again.

Share this:

Twitter Facebook Google Pinterest Reddit Tumblr Pocket Email Print

Like

Be the first to like this.

Posted in Attractions, Quote, Thought and tagged inspirational thoughts, quote on February 26, 2016. Leave a comment

BLOG VISITS

Follow Siddharth Kaushik on Quora

Follow The Nomadic Guy

ABOUT AUTHOR

Siddharth Kaushik

Hello, I am Siddharth Kaushik a Photographer from the lovely state of Haryana. Born in a lively family, I was always an energetic kid who loved adventure. Also like hanging out with my family as well as friends. For me both are important. Family keeps you grounded and friends charge you up for the new adventures in life. As a photographer, I look for perfection, I learnt that

4.12 seconds

We ran three measurement campaigns on eyeorg

1 **PLT metrics**

How well do existing metrics capture user-perceived PLT?

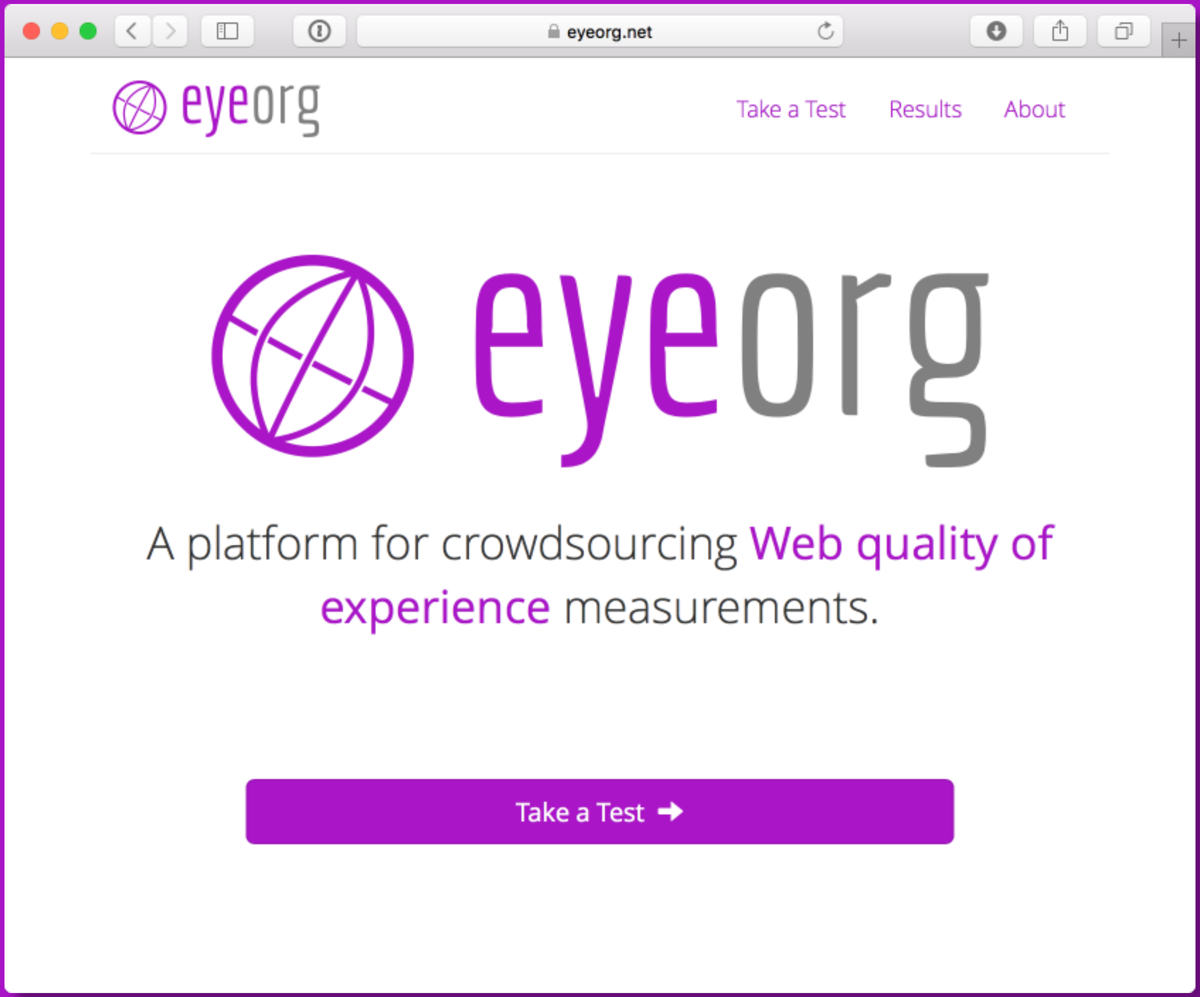
2 HTTP/1.1 vs. HTTP/2

Do users perceive a PLT difference between the two?

See Paper

3 Ad Blockers

Do users perceive a PLT difference between popular ad blockers?



A platform for crowdsourcing Web quality of experience measurements.

Want to use eyeorg?

Get in touch!

https://eyeorg.net