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

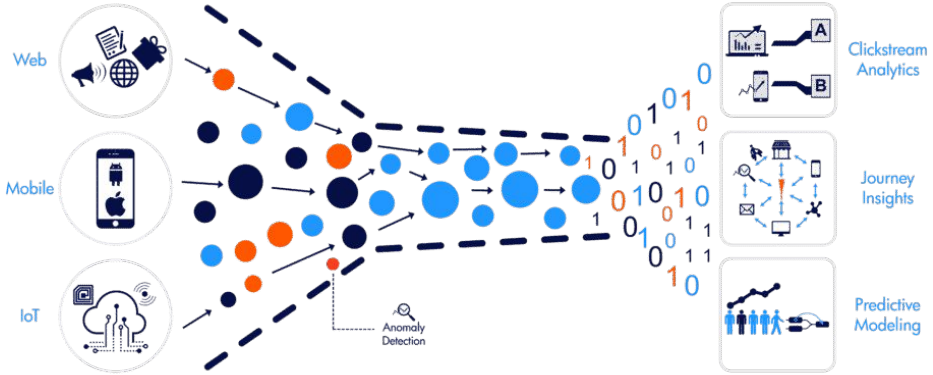
Gunes Acar
Radboud University
g.acar@cs.ru.nl

Mathias Humbert
University of Lausanne
mathias.humbert@unil.ch

Frederik Zuiderveen Borgesius
Radboud University
frederikzb@cs.ru.nl

Background

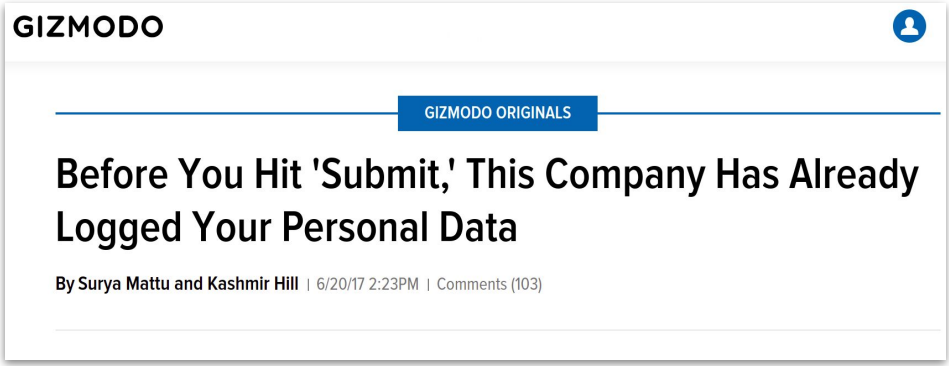
- Websites use advertising and marketing for monetization
  - built-in anti-tracking countermeasures
  - potential third-party cookie phase-out
- Tracking by email addresses
  - persistent, cross-site, cross-platform



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

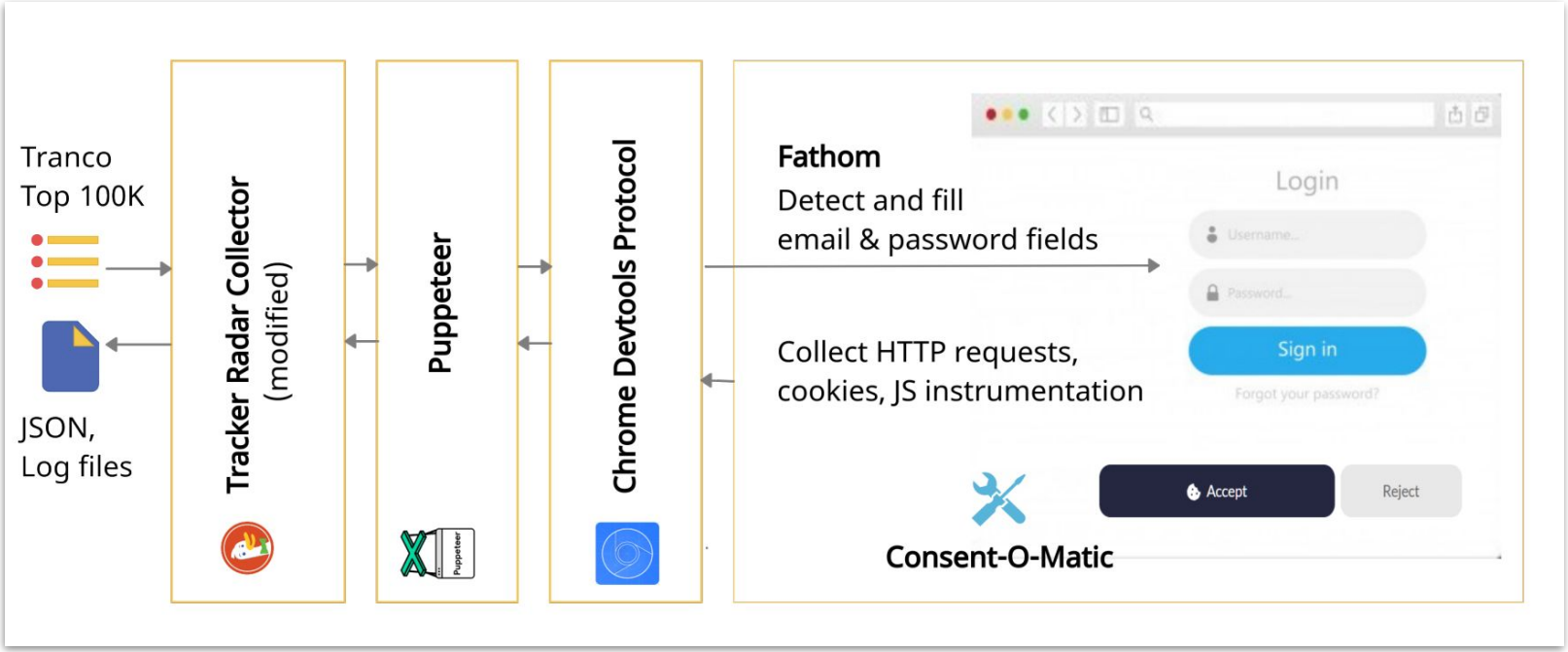


Study Objectives

- 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)



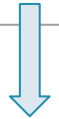
Crawls (May, June 2021)

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	**1,844**	**2,950**	**48**	**49**



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

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.

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*	bizible.com	128	businessinsider.com	taboola.com
243	independent.co.uk*	taboola.com	154	usatoday.com	taboola.com
300	shopify.com	bizible.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*	bizible.com	34	foxnews.com	rlcdn.com
1,153	prothomalo.com	facebook.com	242	trello.com*	bizible.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%)
Pornography	**1,388**	**528**	**0 (0.0%)**	**645**	**0 (0.0%)**

EU vs US

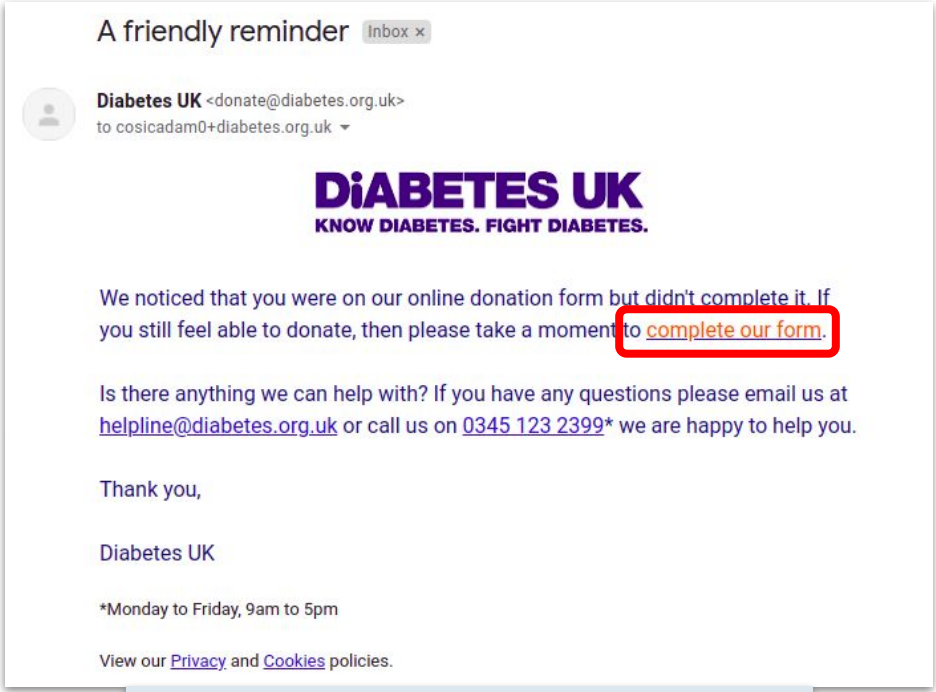
Num distinct websites	EU	US
Visited websites	99,380	99,437
Websites where we filled	52,055	53,038
Emails sent to 1st party	4,395	5,518
Emails sent to 3rd party	**2,633**	**3,790**
Emails sent to trackers	**1,844**	**2,950**

60% difference

addthis.com, yahoo.com, doubleclick.net and criteo.com → Only appear in the US crawl

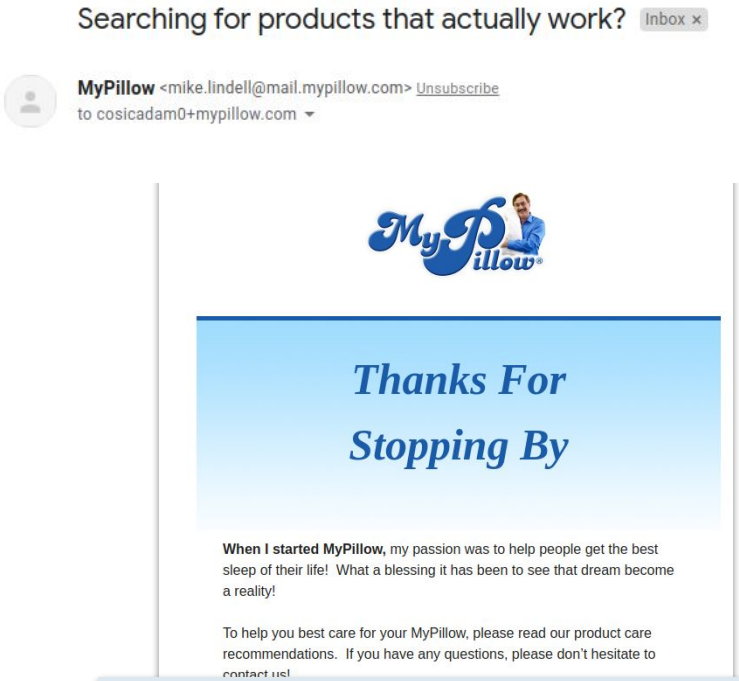
Received Emails

- 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

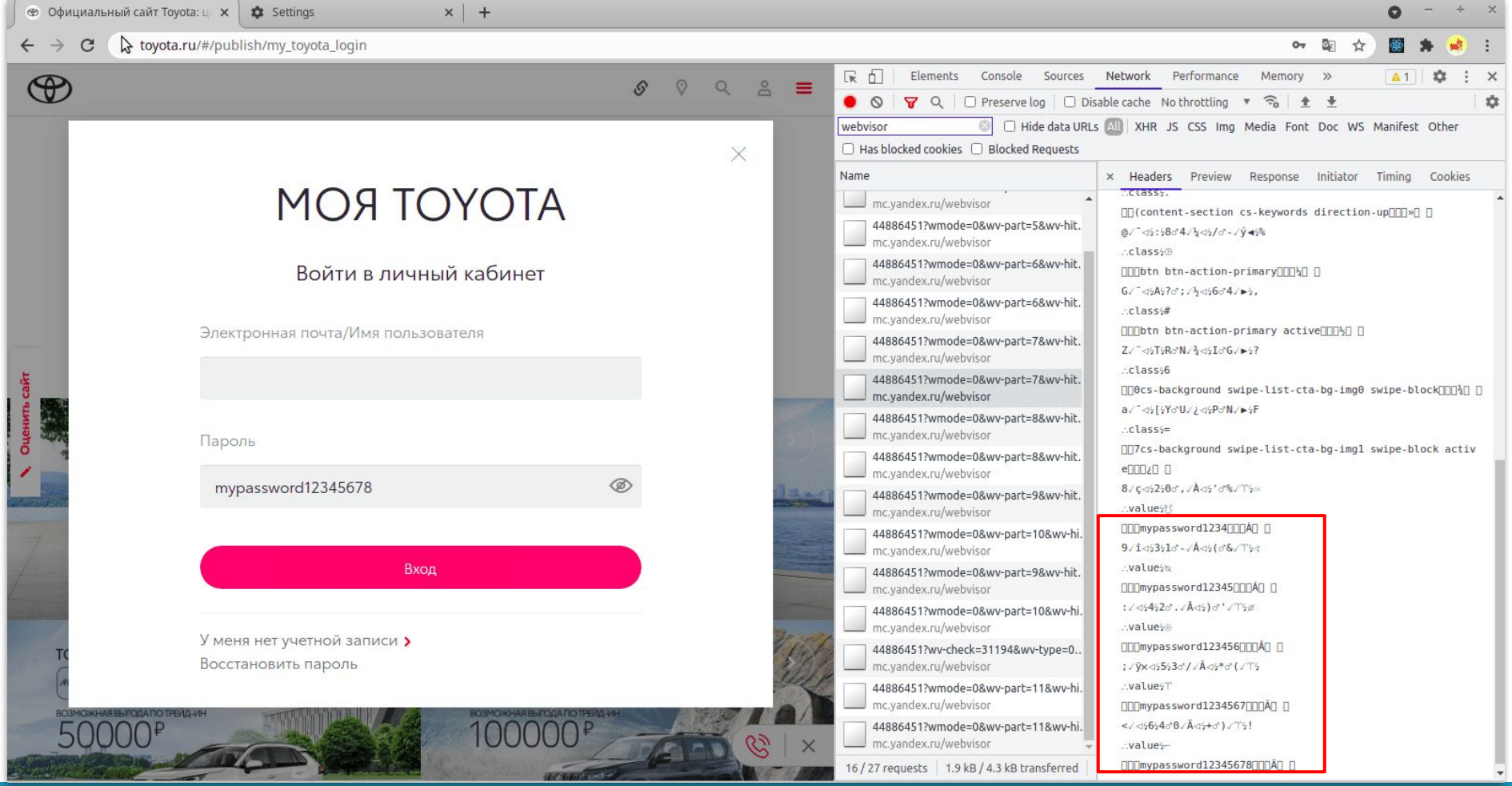


Email from: mypillow.com
Tracker domain: listrakbi.com

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
mypassword12345678
Вход
У меня нет учетной записи
Восстановить пароль
Оценить сайт
Elements
Console
Sources
Network
Performance
Memory
Preserve log
Disable cache
No throttling
webvisor
Hide data URLs
All
XHR
JS
CSS
Img
Media
Font
Doc
WS
Manifest
Other
Has blocked cookies
Blocked Requests
Name
Headers
Preview
Response
Initiator
Timing
Cookies
16 / 27 requests
1.9 kB / 4.3 kB transferred

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

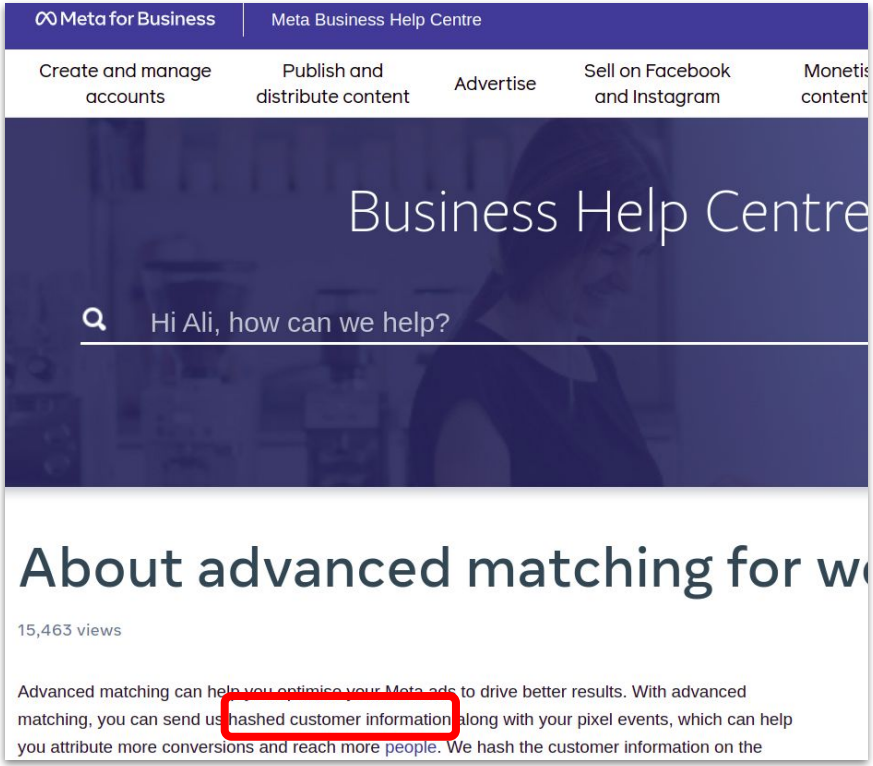
- Adobe and Yandex: Referred to corresponding first parties
- Taboola: ad & content personalization, CMP misconfiguration

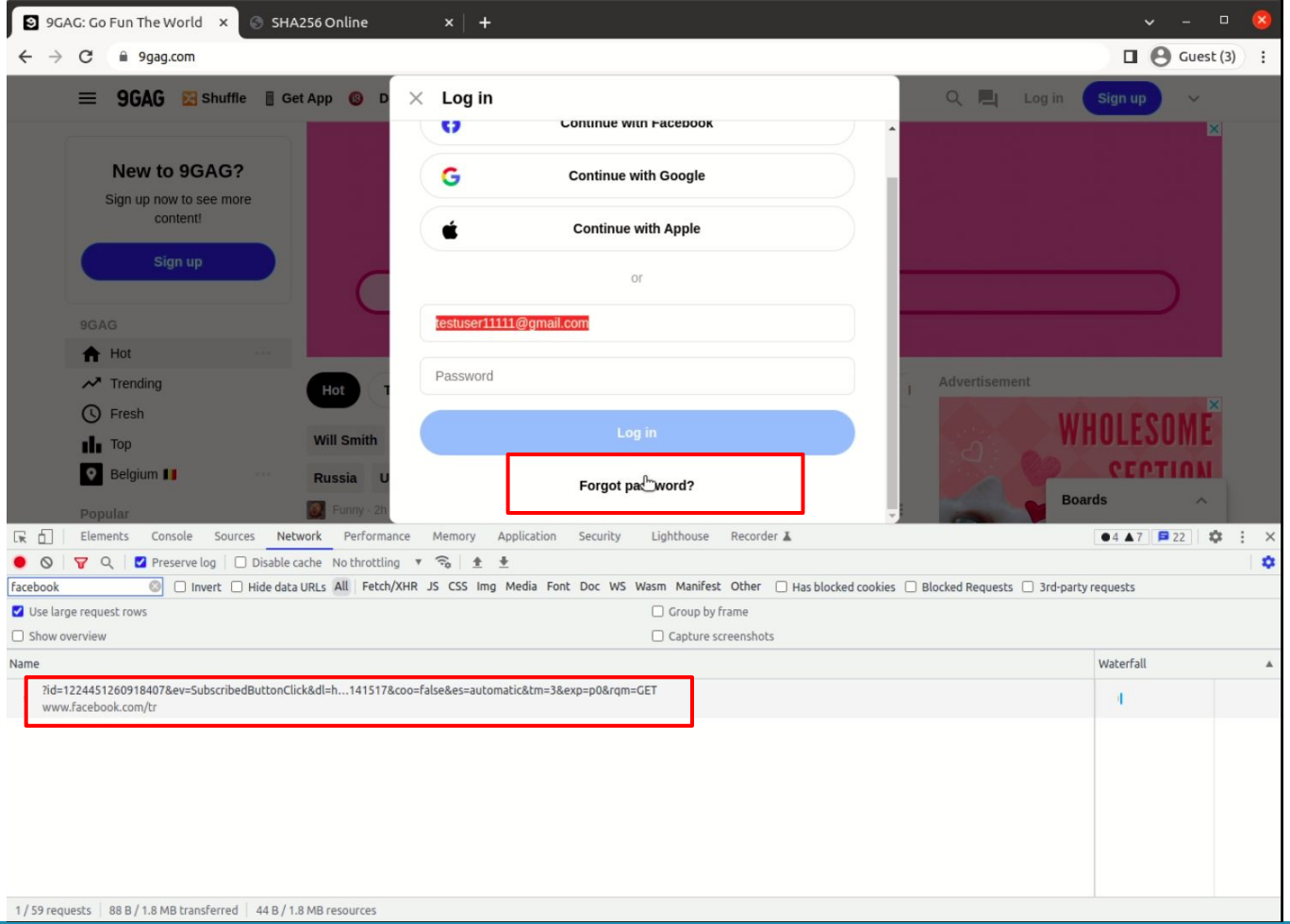
0/33 first parties replied (Websites in the US crawl)

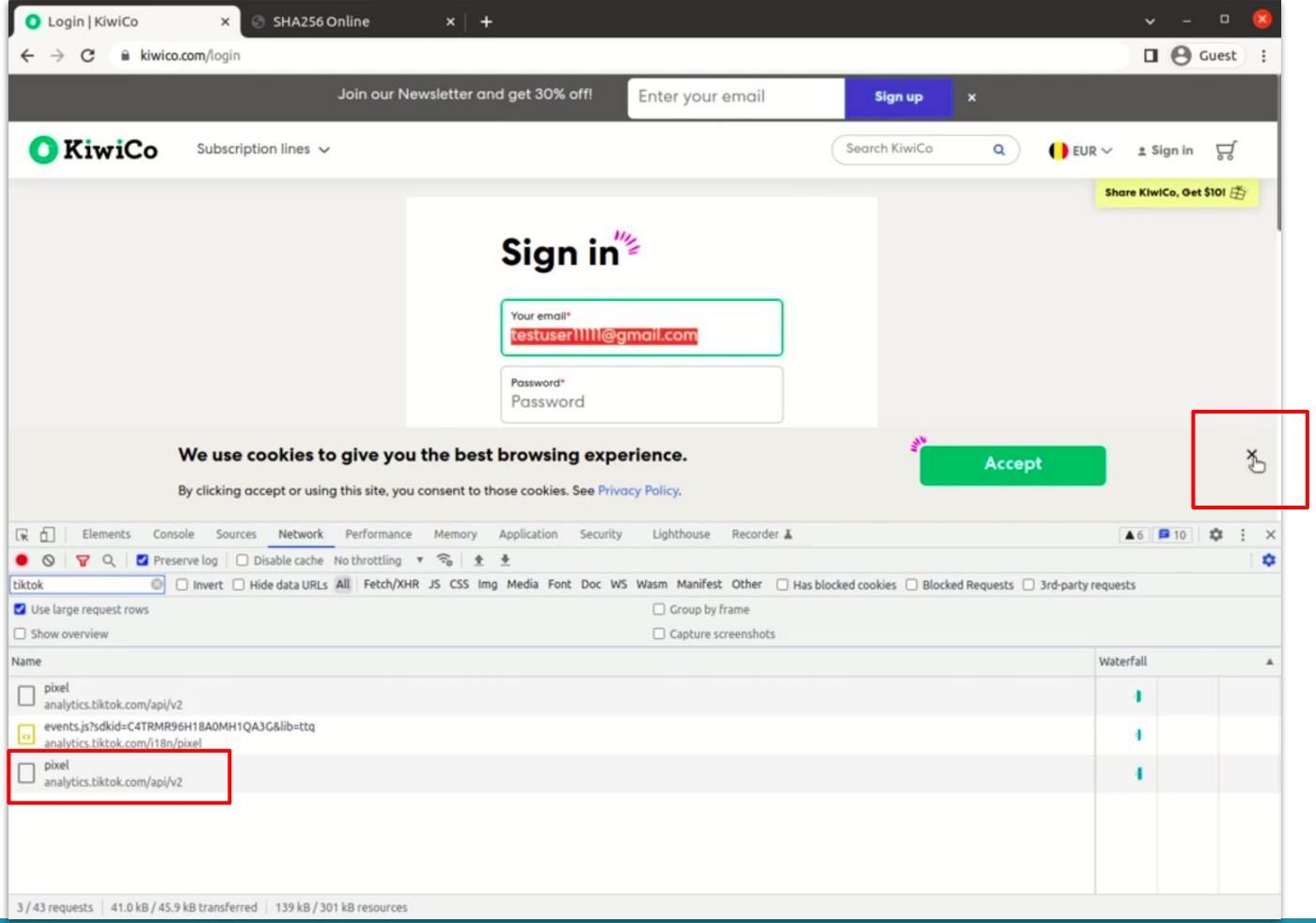
- No response from these 33 websites.

Leaks to Facebook & TikTok

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




9GAG: Go Fun The World
SHA256 Online
9gag.com
Guest (3)
9GAG
Shuffle
Get App
Log in
Sign up
New to 9GAG?
Sign up now to see more content!
Sign up
9GAG
Hot
Trending
Fresh
Top
Belgium
Popular
Hot
Will Smith
Russia
Log in
Continue with Facebook
Continue with Google
Continue with Apple
or
testuser11111@gmail.com
Password
Log in
Forgot password?
Advertisement
WHOLESOME
Boards
Elements
Console
Sources
Network
Performance
Memory
Application
Security
Lighthouse
Recorder
Preserve log
Disable cache
No throttling
facebook
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
?id=1224451260918407&ev=SubscribedButtonClick&dl=h...141517&coo=false&es=automatic&tm=3&exp=p0&rqm=GET
www.facebook.com/tr
1 / 59 requests
88 B / 1.8 MB transferred
44 B / 1.8 MB resources


Login | KiwiCo
SHA256 Online
kiwico.com/login
Guest
Join our Newsletter and get 30% off!
Enter your email
Sign up
KiwiCo
Subscription lines
Search KiwiCo
EUR
Sign in
Share KiwiCo, Get $10!
Sign in
Your email*
testuser11111@gmail.com
Password*
Password
We use cookies to give you the best browsing experience.
By clicking accept or using this site, you consent to those cookies. See Privacy Policy.
Accept
Elements
Console
Sources
Network
Performance
Memory
Application
Security
Lighthouse
Recorder
Preserve log
Disable cache
No throttling
tiktok
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
pixel
analytics.tiktok.com/api/v2
events.js?sdkid=C4TRMR96H18A0MH1QA3G&lib=ttq
analytics.tiktok.com/i18n/pixel
pixel
analytics.tiktok.com/api/v2
3 / 43 requests
41.0 kB / 45.9 kB transferred
139 kB / 301 kB resources

The Facebook company is now Meta. Learn more.

Home Implement Target Optimize Accelerate Scorecard Connect with a Partner

How it Works

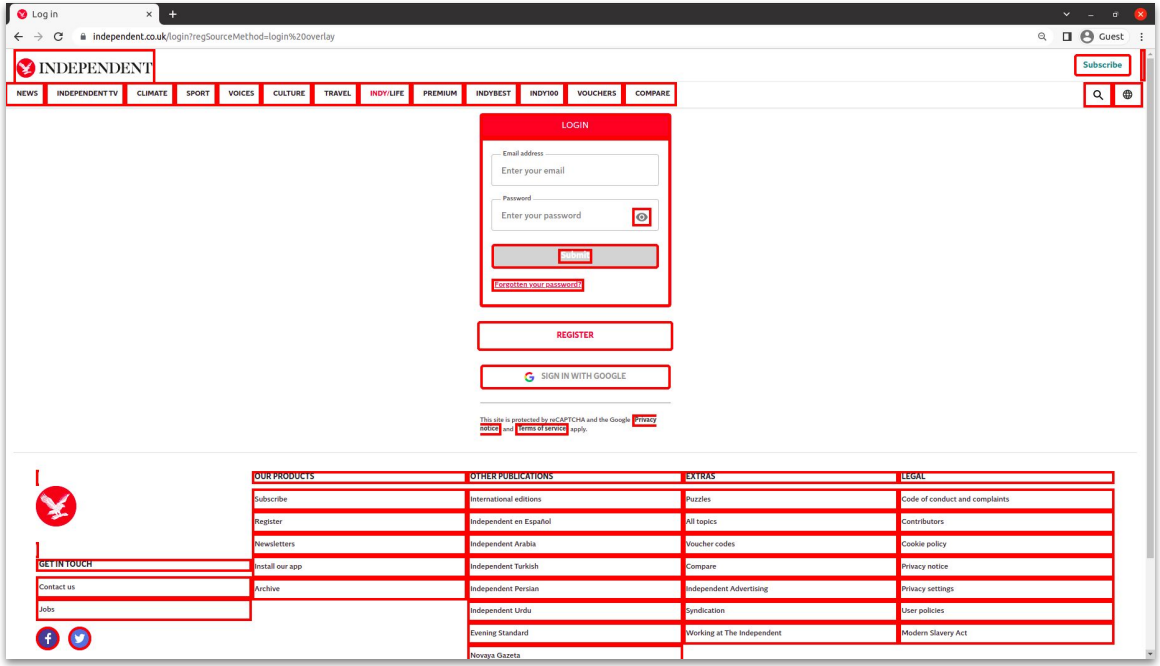
With automatic advanced matching, we can capture the hashed customer data (ex: email addresses) you collect from your website during processes like checkout, account sign-in, or registration. Hashing is the process we use to transform data for security reasons. We can then use hashed identifiers to better 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

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

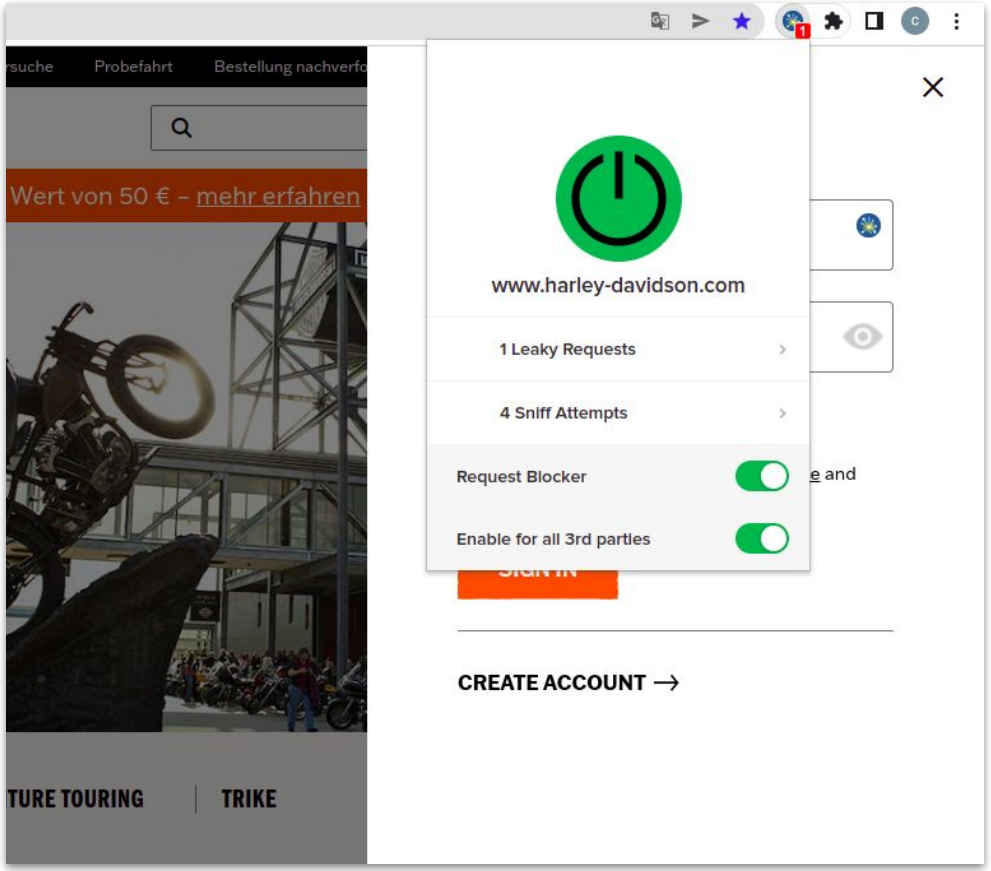


Countermeasures

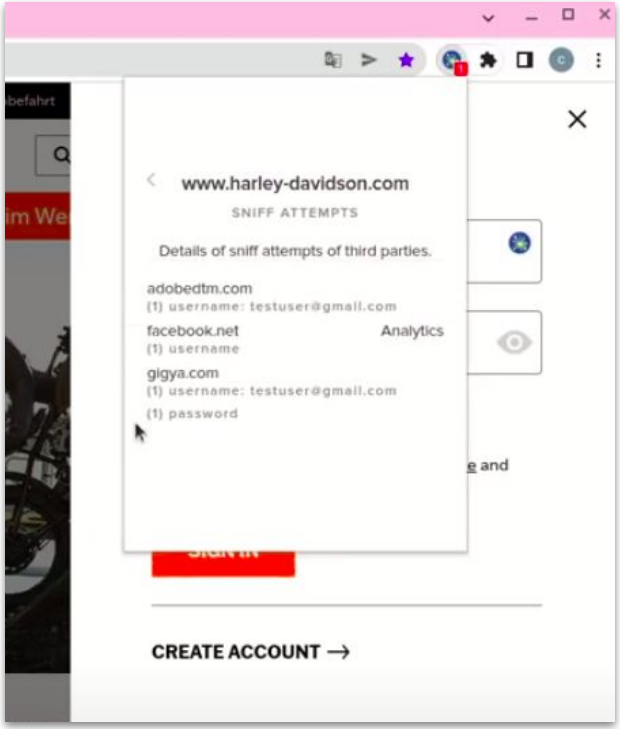
- Adblockers that block requests to tracker domains
- Private email relay services that hide users' emails
  - Apple, Mozilla, DuckDuckGo
  - e.g. testuser@duck.com-> testuser@gmail.com
- NO tool for detection and prevention of sniff & exfiltration on online forms

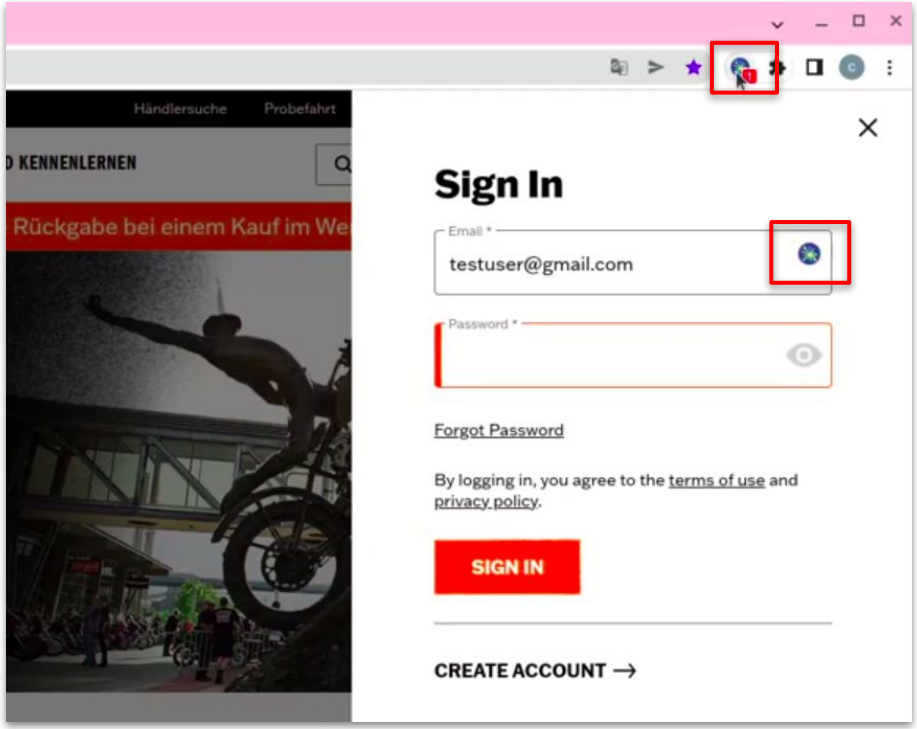
LEAKINSPECTOR

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



LEAKINSPECTOR





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?

- 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