

# 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

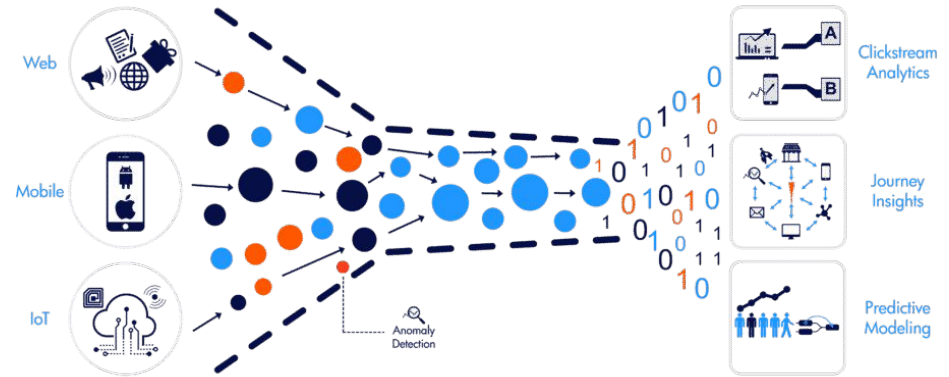
Frederik Zuiderveen Borgesius

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
  - persistent, cross-site, cross-platform



<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 (n=502):
  - 81% abandoned forms at least once
  - 59% abandoned a form in the last month



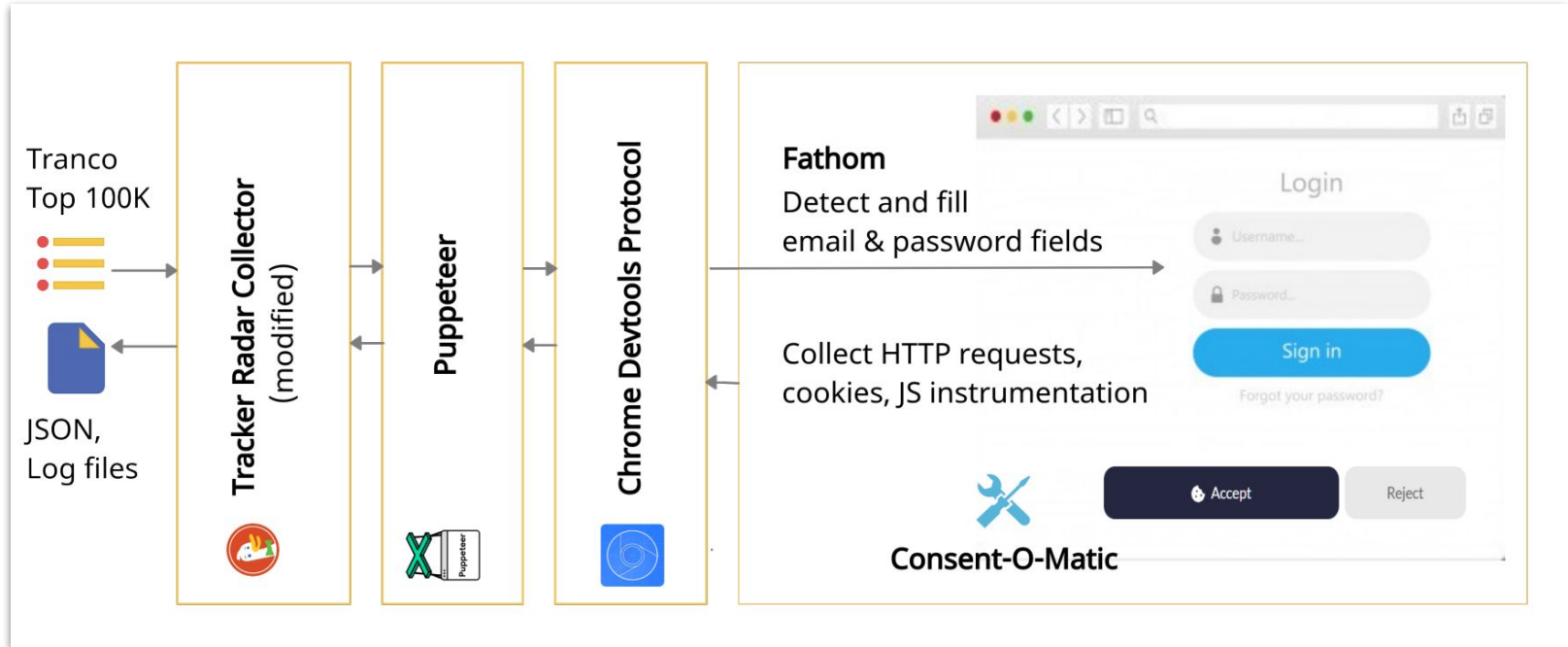
# Study Objectives

---

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

# Method – Web Crawler

- Built on Tracker Radar Collector (developed by DuckDuckGo)



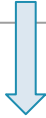
# Crawls (May, June 2021)

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
Visited 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

> 2.8 million pages

# Results

	Email		Password	
	EU	US	EU	US
Websites where we filled	52,055	53,038	31,002	31,324
Leaks to 1st party	4,395	5,518	89	92
Leaks to 3rd party	2,633	3,790	87	87
Leaks to trackers	<b>1,844</b>	<b>2,950</b>	<b>48</b>	<b>49</b>



## How did we label them?

By using blocklists such as Easylist, EasyPrivacy

+ Manual labeling: Discovered **41** unlisted tracker domains

# Prominent Tracker Domains

EU			US		
Name	Domain	Num Sites	Name	Domain	Num Sites
Taboola	taboola.com	327	LiveRamp	rlcdn.com	524
FullStory	fullstory.com	182	Taboola	taboola.com	383
Adobe	bizible.com	160	Adobe	bizible.com	191
Yandex	yandex.com	121	BounceX	bouncex.net	189
Awin	awin1.com	113	Awin	awin1.com	119
	zenaps.com	112		zenaps.com	118





Home > ... > Lookalike Targeting

## 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](#).

## Search

## Getting Started

Glossary of All Terms

Doc Site Tips

## LaunchPad

## Authenticated Traffic Solution

## Registration Manager

## Privacy Manager

## PreferenceLink

## Release Notes and System Information

4 If you selected a method that includes On-page detection, use the Start Detecting Identifier on dropdown to choose the listener event type for when ATS needs to actually detect the identifier on the website:

- **Click event:** [Click event](#) will fire off whenever a specified element is clicked (enter these elements in the Trigger Elements field).
- **Submit Event:** [Submit event](#) will fire off whenever a specified form is submitted (enter these elements in the Trigger Elements field).

### Note

Trigger Elements are CSS selectors to define elements on which the event will be triggered. For examples: #button-id-click or #form-id. As shown in the examples, the given value should start with a hash #. In order to configure Trigger Elements it is recommended to add a CSS ID of html elements to your forms.

- **Blur Event:** [Blur event](#) will fire off whenever a specified input field loses focus for example when a user clicks outside of the input field.

### Warning

The 'Blur Event' method doesn't require human interaction for identifiers to be obtained, while other methods require users to click on a button such as "Submit" or "Ok". To your users, this may give the perception that malicious activities are happening in the background, which is not the case because ATS.js will only start detection with proper consent in place.

Blur Event detection also leaves room for incorrect identifiers because it will not wait for actions from the user like clicking on a login button. For these reasons, we recommend using On Click or On Submit method instead.

# Top ten websites

EU			US		
Rank	Website	3rd-party	Rank	Website	3rd-party
154	usatoday.com*	taboola.com	95	issue.com	taboola.com
242	trello.com*	bizable.com	128	businessinsider.com	taboola.com
243	independent.co.uk*	taboola.com	154	usatoday.com	taboola.com
300	shopify.com	bizable.com	191	time.com	bouncex.net
328	marriott.com	glassboxdigital.io	196	udemy.com udemy.com	awin1.com
567	newsweek.com*	rlcdn.com			zenaps.com
705	prezi.com	taboola.com	217	healthline.com	rlcdn.com
754	branch.io*	bizable.com	34	foxnews.com	rlcdn.com
1,153	prothomalo.com	facebook.com	242	trello.com*	bizable.com
1,311	codecademy.com	fullstory.com	278	theverge.com	rlcdn.com
1,543	azcentral.com*	taboola.com	288	webmd.com	rlcdn.com

\*: Not reproducible anymore as of February 2022.

# Website Categories

	EU/US	EU		US	
Categories	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. %)	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%)
.....	.....	.....	.....	.....	.....
Gov't/Military	3,754	939	3 (0.5%)	974	7 (0.7%)
<b>Pornography</b>	<b>1,388</b>	<b>528</b>	<b>0 (0.0%)</b>	<b>645</b>	<b>0 (0.0%)</b>

# EU vs US

Num distinct websites	EU	US
Visited websites	99,380	99,437
Websites where we filled	52,055	53,038
Emails sent to 1st party	4,395	5,518
Emails sent to 3rd party	<b>2,633</b>	<b>3,790</b>
Emails sent to trackers	<b>1,844</b>	<b>2,950</b>

60% difference

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



Only appear in the US crawl

# EU vs US

The screenshot shows the Vail Mountain Resort website with a sign-in form. The form has fields for 'EMAIL ADDRESS OR USERNAME\*' and 'PASSWORD\*', with a 'SHOW' button for the password. A 'CREATE ACCOUNT' button is also visible. Below the form, the browser's developer console is open, displaying a list of 10 'Failed to load resource: the server responded with a status of 451 ()' errors. A red arrow points from the console to a callout box on the right.

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

# 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=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(){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(!o){o}
return 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="m+"+_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}}j.push.apply(j,k);j.ready()}el
se{j=new a();j.ready()}return j?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=n||[],"string"!=typeof e||e||1!==p&&9!==p&&11!==p)return n;if(!r&&(t?t.ownerDocu
```

in the US

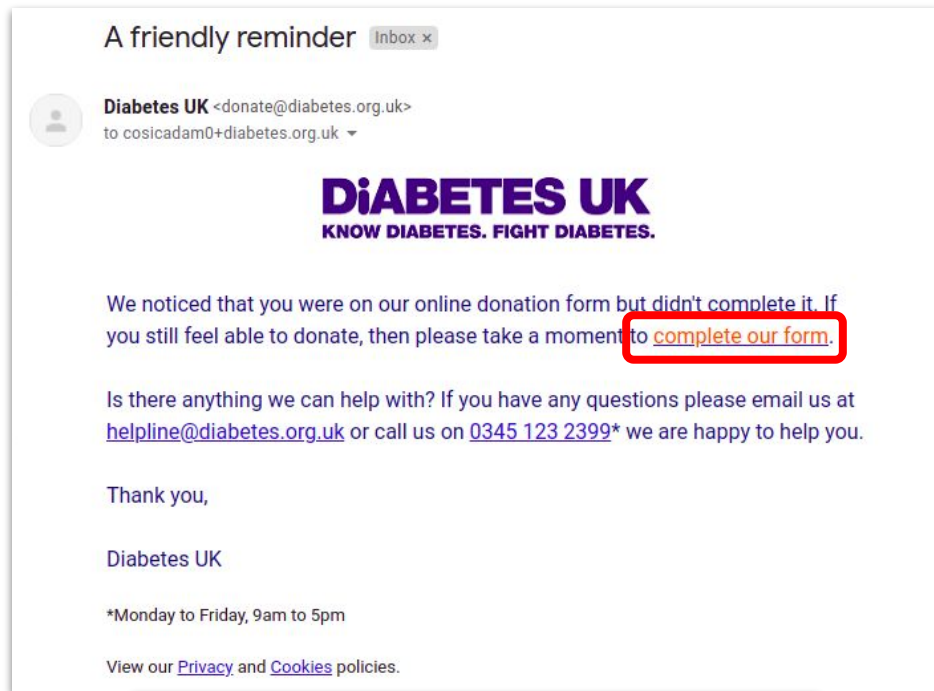
## Results - Mobile

	<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%)



# Received Emails

- 290 emails from 88 distinct sites
  - Offer a discount, or
  - Invite us back to their site



**Email from:** diabetes.org.uk

**Tracker domain:** freshaddress.biz

# Received Emails

Searching for products that actually work? Inbox x



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



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 contact us!

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

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



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

🌐 Spanish ▾ > English ▾ [Translate message](#)



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


# Password Leaks

---

- Incidental collection on 52 sites by
  - Yandex Metrica: due to React framework (50 websites)
  - Mixpanel: due to outdated SDK usage (1 website)
  - LogRocket: No response (1 website)
- Fixed thanks to our disclosures

Официальный сайт Toyota:  Settings

toyota.ru/#/publish/my\_toyota\_login



# МОЯ TOYOTA

Войти в личный кабинет

Электронная почта/Имя пользователя

Пароль

**Вход**

[У меня нет учетной записи >](#)  
[Восстановить пароль](#)

50000P 100000P

Elements Console Sources Network Performance Memory

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

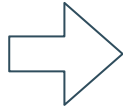
Has blocked cookies Blocked Requests

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

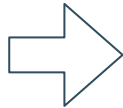
# Outreach Efforts

First parties: 30/58 replied



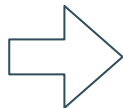
- Were not aware & removed
  - fivethirtyeight.com (via Walt Disney's DPO)
  - trello.com (Atlassian)
- Marriott: Glassbox is used for **customer care, technical support, and fraud prevention**

Third parties: 15/28 replied



- Adobe and Yandex: Referred to corresponding first parties
- Taboola: ad & content personalization, CMP misconfiguration

0/33 first parties replied  
(Websites in the US crawl)



- No response from these 33 websites.

# Leaks to Facebook & TikTok

- Closer look to Facebook
- Due to Automatic Advanced Matching feature of Facebook/TikTok Pixel (scrapes personal information from forms)

Meta for Business | Meta Business Help Centre

Create and manage accounts | Publish and distribute content | Advertise | Sell on Facebook and Instagram | Monetize content

## Business Help Centre

Hi Ali, how can we help?

### About advanced matching for w...

15,463 views

Advanced matching can help you optimize 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

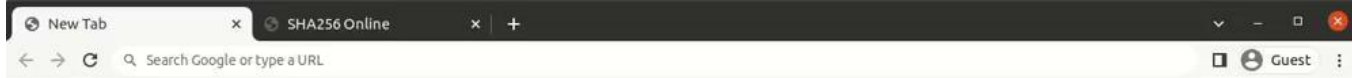


Connect with a Partner

# How it Works

With automatic advanced matching, we can capture the hashed customer data (ex: email addresses) you collect from your website during processes like checkout, account sign-in, or registration. Hashing is the process we use to transform data for security reasons. We can then use hashed identifiers to better match people visiting your website with people on Facebook, which can lead to more attributed conversions for your Facebook campaigns and a larger size of your custom audiences.

- 1 A visitor fills out a form on your website, such as during checkout, account sign-in, or registration.
- 2 After the visitor hits **Submit**, the Pixel's JavaScript code automatically detects and passes the relevant form fields to Facebook. Sensitive data, such as passwords or financial data, is **never** shared with Facebook.
- 3 The form field data is hashed in the visitor's browser before it is sent to Facebook.
- 4 Facebook takes the form data and the action that was taken (for example, a purchase), and matches it to a Facebook user.



## 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](#)

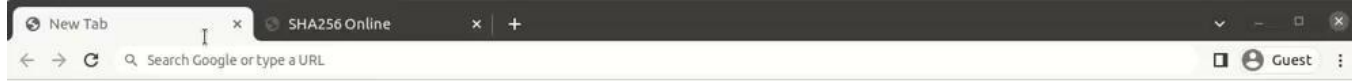


Recording network activity...

Perform a request or hit **Ctrl+R** to record the reload.

[Learn more](#)





## 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](#)



Recording network activity...

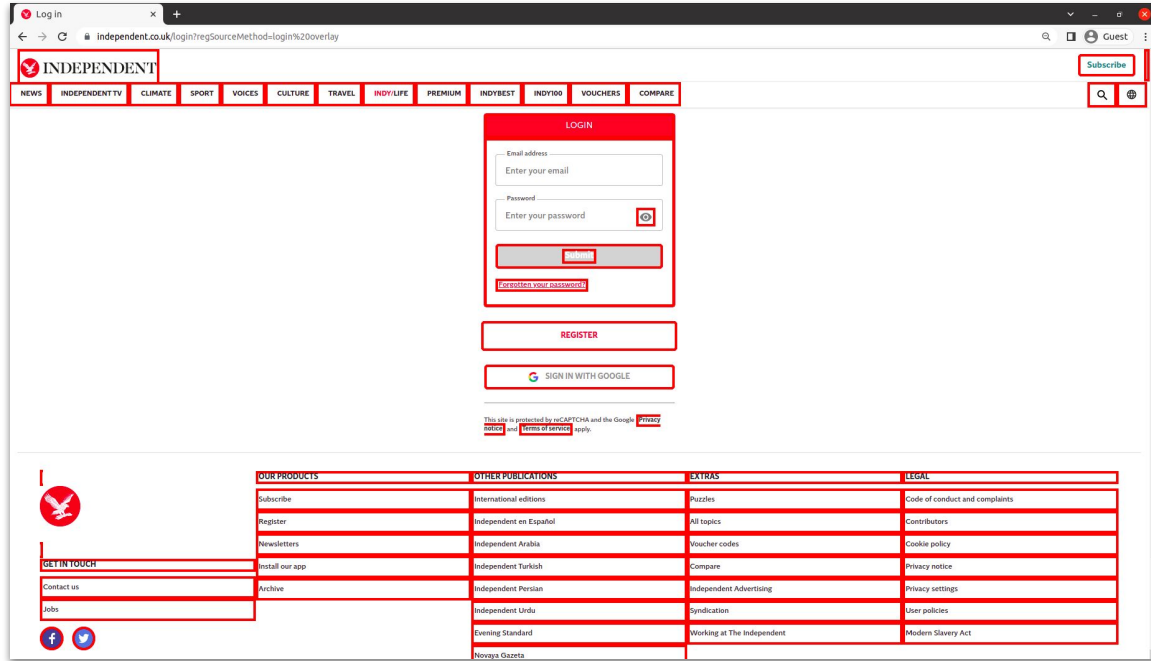
Perform a request or hit **Ctrl+R** to record the reload.

[Learn more](#)

# Leaks to Facebook & TikTok

- Triggered when the user clicks any **link** or **button** on the page

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



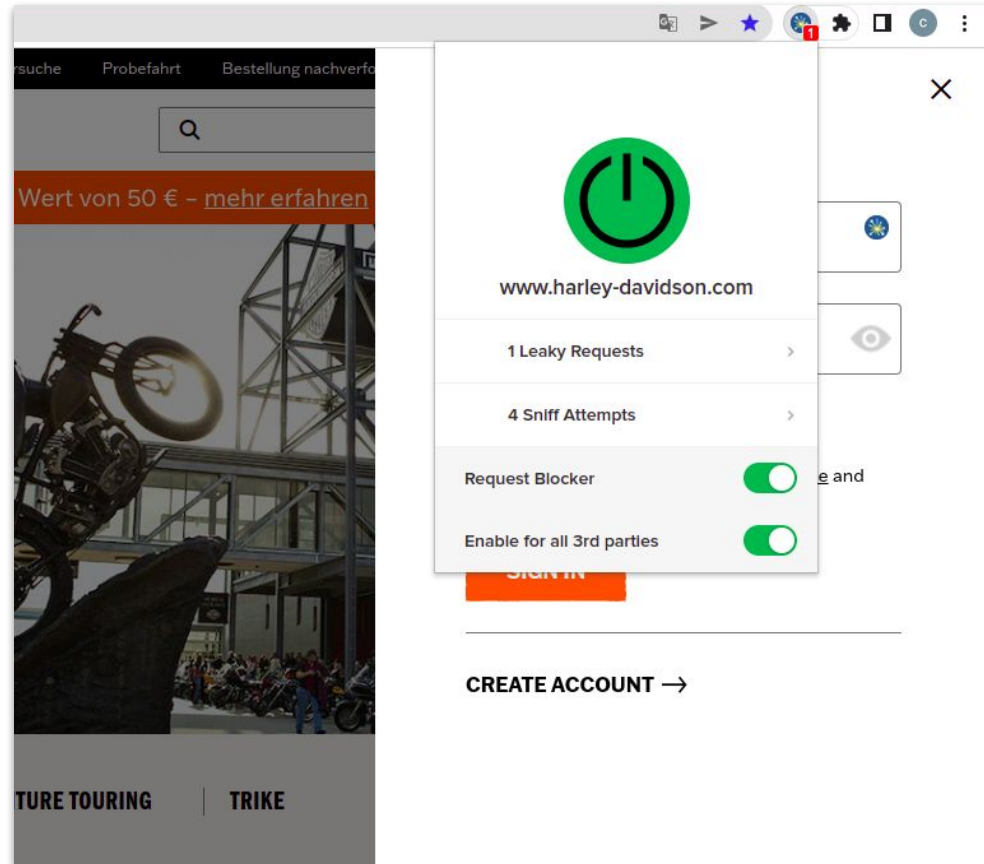
# Countermeasures

---

- Adblockers that block requests to tracker domains
- Private email relay services that hide users' emails
  - Apple, Mozilla, DuckDuckGo
  - e.g. testuser@duck.com-> testuser@gmail.com
- NO tool for detection and prevention of sniff & exfiltration on online forms

# LEAKINSPECTOR

- Proof-of-concept browser add-on
- (<https://github.com/leaky-forms/leak-inspector>)
- Detects sniff attempts
- Blocks leaky requests



NewsWire | Press Release | newswire.com

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

**Leader**  
SUMMER  
2021

**Capterra**  
★★★★★  
USER REVIEWS

**Trustpilot**  
★★★★★  
TrustScore 4.6 | 215 reviews

**1500+**  
RATINGS  
SHOPPER  
Approved

# Summary

---

- Email leaks on 1,844 (EU), 2,950 (US) websites
- Password leaks on 52 websites due to session replay scripts
- Uncovered 41 unlisted tracking domains
- Developed a transparency browser add-on that detects and blocks personal data exfiltration from online forms

# Any Questions?

Leaky Forms   Highlights   Findings   Leaks to Meta & TikTok   Outreach   Follow-up Crawls   LeakInspector   Data and Code   Q&A   Acknowledgments   Corrigendum   Contact

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

To appear at [USENIX Security'22](#)

Email addresses—or identifiers derived from them—are known to be used by data brokers and advertisers for cross-site, cross-platform, and persistent identification of potentially unsuspecting individuals. In order to find out whether access to online forms are misused by online trackers, we present a measurement of *email and password collection that occur before form submission* on the top 100K websites.

[Paper »](#)   [Source code »](#)   [Browser add-on »](#)

- Data: [https://data.ru.nl/collections/ru/icis/usenix-sec-22\\_dsc\\_549](https://data.ru.nl/collections/ru/icis/usenix-sec-22_dsc_549)
- Source code: <https://github.com/leaky-forms/leaky-forms>
- Project website: <https://homes.esat.kuleuven.be/~asenol/leaky-forms>
- Press coverage: [Wired](#), [Fortune](#), [Le Monde](#), [Die Zeit](#)