

# Leaky Forms: A Study of Email and Password Exfiltration Before Form Submission

Asuman Senol

imec-COSIC, KU Leuven

asuman.senol@esat.kuleuven.be

Gunes Acar

Radboud University

g.acar@cs.ru.nl

Mathias Humbert

University of Lausanne

mathias.humbert@unil.ch

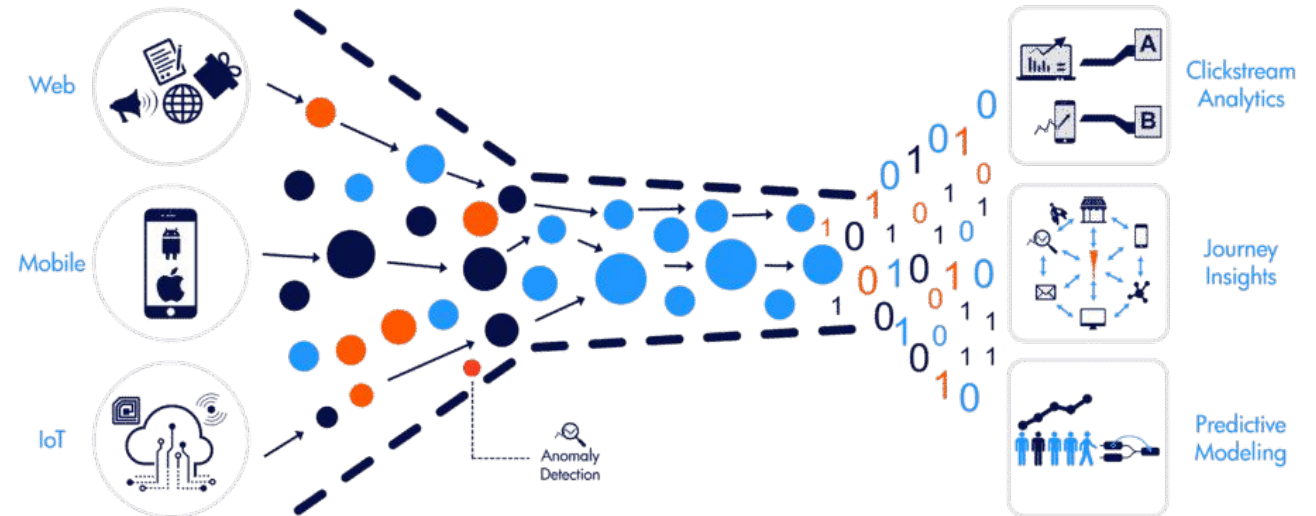
Frederik Zuiderveen Borgesius

Radboud University

frederikzb@cs.ru.nl

# Background

- Websites use advertising and marketing for monetization
  - built-in anti-tracking countermeasures
  - potential third-party cookie phase-out
- Tracking by email addresses
  - enables cross-site, cross-platform, persistent tracking



<https://medium.com/@ugurekmekci/real-time-user-activity-tracking-w-divolte-collector-and-kafka-d8c106313400>

# Motivation

- PII collection before form submission on a mortgage calculator website (Gizmodo, 2017)
- A 2018 survey:
  - 81% of the 502 respondents have abandoned forms at least once
  - 59% abandoned a form in the last month [6].

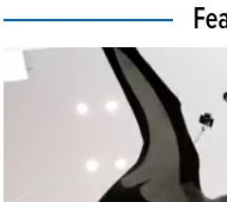
GIZMODO

HOME LATEST TECH NEWS REVIEWS SCIENCE EARTHER 109

GIZMODO ORIGINALS

## Before You Hit 'Submit,' This Company Has Already Logged Your Personal Data

By Surya Mattu and Kashmir Hill | 6/20/17 2:23PM | Comments (103)



# Goals

- Measure email and password collection prior to form submission
  - effect of location: EU vs. US
  - effect of consent
  - mobile vs. desktop

New Tab x +

← → ↻ Search Google or type a URL Guest ⋮

## You're browsing as a Guest

Pages you view in this window won't appear in the browser history and they won't leave other traces, like cookies, on the computer after you close all open Guest windows. Any files you download will be preserved, however.

[Learn more](#)

Elements Console Sources **Network** Performance Memory Application Security Lighthouse Recorder ⚙️ ⋮ ×

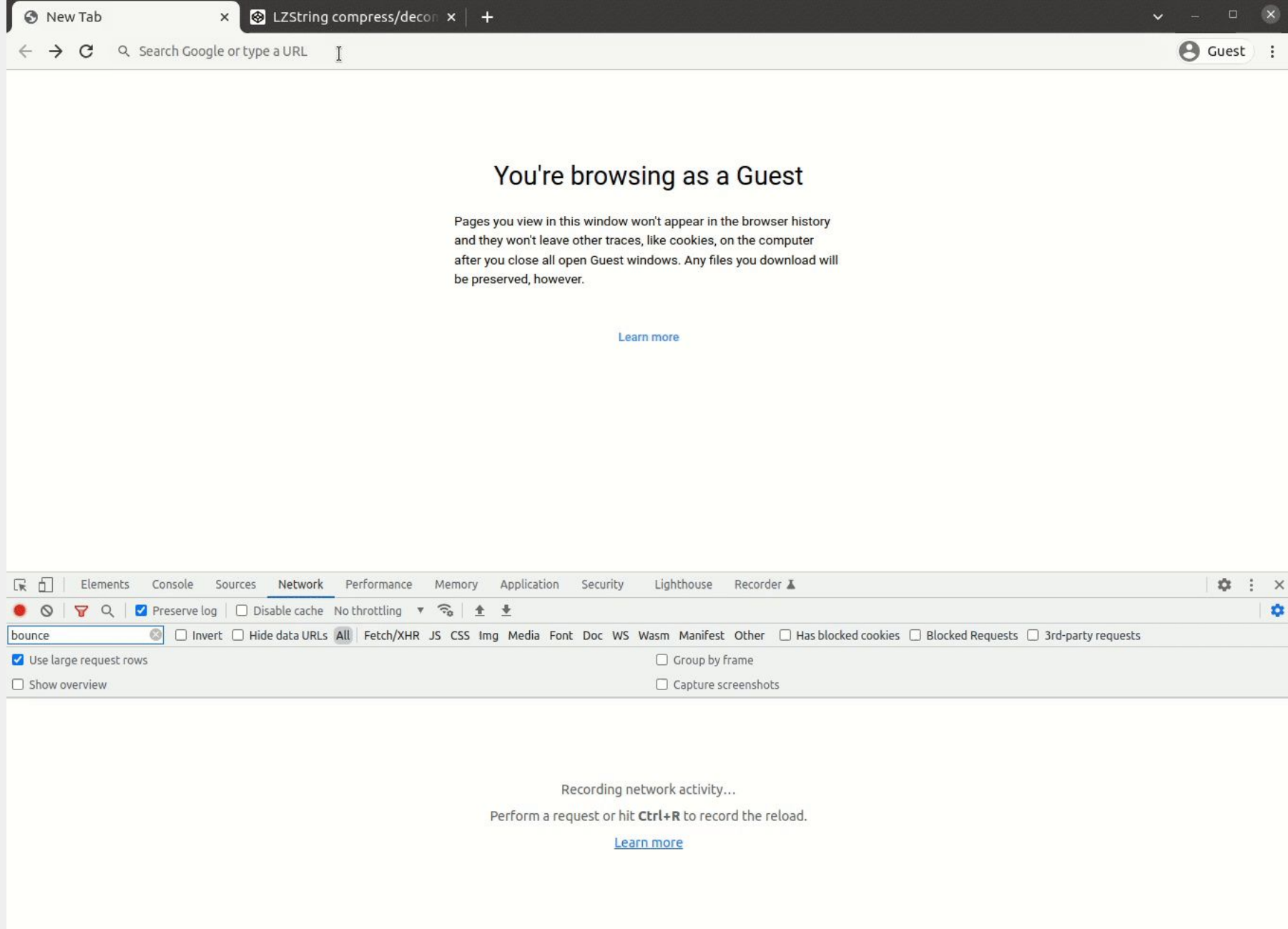
⊙ 🔍  Preserve log  Disable cache No throttling 📶 ⬆️ ⬇️ ⬇️ ⚙️

bizable.com  Invert  Hide data URLs **All** Fetch/XHR JS CSS Img Media Font Doc WS Wasm Manifest Other  Has blocked cookies  Blocked Requests  3rd-party requests

Use large request rows  Group by frame  
 Show overview  Capture screenshots

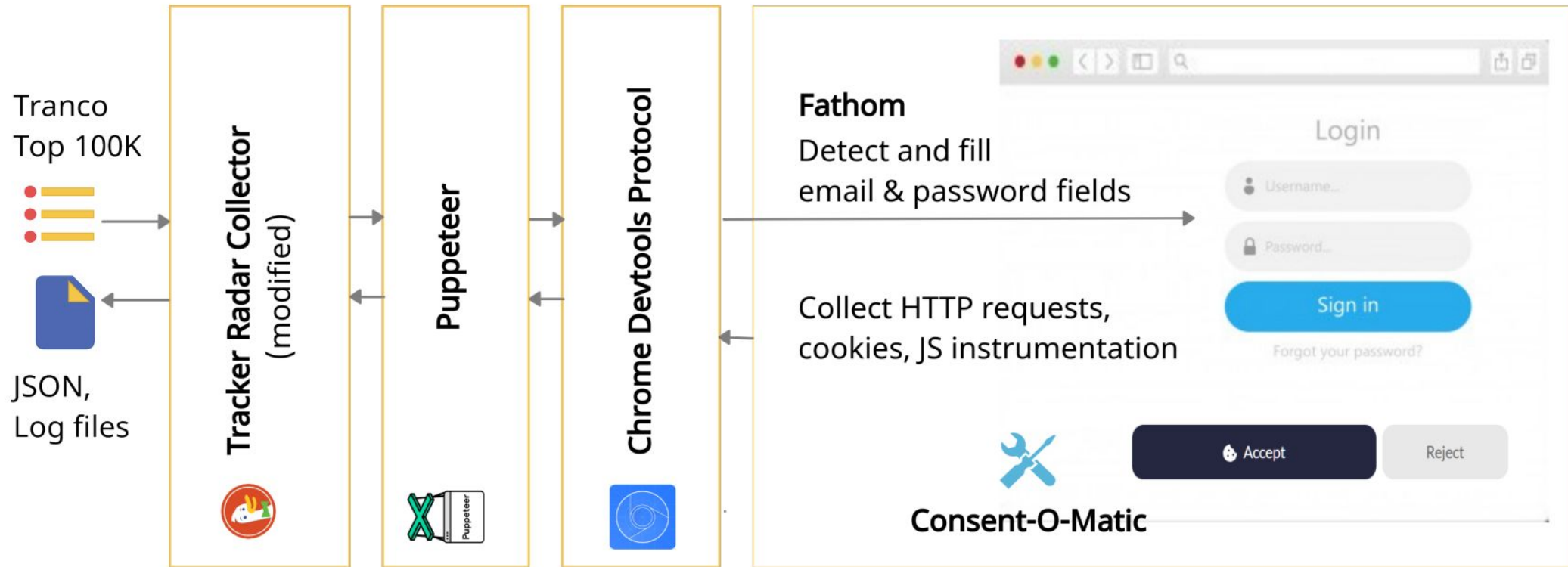
Name	Waterfall

0 / 5 requests | 0 B / 131 kB transferred | 0 B / 131 kB resources | Finish: 49 ms | DOMContentLoaded: 19 ms | Load: 54 ms



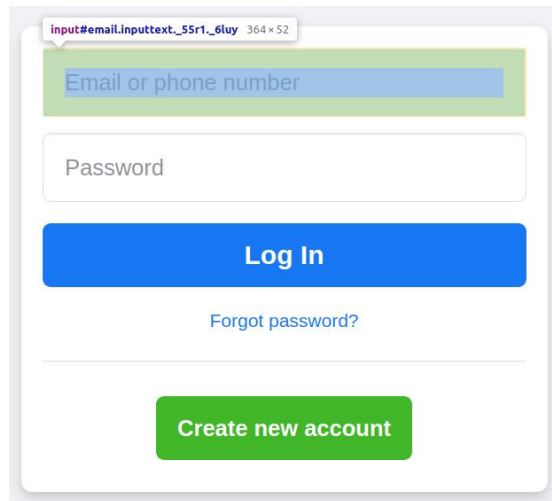
# Method – Web Crawler

- Built on Tracker Radar Collector (developed by DuckDuckGo)



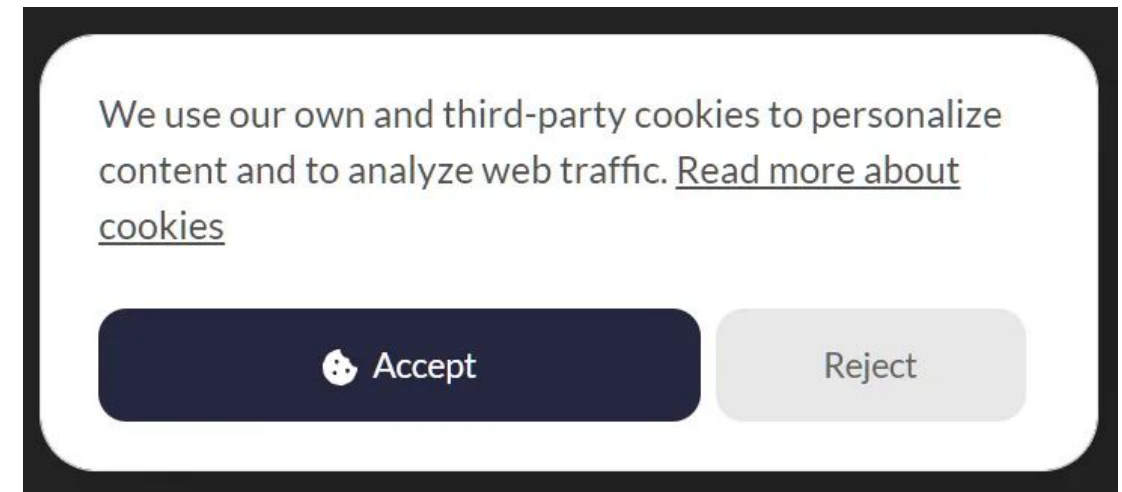
# Method – Email field and CMP detection

- **Fathom:** A supervised learning framework specialized to detect webpage parts [8]]
- **Consent-O-Matic:** browser extension that can recognize and interact with Consent Management Provider (CMP) pop-ups [7]



A screenshot of a login form. At the top, there is a text input field with the placeholder text "Email or phone number". Below it is a password input field. A blue "Log In" button is centered below the password field. Below the button is a link that says "Forgot password?". At the bottom of the form is a green "Create new account" button.

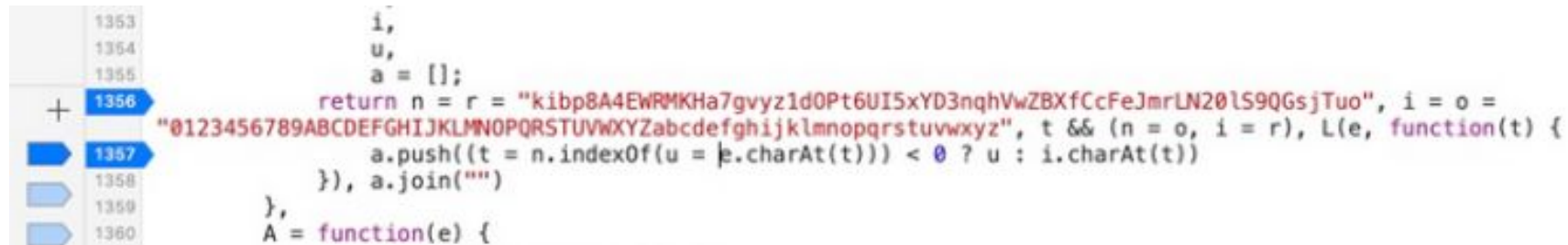
```
▼ <div class="_6lux">  
  <input type="text" class="inputtext_55r1_6luy" name="email" id="email" data-testid="royal_email" placeholder="Email or phone number" autofocus="1" aria-label="Email or phone number"> == $0
```





# Method – Leak Detection

- Based on Englehardt et al.'s method [3]
  - search for different encodings and hashes of search terms
- Identified two new encodings and one hash method
  - LZString, custom mapping, hashing with a fixed salt



```
1353     i,  
1354     u,  
1355     a = [];  
+ 1356     return n = r = "kibp8A4EWRMKHa7gvyz1d0Pt6UI5xYD3nqhVwZBXfCcFeJmrLN20lS9QGsJTuo", i = o =  
1357     "0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz", t && (n = o, i = r), L(e, function(t) {  
1358         a.push((t = n.indexOf(u = |e.charAt(t))) < 0 ? u : i.charAt(t))  
1359     }), a.join("")  
1360 },  
A = function(e) {
```

# Method – Tracker Labeling

- Disconnect
- Whotracks.me
- DuckDuckGo (tds.json)
- uBlock Origin
- + Manual labeling

# Crawls

- 8 Crawls, 100K websites, 2.8M pages, May & June 2021
- 2 locations: EU (Frankfurt), US (NYC)
- 3 consent modes:
  - no action (100K)
  - accept all (7,220)
  - reject all (7,220)
- Mobile
  - EU, US
  - 100K – no action

# Dataset

Crawl Option	EU				US			
	No Action	Accept All	Reject All	Mobile	No Action	Accept All	Reject All	Mobile
Crawled URLs	100K	7,720	7,720	100K	100K	7,720	7,720	100K
Successfully loaded websites	99,380	7,716	7,716	99,363	99,437	7,714	7,716	99,409
Crawled pages	625,143	44,752	40,385	597,791	690,394	51,735	49,260	668,848
Websites where we filled email	52,055	5,076	5,115	47,825	53,038	5,071	5,077	49,615
Websites where we filled password	31,002	2,306	2,342	29,422	31,324	2,263	2,283	30,356

Overview of crawl statistics based on servers located in EU and the USA

# Results - Leaks

	EU			US		
	All	Third party	Tracking related	All	Third party	Tracking related
Email	4,395	2,633	1,844	5,518	3,790	2,950
Password	88	86	46	92	87	48

- Discovered 41 unlisted tracker domains

# Results - Email Leaks

EU							US					
Leak Type	Entity Name	Tracker Domain	Key by key	Num. sites	Prom.	Min. Rank	Entity Name	Tracker Domain	Key by key	Num. sites	Prom.	Min. Rank
email	Taboola	taboola.com	No	327	302.9	154	TowerData	rlcdn.com	No	524	553.8	217
	Adobe	bizable.com	Yes	160	173.0	242	Taboola	taboola.com	No	383	499.0	95
	FullStory	fullstory.com	Yes	182	75.6	1,311	Bounce Exchange	bouncex.net	No	189	224.7	191
	Awin Inc.	zenaps.com*	No	113	48.7	2,043	Adobe	bizable.com	Yes	191	212.0	242
		awin1.com*	No	112	48.5	2,043	Awin	zenaps.com*	No	119	111.2	196
	Yandex	yandex.com	Yes	121	41.9	1,688	Awin	awin1.com*	No	118	110.9	196
	AdRoll	adroll.com	No	117	39.6	3,753	FullStory	fullstory.com	Yes	230	105.6	1,311
	Glassbox	glassboxdigital.io*	Yes	6	31.9	328	Listrak	listrakbi.com	Yes	226	66.0	1,403
	Listrak	listrakbi.com	Yes	91	24.9	2,219	LiveRamp	pippio.com	No	138	65.1	567
	Oracle	bronto.com	Yes	90	24.6	2,332	SmarterHQ	smarterhq.io*	Yes	32	63.8	556
	TowerData	rlcdn.com	No	11	20.0	567	Verizon Media	yahoo.com*	Yes	255	62.3	4,281
	SaleCycle	salecycle.com	Yes	35	17.5	2,577	AdRoll	adroll.com	No	122	48.6	2,343
	Automattic	gravatar.com*	Yes	38	16.7	2,048	Yandex	yandex.ru	Yes	141	48.1	1,648
	Facebook	facebook.com	Yes	21	14.8	1,153	Criteo SA	criteo.com*	No	134	46.0	1,403
	Salesforce	pardot.com*	Yes	36	30.8	2,675	Neustar	agkn.com*	No	133	45.9	1,403
	Oktopost	okt.to*	Yes	31	11.4	6,589	Oracle	addthis.com	No	133	45.9	1,403

## Answers to Your Top Questions About Hashed Email [Video]

November 24, 2015 By [Phil Davis](#)



Did you know your subscribers leave a digital fingerprint behind on all their online activities? Even better, did you know you already have everything you need to access this information?



Email hashing is the perfect way to connect the dots on customer behavior—across



## Taboola Ads

### Getting Started

### Create & Manage Great Campaigns ▼

#### Create A New Campaign

#### Edit Campaigns

### Campaign Targeting Options ▼

Send your 1st Party Audiences  
via DMP or MMP

# Lookalike Targeting

Look-alike models are used to increase scale by finding new people likely to be interested in your business because they resemble existing customers.

Use your CRM data to create lookalike audiences on Taboola! You can upload either a customer list of hashed email addresses, mobile device IDs, or 5 digit US zip codes and Taboola's predictive engine will find similar users that are more likely to convert based on the assumption that these users will be "like" your current customers in your database.

Visit our Advertiser Data Use Policy [here](#).



[Create and manage accounts](#)[Publish and distribute content](#)[Advertise](#)[Sell on Facebook and Instagram](#)[Monetise your content or app](#)

# Business Help Centre



## About advanced matching for web

15,463 views

Advanced matching can help you optimise your Meta ads to drive better results. With advanced matching, you can send us hashed customer information along with your pixel events, which can help you attribute more conversions and reach more [people](#). We hash the customer information on the website before they're sent to Facebook to help protect user privacy.

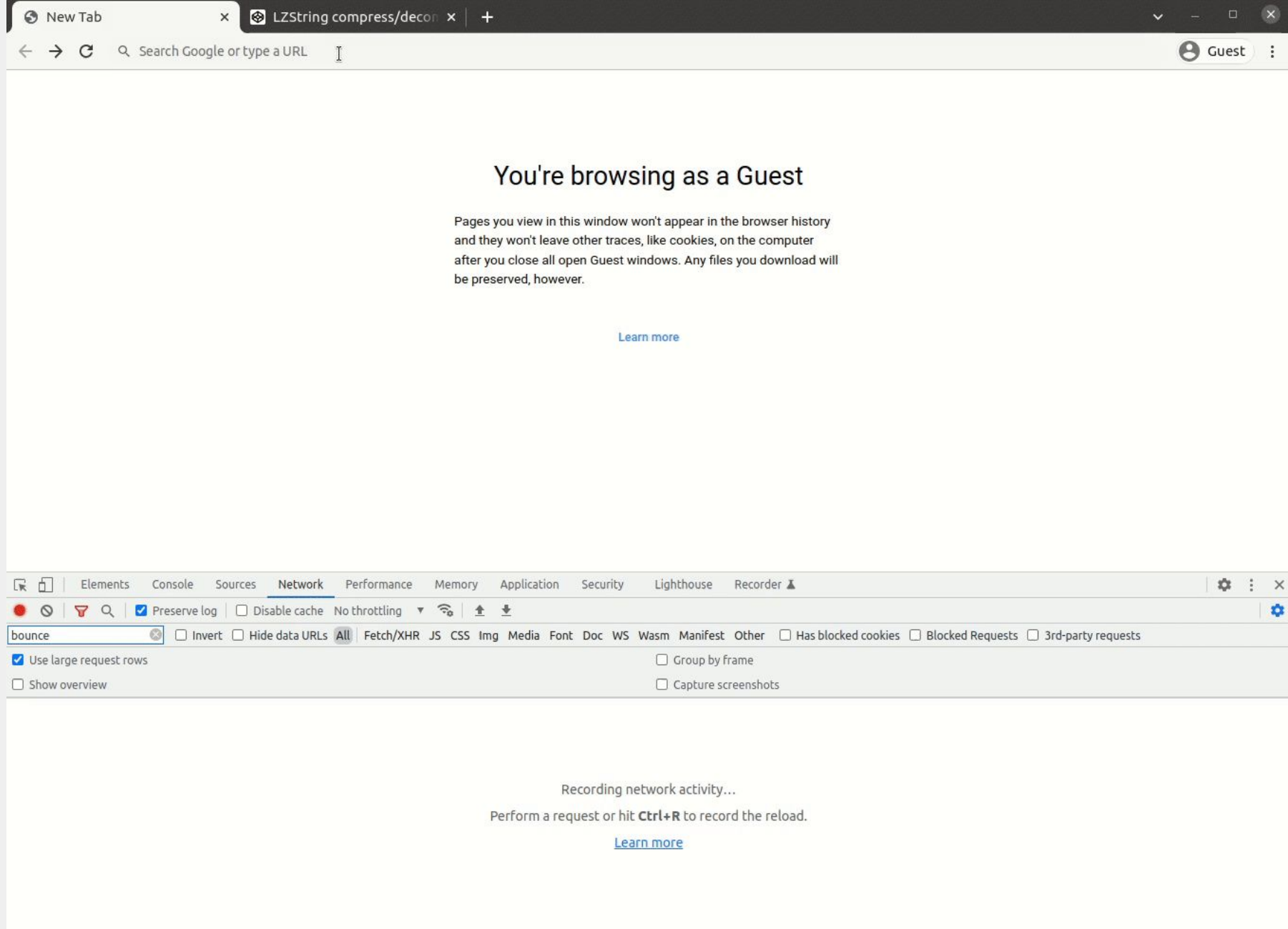
You can use advanced matching to help:



Tru

The image shows a browser window with several tabs: 'New Tab', 'SHA256 Online', 'MD5 Online', and 'SHA1 Online'. The address bar contains the text 'Search Google or type a URL'. The main content area displays a message: 'You're browsing as a Guest' with a subtext explaining that pages won't appear in history and cookies won't be saved. A 'Learn more' link is provided.

Below the main content is the Chrome DevTools interface, with the 'Network' tab selected. The filter dropdown is set to 'All'. The 'Recording network activity...' message is visible, along with instructions to perform a request or hit **Ctrl+R** to record the reload. Another 'Learn more' link is present at the bottom of the DevTools panel.



ABCmouse.com  
Early Learning Academy

Full Online Curriculum  
FOR CHILDREN AGES 2-8

Reading • Math • Science • Art & Colors

Special Offer  
UP TO  
**49% OFF**  
with your Annual Subscription!  
[Learn More](#)

[En Español](#)  
[Create a Shortcut](#)  
[Teachers](#)

[Log In](#)  
[Sign Up](#)  
[Give a Gift](#)

Try it **FREE** for 30 days!  
Click Here to Learn More

Network

Name	Status	Domain	Type	Initiator	Size	Time	Waterfall
task.nextdoor.com				script	0 B	000 ms	
?id=686407008078166&ev=Microdata&dl=https... www.facebook.com/tr	200	www.facebook.com	gif	www.facebook.com/ Redirect	91 B 44 B	29 ms 25 ms	
?id=686407008078166&ev=Microdata&dl=https... www.facebook.com/tr	307 Inter...	www.facebook.com	/Redirect	VM915_main.min.js:6 Script	0 B 0 B	Pending	
<input type="checkbox"/> trace www.cloudflare.com/cdn-cgi	200	www.cloudflare.com	xhr	(index):6 Script	439 B 296 B	242 ms 239 ms	
<input type="checkbox"/> p b.px-cdn.net/api/v1/PXTU0h5FRQ/d	200	b.px-cdn.net	xhr	(index):6 Script	19 B 3 B	98 ms 94 ms	

18 requests | 23.5 kB transferred | 54.7 kB resources

The image shows a browser window with four tabs: 'New Tab', 'SHA256 Online', 'MD5 Online', and 'SHA1 Online'. The address bar contains the text 'Search Google or type a URL'. The main content area displays a message: 'You're browsing as a Guest' followed by a paragraph explaining that pages viewed in this window won't appear in the browser history and won't leave traces like cookies. A 'Learn more' link is provided below the text.

The Chrome DevTools interface is open at the bottom, with the 'Network' tab selected. The toolbar includes options for 'Preserve log', 'Disable cache', 'No throttling', and various filters. The filter dropdown is set to 'All'. Below the toolbar, there are checkboxes for 'Use large request rows', 'Show overview', 'Group by frame', and 'Capture screenshots'. The main area of the DevTools shows the text: 'Recording network activity...' and 'Perform a request or hit **Ctrl+R** to record the reload.', with a 'Learn more' link below.

# Results - Top ten websites

EU				US			
Rank	Website	Third-party	Hash/encoding/compression	Rank	Website	Third-party	Hash/encoding/compression
154	<b>usatoday.com*</b>	taboola.com	Hash (SHA-256)	95	<b>issuu.com</b>	taboola.com	Hash (SHA-256)
242	<b>trello.com*</b>	bizable.com	Encoded (URL)	128	<b>businessinsider.com</b>	taboola.com	Hash (SHA-256)
243	<b>independent.co.uk*</b>	taboola.com	Hash (SHA-256)	154	<b>usatoday.com</b>	taboola.com	Hash (SHA-256)
300	<b>shopify.com</b>	bizable.com	Encoded (URL)	191	<b>time.com</b>	bouncex.net	Compression (LZW)
328	<b>marriott.com</b>	glassboxdigital.io	Encoded (BASE-64)	196	<b>udemy.com</b>	awin1.com	Hash (SHA-256 with salt)
567	<b>newsweek.com*</b>	rlcdn.com	Hash (MD5, SHA-1, SHA-256)			zenaps.com	Hash (SHA-256 with salt)
705	<b>prezi.com*</b>	taboola.com	Hash (SHA-256)	217	<b>healthline.com</b>	rlcdn.com	Hash (MD5, SHA-1, SHA-256)
754	<b>branch.io*</b>	bizable.com	Encoded (URL)	234	<b>foxnews.com</b>	rlcdn.com	Hash (MD5, SHA-1, SHA-256)
1,153	<b>prothomalo.com</b>	facebook.com	Hash (SHA-256)	242	<b>trello.com*</b>	bizable.com	Encoded (URL)
1,311	<b>codecademy.com</b>	fullstory.com	Unencoded	278	<b>theverge.com</b>	rlcdn.com	Hash (MD5, SHA-1, SHA-256)
1,543	<b>azcentral.com*</b>	taboola.com	Hash (SHA-256)	288	<b>webmd.com</b>	rlcdn.com	Hash (MD5, SHA-1, SHA-256)

# Results - Website Categories

Categories	EU/US		EU		US	
	Sites	Filled sites	Leaky sites	Filled sites	Leaky sites	
Fashion/Beauty	1,669	1,176	131 (11.1%)	1,179	224 (19.0%)	
Online Shopping	5,395	3,658	345 (9.4%)	3,744	567 (15.1%)	
General News	7,390	3,579	235 (6.6%)	3,848	392 (10.2%)	
Software/Hardware	4,933	2,834	138 (4.9%)	2,855	162 (5.7%)	
Business	13,462	7,805	377 (4.8%)	7,924	484 (6.1%)	
...	...	...	...	...	...	
Games	2,173	925	9 (1.0%)	896	11 (1.2%)	
Public Information	2,346	1,049	8 (0.8%)	1,084	27 (2.5%)	
Govern.Military	3,754	939	5 (0.5%)	974	7 (0.7%)	
Uncategorized	1,616	636	3 (0.5%)	646	2 (0.3%)	
<b>Pornography</b>	1,388	528	<b>0 (0.0%)</b>	645	<b>0 (0.0%)</b>	



🔍 I



Gmail Images ☰

# Google

🔍 Search Google or type a URL 🔊



Web Store



Add shortcut



# Results - HTTP - WebSocket Usage



- 15 websites in the EU
- 14 websites in the US

- WebSocket
- To four tracker domains:
    - hotjar.com
    - freshrelevance.com
    - noibu.com
    - decibelinsight.net

# Results - EU vs US

	EU			US		
	Distinct websites (All)	Distinct websites (Leaks to 3rd P)	Distinct websites (Tracking related)	Distinct websites (All)	Distinct websites (Leaks to 3rd P)	Distinct websites (Tracking related)
Email	4,395	2,633	1,844	5,518	3,790	2,950

60% difference

addthis.com, yahoo.com,  
doubleclick.net and criteo.com



Only appear in the US crawl

# Results - EU vs US

The screenshot shows the Vail Mountain Resort website with a login form for an Epic account. The form includes fields for 'EMAIL ADDRESS OR USERNAME\*' (containing 'testuser1111111@gmail.com') and 'PASSWORD\*'. Below the form is a 'CREATE ACCOUNT' button. The console window at the bottom shows a list of 8 errors, all of which are 'Failed to load resource: the server responded with a status of 451 ()'. A red arrow points from the console to a text box on the right.

**rlcdn.com sends HTTP 451 error: Unavailable For Legal Reasons**

# Results - EU vs US

Same script (from securedvisit.com) served with **different content**

```
1 /* sv_ea082ada0bf69f160b0bc84078d230c0.js
2 THIS APPLICATION CONTAINS INFORMATION PROPRIETARY TO SECUREDVISIT.COM
3 TO USE THIS SOFTWARE, YOU MUST BE AN AUTHORIZED EMPLOYEE OR AGENT
4 OF SECUREDVISIT.COM.
5 ALL RIGHTS NOT GRANTED TO YOU HEREIN ARE EXPRESSLY AND UNCONDITIONALLY
6 RESERVED. YOU MAY NOT REMOVE ANY PROPRIETARY NOTICE FROM ANY COPY OF THE SOFTWARE.
7 YOU MAY NOT PUBLISH, DISPLAY, DISCLOSE, RENT, LEASE, LICENSE,
8 SUBLICENSE, MODIFY, RENAME, LOAN, DISTRIBUTE, OR CREATE DERIVATIVE WORKS
9 BASED ON ANY PART OF THE SOFTWARE. YOU MAY NOT REVERSE ENGINEER,
10 DECOMPILE, TRANSLATE, ADAPT, OR DISASSEMBLE ANY PART OF THE SOFTWARE,
11 NOR SHALL YOU ATTEMPT TO CREATE THE SOURCE CODE FROM THE OBJECT CODE FOR
12 ANY PART OF THE SOFTWARE.
13 JQuery Sizzle:
14 This software consists of voluntary contributions made by many
15 individuals. For exact contribution history, see the revision history
16 available at https://github.com/jquery/sizzle
17 MD5 (Message-Digest Algorithm):
18 available at http://www.webtoolkit.info/ */
19 window.sv_DNT=true;
20 !function(e){var t,n=!e.sv_DNT,r=e.sv_px,o="https://track.securedvisit.com/citecapture",i=r&&"string"==typeo
f r.url&&"https"==r.url.substring(0,6)?r.url:"https://track.securedvisit.com",u=i+"/identity",a={key:"sv_px_
domain_data",value:r&r.domain_data?r.domain_data:void 0},c={key:"sv_pubid",value:e.sv_pubid},s={key:"sv_ci
d",value:e.sv_cid},l=!e.sv_idq_wait&&[],f=function(e){function t(e,t,n,r){var o,i,u,a,c,l,d,v=t&&t.ownerDocu
ment,p=t?t.nodeType:9;if(n||[],"string"!=typeof e||e||1!==p&&9!==p&&11!==p)return n;if(!r&&((t?t.ownerDocu
```

in the EU

```
1 /* sv_ea082ada0bf69f160b0bc84078d230c0.js
2 THIS APPLICATION CONTAINS INFORMATION PROPRIETARY TO SECUREDVISIT.COM
3 TO USE THIS SOFTWARE, YOU MUST BE AN AUTHORIZED EMPLOYEE OR AGENT
4 OF SECUREDVISIT.COM.
5 ALL RIGHTS NOT GRANTED TO YOU HEREIN ARE EXPRESSLY AND UNCONDITIONALLY
6 RESERVED. YOU MAY NOT REMOVE ANY PROPRIETARY NOTICE FROM ANY COPY OF THE SOFTWARE.
7 YOU MAY NOT PUBLISH, DISPLAY, DISCLOSE, RENT, LEASE, LICENSE,
8 SUBLICENSE, MODIFY, RENAME, LOAN, DISTRIBUTE, OR CREATE DERIVATIVE WORKS
9 BASED ON ANY PART OF THE SOFTWARE. YOU MAY NOT REVERSE ENGINEER,
10 DECOMPILE, TRANSLATE, ADAPT, OR DISASSEMBLE ANY PART OF THE SOFTWARE,
11 NOR SHALL YOU ATTEMPT TO CREATE THE SOURCE CODE FROM THE OBJECT CODE FOR
12 ANY PART OF THE SOFTWARE.
13 JQuery Sizzle:
14 This software consists of voluntary contributions made by many
15 individuals. For exact contribution history, see the revision history
16 available at https://github.com/jquery/sizzle
17 MD5 (Message-Digest Algorithm):
18 available at http://www.webtoolkit.info/ */
19 window.sv_px={"url":"https://track.securedvisit.com","domain_data":{"sid_found":false,"ver":"1.0.0","sid_va
l":""}};
20 window.c=(function(){var i="1.0.1";var e="onready";var g="fire";function f(j){if(Array.isArray){return Array.
isArray(j)}else{return Object.prototype.toString.apply(j)=== "[object Array]"}function b(j,l){var k=l?l:j;if
(typeof k=="object"){k=JSON.stringify(k)}if(l){k+="": "+k"}console.log(k)}function h(j){this.name=j;this.call
backs=[]}h.prototype.registerCallback=function(j){this.callbacks.push(j)};function d(){this.type="";this.deta
il="";this.timeStamp=Date.now()}function a(){this.dataLayer={};this.events={};this.isReady=false}a.prototype.
registerEvent=function(j){this.events[j]=new h(j)};a.prototype.on=a.prototype.addEventListener=function(j,k)
{if(!this.events.hasOwnProperty(j)){this.registerEvent(j)}this.log("listening for",j,k);this.events[j].regist
erCallback(k)};a.prototype.fire=a.prototype.dispatchEvent=function(k,j){if(this.events.hasOwnProperty(k)){for
(var l=0;l<this.events[k].callbacks.length;l++){var n=this.events[k].callbacks[l];if(n){var m=new d();m.type=
k;m.detail=j;this.log("firing",k,j);n.call(this,m)}else{this.log("dispatching event with no callback",k,n)}}}
else{this.log("dispatching an event with no registered listeners",k,j)};a.prototype.ready=function(){this.is
Ready=true;this.dispatchEvent(e,this)};a.prototype.push=function(o){var k=this,j=arguments.length;if(!f(o)){r
eturn}for(var m=0;m<j;m++){try{k[arguments[m][0]].apply(k,arguments[m].slice(1));catch(n){}}};a.prototype.log
=function(m,l,j){var k;if(this.debug){k=" "+_svData.Event<" "+l+">";b(k,j)};function c(){var k=window._svData,
j=null;if(k){if(f(k)){j=new a();if(k.hasOwnProperty("debug")){j.debug=k.debug}}.push.apply(j,k);j.ready()}el
se{j=new a();j.ready()}return j?:k}return c()}());
21 !function(e){var t,n=!e.sv_DNT,r=e.sv_px,o="https://track.securedvisit.com/citecapture",i=r&&"string"==typeo
f r.url&&"https"==r.url.substring(0,6)?r.url:"https://track.securedvisit.com",u=i+"/identity",a={key:"sv_px_
domain_data",value:r&r.domain_data?r.domain_data:void 0},c={key:"sv_pubid",value:e.sv_pubid},s={key:"sv_ci
d",value:e.sv_cid},l=!e.sv_idq_wait&&[],f=function(e){function t(e,t,n,r){var o,i,u,a,c,l,d,v=t&&t.ownerDocu
ment,p=t?t.nodeType:9;if(n||[],"string"!=typeof e||e||1!==p&&9!==p&&11!==p)return n;if(!r&&((t?t.ownerDocu
```

in the US

# Results - The Effect of Consent

<b>Consent modes</b>	<b>EU</b>	<b>US</b>
Accept all	239	242
Reject all	201	199
No action	202	228

0.05%

13%

# Results - Mobile

	<b>Leaky/ Filled Sites EU</b>	<b>Leaky/ Filled Sites US</b>
<b>Desktop</b>	1,844 / 60,008 (3.0%)	2,950 / 60,999 (4.8%)
<b>Mobile</b>	1,745 / 55,738 (3.1%)	2,744 / 57,715 (4.8%)

# Results - Received Emails

- 290 emails from 88 distinct sites

Email from: diabetes.org.uk  
Tracker domain: freshaddress.biz

A friendly reminder Inbox x



**Diabetes UK** <donate@diabetes.org.uk>  
to cosicadam0+diabetes.org.uk ▾

**DiABETES UK**  
KNOW DIABETES. FIGHT DIABETES.

We noticed that you were on our online donation form **but didn't complete it**. If you still feel able to donate, then please take a moment to [complete our form](#).

Is there anything we can help with? If you have any questions please email us at [helpline@diabetes.org.uk](mailto:helpline@diabetes.org.uk) or call us on [0345 123 2399](tel:03451232399)\* we are happy to help you.

Thank you,

Diabetes UK

\*Monday to Friday, 9am to 5pm

View our [Privacy](#) and [Cookies](#) policies.

# Results - Received Emails

Searching for products that actually work? Inbox x



**MyPillow** <mike.lindell@mail.mypillow.com> [Unsubscribe](#)  
to cosicadam0+mypillow.com



*Thanks For  
Stopping By*

When I started MyPillow, my passion was to help people get the best sleep of their life! What a blessing it has been to see that dream become a reality!

To help you best care for your MyPillow, please read our product care recommendations. If you have any questions, please don't hesitate to

**Email from:** mypillow.com  
**Tracker domain:** listrakbi.com

¡Se despide Hot Days! 18 MSI + BONIFICACIÓN Inbox x



**Walmart** <mgnoreply@walmart.com.mx>  
to cosicadam0+walmart.com.mx

🌐 Spanish > English [Translate message](#)

**Walmart.com.mx**



[OUTLET](#) | [TV Y VIDEO](#) | [BEBÉS](#) | [VIDEOJUEGOS](#) | [MUEBLES](#) | [CELULARES](#)



Hasta **18** Meses Sin Intereses + **3 meses de bonificación en Edo. Cta.**

Exclusivo en línea. Válido del 21 al 31 de mayo de 2021. Compra mínima para MSI es de \$1,500 acumulables sin incluir costo de envío. Compra mínima para bonificación en Estado de Cuenta es de \$3,000 a 18 MSI sin incluir costo de envío. Consulta términos y condiciones en: [www.walmart.com.mx/beneficios](http://www.walmart.com.mx/beneficios)



**Email from:** walmart.com.mx  
**Tracker domain:** veinteractive.com



# Results - Password Leaks on 52 websites

Incidental collection by

- Yandex Metrica: due to React framework (41 websites)
- Mixpanel: due to outdated SDK usage (1 website)
- LogRocket: No response (1 website)



# МОЯ TOYOTA

Войти в личный кабинет

Электронная почта/Имя пользователя

Пароль

Вход

У меня нет учетной записи >  
Восстановить пароль

Elements Console Sources **Network** Performance Memory >> 1

webvisor Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest Other

Has blocked cookies  Blocked Requests

Name	Status	Type	Initiator	Size	T...	Waterfall
44886451?wmode=0&wv-part=2&wv... mc.yandex.ru/webvisor	200	xhr	ruxitagentjs_IC... Script	145 B 43 B	7... 7...	
44886451?wmode=0&wv-part=2&wv... mc.yandex.ru/webvisor	200	xhr	ruxitagentjs_IC... Script	73 B 43 B	6... 6...	
44886451?wmode=0&wv-part=3&wv... mc.yandex.ru/webvisor	200	xhr	ruxitagentjs_IC... Script	145 B 43 B	5... 5...	
44886451?wmode=0&wv-part=3&wv... mc.yandex.ru/webvisor	200	xhr	ruxitagentjs_IC... Script	73 B 43 B	5... 5...	

4 requests | 436 B transferred | 172 B resources

Оценить сайт

ВОЗМОЖНАЯ ВЫГОДА ПО ТРЕЙД-ИН  
50000₽

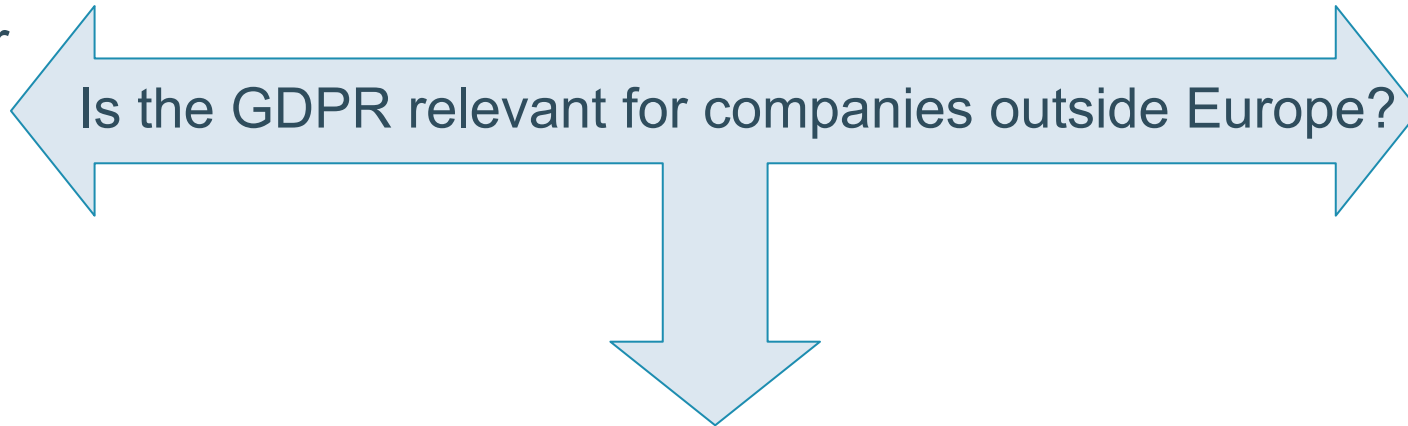
ВОЗМОЖНАЯ ВЫГОДА ПО ТРЕЙД-ИН  
100000₽

# Password collection disclosures

- Notified more than 50 websites about the password collection
- Reached out to all third parties
  - Yandex and MixPanel released an update
  - LogRocket never replied
- Reached out to first parties
  - With Russian translation of the email

# Does Email Exfiltration Comply With the GDPR?

Is the controller  
is based in the  
EU?



Is the GDPR relevant for companies outside Europe?

Do the company  
'monitors' the behavior  
of people in the EU?

Does the company offers goods  
or services to Europeans?

# Does Email Exfiltration Comply With the GDPR?

7

## Transparency principle

Personal data must be processed 'fairly and in a transparent manner'

2

## Purpose limitation principle

Controllers can only collect personal data if they specify a clear purpose in advance

3

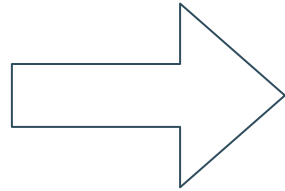
## The requirement for a legal basis such as consent

The controller always needs a 'legal basis' to process personal data

# GDPR Requests

52%

30/58 first parties replied

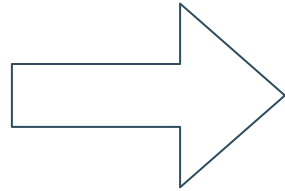


- Were not aware & removed
  - fivethirtyeight.com (via Walt Disney's DPO)
  - trello.com (Atlassian)
  - lever.co, branch.io and cision.com
- Marriott: Glassbox is used for **customer care, technical support, and fraud prevention**
- Tapad, not offering their services to UK & EEA users since August, 2021

# GDPR Requests

53%

15/28 third parties replied

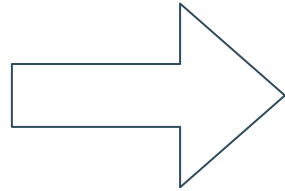


- Adobe, FullStory and Yandex: Suggested our GDPR request to the corresponding first parties
- Taboola: Only collect email hashes after getting user consent

# Outreach to identified websites in the US crawl

0%

0/33 first parties replied



- We sent a friendly notification rather than a formal GDPR request.
- We did not get any response from these 33 websites.

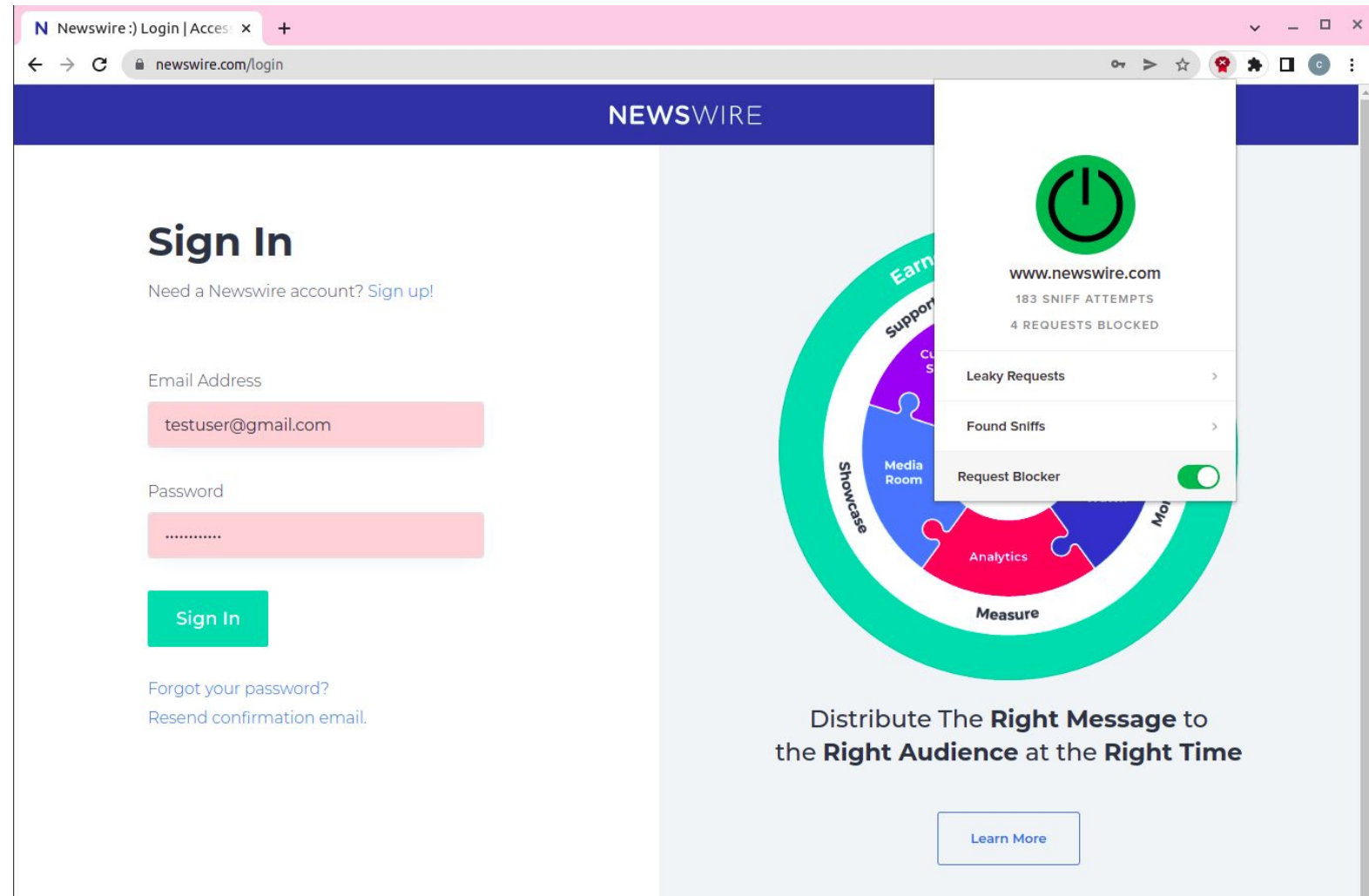


# Countermeasures

- Browser add-ons **block requests** to tracker domains
- Private **email relay** services hide users' emails
  - Apple, Mozilla, DuckDuckGo
- There was **NO** tool for detection, protection of **sniff & exfiltration** on online forms.

# Browser add-on: LEAKINSPECTOR

- Request Blocker
  - Filters & blocks leaky requests to trackers
- Sniff Blocker
  - Detects and logs input sniff attempts by trackers
- The UI is based on DuckDuckGo Privacy Essentials



# Implementation Details - Detection of Sniff Attempts

## Check all interesting input fields

- Modify getter method of HTMLInputElement
- Process the stack trace
- Compose a list of script origins
- Check the script domain is same or not with the domain of the page URL
- Highlight the sniffed input field

The screenshot shows a web browser window displaying the ZSL website. The page title is "Get animals in your inbox" and the URL is "zsl.org/get-the-zsl-email-newsletter". The page features a navigation menu with links for "About ZSL", "Membership", "Support ZSL", "News", "What's on", "Experiences", "Education", "Videos", and "Blogs". The main content area contains a form with the following fields:

- Email Address \* (with a red asterisk indicating a required field) containing "testuser@gmail.com".
- First Name \* (with a red asterisk indicating a required field) containing "test name".
- Last Name (containing "test last name").

Below the form is a "Marketing Permissions" section with the text: "Get the latest updates about exciting animal news from the Zoos, upcoming events, experiences, offers, ZSL's latest work and ways to get involved and". At the bottom of the page, there is a cookie consent banner that reads: "ZSL uses cookies on this website to enhance your user experience. By clicking any link on this page you are giving your consent for us to set cookies. Find out more about cookies." and an "ACCEPT AND CONTINUE" button.

On the right side of the browser window, a sniffing tool overlay is visible, showing a list of detected script origins. The tool title is "www.zsl.org SNIFFS". The list includes:

Script Origin	Category
Third Party facebook.net	Third Party
Facebook, Inc. facebook.net	Analytics
Third Party facebook.net	Third Party
Facebook, Inc. facebook.net	Analytics
Awin AG dwin1.com	Analytics
Awin AG dwin1.com	Analytics
Third Party s3.amazonaws.com	Third Party
Third Party s3.amazonaws.com	Third Party
Third Party	Third Party

# Implementation Details- Detection of Leaky Requests

Before each request was sent to the its end points

- Check the POST body and URL in the request
- Encoded, hashed, obfuscated, cleartext version of the input value is in the request
- Leak detector implemented based on methodology proposed by Englehardt et al. (PETS'18) [3]

The screenshot shows a web browser window with the URL `adespresso.com/join/`. The page features the AdEspresso logo (by Hootsuite) and a promotional banner for a 14-day free trial. Below the banner is a form titled "1. Create your Account" with the following fields:

- Email: `sjdbsjd@gmail.com`
- Password: `.....` (with a note: "The password must have min 8 characters")
- First name: "Your first name..."
- Company: "Your company... (optional)"
- Country: "Choose your country..."

A green "Continue" button is at the bottom of the form. A white overlay window titled "adespresso.com LEAKY REQUESTS" is positioned over the form, displaying a table of detected leaks:

FullStory	Analytics
fullstory.com	

# NEWSWIRE

## Sign In

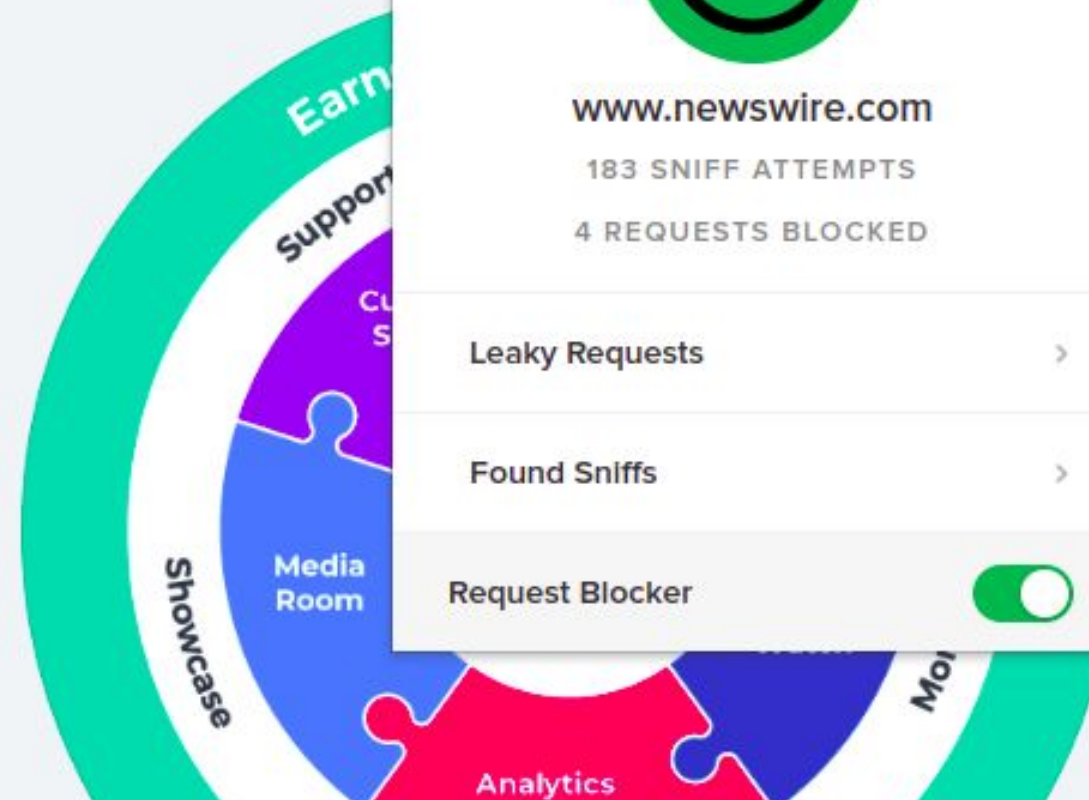
Need a Newswire account? [Sign up!](#)

Email Address

testuser@gmail.com

Password

.....



www.newswire.com

183 SNIFF ATTEMPTS

4 REQUESTS BLOCKED

Leaky Requests >

Found Sniffs >

Request Blocker

NewsWire :) Press Release x +

newsWire.com





NEWSWIRE Products Pricing Resources Blog Customer Stories Newsroom (800) 713-7278 Log In Sign Up

# Press Release Distribution Empowering the Earned Media Advantage

#1 in Customer Satisfaction 2021  
Best in Class - Science, Process and Technology  
Simple, Targeted, Cost-Effective

Distribute the **Right Message** to the **Right Audience** at the **Right Time**

Let's Get Started - Press Release Distribution

# Wrap up

- Email leaks on 1844, 2950 websites, EU, US, resp.
- Password leaks on 54 websites due to session replay scripts
- Uncovered 41 unlisted tracking domains
- Developed a transparency extension that detects PII sniffs and leaks

# Leaks to Facebook & TikTok

- Collection of hashed personal information prior to submission from forms when the user navigates away from the page.

	EU	US
Facebook	7,379	8,438
TikTok	147	154





Any  
Questions?

# References

- [1] E. Papadogiannakis, P. Papadopoulos, N. Kourtellis, and E. P. Markatos. “User Tracking in the Postcookie Era: How Websites Bypass GDPR Consent to Track Users”. In: Proceedings of the Web Conference 2021. 2021, pp. 2130–2141.
- [2] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal. “Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence”. In: Proceedings of the 2020 CHI conference on human factors in computing systems. 2020, pp. 1–13.
- [3] S. Englehardt, J. Han, and A. Narayanan. “I never signed up for this! Privacy implications of email tracking”. In: Proc. Priv. Enhancing Technol. 2018.1 (2018), pp. 109–126.
- [4] Oleksii Starov, Phillipa Gill, and Nick Nikiforakis. FormLock. <https://github.com/ostarov/Formlock>, 2021.
- [5] Surya Mattu and Kashmir Hill. Before You Hit ‘Submit,’ This Company Has Already Logged Your Personal Data. Gizmodo, 2017. <https://gizmodo.com/before-you-hit-submit-this-company-has-already-logge-1795906081>
- [6] 6 Steps for Avoiding Online Form Abandonment. <https://themanifest.com/web-design/blog/6-steps-avoid-online-form-abandonment>, Feb 2022. [Online; accessed 2022-02-12].
- [7] Rolf Bagge, Célestin Matte, Éric Daspét, Kaspar Emanuel, Sam Macbeth, and Steven Roeland. Consent-O-Matic. <https://github.com/cavi-au/Consent-O-Matic/>, 2019.
- [8] Mozilla Fathom documentation. <https://mozilla.github.io/fathom/>. [Online; accessed 2021-06-01].