The State of Web Privacy
Threats & Defenses

Umar Iqbal
University of Washington
@umaarr6
Targeted Ads

https://www.shoeperfect.com > athletics_and_sneakers

Men's Running Shoes - Shoe Carnival
Shop running shoes for men at Shoe Carnival. Shop great deals on men's running shoes, in store & online. Members get FREE shipping!

https://www.coleshaan.com > Men > Men's Performance

Men's Running, Training, & All-Terrain Shoes | Cole Haan
Men's running shoes combine everything there is to love about undoubtedly stylish running-shoes and comfort whether you are going long distance routes or ...

https://www.t3.com > Features > Fitness

Best running shoes 2022 from Adidas, Nike, ASICS and more
Best running shoes 2022 with running trainers from Adidas, Nike, ASICS and more. The best running shoes for jogging, training and racing reviewed and ranked by ...
How are advertisers able to infer your interests?

What technical mechanisms do advertisers use to get that information?

How prevalent are these practices on the internet?

How can users protect their privacy on the internet?
Third-party tracking

Websites load 100s of third-party organizations

- Medical history
- Financial situation
- Weight conscious
- Gambler
- Religious cause donor
- Drug addict
- Single mom

Third party organizations on multiple websites

Infer sensitive information

DataCloudOptOut.oracle.com
User data is sold online!

Online Behavior Advertising

Data Brokers

$120+ billion yearly revenue [IAB]

Sold for less than $1K
The modular web

Add functionality by embedding third-party organizations
Advertising & tracking services

- Google AdSense
- DoubleClick
- Taboola
- MediaMath
- Outbrain
- Bluekai
- Acxiom
- LiveRamp
- OpenX
- theTradeDesk
- PubMatic
- IAS
- Chartbeat
- criteo
- Invitemedia
- DIGILANT
“Trusted” organizations

DoubleClick, criteo, AppNexus, Taboola, inviteMedia, JWPlayer, PubMatic, Quantcast, Google AdSense, acxiom, mopub, BlueKai, Akamai, Outbrain, MediaMath, jQuery, Chartbeat, theTradeDesk, DISQUS, Google Tag Manager, TAPAD, Bootstrap, Advertising, Amazon, LiveRamp, Experian, Choozle, Integral, IAS, Magnite, InMobi, YouTube.
Tracking mechanisms
Cookies

Store state in the browser
Shopping carts, User preferences, Online tracking

Third-party cookies: belong to the domain of embedded content (image, iframes)
Tracking prevalence\textsuperscript{[1]}

82\% of all websites contain at least one tracker

\textbf{Number of trackers per website}

Web Almanac 2022: Privacy

- \(\sim 16\%\) contain one tracker
- \(\sim 10\%\) contain 4 tracker
- Long tail with more than 10 trackers

\textsuperscript{[1]} Measurements conducted on \(\sim 8\) Million websites: [https://almanac.httparchive.org/en/2022/privacy](https://almanac.httparchive.org/en/2022/privacy)
## Tracking prevalence[1]

**Most common trackers**

Web Almanac 2022: Privacy

<table>
<thead>
<tr>
<th>Tracker</th>
<th>Percent of pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>google_analytics (site-analytics)</td>
<td>60%</td>
</tr>
<tr>
<td>google (advertising)</td>
<td>49%</td>
</tr>
<tr>
<td>doubleclick (advertising)</td>
<td>46%</td>
</tr>
<tr>
<td>facebook (advertising)</td>
<td>28%</td>
</tr>
<tr>
<td>google_adservices (advertising)</td>
<td>21%</td>
</tr>
<tr>
<td>google_syndication (advertising)</td>
<td>12%</td>
</tr>
<tr>
<td>wordpress_stats (site-analytics)</td>
<td>6%</td>
</tr>
<tr>
<td>twitter (social_media)</td>
<td>5%</td>
</tr>
<tr>
<td>adobe_audience_manager (advertising)</td>
<td>5%</td>
</tr>
<tr>
<td>rubicon (advertising)</td>
<td>5%</td>
</tr>
</tbody>
</table>

Trackers share data with each other

Even with trackers not present on the website to enhance their coverage!
What are request chains?

A request chain is a sequence of requests and responses between a client and one or more servers. In the diagram, the user requests NEWS.COM, which is redirected to LOCAL-NEWS.COM. LOCAL-NEWS.COM then responds with a 200 OK status code, completing the request chain.
Cookie syncing

Tracking information shared across origin

Same-origin policy
Tracking prevalence[1][2]

doubleclick.net synced its cookies with 118+ third parties (2016)

Privacy protections by main-stream browsers

How can users protect their privacy?
Cookie-based tracking
Impact of cookie blocking
Cookie Blocking

Full Third-Party Cookie Blocking and More

Today’s Firefox Blocks Third-Party Tracking Cookies

Building a more private web: A path towards making third party cookies obsolete
Bounce tracking: bypassing cookie blocking

Has user visited this domain as a website?

Third-party cookie access granted
Bounce tracking prevalence[1]

~14 vendors involved in this practice on top 10K websites

<table>
<thead>
<tr>
<th>Top Domains</th>
<th>Spoofed Safari # of websites</th>
</tr>
</thead>
<tbody>
<tr>
<td>googleadservices.com</td>
<td>3073</td>
</tr>
<tr>
<td>adsrvr.org</td>
<td>1377</td>
</tr>
<tr>
<td>adform.net</td>
<td>322</td>
</tr>
<tr>
<td>flashtalking.com</td>
<td>141</td>
</tr>
<tr>
<td>queue-it.net</td>
<td>9</td>
</tr>
<tr>
<td>ojrq.net</td>
<td>-</td>
</tr>
<tr>
<td>bngpt.com (NSFW)</td>
<td>2</td>
</tr>
</tbody>
</table>

Impact of cookie blocking

Database

DoubleClick

NewYorkPost

FoxNews

Brietbart

DoubleClick

CNN

NewYorkPost

ABC News

DoubleClick

NBC News

CNN

CBS News
Browser fingerprinting: alternative tracking techniques

https://amiunique.org
Browser fingerprinting prevalence[1]

>1/3rd of popular websites deploy fingerprinting scripts

<table>
<thead>
<tr>
<th>Rank Interval</th>
<th>Websites (count)</th>
<th>Websites (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 1K</td>
<td>266</td>
<td>30.60%</td>
</tr>
<tr>
<td>1K to 10K</td>
<td>2,010</td>
<td>24.45%</td>
</tr>
<tr>
<td>10K to 20K</td>
<td>981</td>
<td>11.10%</td>
</tr>
<tr>
<td>20K to 50K</td>
<td>2,378</td>
<td>8.92%</td>
</tr>
<tr>
<td>50K to 100K</td>
<td>3,405</td>
<td>7.70%</td>
</tr>
<tr>
<td>1 to 100K</td>
<td>9,040</td>
<td>10.18%</td>
</tr>
</tbody>
</table>

Browser fingerprinting prevalence

<table>
<thead>
<tr>
<th>Vendor Domain</th>
<th>Tracker</th>
<th>Websites (count)</th>
</tr>
</thead>
<tbody>
<tr>
<td>doubleverify.com</td>
<td>Y</td>
<td>2,130</td>
</tr>
<tr>
<td>adsafeprotected.com</td>
<td>Y</td>
<td>1,363</td>
</tr>
<tr>
<td>alicdn.com</td>
<td>N</td>
<td>523</td>
</tr>
<tr>
<td>adesco.re</td>
<td>N</td>
<td>395</td>
</tr>
<tr>
<td>yimg.com</td>
<td>Y</td>
<td>246</td>
</tr>
<tr>
<td>2,344 others</td>
<td>Y(86)</td>
<td>5,702</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>10,359 (9,040 distinct)</strong></td>
</tr>
</tbody>
</table>

Popular trackers & Ad fraud services

- DoubleVerify
- Integral Ad Service

Fingerprinting services usage

Privacy protections by browser extensions

How can users protect their privacy?
Privacy-enhancing tools

Browser extensions & Privacy focused browsers

EasyList

Filter lists of known advertising and tracking domains

Crowdsourced

Heuristic based

[canvas fingerprinting]
canvas width, height > 15 px calls to save, restore = None [Blocked]

daijiworld.com
#6825 opened 4 days ago by ghajini

indiatimes.com
#6822 opened 4 days ago by ghajini

bikeroar.com
#6790 opened 6 days ago by DandelionSprout

[canvas fingerprinting] canvas width, height > 15 px calls to save, restore = None [Blocked]
Tracker blocking

(a) No Blocking

(b) Filter Lists
Limitations of ad/tracker blockers [1]

Brittle

Mistakes

var n = "UNICODE STRING",
  i = e.getContext("2d");
  i.save(), i.rect(0,0,10,10), i.rect(2,2,6,6),
  t.push(!) === i.isPointInPath(5, 5, "evenodd")
  ? "yes" : "no"), i.restore(), i.save();
  var I = 1.createLinearGradient(0, 0, 200, 0);
  ...;
  i.shadowColor="rgb(85,85,85)", i.shadowBlur=3,
  i.arc(500,15,10,0,2*Math.PI,10), i.stroke(),
  i.closePath(), i.restore(), t.push(e.toDataURL())

Manual

Slow

Redundant

Circumvention

Why is it challenging to protect?

Adversarial Evasions

Scalability
How can we address this problem?

Flawed approach
Cannot possibly put all the trackers in the world in a list!

Fundamentally change in approach
May be detect tracking behavior instead of trackers?
Key insight

Execution traces as a signature of tracking behaviors

Interaction with different elements on a webpage
External network requests initiated by a script

......
Visibility through observing execution behavior

Key insight

```javascript
iframe = document.createElement("iframe");
iframe.src = "tracker.com/load_tracker.html";
document.body.appendChild(iframe);
</script>

<!-- Canvas font fingerprinting script -->
Fonts = ["monospace", ... , "sans-serif"];
CanvasElem = document.createElement("canvas");
CanvasElem.width = "100";
CanvasElem.height = "100";
context = CanvasElem.getContext('2d');
FPDict = {};
for (i = 0; i < Fonts.length; i++) {
    CanvasElem.font = Fonts[i];
    FPDict[Fonts[i]] = context.measureText("example").width;
}

var img = document.createElement("img");
img.src = "tracker.com/track_user/?userId={FPDict}";
</iframe>
```

Visibility of filter lists

Request to a third party with device parameters

Request to a third party to load the iframe markup

Inline Script

Created an iframe HTML

Loaded iframe content

Visibility through observing execution behavior
Browser instrumentation & Machine learning

Instrument browsers + Machine learning based detection

Crucial for robustness
- Capture execution provenance at multiple layers
- Reveal any evasion attempts

Crucial for scalability
- Detect tracking behavior
- Generalize behaviors across trackers
Regulations

Ask for user permission before collecting their data
Regulations

Manage Consent Preferences

+ Strictly Necessary Cookies
  - Necessary
  
  Necessary cookies help make a website usable by enabling basic functions like page navigation and access to secure areas of the website. The website cannot function properly without these cookies.

+ Analytics Cookies

+ Functional Cookies

+ Targeting Cookies

Data Rights Protocol
Standardizing and streamlining consumer data rights requests
The State of Web Privacy
Threats & Defenses

Umar Iqbal
umariqbal.com
@umaarr6