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

Steven Englehardt
@s_englehardt
senginehardt.com

Joint work with:
Gunes Acar and Arvind Narayanan
No boundaries series: Privacy vulnerabilities arising from directly embedded third parties

Session Recording

Autofill abuse

More to come!

“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)
What are session recording scripts?

- Session recording scripts create a “video” of all of a user’s actions on a site.
- Publishers can later review the videos.
Why use session recording scripts?

Answer questions like:

- Who are my most valuable customers?
- Who added items to the cart but didn’t convert?
- Where do users leave the onboarding flow?
- Where are users frustrated?
More than just site optimization: Jornaya (LeadID) uses recordings to “prove consent” for data collection
The problem: recordings require a ton of data

- Full page source and text
- Mouse movements & clicks
- Keypresses
The problem: pages contain a ton of sensitive data

<table>
<thead>
<tr>
<th>Employment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Multi-System Merchant Services</td>
</tr>
<tr>
<td>Occupation</td>
<td>Sales engineer</td>
</tr>
<tr>
<td>Physical characteristics</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>5'9&quot; (174 centimeters)</td>
</tr>
<tr>
<td>Weight</td>
<td>204.8 pounds (92.1 kilograms)</td>
</tr>
<tr>
<td>Blood type</td>
<td>A+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contact Info</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>John</td>
</tr>
<tr>
<td>Email</td>
<td><a href="mailto:john@example.com">john@example.com</a></td>
</tr>
<tr>
<td>Confirm Email</td>
<td><a href="mailto:john@example.com">john@example.com</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shipping</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Address</td>
<td>123 Any Street</td>
</tr>
<tr>
<td>City</td>
<td>New York</td>
</tr>
<tr>
<td>State</td>
<td>NY</td>
</tr>
<tr>
<td>Country</td>
<td>USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Personal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother's maiden name</td>
<td></td>
</tr>
<tr>
<td>SSN</td>
<td></td>
</tr>
<tr>
<td>Birthday</td>
<td>mm/dd/yyyy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Payment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name on card</td>
<td></td>
</tr>
<tr>
<td>Card Number</td>
<td></td>
</tr>
<tr>
<td>Card Number Last Four</td>
<td></td>
</tr>
<tr>
<td>CVC</td>
<td></td>
</tr>
</tbody>
</table>

Live website

Publisher Dashboard

Demo video: https://youtu.be/mh_NpUu0LS4
Scripts have automated redaction...

Excluded Elements

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

```css
input[type="password"]
[autocorrect="off"]
[autocorrect="on"]
```
Scripts have automated redaction...
Scripts also support 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.
Session recording scripts are (too) easy to integrate

A timer counting how long it takes you to embed their code.

Set up Hotjar with one script in a matter of seconds.

Hotjar works out of the box on most popular platforms including:

00:00:26
Tick tock tick tock... every minute you lose a lot of video material!

DONE, SAVE THE RECORDINGS

Instructions for webmasters that manage your site:

Insert the following code into your website between tags `<head>` and `</head>`. The moment you have a visitor on your website, recordings will start to show within a few minutes.

```
<script type="text/javascript">
    window.smartlook||(function(d)
        var o=smartlook=function(){o.api.push(arguments)},h=d.getElementsByTagName('head')[0]
        var c=d.createElement('script');o.api=new Array();c.async=true;c.type='text/javascript';c.charset='utf-8';c.src='https://rec.smartlook.com/recorder.js';h.appendChild(c);
    )();
    smartlook('init', 'd83e658fa9d18b86892b229b95a2f5a4d84a4e7b');
</script>
```
How can things go wrong?

We found session recordings containing:

- Health data (Walgreens)
- Student data (Gradescope)
- Credit Card data (Bonobos)
- Purchase data (Lenovos)

in a relatively small manual review of sites.
Recording redactions miss sensitive data
Recording includes CVV field → Not PCI compliant?

From: https://www.pcisecuritystandards.org/pdfs/pci_fs_data_storage.pdf
What happened?

**Bonobos used:**

```html
<input type="text"></input>
```

**Bonobos should have used:**

```html
<input type="text" autocomplete="cc-number"></input>
```
Unexpected input types can also cause password leaks

Demo from: https://codepen.io/geoffreyrose/pen/HKDkB
Walgreens misses fields during redaction

Walgreens makes thorough use of redaction
Walgreens makes thorough use of redaction

But prescription information is missed!

(the user’s full name included was not redacted on the previous page)
Recording redactions miss sensitive data

“Walgreens does not retain this data and cannot access or view your answers.”
Recording redactions miss sensitive data

“You are just a few minutes away from having full access to your pharmacy records. Your answers are used only to verify your identity. Walgreens does not retain this data and cannot access or view your answers.

1. Which of the following CITIES have you PREVIOUSLY OR CURRENTLY used as your address?
   - NEW YORK
   - LOS ANGELES
   - CHICAGO
   - HOUSTON
   - NONE OF THE ABOVE

2. Which of the following STREETS have you PREVIOUSLY or CURRENTLY used as your address?
   - ADAMS AVE
   - WASHINGTON AVE

Although selection inputs redacted, mouse trace is still recorded.
Gradescope recordings included:

- Student name
- Student emails
- Student grades
- Professor comments

<table>
<thead>
<tr>
<th>NAME</th>
<th>EMAIL</th>
<th>SCORE/3.0</th>
<th>GRADED?</th>
<th>VIEWED?</th>
<th>TIME (EST)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gunes Acar</td>
<td><a href="mailto:guenes@princeton.edu">guenes@princeton.edu</a></td>
<td>2.5</td>
<td>✔️</td>
<td>✔️</td>
<td>Dec 20 7:21pm</td>
</tr>
<tr>
<td>Steven Englehardt</td>
<td><a href="mailto:ste@princeton.edu">ste@princeton.edu</a></td>
<td>3.0</td>
<td>✔️</td>
<td>✔️</td>
<td>Dec 20 5:06pm</td>
</tr>
</tbody>
</table>
FullStory forbids PII sharing?

1.3 "Sensitive Data” means any information that: (a) requires a high degree of protection by law and where loss or unauthorized disclosure would require notification by Customer to government agencies, individuals or law enforcement, (b) any information that, if made public, could expose individuals to a risk of physical harm, fraud, or identity theft. Sensitive Data includes, but is not limited to, social security numbers or other government-issued identification numbers, financial account numbers, credit card or debit card numbers, CVVs, credit report information or other personal financial information, health or medical information or other information that is subject to international, federal, state, or local laws or ordinances now or hereafter enacted regarding data protection or privacy, including, but not limited to, the Health Insurance Portability and Accountability Act, the Health Information Technology for Economic and Clinical Health Act, the Fair Credit Reporting Act, the Children’s Online Privacy Protection Act and the Gramm-Leach-Bliley Act

First party restricted from sharing sensitive data:

“Customer agrees that it will not provide any Sensitive Data to FullStory.”

https://www.fullstory.com/legal/terms-and-conditions/
Session recordings are widespread

- 14+ analytics company 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?
Recording companies don’t always handle data well

Recordings on HTTPS pages played back over HTTP

- At time of measurement: Yandex, Smartlook, and Hotjar were doing this.
- Smartlook has since fixed this

Password length leaked in recording

- At time of measurement: Smartlook and UserReplay were doing this.
- Smartlook has since fixed this

password → ***************  while  abc123 → ******
Takeaways

1. Recordings contain sensitive information

2. Redaction is difficult and brittle.

Example rule:
form[name="financialInfoForm"] >table:last-child>tbody>tr>td>table:first-child>tbody>tr>td>table>tbody>tr>td>table>tbody>tr>td>table>tbody>tr:nth-child(3)

3. Are users comfortable being watched?

Need to ask “Is this legal?” should give you pause.

Data security & legal info

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.
Thank you!

Blog post series:

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

No boundaries for user identities: Web trackers exploit browser login managers

Website operators are in the dark about privacy violations by third-party scripts

Image assets from the Noun Project: recording by Guru, browser windows by DTDesign, HTML File by Burak Kucukparmaksiz, mouse click by Tomas Knopp, Keyboard by Arthur Shlain