Browser Feature Usage on the Modern Web

Peter Snyder - Lara Ansari - Cynthia Taylor - Chris Kanich
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

✗ Default javascript
✗ Browser chrome
✗ Browser extensions
✗ 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

1. Each browser requests the selected page

2. Proxy injects hooks at beginning of `<head>`

3. 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

example.org

example.org/path-1
60 sec
example.org/path-2
60
example.org/path-3
60
Automated Measurement

example.org

example.org/path-1

example.org/path-2

example.org/path-3
Thirty days has September, April, June and November. When short February’s all done, all the rest have thirty-one. #NowYouKnow
Automated Measurement

- 5 times per domain

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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Domains measured</td>
<td>9,733</td>
</tr>
<tr>
<td>Total website interaction time</td>
<td>480 days</td>
</tr>
<tr>
<td>Web pages visited</td>
<td>2,240,484</td>
</tr>
<tr>
<td>Feature invocations recorded</td>
<td>21,511,926,733</td>
</tr>
</tbody>
</table>
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

Portion of all websites

HTML: History

Timing Control
Standard Popularity by Introduction

![Graph showing the standard introduction date and sites using standard across different years with different block rates.]

- Sites using standard by block rate:
  - 33% < block rate < 66%
  - 66% < block rate

Key:
- XMLHttpRequest
- Selectors API
Standard Popularity and Block Rates

- **CSS: Object Model**
- **HTML: Channel Messaging**
Ad vs Tracking Blocking

![Graph showing the relationship between AdBlock Plus and Ghostery blocking rates with corresponding symbols for WebCrypto and WebRTC. The UI Events Specification is also indicated.](Image)
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

<table>
<thead>
<tr>
<th>Domain measured</th>
<th>9,733</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total website interaction time</td>
<td>480 days</td>
</tr>
<tr>
<td>Web pages visited</td>
<td>2,240,484</td>
</tr>
<tr>
<td>Feature invocations recorded</td>
<td>21,511,926,733</td>
</tr>
</tbody>
</table>