Browser Feature Usage on the Modern Web

Peter Snyder - Lara Ansari - Cynthia Taylor - Chris Kanich

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

Motivation

- Web browsers constantly gain new features

- How often are they used?

- What sites use them?

- Are they used user-serving purposes?

Outline

- The Web API

- Data sources

- Crawling technique

- Results

- Discussion

Outline

- **The Web API**

- Data sources

- Crawling technique

- Results

- Discussion

The Web API

✓ Browser provided

✓ Javascript-exposed

✓ Website available

✓ Used to build interactive web applications

x Default javascript

x Browser chrome

x Browser extensions

x Static documents

The Web API

- Large, and growing fast

- ~1.4k functions and properties

- Largely standardized

- Generally available to all sites

The Web API

- Document manipulation

- AJAX / server requests

- Cookies

- Browser navigation

The Web API

- Document manipulation

- AJAX / server requests

- Cookies

- Browser navigation

- Complex graphics animations

- WebGL

- Cryptographic operations

- Parallel operations

- Font operations

The Web API

- Document manipulation

- AJAX / server requests

- Cookies

- Browser navigation

- Complex graphics animations

- WebGL

- Cryptographic operations

- Parallel operations

- Font operations

- Ambient light sensing

- Peer-to-peer networking

- Audio synthesis

- "Beacons"

- Geolocation

- Gamepads

- Vibration

- High resolution timers

- DRM

- SVG Animations

- Much more…

Outline

- The Web API

- **Data sources**

- Crawling technique

- Results

- Discussion

Data Sources

- **Website Rankings**
  Alexa 10,000

- **Web API Features**
  1,392 end points

- **Web API Standards**
  74 categorizations

- **Historical Browser Releases**
  168 Firefox versions

- **Popular Blocking Extensions**
  AdBlock Plus, Ghostery

Blocking Extensions

- **AdBlock Plus**
  - Hides or removes advertising
  - Crowdsourced filter lists
  - 42m downloads

- **Ghostery**
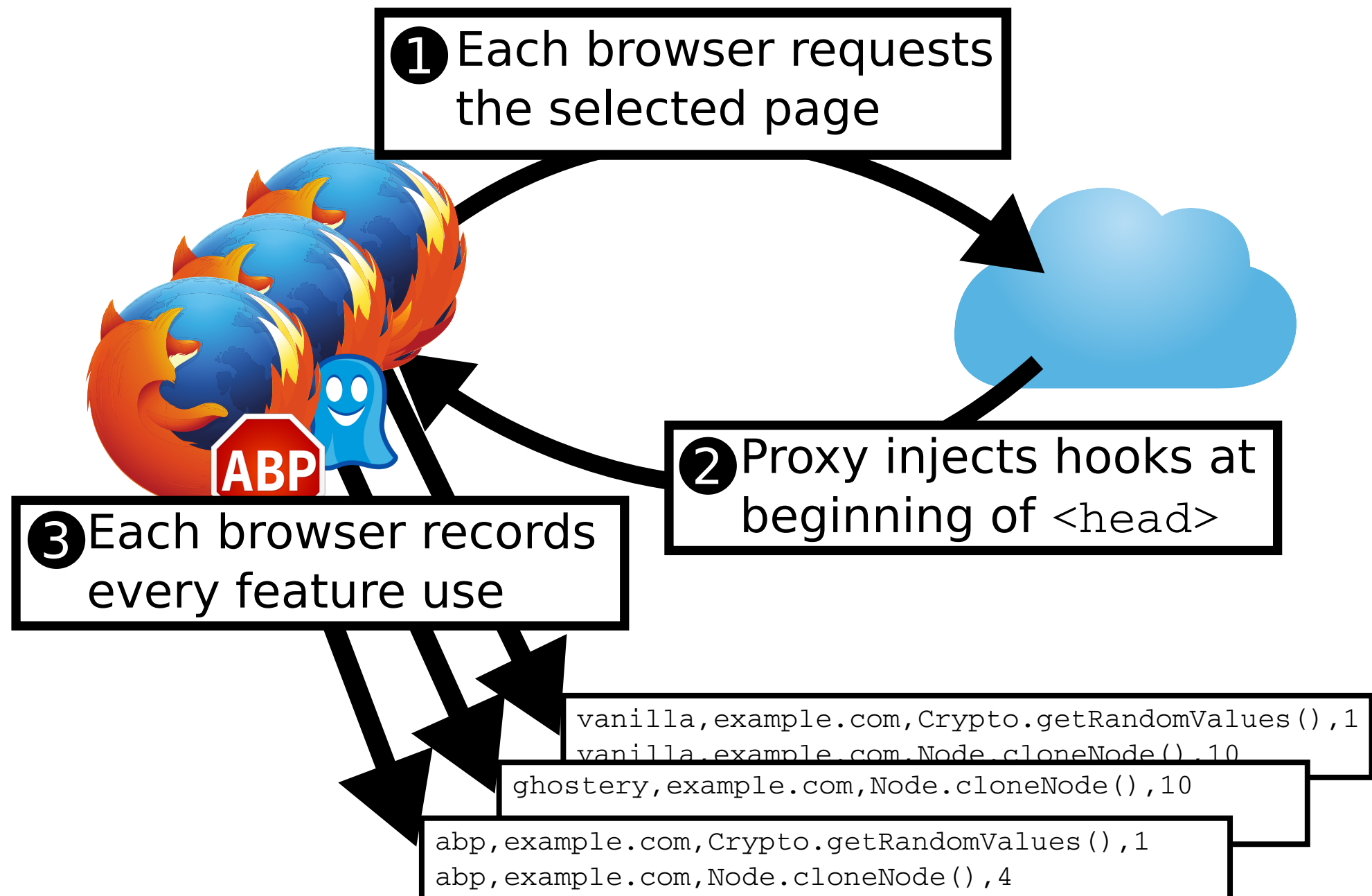 - Blocks online tracking
  - Expert curated
  - 21m downloads

Outline

- The Web API

- Data sources

- **Crawling technique**

- Results

- Discussion

Methodology

- Goal: How is the Web API used?

- 10k too many websites for manual code review

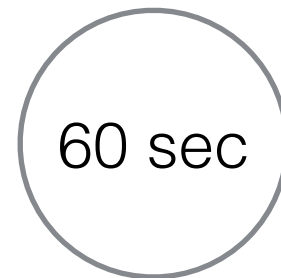- Javascript makes static analysis difficult
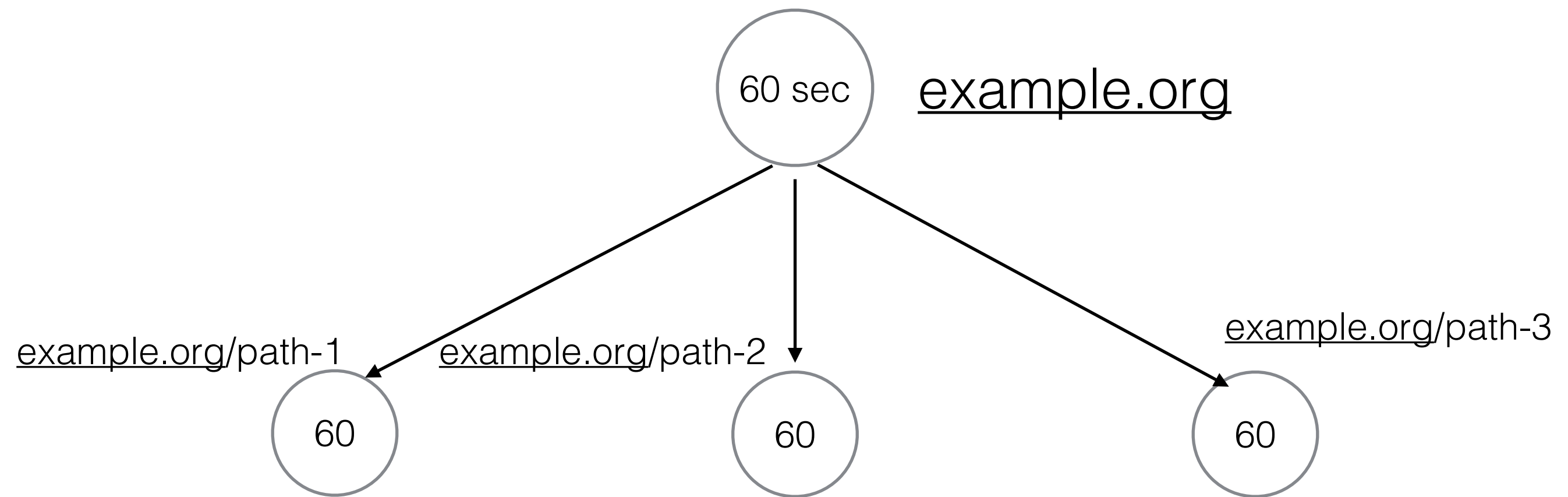
- Automation

Extension Code Injection

❶ Each browser requests the selected page

❷ Proxy injects hooks at beginning of `<head>`

❸ Each browser records every feature use

```
vanilla,example.com,Crypto.getRandomValues(),1
vanilla,example.com,Node.cloneNode(),10
ghostery,example.com,Node.cloneNode(),10
abp,example.com,Crypto.getRandomValues(),1
abp,example.com,Node.cloneNode(),4
```

Automation

- Gremlins.js to simulate human interaction

- 60 seconds per page

- Recurse through the site

Automated Measurement

(60 sec)   example.org
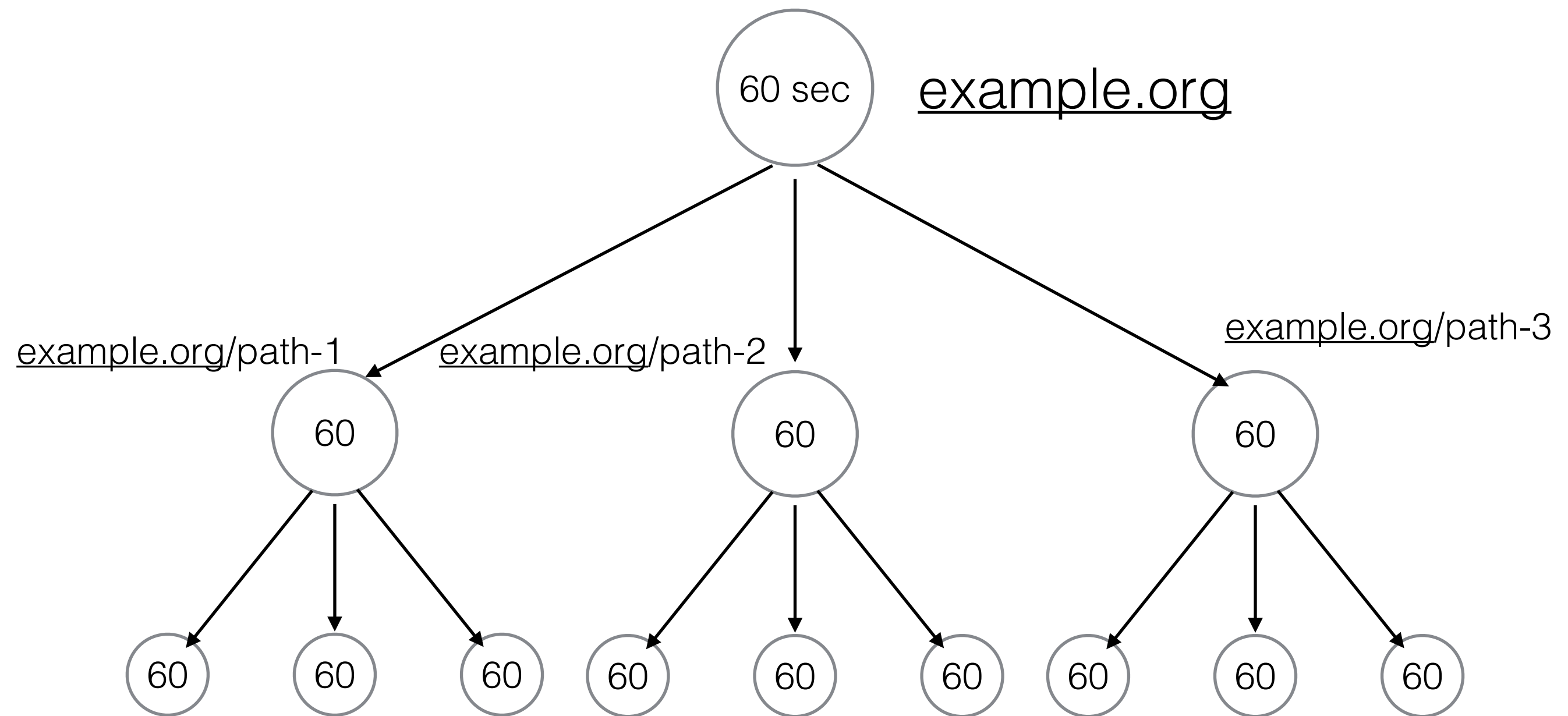
Automated Measurement

60 sec   example.org

example.org/path-1   example.org/path-2   example.org/path-3

60   60   60

Automated Measurement

Automated Measurement

- 5 times per domain

- **4 Test Cases**
  - Default
  - AdBlock Plus
  - Ghostery
  - ABP + Ghostery

Domains measured	9,733
Total website interaction time	480 days
Web pages visited	2,240,484
Feature invocations recorded	21,511,926,733

Validation

- **Internal Validation**
  After 4 measurements, no new standards

- **External Validation**
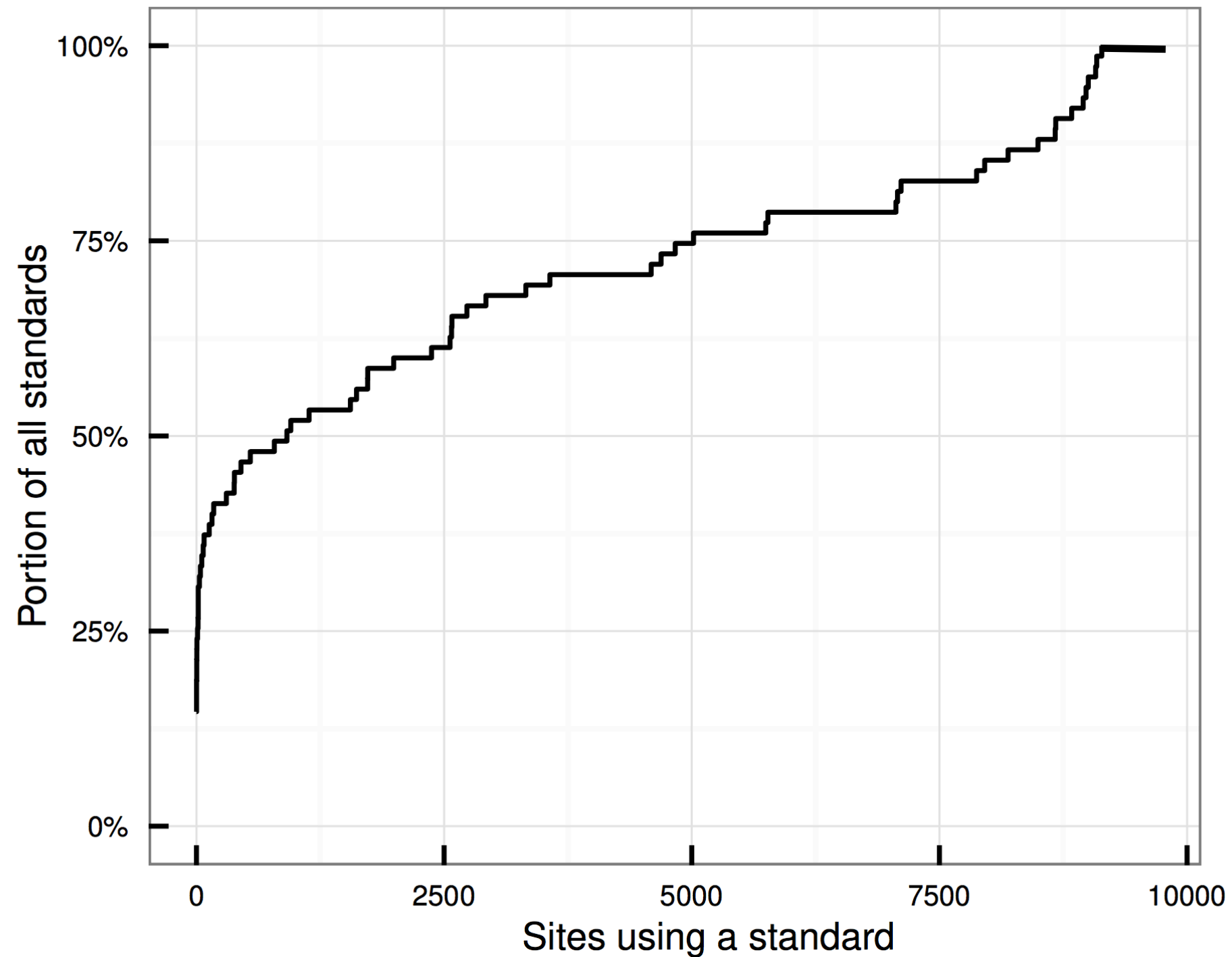 Compared to random sampling of human usage

Outline

- The Web API

- Data sources

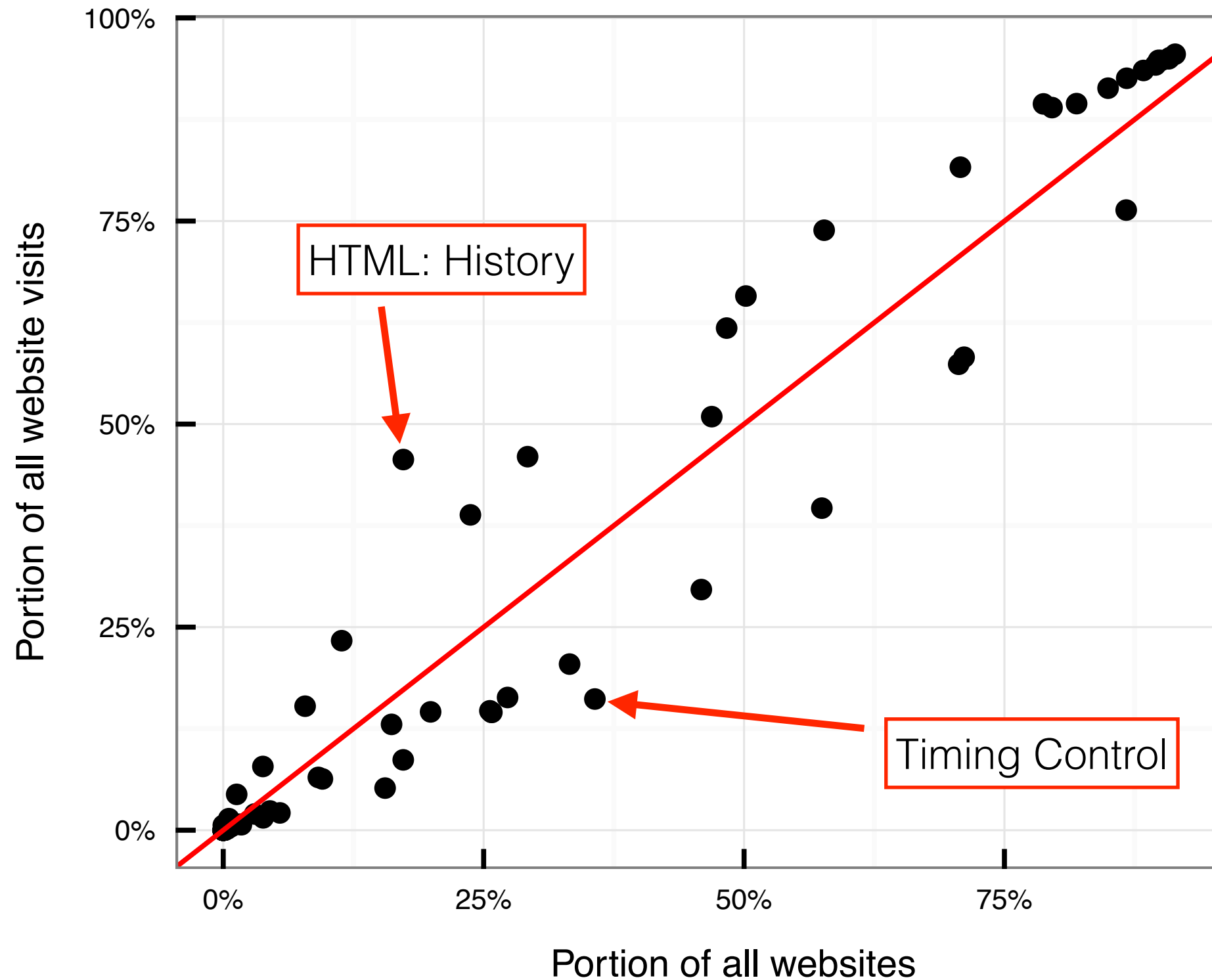- Crawling technique

- **Results**

- Discussion

Results

- **Standard Popularity**
  % of Alexa 10k using any feature from the standard

- **Block Rate**
  How often would a feature be used, except for a blocking extension
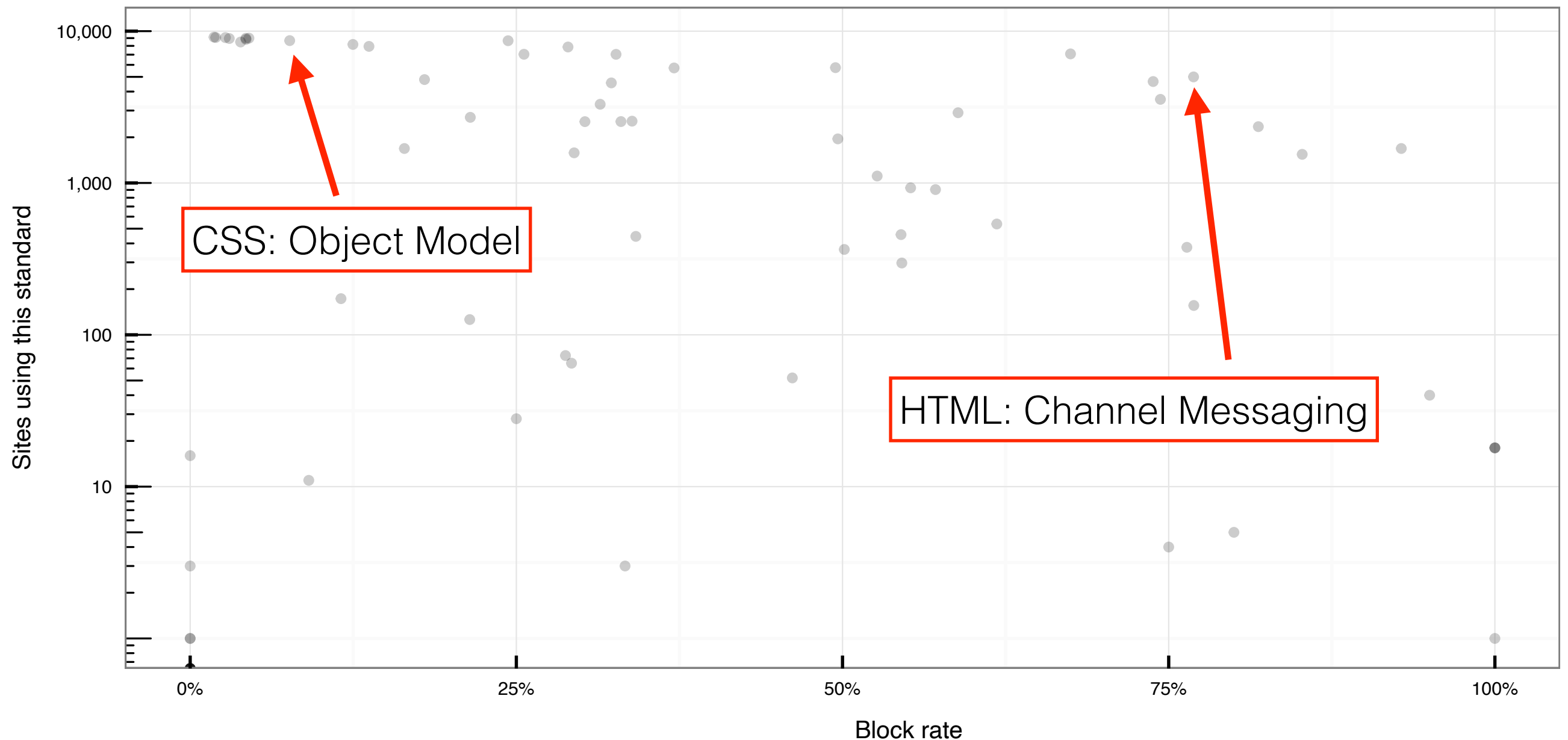
Standard Popularity
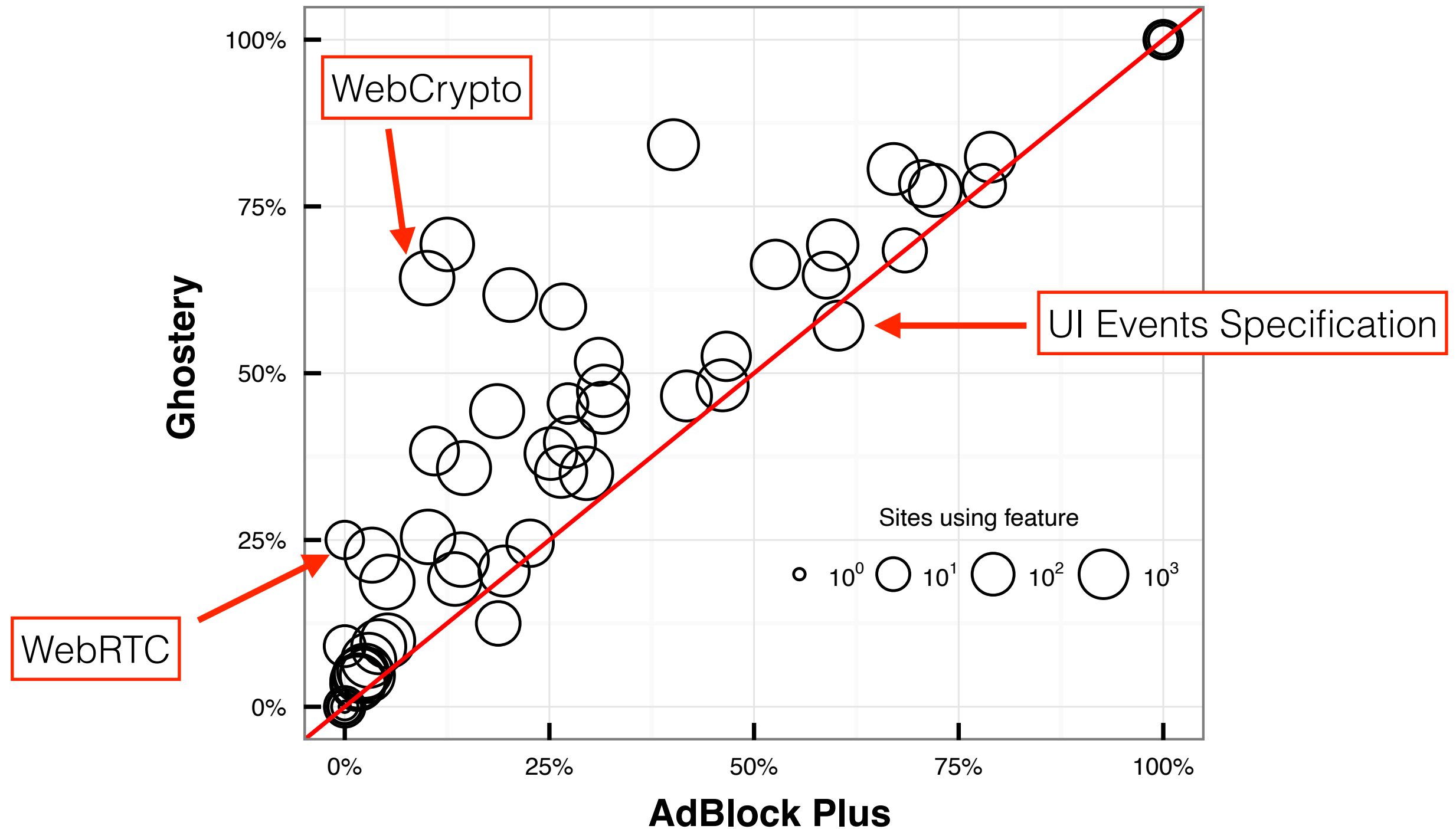
Standard vs. Site Popularity

Portion of all website visits (y-axis), Portion of all websites (x-axis)

- HTML: History
- Timing Control

Standard Popularity by Introduction

Standard Popularity and Block Rates

CSS: Object Model

HTML: Channel Messaging

Sites using this standard

Block rate

Ad vs Tracking Blocking

Outline

- The Web API

- Data sources

- Crawling technique

- Results

- **Discussion**

Discussion

- **Large amounts of the Web API are rarely used**
  37% of standards are used on < 1% of sites

- **Blocking makes more standards rarely used**
  With blocking, 64% of standards are used on < 1%

- **Features makes things more exaggerated**
  83% of features are executed on < 1% of sites

Open Questions / Future Work

- What are the security implications of the growth of the Web API?

- How does "closed" or "authenticated web" differ?

- Even more standards! (speech synthesis, WebVR, etc.)

Also in the Paper

- **Implementation Details**
  Javascript injection / DOM Instrumentation

- **Security Measurements**
  # of CVEs reported against each standard

- **Additional Results**
  Site complexity, block rate by introduction date

Take Aways

- Measurement of 1.4k Web API features in the Alexa 10k

- Web API standards vary wildly in use, many are rarely used

- Many standards appear to be used primarily for advertising or tracking

- 83% of features are executed on < 1% of sites

- Public dataset

http://imdc.datcat.org/collection/1-0723-8=Web-API-usage-in-the-Alexa-10k

Domains measured	9,733
Total website interaction time	480 days
Web pages visited	2,240,484
Feature invocations recorded	21,511,926,733