

# 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.*



**PRIVACYCON**

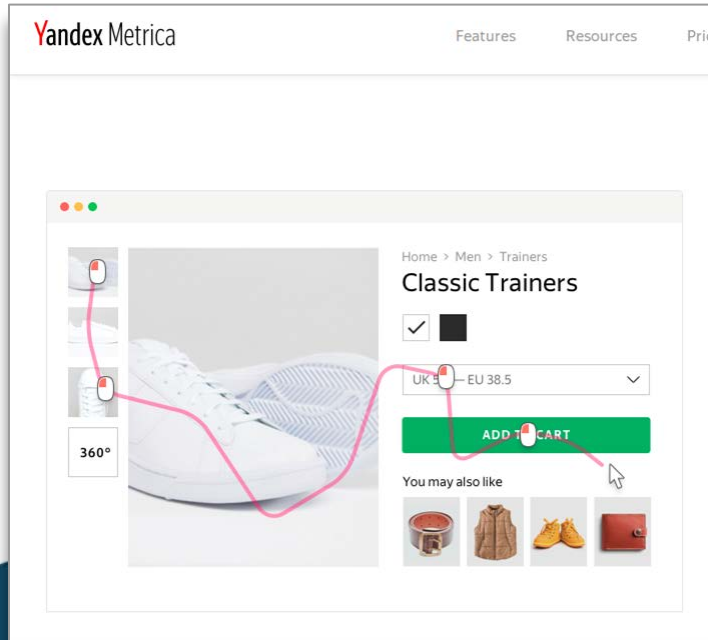
Photo: Dan Saelinger (Prop Stylist: Dominique Baynes)

<https://spectrum.ieee.org/computing/software/browser-fingerprinting-and-the-onlinetracking-arms-race>

# 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...

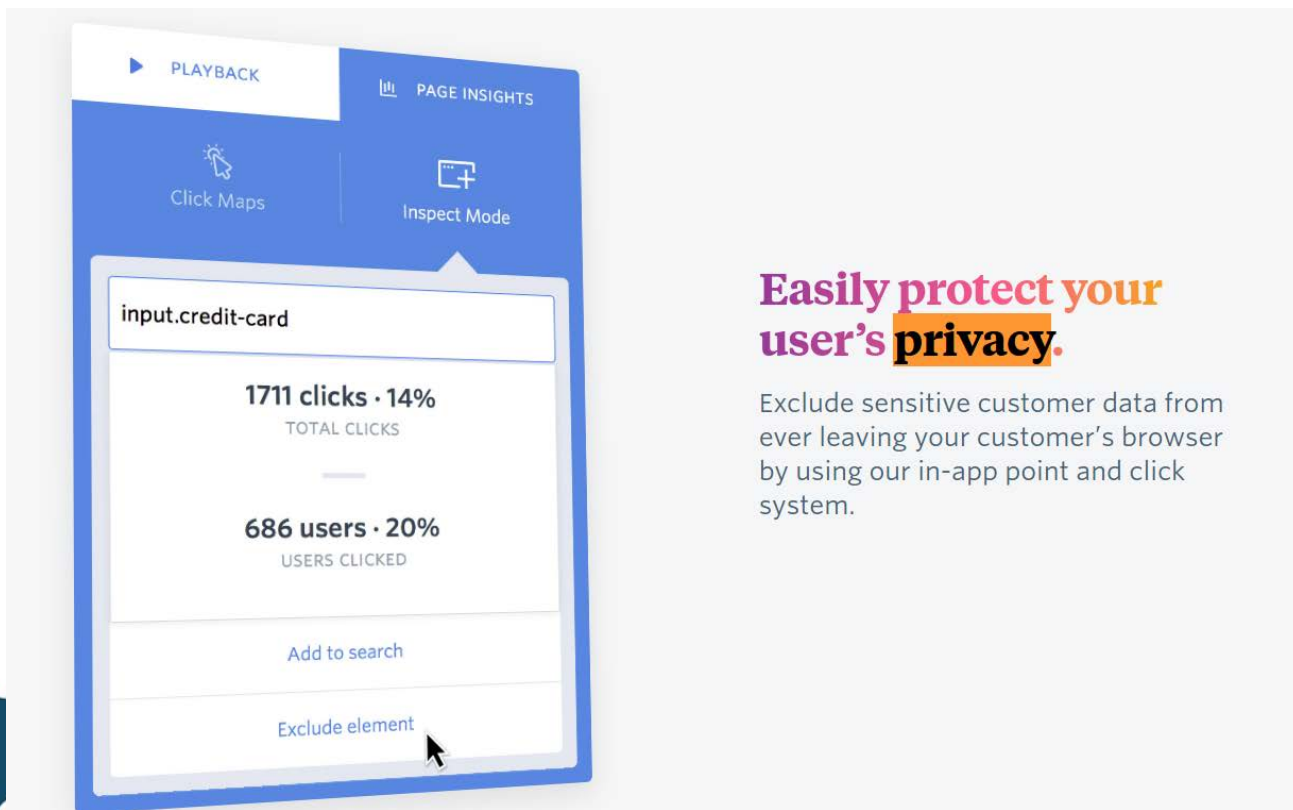
## Excluded Elements

Enter CSS selectors on separate lines, describing elements to be excluded from capture. [Learn more](#)

```
input[type="password"]
[autocomplete="cc-name"]
[autocomplete="cc-given-name"]
[autocomplete="cc-additional-name"]
[autocomplete="cc-family-name"]
[autocomplete="cc-number"]
[autocomplete="cc-exp"]
[autocomplete="cc-exp-month"]
[autocomplete="cc-exp-year"]
[autocomplete="cc-csc"]
[autocomplete="cc-type"]
```

Redacted Field	FullStory	UserReplay	SessionCam	Hotjar	Yandex	Smartlook
Name	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Email	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Phone	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Address	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/> †	<input type="radio"/>	<input type="radio"/>
SSN	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
DOB	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Password	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
CC Number	<input checked="" type="radio"/>	<input checked="" type="radio"/> *	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
CC CVC	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
CC Expiry	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

# 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)



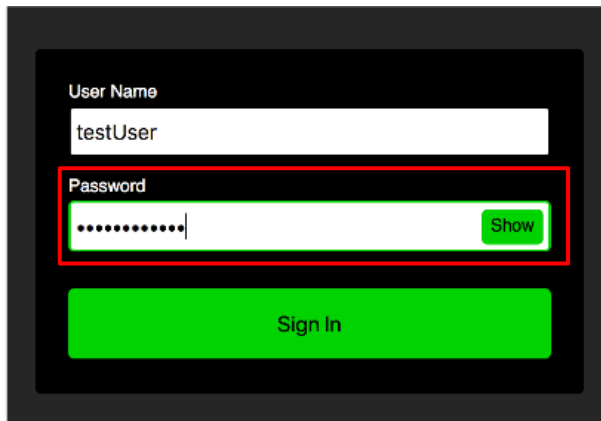
# Passwords leak to Fullstory on PropellerAds

The image shows a web browser window displaying the PropellerAds login page. The page has a logo, a "Welcome back!" message, and a login form with fields for "Username" and "Password". The password field contains the text "1234567890" and is highlighted with a red box. Below the password field is a "Forgot password?" link and a blue "Log in" button. At the bottom, there is a "Don't have an account? Register" link and a "chat" button.

The browser's developer tools are open to the Network tab, showing a list of requests to rs.fullstory.com. The "Response" column for several requests is highlighted with a red box, showing JSON data that includes the password "1234567890".

Name	Response
rs.fullstory.com/rec	7, Args: [2,...]
bundle?OrgId=5T8P0&UserId=5..	18, Args: [122, "12345"]}
rs.fullstory.com/rec	15, Args: [53]}
rs.fullstory.com/rec	14, Args: [54]}
bundle?OrgId=5T8P0&UserId=5..	18, Args: [122, "123456"]}
rs.fullstory.com/rec	15, Args: [54]}
rs.fullstory.com/rec	14, Args: [55]}
bundle?OrgId=5T8P0&UserId=5..	18, Args: [122, "1234567"]}
rs.fullstory.com/rec	15, Args: [55]}
rs.fullstory.com/rec	14, Args: [56]}
bundle?OrgId=5T8P0&UserId=5..	18, Args: [122, "12345678"]}
rs.fullstory.com/rec	15, Args: [56]}
rs.fullstory.com/rec	14, Args: [57]}
bundle?OrgId=5T8P0&UserId=5..	18, Args: [122, "123456789"]}
rs.fullstory.com/rec	15, Args: [57]}
rs.fullstory.com/rec	14, Args: [48]}
rs.fullstory.com/rec	18, Args: [122, "1234567890"]}
rs.fullstory.com/rec	15, Args: [48]}
rs.fullstory.com/rec	3, Args: [253, 231, 98]}
rs.fullstory.com/rec	9, ...]
rs.fullstory.com/rec	9, ...]

# Filter bypass due to unexpected input types

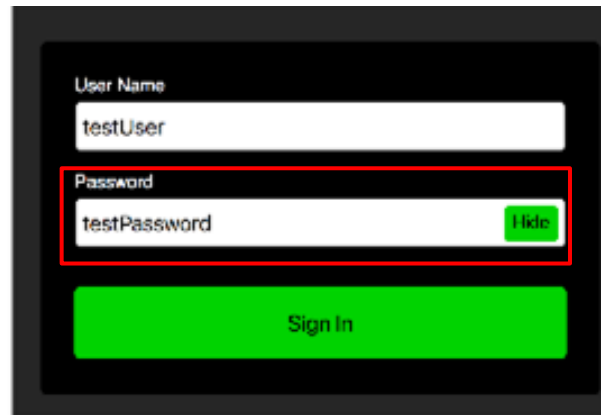


A screenshot of a login form. The 'User Name' field contains 'testUser'. The 'Password' field contains a series of dots, indicating redaction. A green 'Show' button is located to the right of the password field. Below the password field is a large green 'Sign In' button.

`<input type="password">`

(automatically redacted with [type="password"] rule)

Show password  
clicked...



A screenshot of the same login form after the 'Show password' button was clicked. The 'User Name' field still contains 'testUser'. The 'Password' field now displays 'testPassword' in plain text. The green button to the right of the password field is now labeled 'Hide'. The 'Sign In' button remains green.

`<input type="text">`

(automatic redaction fails)

# Passwords leak to Userreplay on Capella University

The image shows a web browser window displaying the Capella University login page. The page has a dark header with the university logo and the text "APPLY FOR ADMISSION Log In". Below the header, there is a login form with fields for "User ID" and a password field (masked with dots). There are links for "Forgot User ID or Resend User ID?" and "Forgot your password?". A blue "LOG IN" button is at the bottom of the form.

The browser's developer tools are open to the Network tab, showing a list of requests. The selected request is from "userreplay" to "userreplay.net". The headers of this request are visible in the right pane:

```
ac-8cad-37d12107e0ec; PS_TOKENEXPIRE=17_Feb_2018_03:10:02_GMT; psback=%22%22url%22%3A%22https%3A%2F%2Fcs.capella.edu%2Fpsc%2Fcsr%2FEMPLOYEE%2F%2F%2Fcu_nac_reg_menu.cu_nac_usr_login.gbl%3Fpage%3DCU_NAC_USR_LOGIN%22%20%22label%22%3A%22Login%20Page%22%20%22origin%22%3A%22PIA%22%20%22layout%22%3A%221%22%22; s_gpvcn=vc%3AAAAdmissions%3AHome; bc_pv_end=; s_getNewRepeat=1518837219871-New; s_sq=capellaproduct%252Ccapellavc%252CcapellavcportalUsers%3D%2526c.%2526a.%2526activitymap.%2526page%2526dvc%25253AAAAdmissions%25253AHome%2526link%25253Dpasswordleak12345678%2526region%25253Dwin0divCU_NAC_USR_WRK_PASSWORD_MIXEDctrl%2526pageIDType%25253D1%2526.activitymap%2526.a%2526c.%2526pid%25253Dvc%25253AAAAdmissions%25253AHome%2526pid%25253D1%2526oid%25253Dfunctiononclick%252528event%252529%25257Bjavascript%25253AcancelBubble%252528event%252529%252538%25257D%2526oid%25253D%2526ot%25253DPASSWORDptoken: 161a1bd7252a9380url: https://cs.capella.edu/psc/cslr/EMPLOYEE/CU_CS/c/CU_NAC_REG_MENU.CU_NAC_USR_LOGIN.GBL?Page=CU_NAC_USR_LOGIN&Action=NrequestStart: 1518837001401currentTime: 1518837219882
```

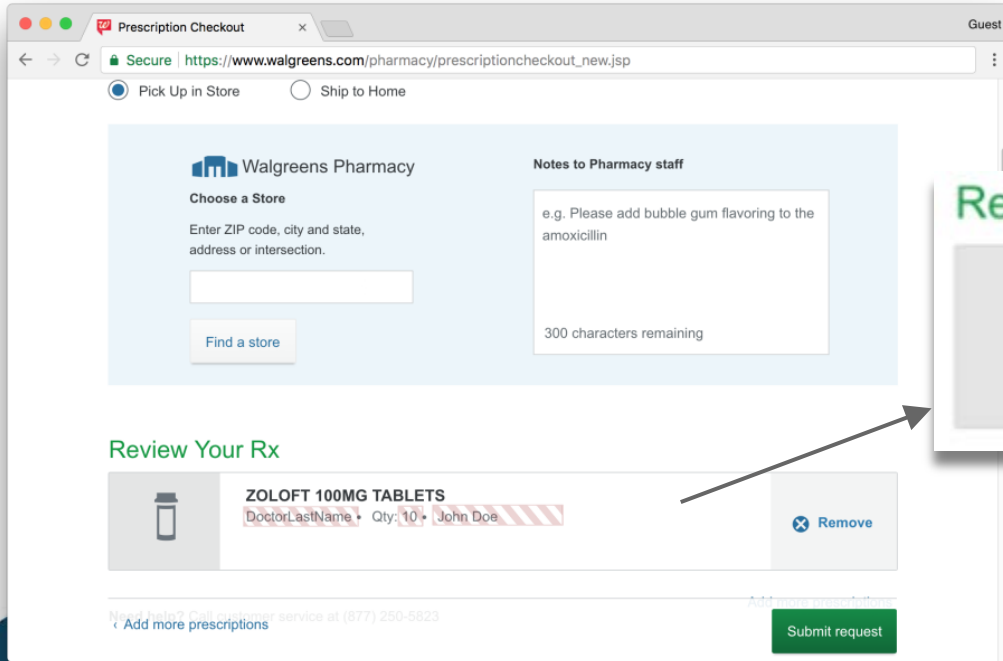
# CC#, CVV and name leak to Fullstory on Bonobos

The screenshot shows the Bonobos account wallet page with a form for adding a new card. The form fields are: NAME (As it appears on your card) with the value "John Doe", CARD NUMBER with the value "4111111111111111 VISA", MONTH with "10", YEAR with "2020", and CVV with "456". The network tab shows a list of requests to fullstory.com. Several requests have their headers highlighted with blue boxes, showing the following values: "John Do", "John Doe", and "4111111111111111".

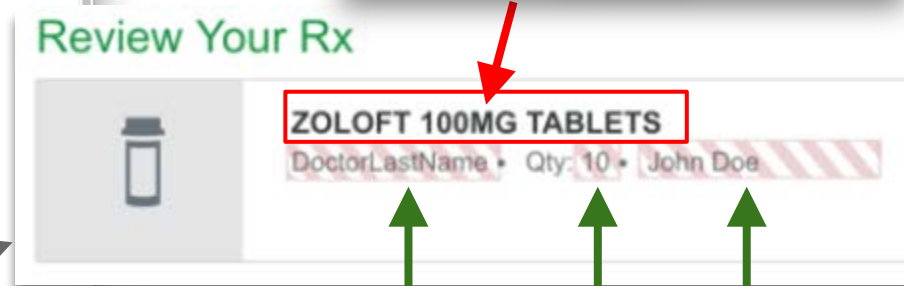
```
response Cookies Timing  
ind: 18, Args: [104, "John Do"]  
ind: 15, Args: [79]  
385442, Args: [1072, "value", "2017-11-11"]  
ind: 14, Args: [69]  
ind: 18, Args: [104, "John Doe"]  
ind: 15, Args: [69]  
385692, Args: [1072, "value", "2017-11-11"]
```

```
args: [72, 313]}  
args: [1047, 72, 313, 45, 29]  
args: [52]  
args: [1047, "4"]}  
args: [52]  
args: [49]  
args: [1047, "41"]}  
args: [49]  
args: [49]  
args: [1047, "411"]}  
args: [49]  
args: [1048, "class",...]}  
args: [1072, "value", "2017-11-11"]  
args: [49]  
args: [1047, "4111"]}  
args: [49]  
args: [1047, "41111"]}  
args: [49]  
args: [1047, "411111"]}  
args: [49]
```

# 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

Gradescope | Review Grades - Mozilla Firefox

Gradescope | Review Gra X +

https://gradescope.com/courses/13530/assignm

gradescope

Review Grades for **Homework1Assignment**

● REGRADE REQUESTS OPEN ● GRADES NOT PUBLISHED

MINIMUM	MEDIAN	MAXIMUM	MEAN	STD DEV
2.5	2.75	3.0	2.75	0.35

2 Students

Search

NAME	EMAIL	SCORE/3.0	GRADED?	VIEWED?	TIME (EST)
Gunes Acar	gunes@princeton.edu	2.5	✓	👁	Dec 20 7:21pm
Steven Englehardt	ste@princeton.edu	3.0	✓	👁	Dec 20 5:06pm

# 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?**

# Takeaways

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

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



Smartlook User Guide

Legal questions

Data security & legal info

GDPR

Opt out of Smartlook

Technical info

## 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.



# Updates

- Walgreens, Gradescope and Bonobos removed session replay script

The screenshot shows the top portion of a BBC News article. At the top left is the BBC logo, followed by a 'Sign in' button and navigation links for News, Sport, Weather, Shop, Earth, and Travel. Below this is a red banner with the word 'NEWS' in white. Underneath the banner is a secondary navigation bar with links for Home, Video, World, US & Canada, UK, Business, Tech, Science, Stories, and Enter. The main content area features the sub-header 'Technology' and the article title 'More than 480 web firms record 'every keystroke'' in large, bold black text. Below the title, it says 'By Jane Wakefield, Technology reporter'. At the bottom left of the article preview, there is a copyright notice '© 21 November 2017'. At the bottom right, there are social media sharing icons for Facebook, Twitter, Messenger, Email, and a 'Share' button.



NITASHA TIKU BUSINESS 11.16.17 06:00 AM

## THE DARK SIDE OF 'REPLAY SESSIONS' THAT RECORD YOUR EVERY MOVE ONLINE

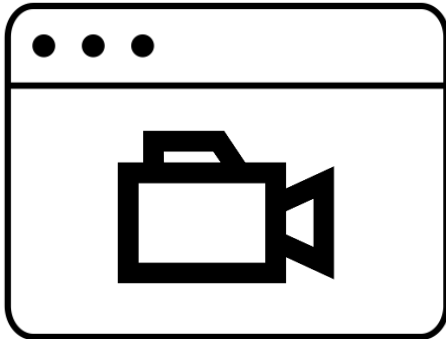
MOTHERBOARD

Over 400 of the World's Most Popular Websites Record Your Every Keystroke, Princeton Researchers Find

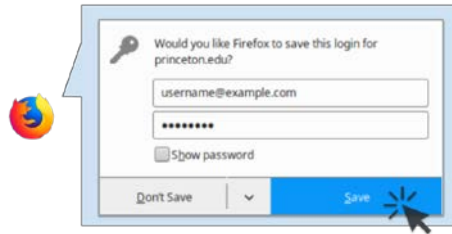
# No boundaries series: Privacy vulnerabilities arising from directly embedded third parties

<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>**

Blog post: <https://freedom-to-tinker.com/2017/11/15/no-boundaries-exfiltration-of-personal-data-by-session-replay-scripts/>

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

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



citp / **OpenWPM**

**PRIVACYCON**

*Image assets from the Noun Project: mouse click by Tomas Knopp, Keyboard by Arthur Shlain, recording by Guru, browser windows by DTDesign, HTML File by Burak Kucukparksiz, Credit by Por Suppasit, Student by Nimal Raj, Password by BomSymbols, Medicine by UNiCORN, Purchase by Ayub Irawan*