

# No boundaries: Data exfiltration by third-party tracking scripts

***Steven Englehardt***

Joint work with:

Güneş Acar, Jeffrey Han, and  
Arvind Narayanan



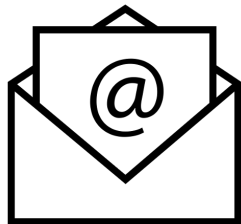
*Disclaimer: I am employed by Mozilla, this talk is in my Princeton capacity and does not necessarily represent Mozilla's views*



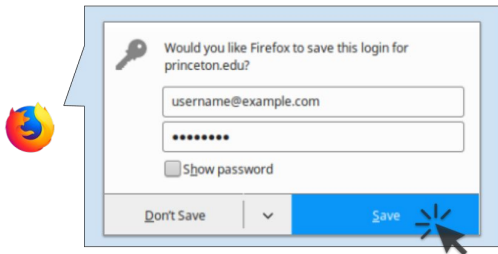
# Third parties collecting PII on the web and in emails

## Email Tracking

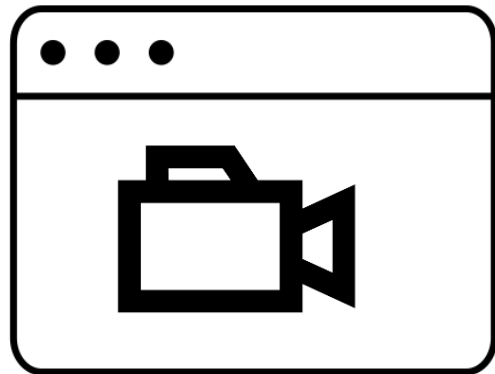
```
UUID = {  
  MD5(bob@example.com),  
  SHA1(bob@example.com),  
  SHA256(bob@example.com)  
}
```



## Autofill abuse



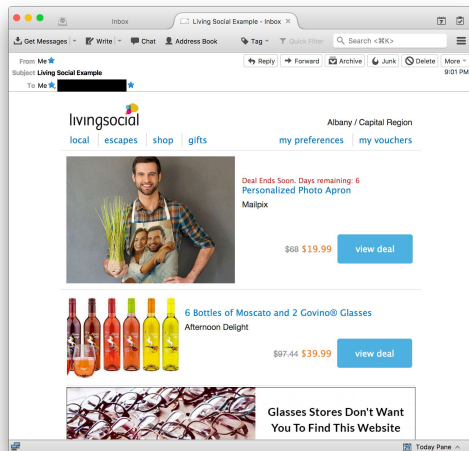
## Session Recording



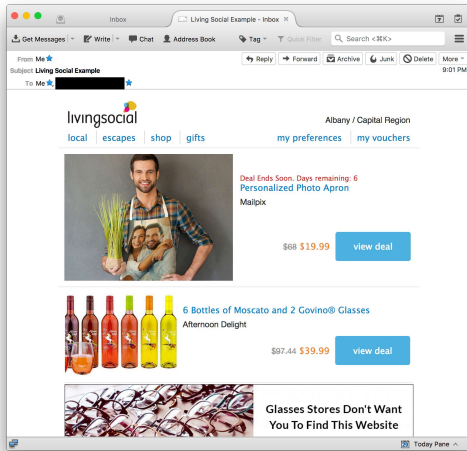
*“No boundaries: Exfiltration of personal data by session-replay scripts” (freedom-to-tinker.com)*

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

*Englehardt, Han, and Narayanan, “I never signed up for this! Privacy implications of email tracking” (PETS 2018)*



# What happens when you load remote content in an email?



Your device contacts 24 companies  
→ 20 can track you (if supported)  
→ 10 receive an MD5 hash of your email address

### Receives MD5(email address) & Sets a Cookie

**American List Counsel** (alcmpn.com)  
**LivelIntent** (liadm.com)  
**Oracle** (nexac.com)  
**Axiom** (rlcdn.com, pippio.com, axiom-online.com)  
**Criteo** (criteo.com)  
**Conversant Media** (dotomi.com)  
**V12 Data** (v12group.com)  
**VideoAmp** (videoamp.com)  
<Unknown> (alocdn.com)

### Sets a Cookie

**OpenX** (openx.net)  
**comScore** (scorecardresearch.com, voicefive.com)  
**Oracle** (bluekai.com)  
**Google** (doubleclick.net)  
**Realtime Targeting Aps** (mojn.com)

**MediaMath** (mathtag.com)  
**TapAd** (tapad.com)  
**IPONWEB** (bidswitch.net)  
**AOL** (advertising.com)  
**Centro** (sitescout.com)  
**The Trade Desk** (adsrvr.org)  
**Adobe** (demdex.net)

### Receives MD5(email addr.)

**Criteo** (emailretargeting.com)  
**Neustar** (agkn.com)

### Receives Bare Request

**LivelIntent** (licasd.com)  
**Google** (2mdn.net)  
**Akamai** (akamai.net)

# LiveIntent Blog Post

Source: <https://blog.liveintent.com/people-based-marketing-not-complicated/>

As an identifier, **email is both deterministic and persistent**. That is, when a consumer gives out a verified email, it usually belongs to only that consumer. That can't be said of all typical advertising identifiers. Cookies, for example, live on desktop browsers that are often shared with no way to distinguish who's using it. And whereas **email is cross-device**, cookies aren't.

# LiveIntent Privacy Policy

Source: <https://liveintent.com/services-privacy-policy>

...we collect non-personal information that does not reveal your specific identity. LiveIntent may also receive non-personal information from **online and offline sources**, including the types described below, from our business partners

# How it works: A redirect chain for a single pixel

Row	Request URL
0	<a href="http://inbox.washingtonexaminer.com/imp?s=...&amp;e=&lt;EMAIL&gt;&amp;p=0">http://inbox.washingtonexaminer.com/imp?s=...&amp;e=&lt;EMAIL&gt;&amp;p=0</a>
1	<a href="http://p.liadm.com/imp?...&amp;m=&lt;MD5&gt;&amp;sh=&lt;SHA1&gt;&amp;sh2=&lt;SHA256&gt;&amp;dom=&lt;EMAIL_DOMAIN&gt;">http://p.liadm.com/imp?...&amp;m=&lt;MD5&gt;&amp;sh=&lt;SHA1&gt;&amp;sh2=&lt;SHA256&gt;&amp;dom=&lt;EMAIL_DOMAIN&gt;</a>
2	<a href="http://x.bidswitch.net/sync?ssp=liveintent&amp;bidder_id=5298&amp;licd=3357&amp;x=EGF.M...">http://x.bidswitch.net/sync?ssp=liveintent&amp;bidder_id=5298&amp;licd=3357&amp;x=EGF.M...</a>
3	<a href="http://x.bidswitch.net/ul_cb/sync?ssp=liveintent&amp;bidder_id=5298&amp;licd=3357&amp;x=EGF.M...">http://x.bidswitch.net/ul_cb/sync?ssp=liveintent&amp;bidder_id=5298&amp;licd=3357&amp;x=EGF.M...</a>
4	<a href="http://p.adsymptotic.com/d/px/?_pid=12688&amp;_psign=d3e69...&amp;bidswitch_ssp_id=liveintent&amp;_redirect=...">http://p.adsymptotic.com/d/px/?_pid=12688&amp;_psign=d3e69...&amp;bidswitch_ssp_id=liveintent&amp;_redirect=...</a>
5	<a href="http://p.adsymptotic.com/d/px/?_pid=12688&amp;_psign=d3e69...&amp;bidswit...&amp;_redirect=...&amp;_expected_cookie=...">http://p.adsymptotic.com/d/px/?_pid=12688&amp;_psign=d3e69...&amp;bidswit...&amp;_redirect=...&amp;_expected_cookie=...</a>
6	<a href="http://x.bidswitch.net/sync?dsp_id=126&amp;user_id=84f3...&amp;ssp=liveintent">http://x.bidswitch.net/sync?dsp_id=126&amp;user_id=84f3...&amp;ssp=liveintent</a>
7	<a href="http://i.liadm.com/s/19751?bidder_id=5298&amp;licd=3357&amp;bidder_uuid=&lt;UUID_1&gt;">http://i.liadm.com/s/19751?bidder_id=5298&amp;licd=3357&amp;bidder_uuid=&lt;UUID_1&gt;</a>
8	<a href="http://cm.g.doubleclick.net/pixel?google_nid=liveintent_dbm&amp;google_cm&amp;google_sc">http://cm.g.doubleclick.net/pixel?google_nid=liveintent_dbm&amp;google_cm&amp;google_sc</a>
9	<a href="http://cm.g.doubleclick.net/pixel?google_nid=liveintent_dbm&amp;google_cm=&amp;google_sc=&amp;google_tc=">http://cm.g.doubleclick.net/pixel?google_nid=liveintent_dbm&amp;google_cm=&amp;google_sc=&amp;google_tc=</a>
10	<a href="http://p.liadm.com/match_g?bidder_id=24314&amp;bidder_uuid=&lt;UUID_2&gt;&amp;google_cver=1">http://p.liadm.com/match_g?bidder_id=24314&amp;bidder_uuid=&lt;UUID_2&gt;&amp;google_cver=1</a>
11	<a href="http://x.bidswitch.net/sync?ssp=liveintent&amp;bidder_id=5298&amp;licd=">http://x.bidswitch.net/sync?ssp=liveintent&amp;bidder_id=5298&amp;licd=</a>
12	<a href="http://pool.udsp.iponweb.net/sync?ssp=bidswitch&amp;bidswitch_ssp_id=liveintent">http://pool.udsp.iponweb.net/sync?ssp=bidswitch&amp;bidswitch_ssp_id=liveintent</a>

# Measuring email tracking at scale

**Sign up for email & get 25% off\***

Email, please

---

Confirm your email

---

**SIGN UP NOW**

\*Valid for first-time registrants only & applies to reg. price items only. [Privacy Policy](#)

Received nearly 13,000  
emails from ~900 sites

# Top recipients of leaked email addresses

29% of emails ( from  
19% of senders) leak  
the email address to  
third parties

Recipient Organization	# of Senders
LiveIntent	68
Acxiom	46
Litmus Software	28
Conversant Media	26
Neustar	24
apxl.com	18
54.211.147.17	18
Tranco	17
WPP	17
54.82.61.160	16



# Why collect email address?

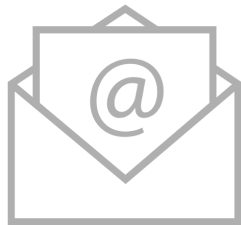
	Recipient Organization	# of Senders
→	LiveIntent	68
→	Acxiom	46
	Litmus Software	28
→	Conversant Media	26
→	Neustar	24
	apxlvr.com	18
	54.211.147.17	18
	Trancos	17
	WPP	17
	54.82.61.160	16

The top email collectors all sell “identity-based” marketing. Allowing advertisers to reach individuals on any device and connect with individual purchase data and other offline data.

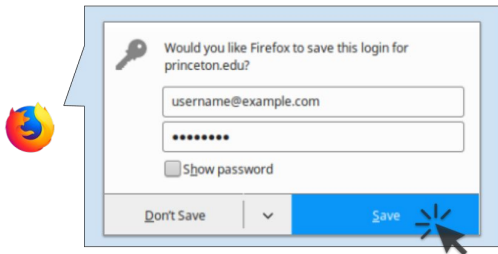
# Third parties collecting PII on the web and in emails

## Email Tracking

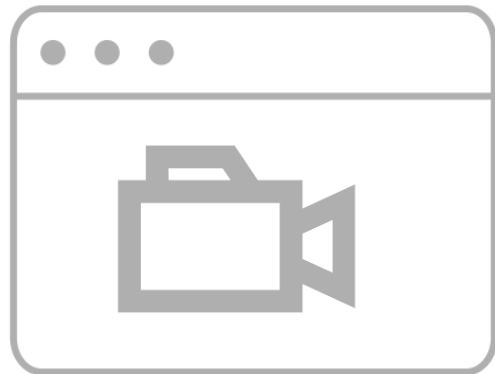
```
UUID = {  
  MD5(bob@example.com),  
  SHA1(bob@example.com),  
  SHA256(bob@example.com)  
}
```



## Autofill abuse



## Session Recording

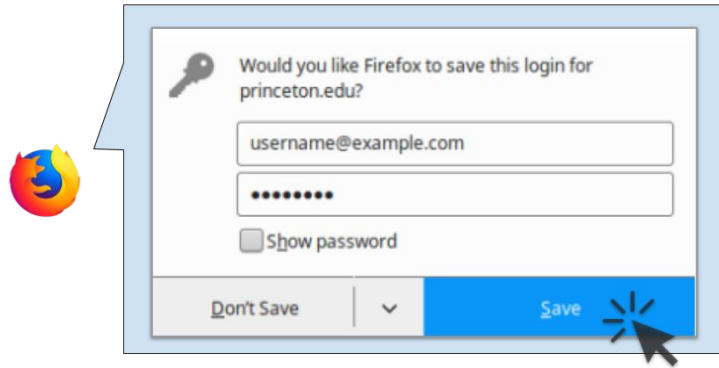


*“No boundaries: Exfiltration of personal data by session-replay scripts” (freedom-to-tinker.com)*

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

*Englehardt, Han, and Narayanan, “I never signed up for this! Privacy implications of email tracking” (PETS 2018)*

# Login manager abuse for web tracking



# Built-in login managers

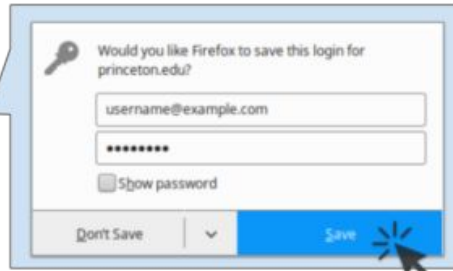
- Remembers username & passwords (opt-in)
- Autofills login forms
- Different than CC and address autofill



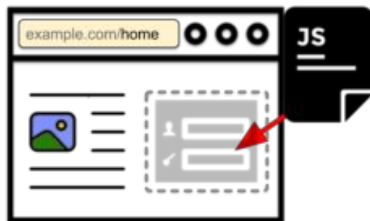
User submits a login or registration form, clicks "Save" to store the credentials.



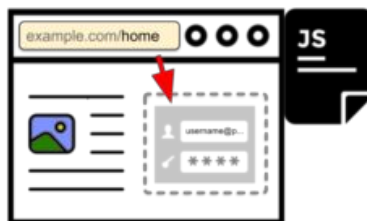
Third-party script  
is **not** present on  
the login page



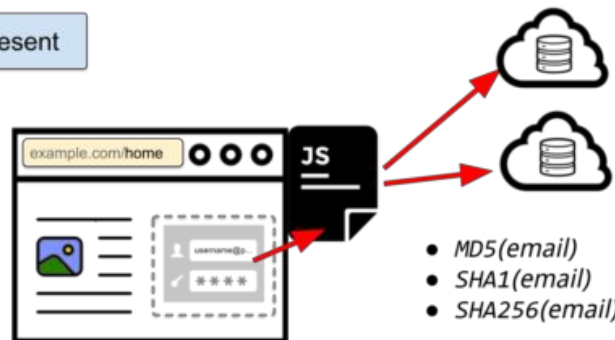
User visits a non-login page on the same site; this time the third party script is present



1. Third-party script injects an invisible login form

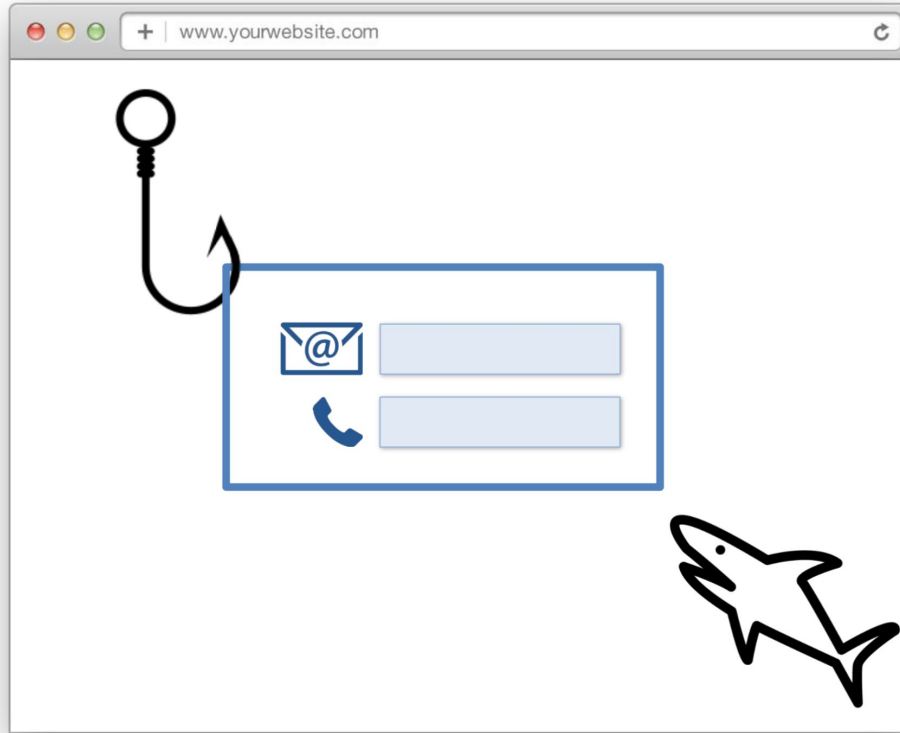


2. Login manager fills in user's email and password



3. The script reads the email address from the form and sends it hashes to third-party servers

# Injecting PII into the web: bait technique



# Instrumentation and measurement



- Crawl 50K sites with OpenWPM, main page and 5 inner pages
- Firefox's nsILoginManager interface to add login credentials
- Mutation events to monitor element insertion (e.g. forms)
- HTMLInputElement instrumentation to intercept access to form input fields
- HTTP instrumentation: request and response headers, POST payloads

OpenWPM

# Findings

Company	Script address	No of sites
Adthink	<a href="https://static.audienceinsights.net/t.js">https://static.audienceinsights.net/t.js</a>	1047
OnAudience	<a href="http://api.behavioralengine.com/scripts/be-init.js">http://api.behavioralengine.com/scripts/be-init.js</a>	63



# Adthink (audienceinsights.net)

- sends MD5, SHA1 and SHA256 hashes of the email address to its server (secure.audienceinsights.net)
- triggers another request containing the MD5 hash of the email to data broker Acxiom (p-eu.acxiom-online.com)

# OnAudience (behavioralengine.com)

- sends the MD5 hash of the email to its server
- also collects hash of browser plugins, MIME types, screen dimensions, language, timezone information, user agent string, OS and CPU information
- 45 of the 63 sites that contain OnAudience script have “.pl” ccTLD



## **BUY** BILLIONS OF USER PROFILES

Mailing Exchange combines programmatic buying with e-mail marketing for fully automatic one-to-one advertising. Moreover, it involves 3rd party data including general interests, purchase intentions, geolocalization, demographics and much more.



## **SELL** MONETIZE YOUR DATABASE

Mailing Exchange gives you a new revenue stream and keeps your business model intact. You have the full control over trading terms and conditions.



### Non-PII data only

We do not collect any personally identifiable information



### Do-Not-Track

We respect DNT headers sent by web browsers



### Opt-Out

We support IAB recommendations

## FIGURES

We own one of the biggest data warehouses in the world.



# LiveIntent Privacy Policy

Source: <https://liveintent.com/services-privacy-policy>

To de-identify this information, either we or our business partners [hash it].

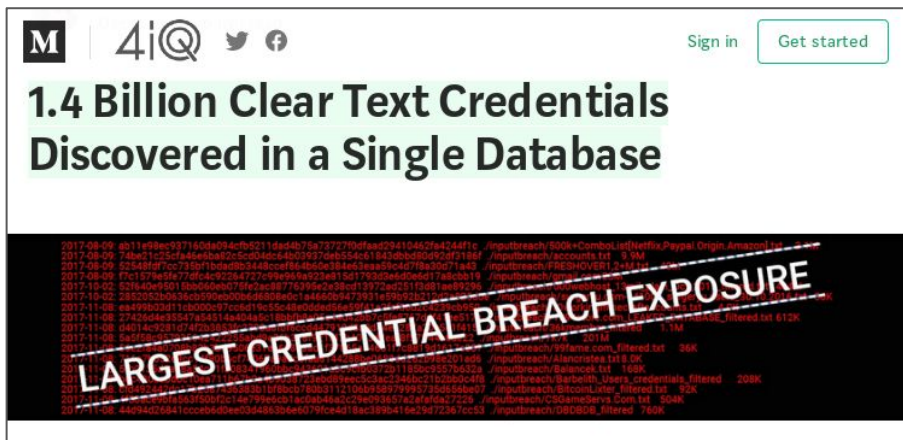
# Criteo Privacy Policy

Source: <https://www.criteo.com/privacy/>

we use a double hashing method ... to ensure the non-reversibility of your information. A hash of your email corresponds to a series of characters that does not permit your identification.

# Email addresses aren't secrets!

Use email database leaks...



The screenshot shows a news article header with a navigation bar containing an 'M' logo, a magnifying glass icon, and social media icons for Twitter and Facebook. The title '1.4 Billion Clear Text Credentials Discovered in a Single Database' is highlighted in green. Below the title is a large red diagonal banner with the text 'LARGEST CREDENTIAL BREACH EXPOSURE' in white. The background of the banner is a dark image with a grid of small, colorful text fragments, likely representing leaked credentials.

...and just guess the rest.


GPU cloud computer: \$24.48 / hour  
→ 450 billion MD5 hashes / second

~4.7 billion email addresses total. If we generate a real address every 1 in 1 million guesses, **we can generate the entire space for less than \$75.**

Past research recovered 45-70% of emails.

# Don't want to guess? Reverse hashes for \$0.04/email

theleadswarehouse.com



String / Original →

theleadswarehouse.com

21ae531dbdb3a09fc726d4e88e965d14

← MD5 Hash

**The Leads Warehouse Does MD5 Reverse**

**Email Encryption:**

- Quickly
- Securely
- Cost-Effectively

infutor.com



## Data Snapshot: MD5 and SHA1 Email Identification and Use Cases

### What is MD5 and SHA1?

MD5 and SHA1 are algorithms used to verify data integrity. Originally created for online security applications to verify data integrity, the MD5 (Message Digest 5) and SHA1 (Secure

datafinder.com



**Datafinder**  
Automated Data Intelligence

Login Signup

## Recover Encrypted Email Addresses

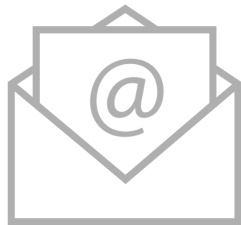
Versium's Email Decryption, starting at \$0.04 per email or \$0.08 with consumer data append

Recover email addresses that have been encrypted using the most common hashing and encryption protocols, with more than a 70% success rate.

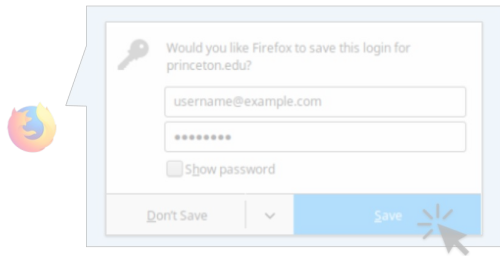
# Third parties collecting PII on the web and in emails

## Email Tracking

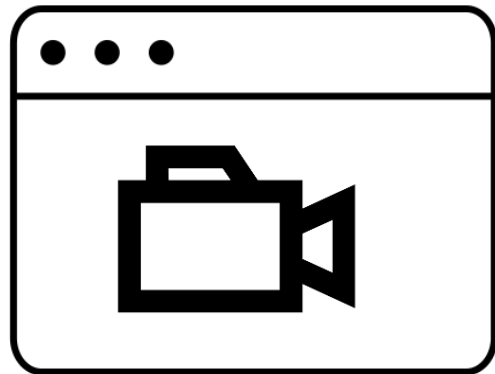
```
UUID = {  
  MD5(bob@example.com),  
  SHA1(bob@example.com),  
  SHA256(bob@example.com)  
}
```



## Autofill abuse



## Session Recording



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

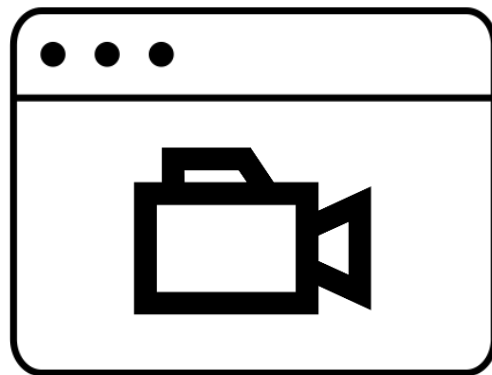
*“No boundaries for user identities: Web trackers exploit browser login managers” (freedom-to-tinker.com)*

*Englehardt, Han, and Narayanan, “I never signed up for this! Privacy implications of email tracking” (PETS 2018)*

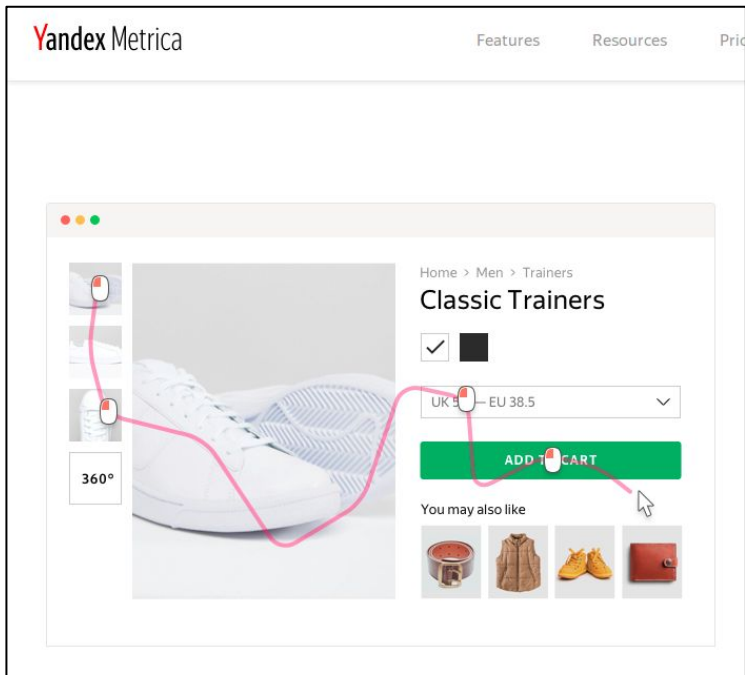


# What are session recording scripts?

- Session recording scripts create a “video” of all of a user’s actions on a site.
  - Key presses
  - Mouse clicks, mouse movements
  - Scrolling behavior...
- 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?

# Exfiltration of personal data by session recordings

FullStory

←

→

localhost:8000

Guest

# Employment

Company

Multi-Systems Merchant Services

Occupation

Sales engineer

## Physical characteristics

Height

5' 9" (174 centimeters)

Weight

204.8 pounds (93.1 kilograms)

Blood type

A+

## Contact Info

Name

John

Email

john@example.com

Confirm Email

john@example.com

Phone

+1-650-450-1212

## Shipping

Address

123 Any Street

City

New York

State

NY

Zip

10011

Country

USA

## Personal

Mother's maiden name

SSN

Birthday

mm/dd/yyyy

Username

Password

Website

Profession

## Payment

Name on card

Full Name

Card Number

Card Number Last Four

CVC

Expiry

MM-YYYY

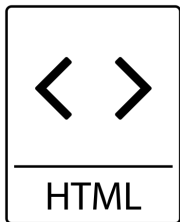
## Live website

[illegible]

# Publisher Dashboard

Demo video: [https://youtu.be/mh\\_NpUu0LS4](https://youtu.be/mh_NpUu0LS4)

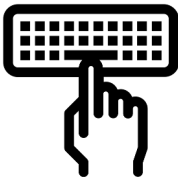
**The problem:** recordings require a **ton** of data



Full page source and text

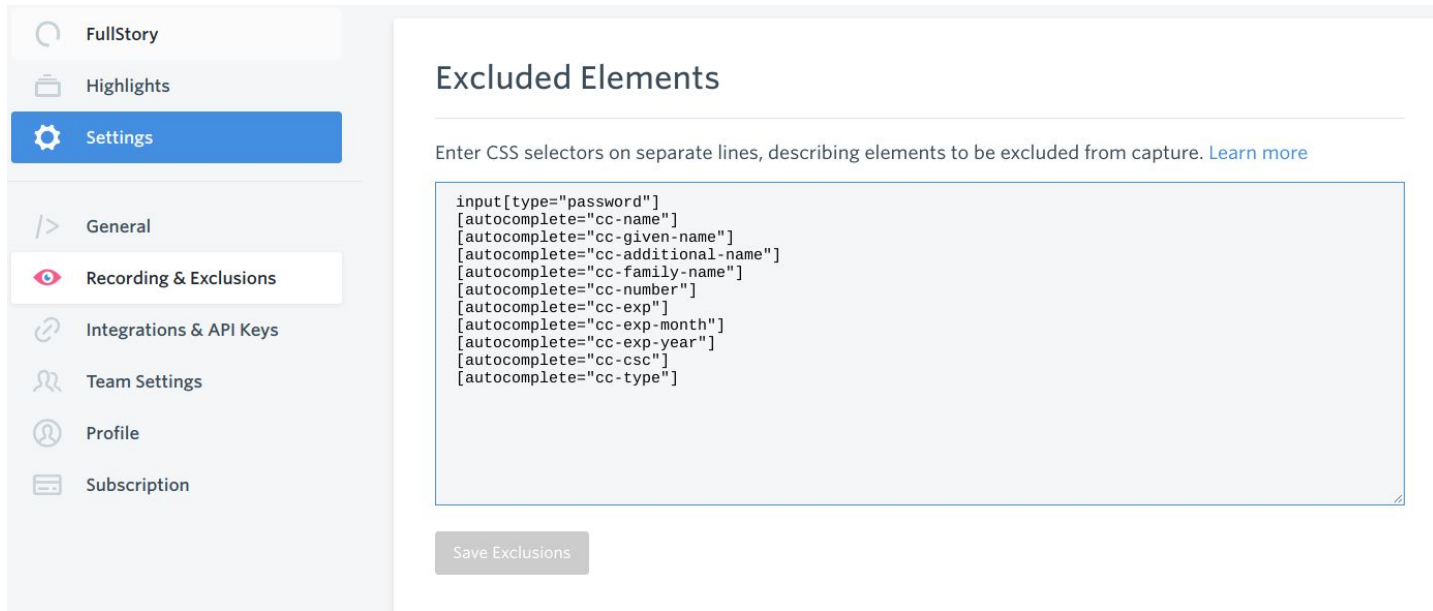


Mouse movements & clicks



Keypresses

# Scripts have automated exclusion...



The screenshot displays the FullStory user interface. On the left is a sidebar with navigation links: FullStory, Highlights, Settings (highlighted in blue), General, Recording & Exclusions (highlighted with a red eye icon), Integrations & API Keys, Team Settings, Profile, and Subscription. The main content area is titled 'Excluded Elements'. Below the title, there is a text prompt: 'Enter CSS selectors on separate lines, describing elements to be excluded from capture. [Learn more](#)'. A large text input field contains the following CSS selectors: `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"]`, and `[autocomplete="cc-type"]`. At the bottom of the main area is a 'Save Exclusions' button.

FullStory

Highlights

Settings

General

Recording & Exclusions

Integrations & API Keys

Team Settings

Profile

Subscription

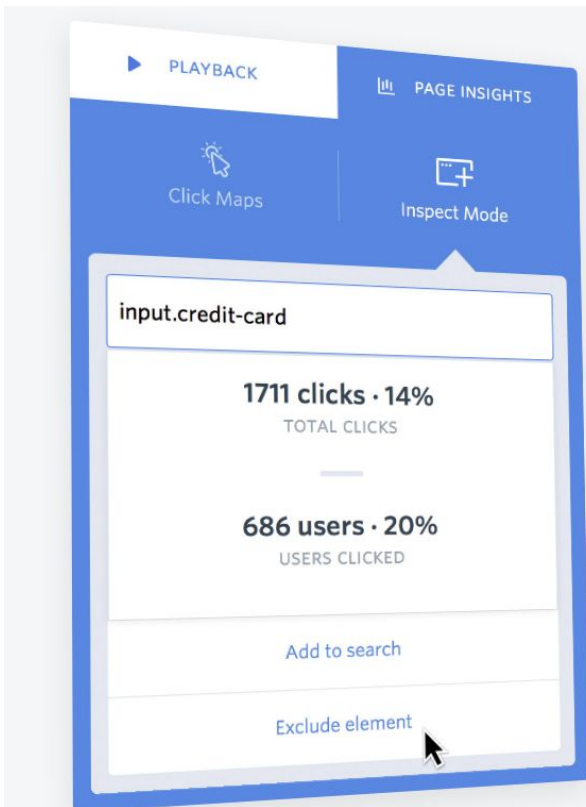
## Excluded Elements

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"]
```

Save Exclusions

# Scripts also support manual redaction



The screenshot shows the Click Maps interface. At the top, there are tabs for 'PLAYBACK' and 'PAGE INSIGHTS'. Below these are icons for 'Click Maps' and 'Inspect Mode'. A white box highlights a specific element, 'input.credit-card'. Below the element name, the statistics are displayed: '1711 clicks · 14%' and 'TOTAL CLICKS'. Below that, '686 users · 20%' and 'USERS CLICKED'. At the bottom of the white box, there are two buttons: 'Add to search' and 'Exclude element'. A mouse cursor is pointing at the 'Exclude element' button.

**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?

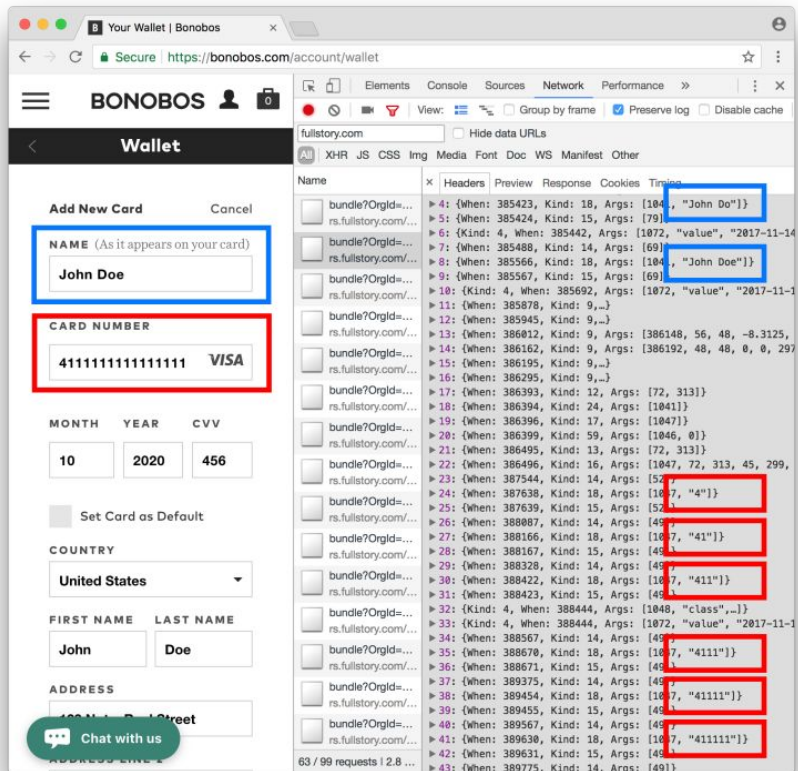
- Name
- CC #
- CVV

- Name
- CC #
- CVV

```
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", "2017-1]
args: [49, ]
args: [1047, "4111"]
args: [49, ]
args: [49, ]
args: [1047, "41111"]
args: [49, ]
args: [49, ]
args: [1047, "411111"]
args: [49, ]
```



# What happened?



Bonobos used:

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

Bonobos should have used:

```
<input type="text"  
autocomplete="cc-number"></input>
```

# Walgreens misses fields during redaction

The screenshot shows a web browser window titled "Prescription Checkout" with the URL [https://www.walgreens.com/pharmacy/prescriptioncheckout\\_new.jsp](https://www.walgreens.com/pharmacy/prescriptioncheckout_new.jsp). The page has a "Guest" user and two radio buttons for "Pick Up in Store" (selected) and "Ship to Home".

The main content area is divided into two sections:

- Walgreens Pharmacy**: Includes a "Choose a Store" section with a text input field for "Enter ZIP code, city and state, address or intersection." and a "Find a store" button.
- Notes to Pharmacy staff**: Includes a text area with the example text "e.g. Please add bubble gum flavoring to the amoxicillin" and a "300 characters remaining" indicator.

Below these sections is the **Review Your Rx** section, which displays a prescription for **ZOLOFT 100MG TABLETS**. The prescription details are shown as "DoctorLastName • Qty: 10 • John Doe", where "DoctorLastName" and "John Doe" are redacted with a red and white striped pattern. A "Remove" button is located to the right of the prescription details.

At the bottom of the page, there is a "Submit request" button and a link to "Add more prescriptions".

Three green arrows point to the redacted fields: "DoctorLastName", "Qty: 10", and "John Doe".

Walgreens makes thorough use of redaction

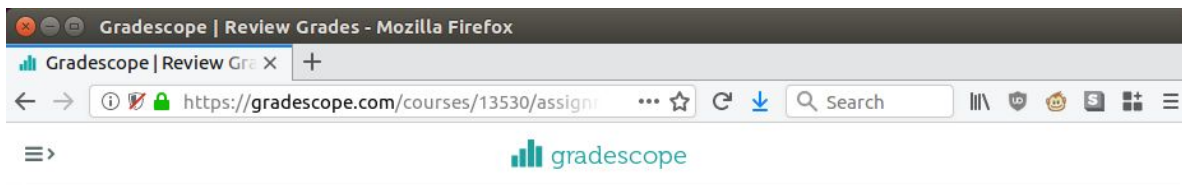
# Walgreens misses fields during redaction

The screenshot shows a web browser window titled "Prescription Checkout" with the URL [https://www.walgreens.com/pharmacy/prescriptioncheckout\\_new.jsp](https://www.walgreens.com/pharmacy/prescriptioncheckout_new.jsp). The page has two radio buttons: "Pick Up in Store" (selected) and "Ship to Home". Below this is a "Walgreens Pharmacy" section with a "Choose a Store" heading. It includes a text input field for "Enter ZIP code, city and state, address or intersection." and a "Find a store" button. To the right is a "Notes to Pharmacy staff" section with a text area containing "e.g. Please add bubble gum flavoring to the amoxicillin" and a "300 characters remaining" indicator. The main section is "Review Your Rx", which displays a prescription card for "ZOLOFT 100MG TABLETS". The card shows "Doctor: LastName" and "Qty: 10" followed by "John Doe". A red arrow points from the text "But prescription information is missed!" to the redacted name "John Doe". Three green arrows point from the bottom to the fields "ZOLOFT 100MG TABLETS", "Doctor: LastName", and "Qty: 10". A "Remove" button is to the right of the prescription card. At the bottom, there is a "Submit request" button and a link to "Add more prescriptions".

But prescription information is missed!

(the user's full name was not redacted on the previous page)

Walgreens makes thorough use of redaction



## Review Grades for Homework1Assignment

● REGRADE REQUESTS OPEN

● GRADES NOT PUBLISHED

MINIMUM

**2.5**

MEDIAN

**2.75**

MAXIMUM

**3.0**

MEAN

**2.75**

STD DEV

**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

Gradescope recordings included:

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

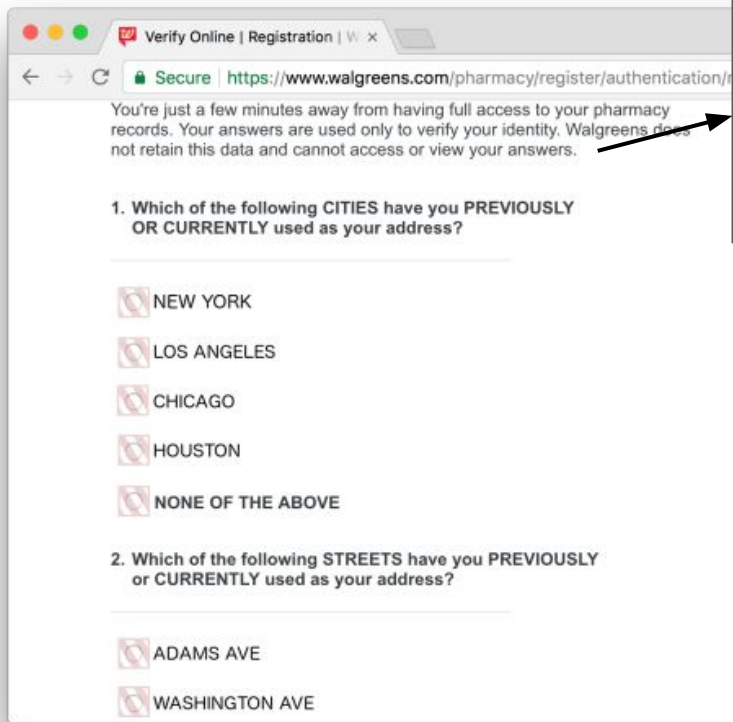
Download Grades

Export Evaluations

Export Submissions

Publish Grades >

# Redactions miss sensitive data



Verify Online | Registration | W x

Secure | <https://www.walgreens.com/pharmacy/register/authentication/>

You're 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?

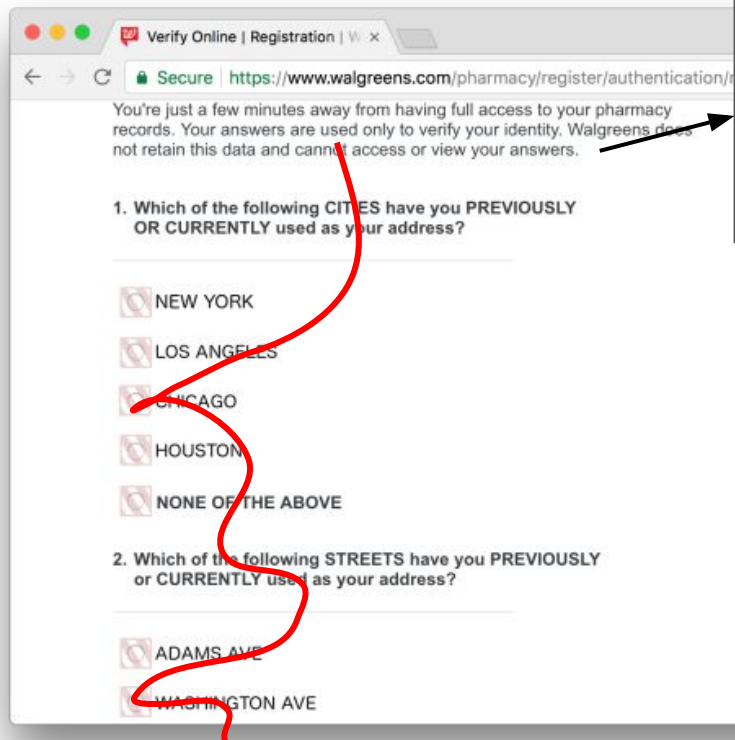
- ☐ NEW YORK
- ☐ LOS ANGELES
- ☐ CHICAGO
- ☐ HOUSTON
- ☐ NONE OF THE ABOVE

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

- ☐ ADAMS AVE
- ☐ WASHINGTON AVE

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

# Redactions miss sensitive data



Verify Online | Registration | W x

Secure | <https://www.walgreens.com/pharmacy/register/authentication/>

You're 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

A red mouse trace is visible, starting from the top left, moving down to the 'CHICAGO' radio button, then curving down to the 'WASHINGTON AVE' radio button. A black arrow points from the text 'Walgreens does not retain this data and cannot access or view your answers.' to the registration form.

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

Although selection inputs redacted, mouse trace is still recorded.

# Session recordings are widespread

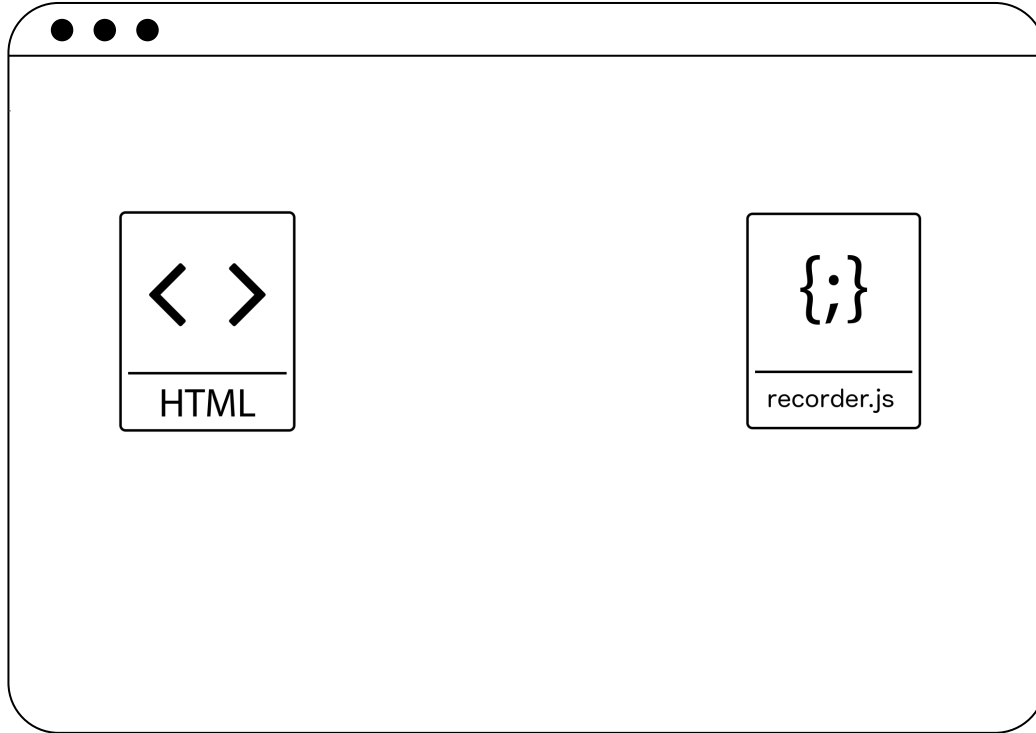
- 14+ analytics company offer recording services
  - Present on 99,174 of the top 1 million sites
- Evidence of recording on 7,918 sites.
  - Likely a lower bound as recording scripts sample users

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

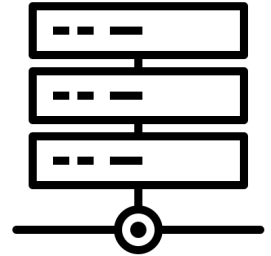
→ **How many more leaks are out there?**

# Detecting session recording

User's browser



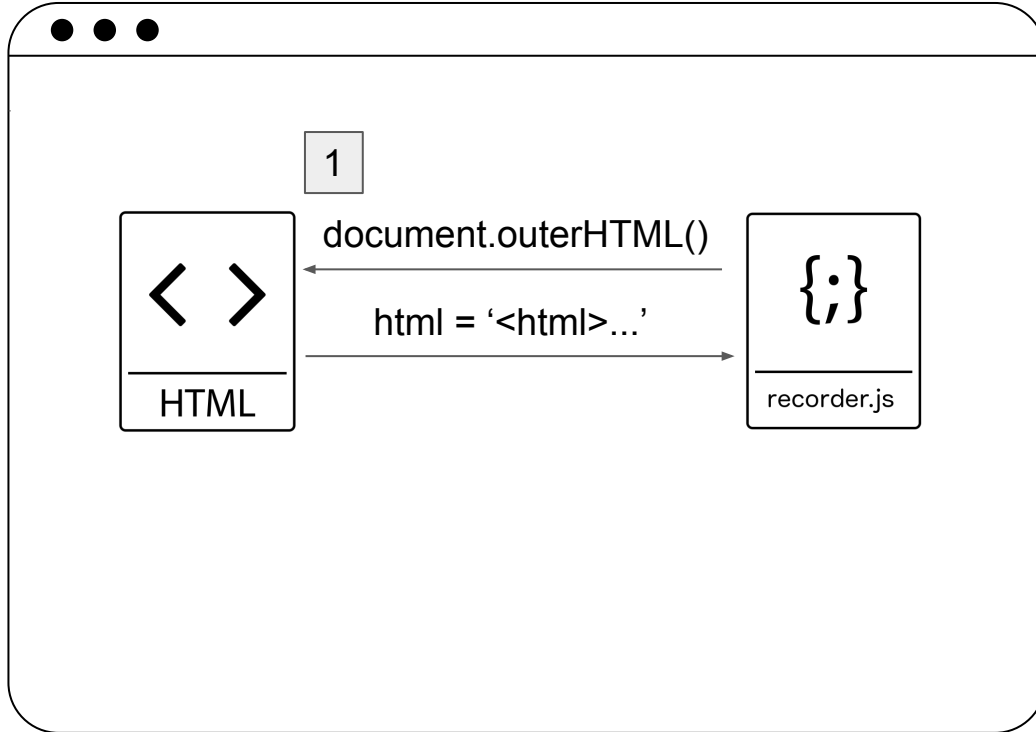
Third-party  
Server



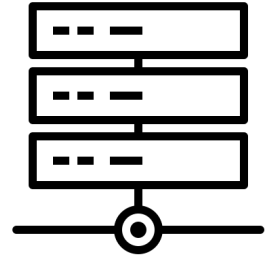


# Detecting session recording

User's browser

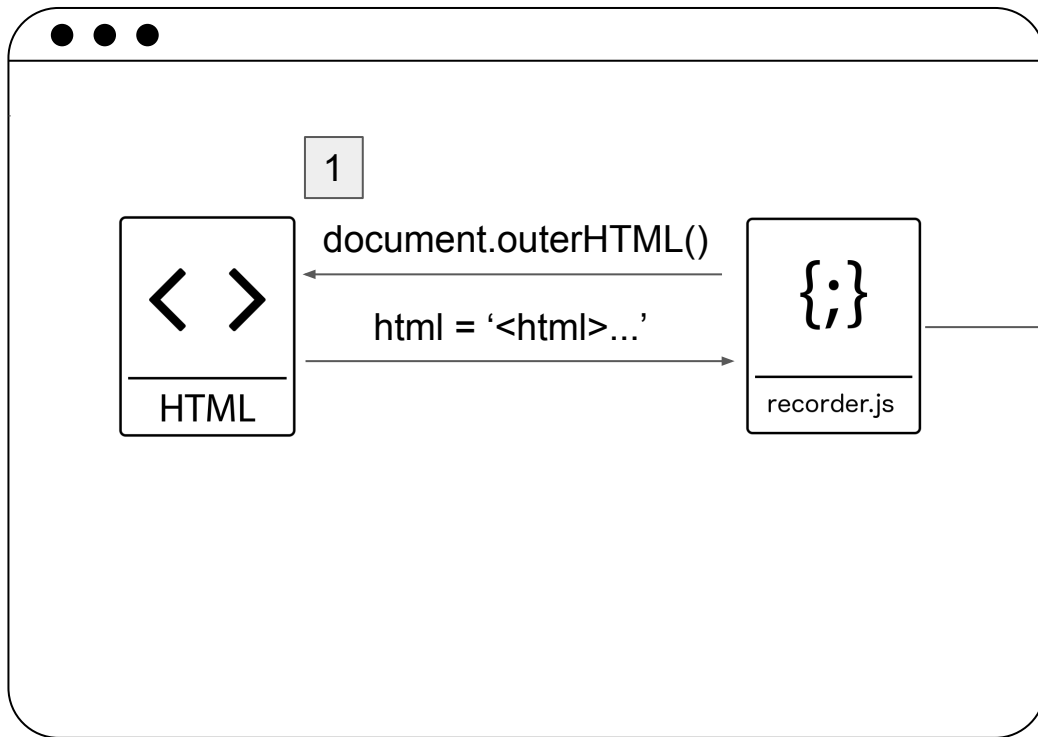


Third-party  
Server



# Detecting session recording

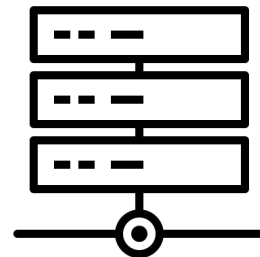
User's browser



2

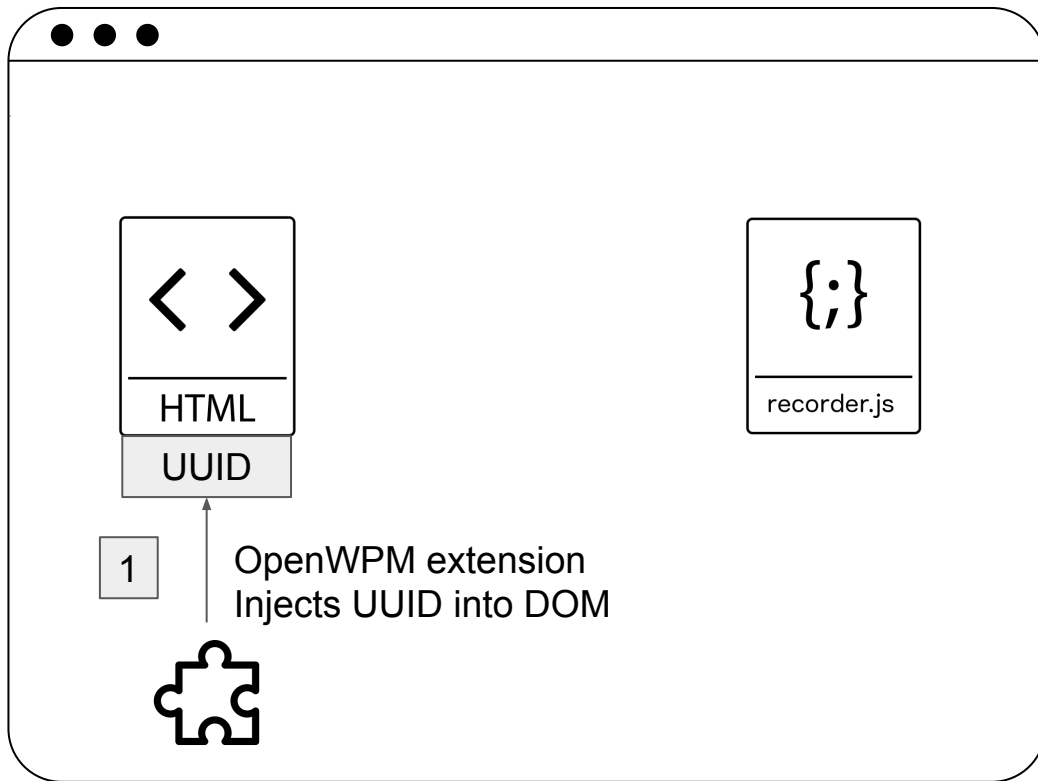
html = '<html>...'

Third-party  
Server

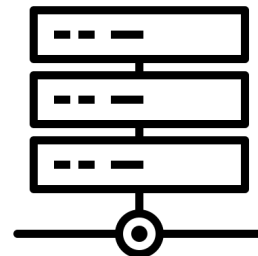


# Detecting session recording

User's browser

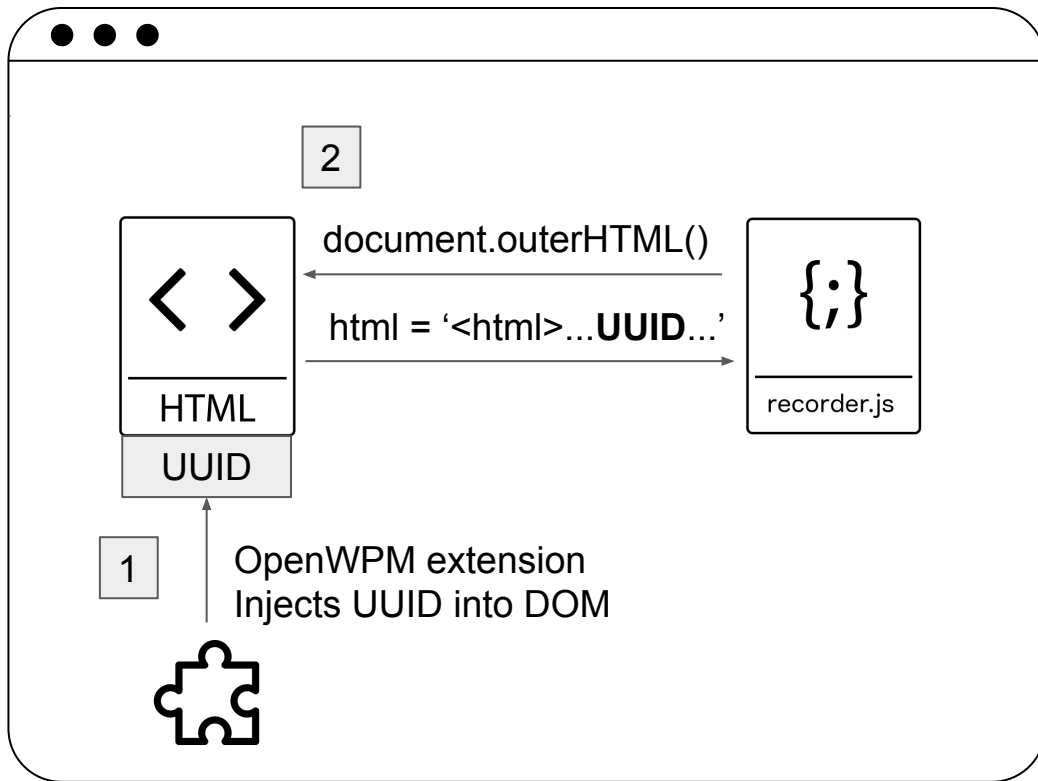


Third-party  
Server

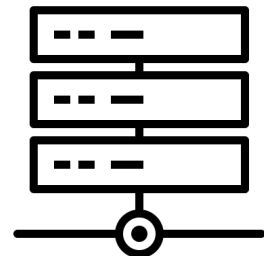


# Detecting session recording

User's browser

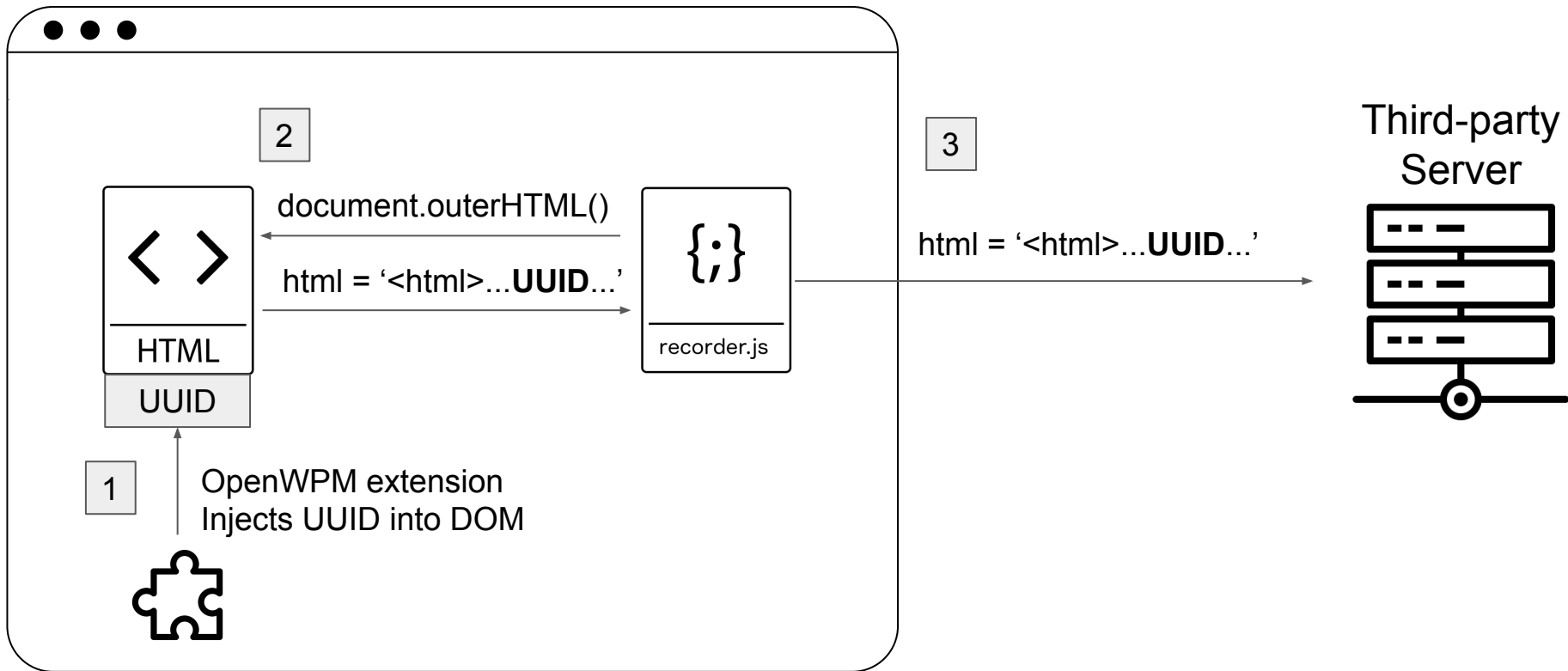


Third-party  
Server



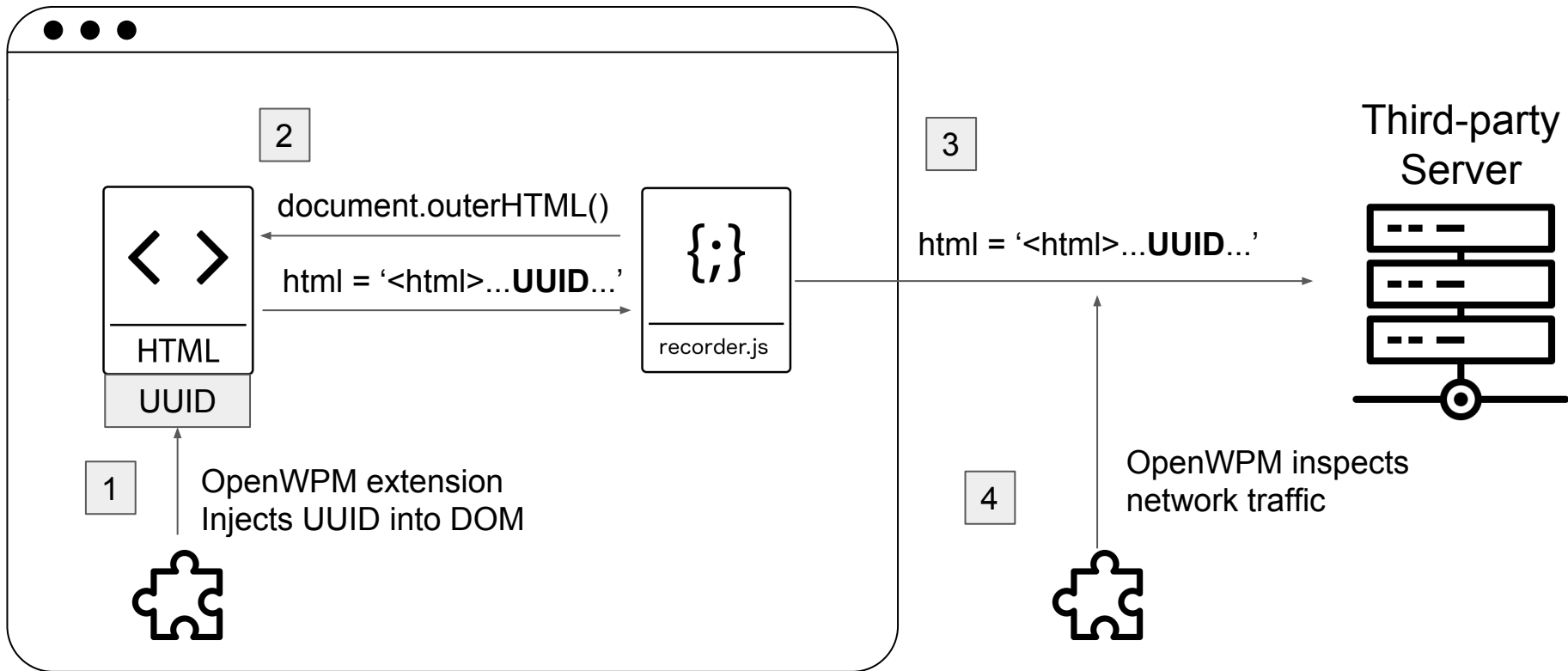
# Detecting session recording

User's browser



# Detecting session recording

User's browser



# Problem: Payload compression in JavaScript!

## Before

```
e.type = a, a = "multi"
this.setPreparedParam(a, e),
!k(c) && (e.data = c, c = "[{"type":"patch","data":{"content":["{\\"r
e.dataLength = c.length,
!e.noBase))
try {
  e.data = this.compress(e.data)
} catch (f) {
  rb()
}
this.pBLT = J(),
Cb.boolTrue.test(ua(b.body, "di-heatmap")) || (d.sendBeacon && e.asyr
e.async = !0,
```

"[{"type":"patch","data":{"content":["{\\"mX\\"":447,\\"mY\\"":2038,\\"pN\\"":11,\\"t\\"":252798}]}","offset":261}]"tsuite.com"]

```
cdList: [{...}]
coll: true
cssChanged: false
dAttr: "on,ng-,data-"
dC: 7
d0: 0
diLoc: "https://cdn.decibelinsight.net/i/13733/95994/di.js"
eC: 0
f: true
```

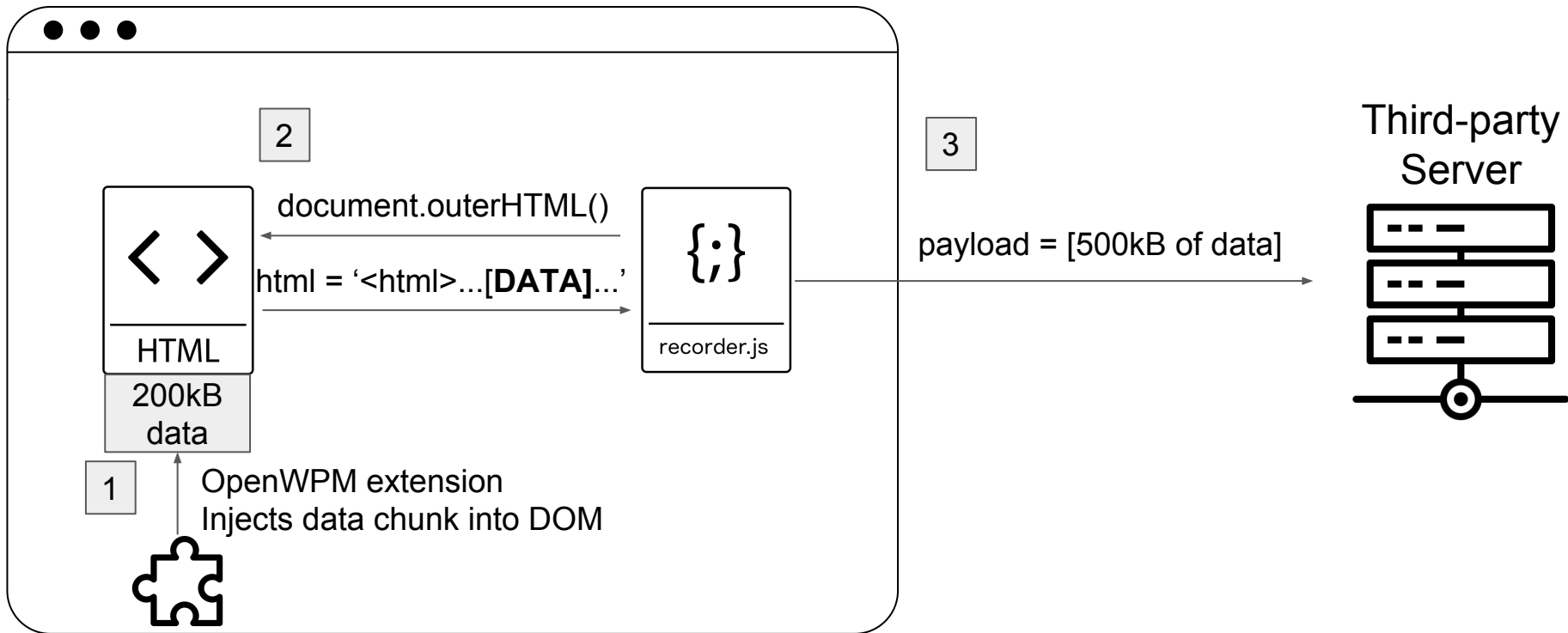
## After

```
e.type = a, a = "multi"
this.setPreparedParam(a, e),
!k(c) && (e.data = c, c = "[{"type":"patch","data":{"content":["{\\"r
e.dataLength = c.length
!e.noBase "DIP1;\|#uzqf#;\#qbudi#-#ebub#;\#dpoufou#;\#\|]#nY]#;558-]#nZ]#;3149-]#q0]#;22-]#u]#;3638:9~^#~-#pggtfu#;372~^"
try {
  e.data = this.compress(e.data)
} catch (f) {
  rb()
}
this.pBLT = J(),
Cb.boolTrue.test(ua(b.body, "di-heatmap")) || (d.sendBeacon && e.asyr
e.async = !0,
```

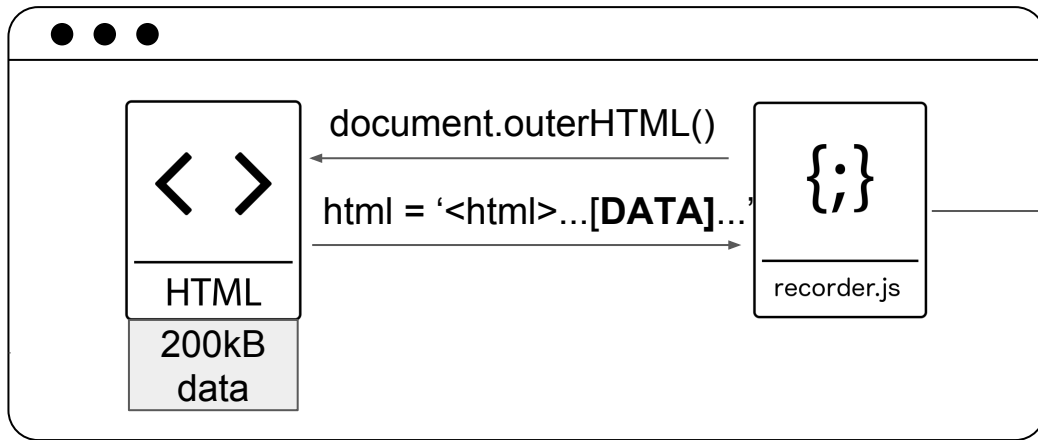
```
cdList: [{...}]
coll: true
cssChanged: false
dAttr: "on,ng-,data-"
dC: 7
d0: 0
diLoc: "https://cdn.decibelinsight.net/i/13733/95994/di.js"
eC: 0
f: true
```

# Detecting session recording

User's browser

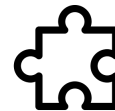
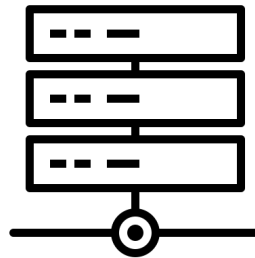




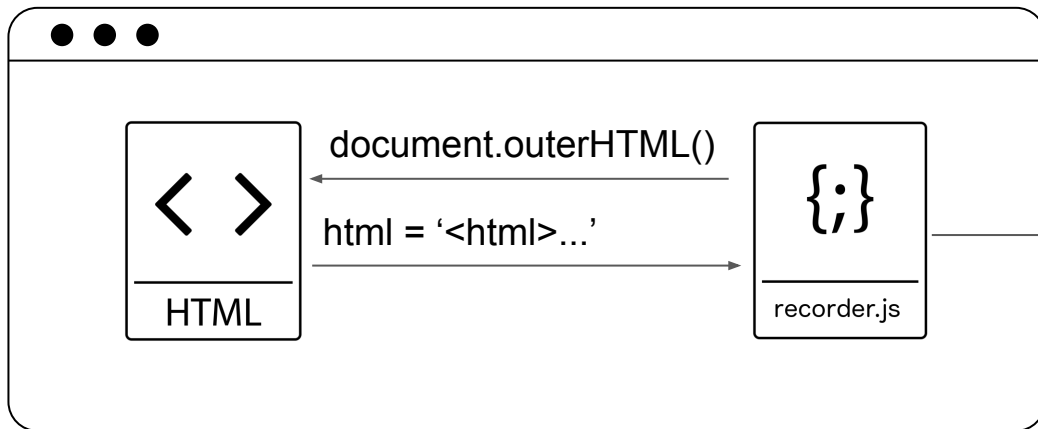


payload = [500kB of data]

Third-party  
Server

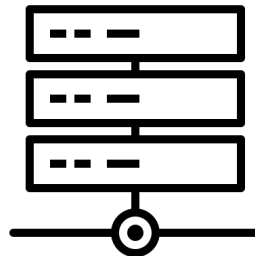


Diff payload size



payload = [300kB of data]

Third-party  
Server



# Methodological details

- Use stack traces to attribute DOM access
- Instrument *addEventListener*:
  - which script listens to which events (mousemove, mouseover, keydown...)
- HTTP instrumentation + JS stacktrace
  - which script initiated the HTTP request
- Manual analysis and debugging for confirmation

# Who is at fault?

- Browser vendors?
- Technology, e.g. Same Origin Policy?
- Publishers?

# What can we do?

- Just keep measuring?
- Try to plug holes in browsers?
- Push for regulation?

# Updates

- Walgreens, Gradescope, Bonobos removed Fullstory
- Adthink and OnAudience removed login manager abuse code
- Brave and Safari disabled password manager autofill

THE VERGE

TECH

SCIENCE

CULTURE

MORE

Ad targeters are pulling data from your browser's password manager

New research shows an alarming new way to track web

BBC

Sign in

News

Sport

Weather

Shop

Earth

NEWS

Home

Video

World

US & Canada

UK

Business

Tech

Science

Stories

Technology

More than 480 web firms record 'every keystroke'

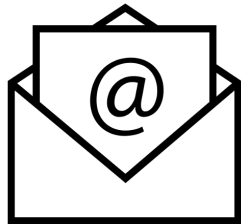
MOTHERBOARD

Over 400 of the World's Most Popular Websites Record Your

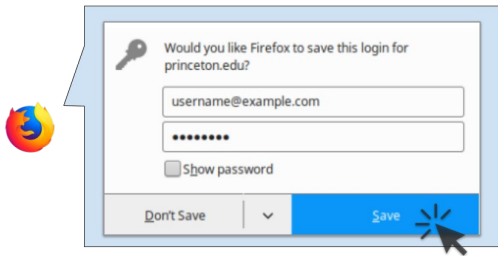
# Third parties collecting PII on the web and in emails

## Email Tracking

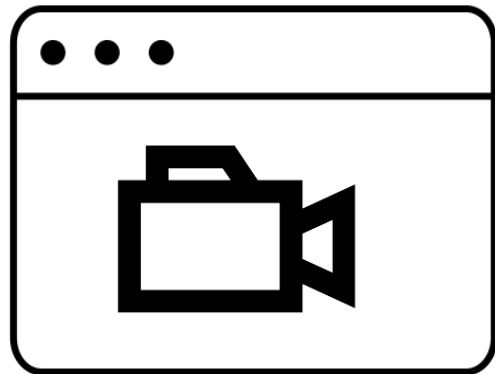
```
UUID = {  
  MD5(bob@example.com),  
  SHA1(bob@example.com),  
  SHA256(bob@example.com)  
}
```



## Autofill abuse



## Session Recording



*“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)*

*Englehardt, Han, and Narayanan, “I never signed up for this! Privacy implications of email tracking” (PETS 2018)*

# Questions?

- Credit for the autofill slides: Güneş Acar
- Contact: [ste@princeton.edu](mailto:ste@princeton.edu)
- Image assets from the Noun Project