No boundaries: Exfiltration of personal data by session-replay scripts

Güneş Acar
Steven Englehardt
Arvind Narayanan

This research is funded by NSF grant CNS 1526353 and by an Amazon AWS Cloud Credits for Research grant.
What are session replay scripts?

“Watch recordings of your visitors’ sessions. Discover how they browse as if you're looking over their shoulder!” (clicktale.com)
Why use session replay scripts?

- Who are my most valuable customers?
- Who added items to the cart but didn’t convert?
- Where are users frustrated?
The problem: recordings require a ton of data

- Full page source and text
- Mouse movements & clicks
- Keypresses
Automated redaction...

Excluded Elements

Enter CSS selectors on separate lines, describing elements to be excluded from capture. Learn more

<table>
<thead>
<tr>
<th>Redacted Field</th>
<th>FullStory</th>
<th>UserReplay</th>
<th>SessionCam</th>
<th>Hotjar</th>
<th>Yandex</th>
<th>Smartlook</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Email</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Phone</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Address</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>SSN</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>DOB</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Password</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>CC Number</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>CC CVC</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>CC Expiry</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

*CC Number is represented by a placeholder image in the document.*
Manual redaction

Easily protect your user’s privacy.

Exclude sensitive customer data from ever leaving your customer’s browser by using our in-app point and click system.
How can things go wrong?

We found session recordings containing:

- Passwords (Propeller Ads, …)
- Credit Card details (Bonobos)
- Student data (Gradescope)
- Health data (Walgreens)
- Purchase details (Lenovo)
Passwords leak to Fullstory on PropellerAds
Filter bypass due to unexpected input types

Demo from: https://codepen.io/geoffreyrose/pen/HKDkB
Passwords leak to Userreplay on Capella University
CC#, CVV and name leak to Fullstory on Bonobos
Prescription details leak to Fullstory on Walgreens

Prescribed drug name is leaked

Redacted
Student details leak to Fullstory on Gradescope

Recordings included:

- Student names
- Student emails
- Student grades
- Professor comments
Session recordings are widespread

- 14+ analytics companies offer recording services
  - Present on 99,174 of the top 1 million sites
- Evidence of recording on 7,918 sites.
  - Likely a lower bound as recording scripts sample users

Session recording present on ~1 - 10% of the top 1 million sites. We found several severe PII leaks after manually reviewing ~30 sites.

→ How many more leaks are out there?
Takeaways

• Recordings contain sensitive information
  – incl. HIPAA, FERPA protected data
• Redaction is difficult and brittle
• Are users comfortable being watched?

Is recording visitors legal?

Yes. Tracking behavior and movement of your visitors using Smartlook is legal, just like using Google Analytics or other services for tracking visitors on your website.

Needing to ask “Is this legal?” should give you pause.
Over 400 of the World's Most Popular Websites Record Your Every Keystroke, Princeton Researchers Find
No boundaries series: Privacy vulnerabilities arising from directly embedded third parties

https://freedom-to-tinker.com/tag/noboundaries/

Session Replay

Autofill abuse

And more....

“No boundaries: Exfiltration of personal data by session-replay scripts” (freedom-to-tinker.com)

“No boundaries for user identities: Web trackers exploit browser login managers” (freedom-to-tinker.com)
Thanks for listening!

Contact:
gunes@princeton.edu
https://gunesacar.net


No boundaries series: https://freedom-to-tinker.com/tag/noboundaries/

Princeton Web Transparency & Accountability Project: https://webtap.princeton.edu/