KU LEUVEN

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

Asuman Senol
imec-COSIC, KU Leuven
asuman.senol@esat.kuleuven.be

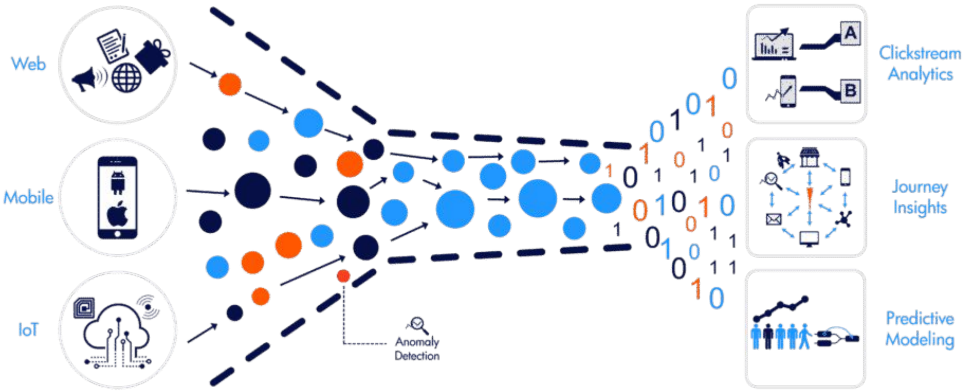
Gunes Acar
Radboud University
g.acar@cs.ru.nl

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



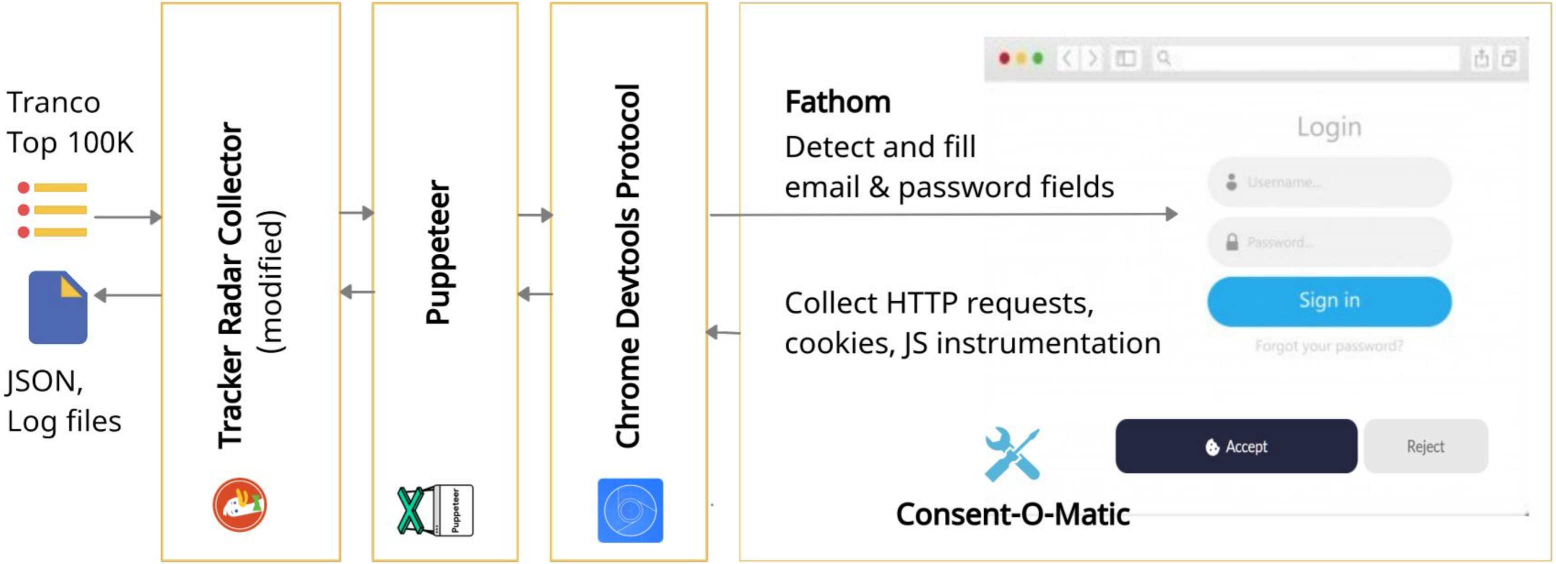
https://medium.com/@ugurekmekci/real-time-user-activity-tracking-w-divolte-collector-and-kafka-d8c106313400

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)



Tranco
Top 100K

JSON,
Log files

Tracker Radar Collector
(modified)

Puppeteer

Chrome Devtools Protocol

Fathom
Detect and fill
email & password fields

Collect HTTP requests,
cookies, JS instrumentation

Consent-O-Matic

Dataset & Results (No interaction with the consent dialogs)

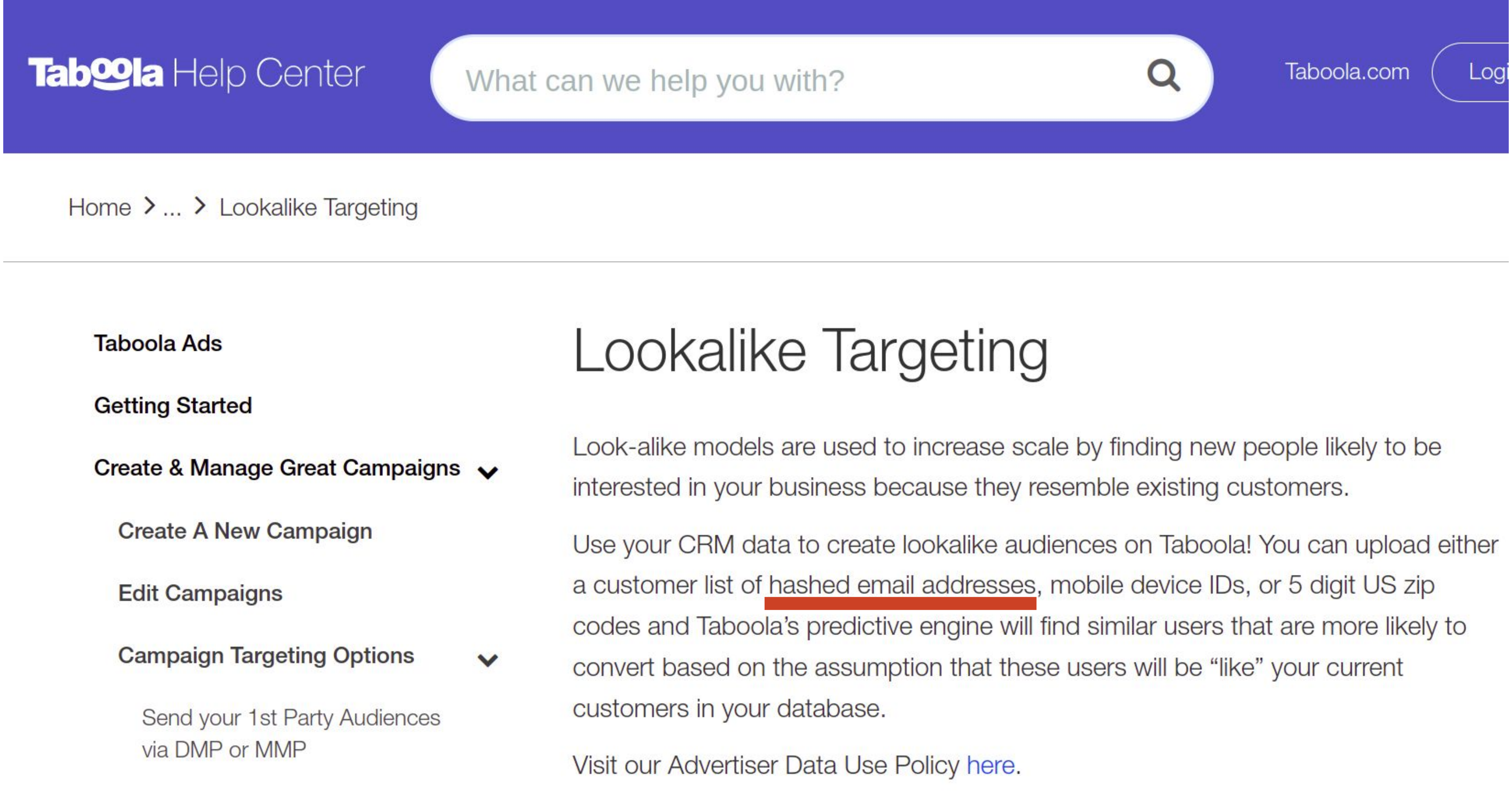
	Email		Password	
	EU	US	EU	US
Visited websites	99,380	99,437	99,380	99,437
Websites where we filled	52,055	53,038	31,002	31,324
Leaks to 1st party	4,395	5,518	89	92
Leaks to 3rd party	2,633	3,790	87	87
Leaks to trackers	**1,844**	**2,950**	**48**	**49**

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

Results - Email Leaks

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

Top 5 tracker domain receiving emails



Taboola Help Center
What can we help you with?
Taboola.com
Logi

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.

email_leak_shopify.com.mp4

Results - EU vs US

	EU	US
Visited websites	99,380	99,437
Websites where we filed	52,055	53,038
Emails sent to 1st party	4,395	5,518
Emails sent to 3rd party	**2,633**	**3,790**
Emails sent to trackers	**1,844**	**2,950**

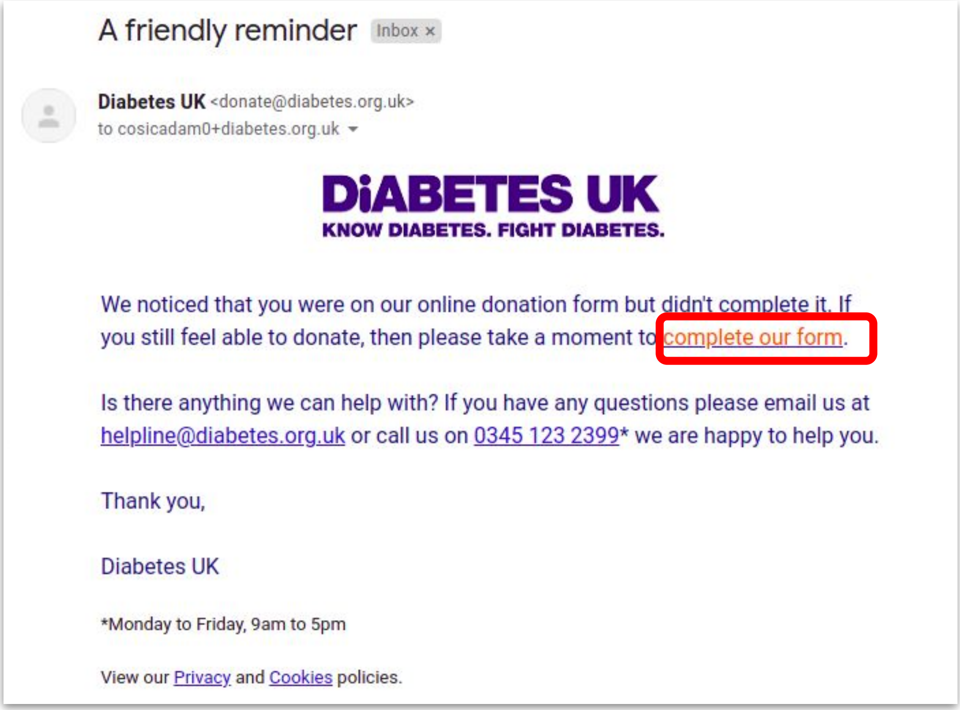
60% difference

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

Only appear in the US crawl

Results - Received Emails

- 290 emails from 88 distinct sites



A friendly reminder

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

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 or call us on 0345 123 2399* we are happy to help you.

Thank you,

Diabetes UK

*Monday to Friday, 9am to 5pm

View our Privacy and Cookies policies.

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

Results - Password Leaks on 52 websites

Incidental collection by

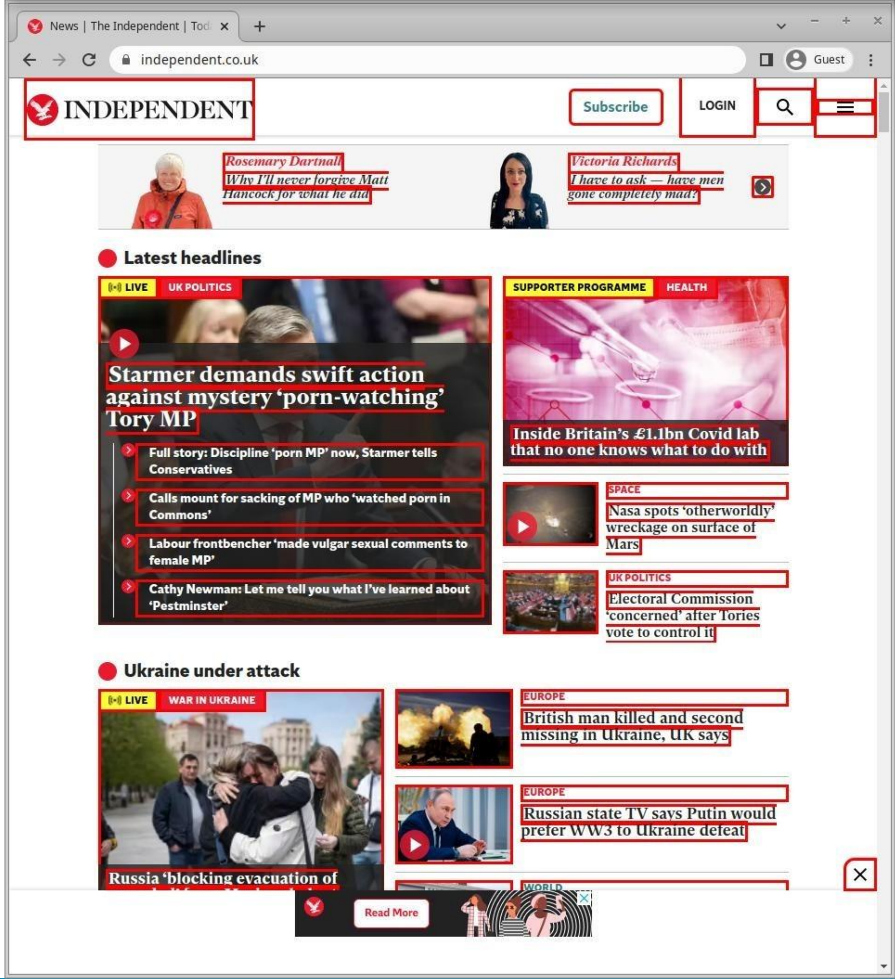
- Yandex Metrica: due to React framework (50 websites)
- Mixpanel: due to outdated SDK usage (1 website)
- LogRocket: No response (1 website)

toyota.ru.password_leak.mp4

Leaks to Facebook & TikTok

- Triggered when the user clicks any link (Facebook) or button on the page.

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



Rosemary Dartnall
Why I'll never forgive Matt Hancock for what he did

Victoria Richards
I have to ask — have men gone completely mad?

Latest headlines

LIVE UK POLITICS

Starmer demands swift action against mystery 'porn-watching' Tory MP

Full story: Discipline 'porn MP' now, Starmer tells Conservatives

Calls mount for sacking of MP who 'watched porn in Commons'

Labour frontbencher 'made vulgar sexual comments to female MP'

Cathy Newman: Let me tell you what I've learned about 'Pestminster'

SUPPORTER PROGRAMME HEALTH

Inside Britain's £1.1bn Covid lab that no one knows what to do with

SPACE

Nasa spots 'otherworldly' wreckage on surface of Mars

UK POLITICS

Electoral Commission 'concerned' after Tories vote to control it

Ukraine under attack

LIVE WAR IN UKRAINE

Russia 'blocking evacuation of

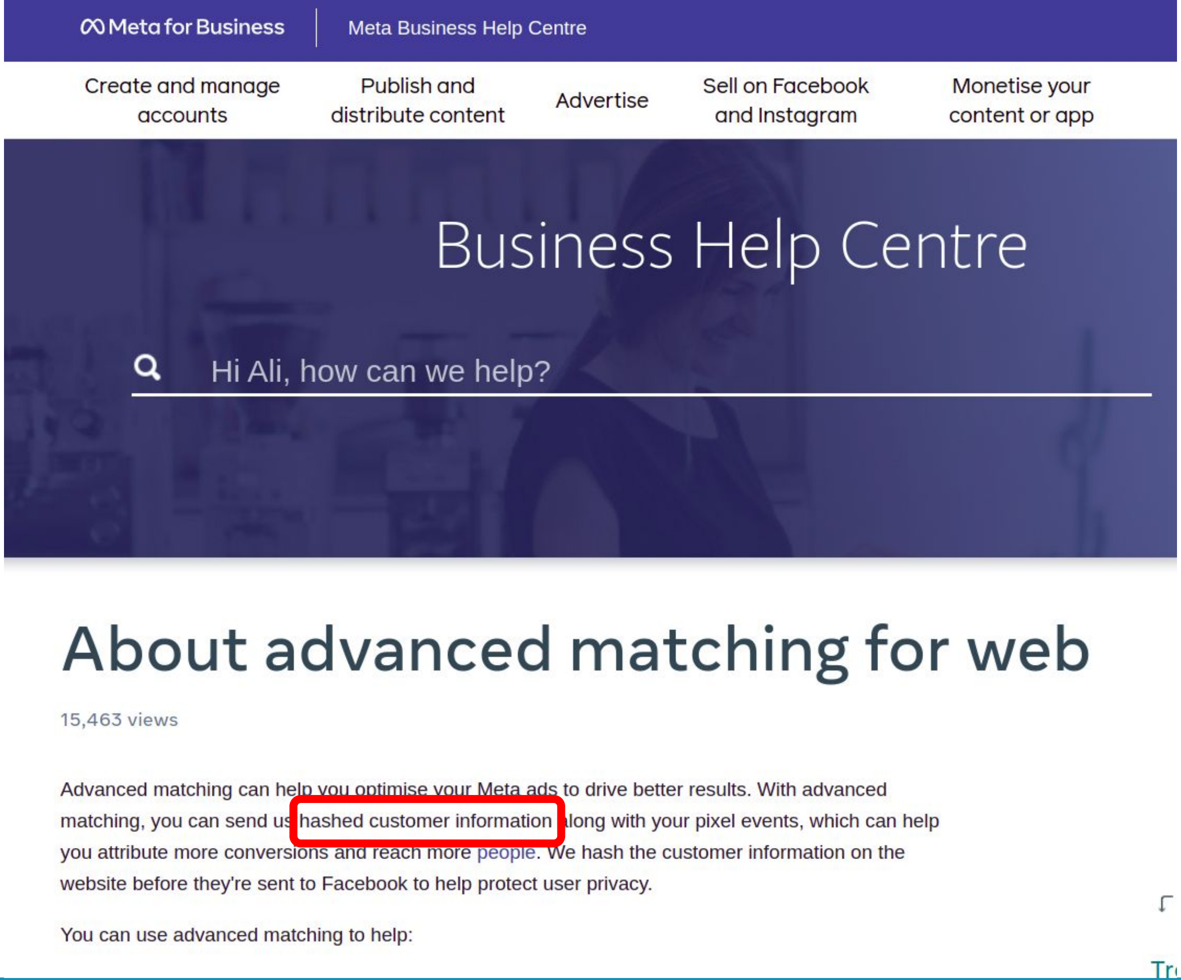
EUROPE

British man killed and second missing in Ukraine, UK says

EUROPE

Russian state TV says Putin would prefer WW3 to Ukraine defeat

WORLD


Meta for Business
Meta Business Help Centre
Create and manage accounts
Publish and distribute content
Advertise
Sell on Facebook and Instagram
Monetise your content or app
Business Help Centre
Hi Ali, how can we help?

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:

GDPR Requests

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

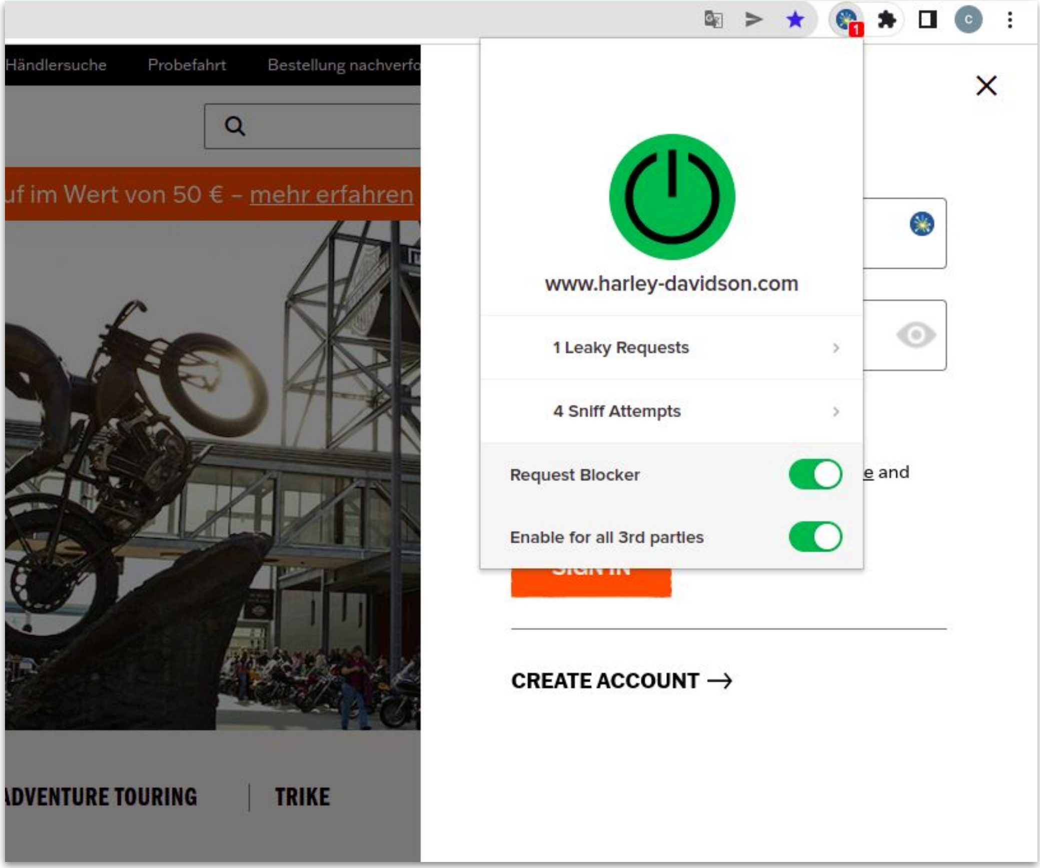
- Adobe and Yandex: Referred to corresponding first parties
- Taboola: ad & content personalization, CMP misconfiguration

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

- No response from these 33 websites.

Browser add-on: LEAKINSPECTOR

- Detects sniff attempts
- Blocks leaky requests
- https://github.com/leaky-forms/leak-inspector



www.harley-davidson.com

1 Leaky Requests ›

4 Sniff Attempts ›

Request Blocker ●

Enable for all 3rd parties ●

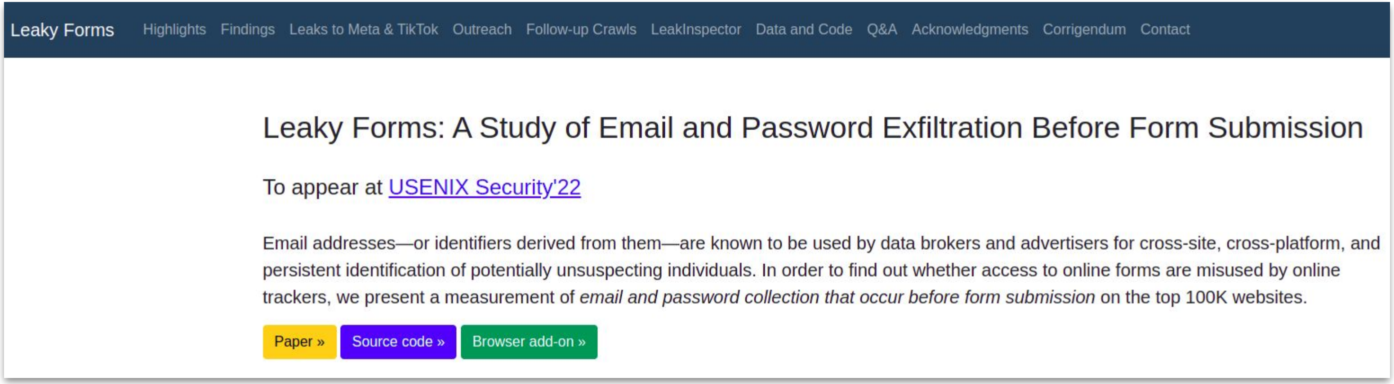
leak_inspector_demo.mp4

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?

Project website: https://homes.esat.kuleuven.be/~asenol/leaky-forms



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.

Source code: https://github.com/leaky-forms/leaky-forms