I know what you did last summer: New persistent tracking mechanisms in the wild

Stefano Belloro & Dr Alexios Mylonas
$\text{whoami}$

Currently:
- Lecturer, Cyber Security @BU

Previously:
- PhD in Cyber Security & BSc @AUEB
- MSc Information Security @RHUL
- Security Consultant
Currently:

- Software Engineering Manager @BBC

Previously:

- M.Sc. in software engineering and Internet architecture
Web

- The world wide web (www) has changed our lives

- We spend more than 34h per week accessing online content
Mobile devices are the primary means used to access the web.
Web Threats?

- Malware
- Phishing
- Malvertising
- Watering hole attacks
- Profiling/tracking
- Browser exploitation kits
Protection from web threats?

Can (mobile|desktop) browsers protect us from web threats?
Protection from web threats?

**Control Availability**
- Popular controls absent from mobile browsers (September 2013)
- Multiple usability issues in the GUI

**Blacklists**
- Blacklist unavailable on mobile browsers or ineffective (July 2014)
- Blacklist ineffective (December 2016 & June 2018)

**Private browsing**
- Artefacts can be recovered after a private session (April 2016)

**Tracking**
- November 2017 & May 2018
- New tracking vectors
Protection from web threats?

Control Availability
- Popular controls absent from mobile browsers (September 2013)
- Multiple usability issues in the GUI

Blacklists
- Blacklist unavailable on mobile browsers or ineffective (July 2014)
- Blacklist ineffective (December 2016 & June 2018)

Private browsing
- Artefacts can be recovered after a private session (April 2016)

Tracking
- November 2017 & May 2018
- New tracking vectors
Protection from web threats?

<table>
<thead>
<tr>
<th>Control Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Popular controls absent from mobile browsers (September 2013)</td>
</tr>
<tr>
<td>• Multiple usability issues in the GUI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Blacklists</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Blacklist unavailable on mobile browsers or ineffective (July 2014)</td>
</tr>
<tr>
<td>• Blacklist ineffective (December 2016 &amp; June 2018)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Private browsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Artefacts can be recovered after a private session (April 2016)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>• November 2017 &amp; May 2018</td>
</tr>
<tr>
<td>• New tracking vectors</td>
</tr>
</tbody>
</table>
Protection from web threats?

Control Availability
- Popular controls absent from mobile browsers (September 2013)
- Multiple usability issues in the GUI

Blacklists
- Blacklist unavailable on mobile browsers or ineffective (July 2014)
- Blacklist ineffective (December 2016 & June 2018)

Private browsing
- Artefacts can be recovered after a private session (April 2016)

Tracking
- November 2017 & May 2018
- New tracking vectors
Tracking

• Web tracking is not new
  – Madrigal. I'm Being Followed: How Google—and 104 Other Companies—are Tracking Me on the Web, link

• Today?
Tracking

DATA GATHERED SINCE: NOV 18 2018
YOU HAVE VISITED: 15 SITES
YOU HAVE CONNECTED WITH: 152 THIRD PARTY SITES

Recent Site
GRAPH VIEW
Tracking

• Client-side tracking is not new
  – Madrigal. I'm Being Followed: How Google—and 104 Other Companies—are Tracking Me on the Web, link

• Different tracking vectors
  – Cookies, Flash cookies, Silverlight, ...
  – HTML 5.0 storage
HTML 5.0 client-side technologies

• Focus
  – Web Storage, Web SQL Database, Indexed Database API

• Have not received the same level of attention
  – Infrequent use or no use as tracking vector
  – Should be treated as cookies
Used for tracking?

1. Frequency of their use?

2. How often used for tracking?
Methodology
Methodology: Architecture

Diagram showing the flow of data from the World Wide Web, Alexa's List, and HTTP Archive, leading to a blacklist merger, tracking domains, Google BigQuery, and an SQL query console.
Frequency of use

APIs often found as 3\textsuperscript{rd} party subresource ($N=460K$)

<table>
<thead>
<tr>
<th>Client-side storage API</th>
<th>Websites with construct in subresource (%)</th>
<th>Websites with construct in 3rd party subresource (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Storage</td>
<td>71.66</td>
<td>65.39</td>
</tr>
<tr>
<td>IndexedDB</td>
<td>5.56</td>
<td>5.15</td>
</tr>
<tr>
<td>Web SQL DB</td>
<td>1.34</td>
<td>1.18</td>
</tr>
</tbody>
</table>
Tracking?

Tracking is their main use case

<table>
<thead>
<tr>
<th>API/ Subresources using the API that are flagged as ‘tracker’ (%)</th>
<th>Whole Dataset (May 2018)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Storage</td>
<td>71.18</td>
</tr>
<tr>
<td>IndexedDB</td>
<td>31.87</td>
</tr>
<tr>
<td>Web SQL DB</td>
<td>53.59</td>
</tr>
</tbody>
</table>
Pervasiveness?

High percentage of websites containing at least one tracking subresource \((N=460K)\)

<table>
<thead>
<tr>
<th>API / Websites with at least one tracking subresource using API (%)</th>
<th>Whole Dataset (May 2018)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Storage</td>
<td>57.72</td>
</tr>
<tr>
<td>IndexedDB</td>
<td>1.68</td>
</tr>
<tr>
<td>Web SQL DB</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Browser Protection

• Can I erase them like cookies?
  – Tested all popular desktop and mobile browsers
  – Windows, Mac OS
  – Android, iOS, Windows Phone
Methodology

https://github.com/stefano-belloro/storage-watcher
Clearing browsing data might not be enough

1. Data from these APIs might not be removed
2. Extra step in the GUI is required
<table>
<thead>
<tr>
<th>Issue</th>
<th>OS</th>
<th>Browser</th>
<th>APIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data persists after clearing local data</td>
<td>iOS 10.2.1</td>
<td>Safari, Chrome 62.0</td>
<td>IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Android 6</td>
<td>Firefox 57, Firefox 60</td>
<td>IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Android 7</td>
<td>Firefox 54, Firefox 57</td>
<td>IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Android 8</td>
<td>Firefox 60</td>
<td>IndexedDB</td>
</tr>
<tr>
<td>Data deletion requires extra step in the UI</td>
<td>Windows Phone 8.10 by HTC</td>
<td>Internet Explorer</td>
<td>IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Mac OS 10.12.5</td>
<td>Firefox 57.0 (quantum), Firefox 56.0</td>
<td>IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Windows 10</td>
<td>Firefox 56</td>
<td>IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Windows XP</td>
<td>Firefox 47</td>
<td>LocalStorage, IndexedDB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Firefox 56, 57</td>
<td>IndexedDB</td>
</tr>
</tbody>
</table>
Private session might not be enough

1. Data persists after closing private mode or guest mode
2. Data from a private session leaked to normal session
<table>
<thead>
<tr>
<th>Issue</th>
<th>OS</th>
<th>Browser</th>
<th>APIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data persists after closing private session</td>
<td>iOS 11.1.2</td>
<td>Opera 16</td>
<td>LocalStorage</td>
</tr>
<tr>
<td></td>
<td>Android 6</td>
<td>Opera 43.0</td>
<td>IndexedDB, Web SQL</td>
</tr>
<tr>
<td></td>
<td>MiuiBrowser 9.1.3</td>
<td></td>
<td>LocalStorage, IndexedDB</td>
</tr>
<tr>
<td></td>
<td>Android 7</td>
<td>Opera 42.7, Opera 43.0</td>
<td>IndexedDB, Web SQL</td>
</tr>
<tr>
<td></td>
<td>Android 8</td>
<td>Opera 46.3</td>
<td>IndexedDB, Web SQL</td>
</tr>
<tr>
<td>Values from non-private session are leaked</td>
<td>Android 6</td>
<td>MiuiBrowser 9.1.3</td>
<td>IndexedDB</td>
</tr>
<tr>
<td>Data stored in guest mode is deleted only after quitting the browser</td>
<td>Mac OS 10.10.5, Windows 10</td>
<td>Chrome 62</td>
<td>localStorage, IndexedDB, Web SQL</td>
</tr>
</tbody>
</table>
Submitted bugs...

• Most of the bugs that we found have been patched 😊
  – Users might not update their OS or app 😞

• Newer versions of the browser introduce other bugs 😞
  – Noticed this in our experiments
  – Bugs appear and disappear in newer versions! 😞
Demo

Android 8
• Firefox 65.0
• Opera 49.2
More info

Questions

Now!

Later:

• Alexios Mylonas, amylonas@bournemouth.ac.uk, alexios.mylonas@gmail.com

• Steafano Belloro, stefano.belloro@gmail.com