I Do Not Know What You Visited Last Summer: Protecting users from stateful third-party web tracking with TrackingFree Browser

Xiang Pan§, Yinzhi Cao†, Yan Chen§

§ Northwestern University
† Columbia University
<table>
<thead>
<tr>
<th>Roadmap</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Introduction &amp; Background</td>
</tr>
<tr>
<td>• System Design</td>
</tr>
<tr>
<td>• Evaluation</td>
</tr>
<tr>
<td>• Conclusion</td>
</tr>
</tbody>
</table>
Web Tracking

User

Referer: http://online.wsj.com/
Cookie: id = 12345

Referer: http://www.cnn.com/
Cookie: id = 12345

Tracking server
Web Tracking is Prevalent and Serious

- Prevalent
  - More than 90% of Alexa Top 500 websites [Roesner, NSDI 2012].
  - A web page usually has multiple tracking elements.
Web Tracking is Prevalent and Serious

- Prevalent
- More than 90% of Alexa Top 500 websites [Roesner, NSDI 2012].

- A web page usually has multiple tracking elements.

- Jobs Growth Jumps as Economy Gains Steam

- GM Weighs Equal Treatment for Victims in Recall

- Berkshire Heir Apparent: Buffett's Eldest Son


- How Exposed Are You to the New Russian Sanctions?

- Risk & Compliance

- How Exposed Are You to the New Russian Sanctions?

- Risk & Compliance

- How Exposed Are You to the New Russian Sanctions?

- Risk & Compliance

- How Exposed Are You to the New Russian Sanctions?
Web Tracking is Prevalent and Serious

- **Prevalent**
  - More than 90% of Alexa Top 500 websites [Roesner, NSDI 2012].
  - A web page usually has multiple tracking elements.

- **Serious**
  - Not only browsing history, but also other sensitive information such as location, name and email, will be leaked out.
  - Potential for abuse is enormous.
No Effective Defense Approach

- Disable third-party cookie
  - Many other storages to store user’s identifier.

- Blacklist-based anti-tracking tools
  - Priori knowledge of tracking servers.

- *Do-Not-Track* header
  - No enforcement.
Core Idea: TrackingFree partitions client-side states into multiple isolation units so that the identifiers still exist but not unique any more!
Roadmap

• Introduction & Background
• System Design
• Evaluation
• Conclusion
Regular Browser Architecture

Persistent Storage

Cookie
HTML5 Local Storage
Plugins
User Configuration
Cache

books.ebay.com
books.amazon.com

toys.ebay.com
toys.amazon.com

deals.ebay.com
deals.amazon.com

deals.ebay.com
deals.amazon.com

10
TrackingFree Architecture

Principal Kernel Interface

Kernel

Principal Backend

Message Policy Enforcer
Principal Manager
Public History Manager
Preference Configuration Manager
Domain Data Manager

Persistent Storage
Cookie
HTML5 Local Storage
Cache
Plugins
User Configuration

Principal

books.amazon.com
books.ebay.com

toys.amazon.com
toys.ebay.com

deals.amazon.com
deals.ebay.com

deals.amazon.com

deals.ebay.com

toys.amazon.com

toys.ebay.com

deals.amazon.com

deals.ebay.com

toys.amazon.com

tenant.

tenant.

tenant.

tenant.

tenant.
Contents Allocation Mechanism

- **Initial Contents Allocation**
  - Handles those top frames that are navigated by users directly.
  - Based on registered domain name (e.g. google.com, sina.com.cn).

- **Derivative Contents Allocation**
  - Handles those frames that are generated due to the contents on other frames, which we call child frame.
Derivative Contents Allocation

- **Principal Switch**
  - Cross-domain
  - User-triggered

- **Principal Selection**
  - Maintains an in-degree-bounded graph for principals.
  - The in-degree of the graph is set to two.
Derivative Contents Allocation

- Principal Switch
- Cross-domain Navigation
- User-triggered Operations
- Principal Selection

Maintains an in-degree-bounded graph for principals. The in-degree of the graph is set to two.

Diagram:
- Principal Backend
- Connections to:
  - (1) a.com
  - (2) b.com
  - (3) c.com
  - (4) d.com
  - (5) b.com
  - (6) c.com
  - (7) b.com
  - (10) b.com
  - (11) c.com

Note: Connections marked with an "X" indicate some form of restriction or oversight in the allocation process.
Roadmap

- Introduction & Background
- System Design
- Evaluation
- Conclusion
Evaluation

- Anti-tracking capability
  - Formal proof
  - Experiments with real world websites

- Compatibility

- Performance
  - Latency
  - Memory usage
  - Disk usage
Formal Proof

• Methodology
  • Use Alloy to formally analyze TrackingFree’s anti-tracking ability.
  • Describe TrackingFree’s behaviors on an existing Alloy web model [Akhawe et al. CSF 2010].

• Results
  • Formally verified that trackers can correlate TrackingFree user’s activities up to three principals without site collaboration.
Anti-tracking Capability with Real World Websites

• Re-implemented an in-complete but accurate tracking token detection approach proposed on [Roesner et al. NSDI 2012].

• The approach is based on the observation that each tracking request must contain the user’s globally unique identifier.
## Anti-tracking Capability with Real World Web Sites

<table>
<thead>
<tr>
<th>Tracking Host</th>
<th>Prevalence (# Domains)</th>
<th>Tracking Token(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>b.scorecardresearch.com</td>
<td>133</td>
<td>UIDR</td>
</tr>
<tr>
<td>ad.doubleclick.net</td>
<td>117</td>
<td>id, __gads</td>
</tr>
<tr>
<td>ib.adnxs.com</td>
<td>75</td>
<td>anj</td>
</tr>
<tr>
<td>p.twitter.com</td>
<td>70</td>
<td>__utma</td>
</tr>
<tr>
<td>cm.g.doubleclick.net</td>
<td>56</td>
<td>id</td>
</tr>
<tr>
<td>ad.yieldmanager.com</td>
<td>52</td>
<td>bx</td>
</tr>
<tr>
<td>bs.serving-sys.com</td>
<td>40</td>
<td>A4</td>
</tr>
<tr>
<td>cdn.api.twitter.com</td>
<td>40</td>
<td>__utmz</td>
</tr>
<tr>
<td>secure-us.imrworldwide.com</td>
<td>38</td>
<td>IMRID</td>
</tr>
<tr>
<td>adfarm.mediaplex.com</td>
<td>31</td>
<td>svid</td>
</tr>
</tbody>
</table>

Top 10 Tracking Hosts
Anti-tracking Capability with Real World Web Sites

<table>
<thead>
<tr>
<th>Tracking Host</th>
<th>Prevalence (# Domains)</th>
<th>Tracking Token(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>b.scorecardresearch.com</td>
<td>133</td>
<td>UIDR</td>
</tr>
<tr>
<td>ad.doubleclick.net</td>
<td>117</td>
<td>id, __gads</td>
</tr>
<tr>
<td>ib.adnxs.com</td>
<td>75</td>
<td>anj</td>
</tr>
<tr>
<td>p.twitter.com</td>
<td>70</td>
<td>__utma</td>
</tr>
<tr>
<td>bs.serving-sys.com</td>
<td>40</td>
<td>A4</td>
</tr>
<tr>
<td>cdn.api.twitter.com</td>
<td>40</td>
<td>__utmz</td>
</tr>
<tr>
<td>secure-us.imrworldwide.com</td>
<td>38</td>
<td>IMRID</td>
</tr>
<tr>
<td>adfarm.mediaplex.com</td>
<td>31</td>
<td>svid</td>
</tr>
</tbody>
</table>

- TrackingFree eliminated all of them.

Top 10 Tracking Hosts
Compatibility

- Manually tested TrackingFree’s compatibility on 69 third-party services from Alexa Top 50 websites.

<table>
<thead>
<tr>
<th>Name</th>
<th>Example</th>
<th># Succeeded Instance</th>
<th># Total Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-site online payment</td>
<td>Purchase on Ebay and make payment on Paypal</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cross-site content sharing</td>
<td>Share Youtube video to Facebook account</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Single sign-on</td>
<td>Using Facebook account to login Yahoo</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>Overall Results</td>
<td></td>
<td>68</td>
<td>69</td>
</tr>
</tbody>
</table>
## Performance

<table>
<thead>
<tr>
<th>Source</th>
<th>Overhead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>~3%–~20%</td>
</tr>
<tr>
<td>Memory</td>
<td>~25MB/principal</td>
</tr>
<tr>
<td>Disk</td>
<td>~0.6MB/principal</td>
</tr>
</tbody>
</table>
Roadmap

• Introduction & Background
• System Design
• Evaluation
• Conclusion
Conclusion

• We designed and implemented TrackingFree browser which completely protects users from third-party web tracking by isolating web contents.

• We theoretically and experimentally proved TrackingFree’s anti-tracking capability.

• TrackingFree is backward compatible with existing websites.
Thanks & Questions?
http://list.cs.northwestern.edu/WebSecurity
Backup slides …
Out-of-scope threats

- Within-site Tracking
- Tracking by exploiting browser vulnerabilities
- Stateless tracking
Preference Configuration Manager

- User preference can be abused to store tracking identifier. (e.g. strict transport security)

- Completely isolating user preference affects user preference.

- Our solution:
  - Isolate user preference.
  - Apply user-initiated changes to all of the principals.
  - Monitor GUI message to determine user-initiated preference change.
## Related Work

<table>
<thead>
<tr>
<th>Browser</th>
<th>Isolation Mechanism</th>
<th>Contents Allocation Mechanism</th>
<th>Anti-tracking Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>IE8</td>
<td>In-memory Isolation</td>
<td>Tab based</td>
<td>No</td>
</tr>
<tr>
<td>Chromium</td>
<td>In-memory Isolation</td>
<td>Top-frame based</td>
<td>No</td>
</tr>
<tr>
<td>Gazelle</td>
<td>In-memory Isolation</td>
<td>SOP based</td>
<td>No</td>
</tr>
<tr>
<td>OP</td>
<td>In-memory Isolation</td>
<td>Web Page based</td>
<td>No</td>
</tr>
<tr>
<td>AppIsolation</td>
<td>Technique-specific Storage</td>
<td>User Configuration based</td>
<td>Not complete</td>
</tr>
<tr>
<td>Tahoma</td>
<td>Virtual Machine</td>
<td>User Configuration based</td>
<td>Not complete</td>
</tr>
<tr>
<td>Stainless</td>
<td>Technique-specific Storage</td>
<td>User Configuration based</td>
<td>Not complete</td>
</tr>
<tr>
<td>Fluid, MultiFirefox</td>
<td>Profile</td>
<td>User Configuration based</td>
<td>Not complete</td>
</tr>
<tr>
<td>TrackingFree</td>
<td>Profile</td>
<td>Indegree-bounded Principal Graph based</td>
<td>Complete</td>
</tr>
</tbody>
</table>
Principal Switch

- Two intuitive yet extreme policies:
  - Not privacy-preserving (no switch)
  - Unnecessary overhead (switch all the time)

- Our solution: switch principal only if the following two conditions are met:
  - Cross-site
  - User-triggered
Principal Switch

Same principal

Different principal
Principal Selection

• Two intuitive yet extreme policies:
  • Break compatibility (always create new principal)
  • Break anti-tracking capacity (create at most one principal for each domain)
Principal Selection

- Two intuitive yet extreme policies:
  - Break compatibility (always create new principal)
  - Break anti-tracking capacity (create at most one principal for each domain)
Principal Selection

• Two intuitive yet extreme policies:
  • Break compatibility (always create new principal)
  • Break anti-tracking capacity (create at most one principal for each domain)

• Our solution:
  • Maintains an in-degree-bounded graph for principals.
  • The in-degree of the graph is set to two.
Principal Selection

• Two intuitive yet extreme policies:
  • Break compatibility (always create new principal)
  • Break anti-tracking capacity (create at most one principal per domain)

• Our solution:
  • Maintains an in-degree-bounded graph for principals.
  • The in-degree of the graph is set to two.

(1) a.com
(2) b.com
(3) c.com
(4) d.com
(5) b.com
(6) c.com
(7) b.com
(8) 
(9) 
(10) b.com
(11) c.com

Principal Backend
Principal Communication

- Explicit communication is widely used, but break the isolation mechanism.
- Our solution: we restrict the use of explicit communication as follows:
  - Third-party elements in one principle can not explicitly communicate with other principals.
  - First-party elements can only explicitly communicate with the first-party elements placed in its neighbor principals.
Explicit communication is widely used, but breaks...
Principal Communication

- Implicit Communication
  - History sharing (e.g. history, bookmarks)
  - User preference sharing
  - Communication through navigation URL parameters
Case study: Logging Yahoo using Facebook Account

Server-side

(1). Logging using FB
(2). Data (Y) Carried on URL
(3). Data (Y) Carried on URL

Client-side

Principal Yahoo

Local Storage

Principal Facebook

Local Storage

(4). Data (Y) Carried on URL
(5). Logging FB
(6). Data (F) Carried on URL
(7). Data (F) Carried on URL
(8). Data (F) Carried on URL
Performance

(1). Address Bar Navigation without Principal
Avg. Overhead 8.29%
(2). Address Bar Navigation with Principal
Avg. Overhead 3.36%
Performance

(3). Cross Site Navigation
Avg. Overhead 19.43%
Performance

(4). Within-site Navigation
Avg. Overhead 4.70%
## Performance

<table>
<thead>
<tr>
<th>Latency Overhead Source</th>
<th>Cost (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal Construction</td>
<td>322.36</td>
</tr>
<tr>
<td>Extra IPC</td>
<td>349.06</td>
</tr>
<tr>
<td>Render/JS Engine Instrumentation</td>
<td>139.21</td>
</tr>
</tbody>
</table>

Overall Overhead: ~3% - ~20%
## Memory/Disk Overhead

### Memory Overhead on 12 Web Pages (~25MB/Principal)

<table>
<thead>
<tr>
<th>Memory</th>
<th>Chromium</th>
<th>TrackingFree</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Principal</td>
<td>477.1 MB</td>
<td>505 MB</td>
<td>27.9 MB</td>
</tr>
<tr>
<td>4 Principals</td>
<td>623.6 MB</td>
<td>702.8 MB</td>
<td>79.2 MB</td>
</tr>
<tr>
<td>12 Principals</td>
<td>434.6 MB</td>
<td>642.5 MB</td>
<td>297.9 MB</td>
</tr>
</tbody>
</table>

### Disk Overhead on 12 Web Pages (~0.6MB/Principal)

<table>
<thead>
<tr>
<th>Memory</th>
<th>Chromium</th>
<th>TrackingFree</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Principal</td>
<td>21.3 MB</td>
<td>21.8 MB</td>
<td>0.5 MB</td>
</tr>
<tr>
<td>4 Principals</td>
<td>22.5 MB</td>
<td>25.9 MB</td>
<td>3.4 MB</td>
</tr>
<tr>
<td>12 Principals</td>
<td>23.7 MB</td>
<td>29.4 MB</td>
<td>5.7 MB</td>
</tr>
</tbody>
</table>