Exploring the Cookieverse: A Multi-Perspective Analysis of Web Cookies

Passive and Active Measurement Conference 2023

Ali Rasaii
Shivani Singh
Devashish Gosain
Oliver Gasser
Washington’s bank rescue fails to erase all doubts
Background
Background
Background
Background

- General Data Protection Regulation (GDPR)
  - EU law
  - May 25, 2018
  - Enhance individuals' control and rights over their personal data
Background

● General Data Protection Regulation (GDPR)
  • EU law
  • May 25, 2018
  • Enhance individuals' control and rights over their personal data

● California Consumer Privacy Act (CCPA)
  • California State law
  • January 1, 2020
Background

● General Data Protection Regulation (GDPR)
  • EU law
  • May 25, 2018
  • Enhance individuals' control and rights over their personal data

● California Consumer Privacy Act (CCPA)
  • California State law
  • January 1, 2020

● Enterprises and vendors react
  • Cookies banners
Research Questions
Research Questions

● How many websites do show the banners?
Research Questions

- How many websites do show the banners?
  - Does the banner give users the options to explicitly agree or decline?
Research Questions

- How many websites do show the banners?
  - Does the banner give users the options to explicitly agree or decline?

- Do they respect the users preferences?
Research Questions

- How many websites do show the banners?
  - Does the banner give users the options to explicitly agree or decline?

- Do they respect the users preferences?

- Do websites exhibit different behavior?
Research Questions

● How many websites do show the banners?
  • Does the banner give users the options to explicitly agree or decline?

● Do they respect the users preferences?

● Do websites exhibit different behavior?
  • Geographic location (EU vs. non-EU)
Research Questions

● How many websites do show the banners?
  • Does the banner give users the options to explicitly agree or decline?

● Do they respect the users preferences?

● Do websites exhibit different behavior?
  • Geographic location (EU vs. non-EU)
  • User agent (mobile vs. desktop)
Related works
Related works

- Not considering users’ characteristics
- Manual or semi-automated inspection
Related works

- Not considering users’ characteristics
- Manual or semi-automated inspection
- Hard-to-scale automated tool
- Just “accept” option
Related works

- Not considering users’ characteristics
- Manual or semi-automated inspection
- Hard-to-scale automated tool
- Just “accept” option

Goal: analysis of cookie landscape from different perspectives in a automated way
Methodology

START  open next domain  keep collecting cookies for fixed time

interact with banner

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

collect cookies

eotarget List?

FINISH  attempt to detect banner
Methodology

START → open next domain → EOTarget List?

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

FINISH
Methodology

START \[\rightarrow\] open next domain \[\rightarrow\] EOTarget List? \[\rightarrow\] collect cookies \[\rightarrow\] FINISH

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...
Methodology

open next domain

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

collect cookies

attempt to detect banner
Methodology
Methodology

START

open next domain

keep collecting cookies for fixed time

interact with banner

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

collect cookies

attempt to detect banner

EOTarget List?

FINISH
Methodology

- Open next domain

Keep collecting cookies for fixed time

Interact with banner

Target List:
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

Collect cookies

Attempt to detect banner

EOTarget List?

Start

Finish
Methodology

START

open next domain

keep collecting cookies for fixed time

interact with banner

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

collect cookies

attempt to detect banner

FINISH

EOTarget List?
Methodology

1. Start
2. Open next domain
3. Collect cookies
4. Keep collecting cookies for fixed time
5. Attempt to detect banner
6. Interact with banner

Target List:
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...
Methodology

**Open next domain**

- keep collecting cookies for fixed time
- interact with banner

**Collect cookies**

- attempt to detect banner

**Target List**

1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

**Start**

**Finish**

BannerClick

Results

Conclusion
Methodology

OpenWPM

START

open next domain

keep collecting cookies for fixed time

Target List
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

FINISH

EOTarget List?

collect cookies

interact with banner

attempt to detect banner
Methodology

**OpenWPM**

- **START**
- **open next domain**
- **FINISH**
- **collect cookies**
- **EOTarget List?**
- **keep collecting cookies for fixed time**

**BannerClick**

- **interact with banner**
- **attempt to detect banner**

**Target List**
1. google.com
2. youtube.com
3. facebook.com
4. netflix.com
...

---

**Background**

**Methodology**

**BannerClick**

**Results**

**Conclusion**
BannerClick
BannerClick
BannerClick – Detection

Background
Methodology
BannerClick
Results
Conclusion
BannerClick – Detection

Background
Methodology
BannerClick
Results
Conclusion
BannerClick – Detection

Cookies Related Text
BannerClick – Detection
BannerClick – Detection

“Cookies”
“Accept” + “Legitimate Interest”
“Accept” + “Privacy Policy”

End Node
BannerClick – Detection

```
div.firstnode {
  position: fixed;
  width: 300px;
  border: 3px
}
```

End Node

“Cookies”
“Accept” + “Legitimate Interest”
“Accept” + “Privacy Policy”
BannerClick – Detection

```
div.firstnode {
    position: fixed;
    width: 300px;
    border: 3px
}
```

“Cookies”
“Accept” + “Legitimate Interest”
“Accept” + “Privacy Policy”
BannerClick – Detection

```
div.firstnode {
  position: fixed;
  width: 300px;
  border: 3px
}
```

---

**Background**

**Methodology**

**BannerClick**

**Results**

**Conclusion**

---

“Cookies”
“Accept” + “Legitimate Interest”
“Accept” + “Privacy Policy”

---

End Node

Optimal Node
BannerClick – Detection

```css
div.firstnode {
  position: fixed;
  width: 300px;
  border: 3px
}
```

---

**“Cookies”**

**“Accept” + “Legitimate Interest”**

**“Accept” + “Privacy Policy”**

---

---

---

---

---

---

---

---

---

---

---
BannerClick – Detection

div.firstnode {
  position: fixed;
  width: 300px;
  border: 3px
}

OPTIMAL NODE

End Node

“Cookies”
“Accept” + “Legitimate Interest”
“Accept” + “Privacy Policy”

99% Accuracy
BannerClick – Interaction
BannerClick – Interaction

Word sets:
acc: ["accept","agree","confirm"]
BannerClick – Interaction

Word sets:
acc: ["accept","agree","confirm"]
rej: ["reject","refuse","decline"]
BannerClick – Interaction

Word sets:
acc: ["accept","agree","confirm"]
rej: ["reject","refuse","decline"]
set: ["setting","option","choice"]
Measurement Setup
Measurement Setup

- **8 vantage points:** Germany, Sweden, US West, US East, India, Brazil, South Africa, Australia
Measurement Setup

- **8 vantage points:** Germany, Sweden, US West, US East, India, Brazil, South Africa, Australia
- **Target list:** Tranco Top 10k domains
Banners detected, accepted, rejected
Banners detected, accepted, rejected

---

**EU**

- **Det.**
- **Det. △ Acc.**
- **Det. △ Acc. △ Rej.**
More banners in EU compared to non-EU countries
Banners detected, accepted, rejected

![Bar chart showing the fraction of websites in different regions for detected (Det.), detected and accepted (Det. ∩ Acc.), and detected, accepted, and rejected (Det. ∩ Acc. ∩ Rej.) banners.](chart.png)
Banners detected, accepted, rejected

Half of the websites do not show any reject option
Banners detected, accepted, rejected
Cookies differences after interaction

![Graph showing differences in average number of cookies across interaction states: No interaction, Accepted, and Rejected. The y-axis represents the average number of cookies on a logarithmic scale ranging from $10^0$ to $10^2$, and the x-axis represents the interaction states. The bars indicate the range of values with error bars for each category.]
Cookies differences after interaction

Interacting with banners impacts cookie distribution
Cookies differences after interaction

Interacting with banners impacts cookie distribution
Cookies differences after interaction

Interacting with banners impacts cookie distribution
Cookies differences after interaction

Interacting with banners impacts cookie distribution
Cookies differences after interaction

Interacting with banners impacts cookie distribution
Cookies differences after interaction

Interacting with banners impacts cookie distribution
Cookies – EU vs. non-EU

![Graph showing the fraction of websites vs. average number of cookies for EU and non-EU regions.](image)

**Y-axis:** Fraction of websites

**X-axis:** Average number of cookies

Legend:
- Blue line: No interaction EU
- Orange line: No interaction non-EU
Cookies – EU vs. non-EU
Cookies – EU vs. non-EU

Fewer cookies in EU compared to non-EU
Cookies – EU vs. non-EU

Fewer cookies in EU compared to non-EU
Cookies – EU vs. non-EU

Fewer cookies in EU compared to non-EU
Conclusion
Conclusion

- Impact of
  - Geographical location of users
  - Interaction with banner
Conclusion

- Impact of
  - Geographical location of users
  - Interaction with banner

- Check out the paper
  - CCPA impact
  - Landing vs. Inner pages
  - Mobile vs. Desktop
  - Consistency analysis
Conclusion

- Impact of
  - Geographical location of users
  - Interaction with banner

- Check out the paper
  - CCPA impact
  - Landing vs. Inner pages
  - Mobile vs. Desktop
  - Consistency analysis

- Source code available
  - BannerClick
  - Analysis data for reproducibility

bannerclick.github.io
Backup: Measurement Setup

● Main run
  • TrancoTop10kList domains ➔ 9020 totally reachable
  • 5 iteration for each mode of interaction

● Desktop vs. mobile (and others)
  • Tranco tiered 300 ➔ top-100, 1001–1100, and 9901–10k
  • User Agent
    ◦ Mobile: “Mozilla/5.0 (Android 12; Mobile; rv:68.0) Gecko/68.0 Firefox/93.0”
    ◦ Desktop: “Mozilla/5.0 (X11; Linux x86_64; rv:95.0) Gecko/20100101 Firefox/95.0”
  • Screen size
    ◦ Mobile: 340x695
    ◦ Desktop: 1366x768
Backup: Measurements

Table 1. Overview of different measurement types.

<table>
<thead>
<tr>
<th>Measurement Type</th>
<th>Start Date</th>
<th>Duration</th>
<th>Target</th>
<th>Websites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banner Interaction</td>
<td>Jan 20, 2022</td>
<td>20 days</td>
<td>Tranco Top 10k</td>
<td></td>
</tr>
<tr>
<td>Consistency Tests</td>
<td>Feb 9, 2022</td>
<td>10 days</td>
<td>Tranco tiered 300</td>
<td></td>
</tr>
<tr>
<td>Landing vs. Inner</td>
<td>Mar 8, 2022</td>
<td>4 days</td>
<td>Tranco tiered 300</td>
<td></td>
</tr>
<tr>
<td>Desktop vs. Mobile</td>
<td>Feb 27, 2022</td>
<td>10 hours</td>
<td>Tranco tiered 300</td>
<td></td>
</tr>
<tr>
<td>Impact of CCPA</td>
<td>Mar 13, 2022</td>
<td>10 hours</td>
<td>Tranco tiered 300</td>
<td></td>
</tr>
</tbody>
</table>
Backup: Words Frequency
Backup: Inner vs Landing

![Graphs showing the comparison between Inner and Landing for Third Party and Tracking metrics across Germany, US East, and Brazil.](image-url)
Backup: Mobile vs Desktop

User agent would impact the number of cookies.
Backup: Cookies – EU vs non-EU

Less cookies in EU compared to non-EU
Backup: Consistency

(a)

Coefficient of variation of TP cookies

Fraction of website-interaction tuples

(b)

Germany  Sweden  Australia  Brazil  India  US East  US West  South Africa

0.0  0.1  0.2  0.3  0.4  0.5
Backup: Consent Management Platform

![Chart showing market share distribution for various consent management platforms.](image-url)
Backup: Explicit vs. Implicit