It’s not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

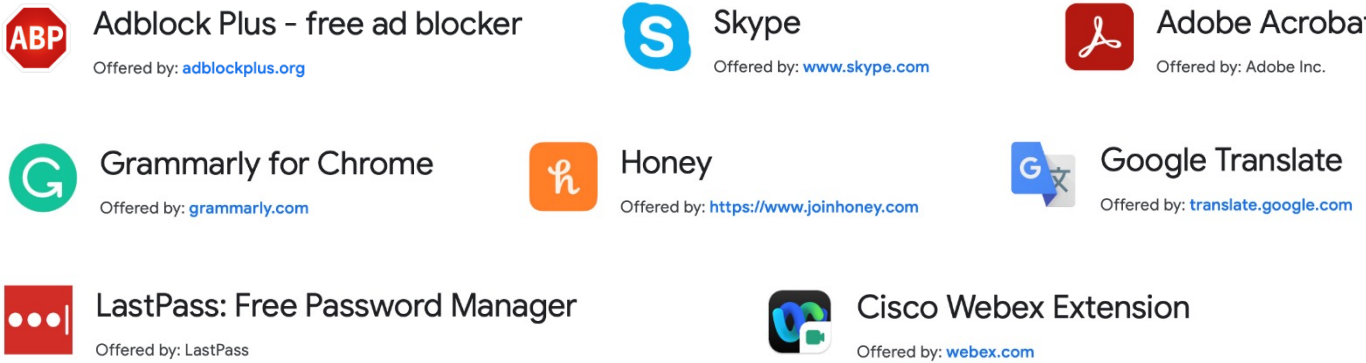
Ben Rosenzweig[DE], Valentino Dalla Valle[DE], Giovanni Apruzzese[IS], and Aurore Fass[FR]

Saarland University (DE), CISPA (DE), Reykjavik University (IS), Inria Centre at Université Côte d’Azur (FR)

Published at ACM Transactions on the Web (TWEB) 2026, presentation at TheWebConf 2026 (July 3rd)

What are Browser Extensions?

- Third-party programs to **improve user browsing experience**



- **~270k** Chrome extensions totaling **~3B** active users
  - ➢ Necessitates a thorough **security and privacy assessment**

How Secure are Browser Extensions?

- Browser extensions provide additional functionality…
- … so browser extensions need additional & elevated privileges compared to web pages
  - e.g., ad-blockers need to modify web page content or intercept network requests

➢ Browser extensions are an attractive target for attackers 😈

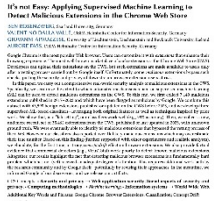
Detecting Malicious Extensions — Lab Setting


It's not Easy: Applying Supervised Machine Learning to Detect Malicious Extensions in the Chrome Web Store

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

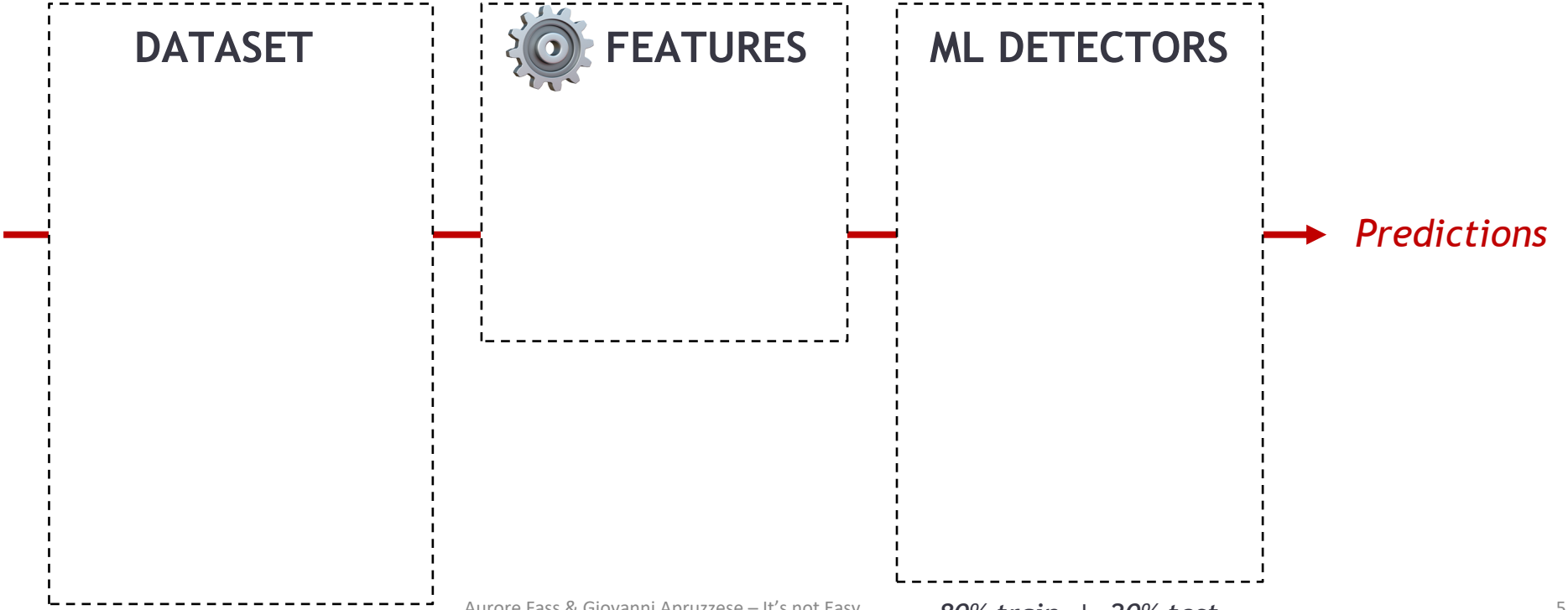
In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

Detecting Malicious Extensions — Lab Setting

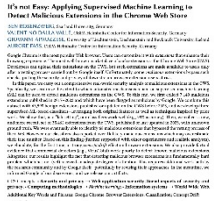

It's not Easy: Applying Supervised Machine Learning to Detect Malicious Extensions in the Chrome Web Store

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass



Detecting Malicious Extensions — Lab Setting


It's not Easy: Applying Supervised Machine Learning to Detect Malicious Extensions in the Chrome Web Store

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

DATASET

Chrome-Stats (malware)

+

Chrome Web Store (benign)

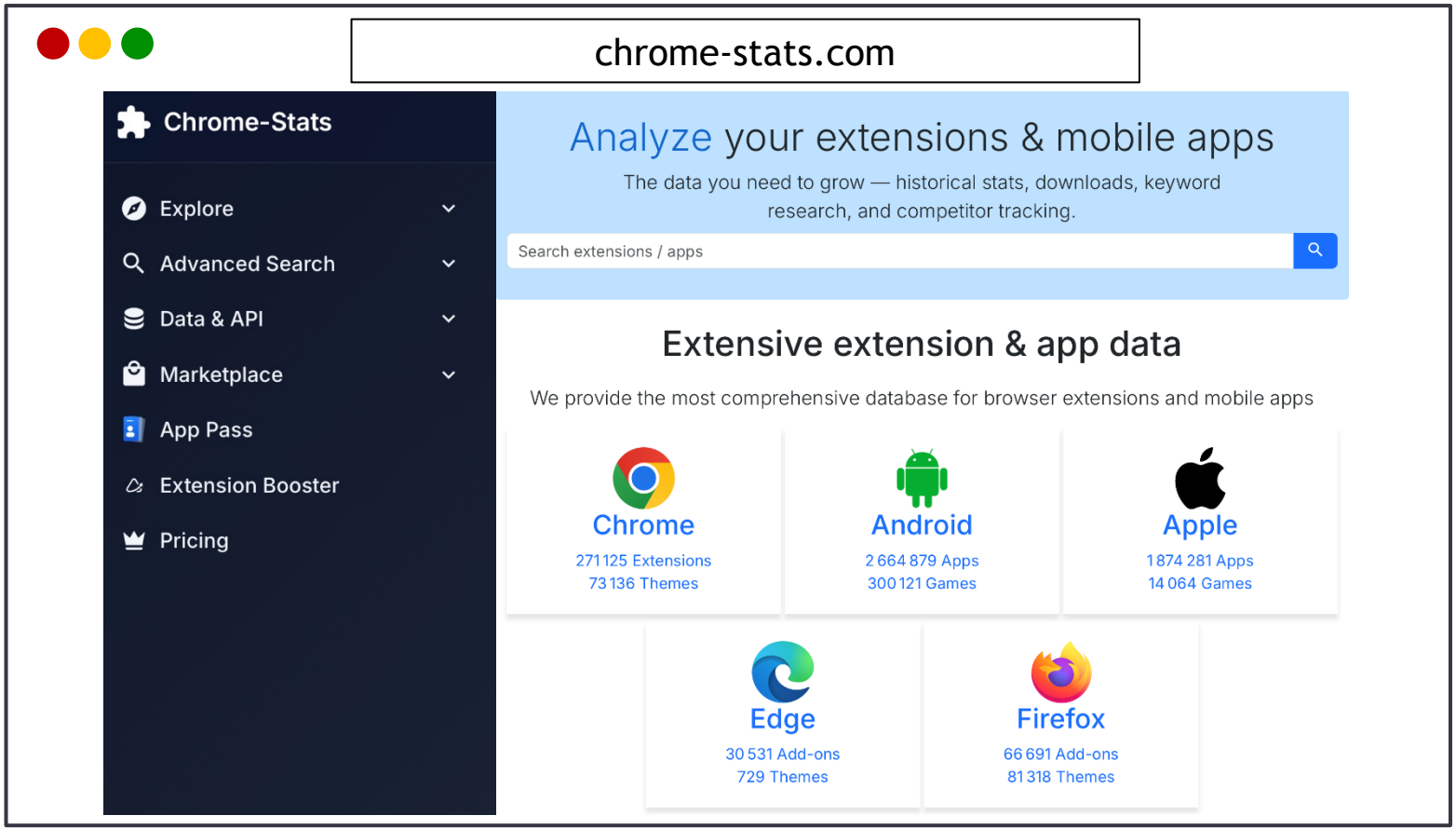
FEATURES

ML DETECTORS

Predictions

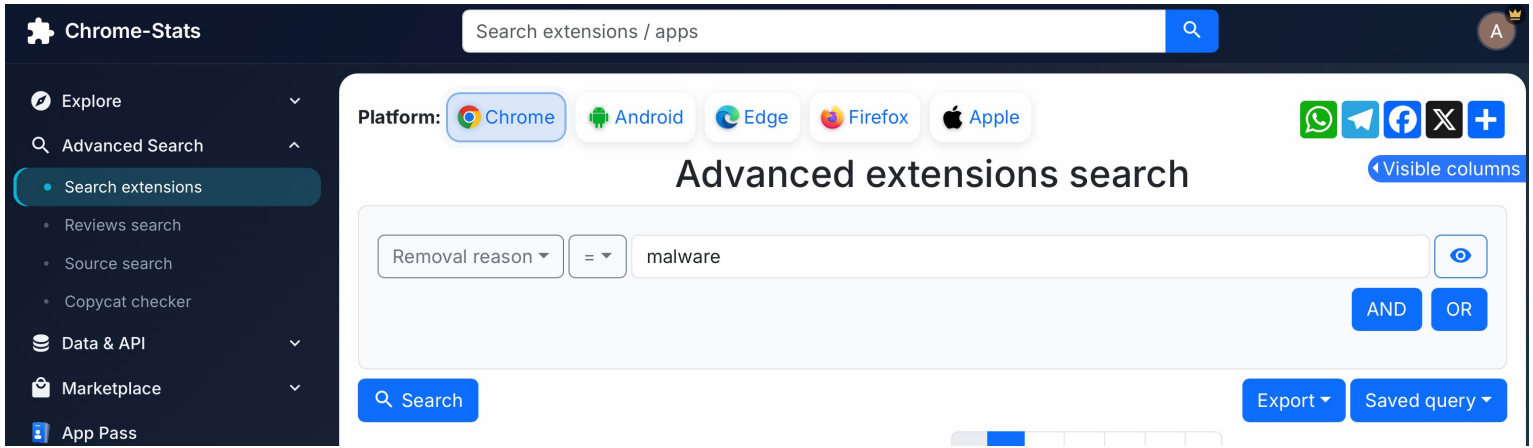
80% train | 20% test

Browser Extension Collection: Chrome-Stats



Browser Extension Collection: Chrome-Stats

chrome-stats.com



Detecting Malicious Extensions — Lab Setting


It's not Easy: Applying Supervised Machine Learning to Detect Malicious Extensions in the Chrome Web Store

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

DATASET

7,140 malicious extensions

Last update 2017—23

63,598 "benign" extensions

Last update 2012—22

FEATURES

ML DETECTORS

Predictions

80% train | 20% test

Detecting Malicious Extensions — Lab Setting


It's not Easy: Applying Supervised Machine Learning to Detect Malicious Extensions in the Chrome Web Store

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

DATASET

7,140 malicious extensions

Last update 2017—23

63,598 "benign" extensions

Last update 2012—22

FEATURES

Source code [A]

Metadata

[A] Fass et al.; *ACSAC* 2018

ML DETECTORS

80% train | 20% test

Predictions

Extension Features

Feature name	Feature length	Feature origin
Permissions	70	`manifest`
Host Permissions	400	`manifest`
Content Script Matches	400	`manifest`
Number of Content Scripts	1	`manifest`
Number of Service Workers	1	`manifest`
Number of Users	1	Chrome-Stats
Average Rating Score	1	Chrome-Stats
Number of Ratings	1	Chrome-Stats
Description Keywords	400	Chrome-Stats
Summary Keywords	400	Chrome-Stats
Review Keywords	400	Chrome-Stats
Same Developer Count	1	Chrome-Stats
CRX Size	1	File attributes
File Count	1	File attributes
JavaScript File Count	1	File attributes
JavaScript Size	1	File attribute
Related Permissions	70	`manifest` + Chrome-Stats
Source Code (AST)	2,457	JaSt [Fass et al.; *ACSAC* 2018]

Detecting Malicious Extensions — Lab Setting


It's not Easy: Applying Supervised Machine Learning to Detect Malicious Extensions in the Chrome Web Store

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

DATASET

7,140 malicious extensions

Last update 2017—23

63,598 "benign" extensions

Last update 2012—22

FEATURES

Source code [A]

Metadata

[A] Fass et al.; *ACSAC* 2018

(takes <1s per extension)

ML DETECTORS

(baseline) Source code classifier

Metadata classifier

Combined classifier

80% train | 20% test

Predictions

Detecting Malicious Extensions — Lab Setting

> It's not Easy: Applying Supervised ML to Detect Malicious Extensions in the CWS

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

DATASET

7,140 malicious extensions

Last update 2017—23

63,598 "benign" extensions

Last update 2012—22

FEATURES

Source code [A]

Metadata

[A] Fass et al.; *ACSAC* 2018

(takes <1s per extension)

ML DETECTORS

(baseline) Source code classifier

Metadata classifier

Combined classifier

80% train | 20% test

Predictions

Acc: 98.37%

TNR: 98.71%

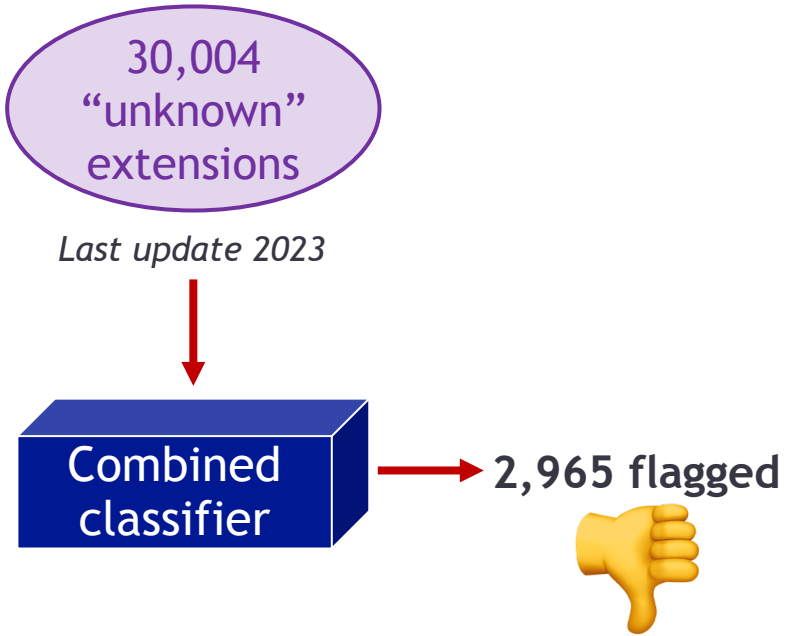
TPR: 95.39%

Detecting Malicious Extensions — Real World

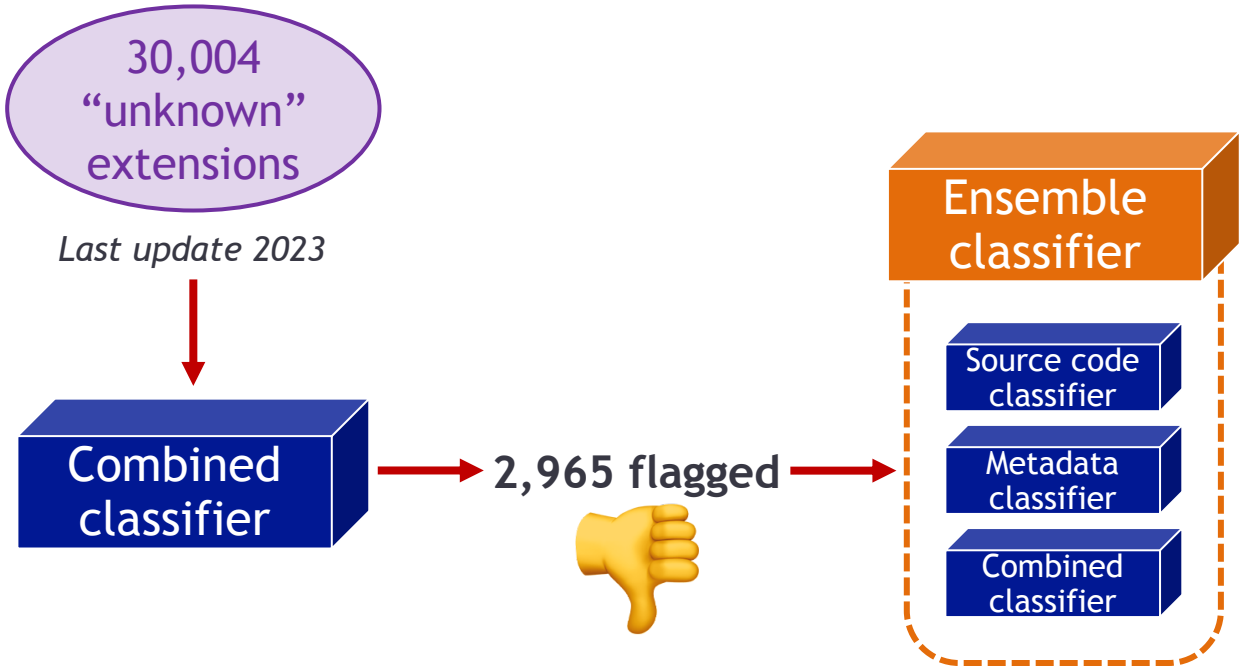
30,004 “unknown” extensions

Last update 2023

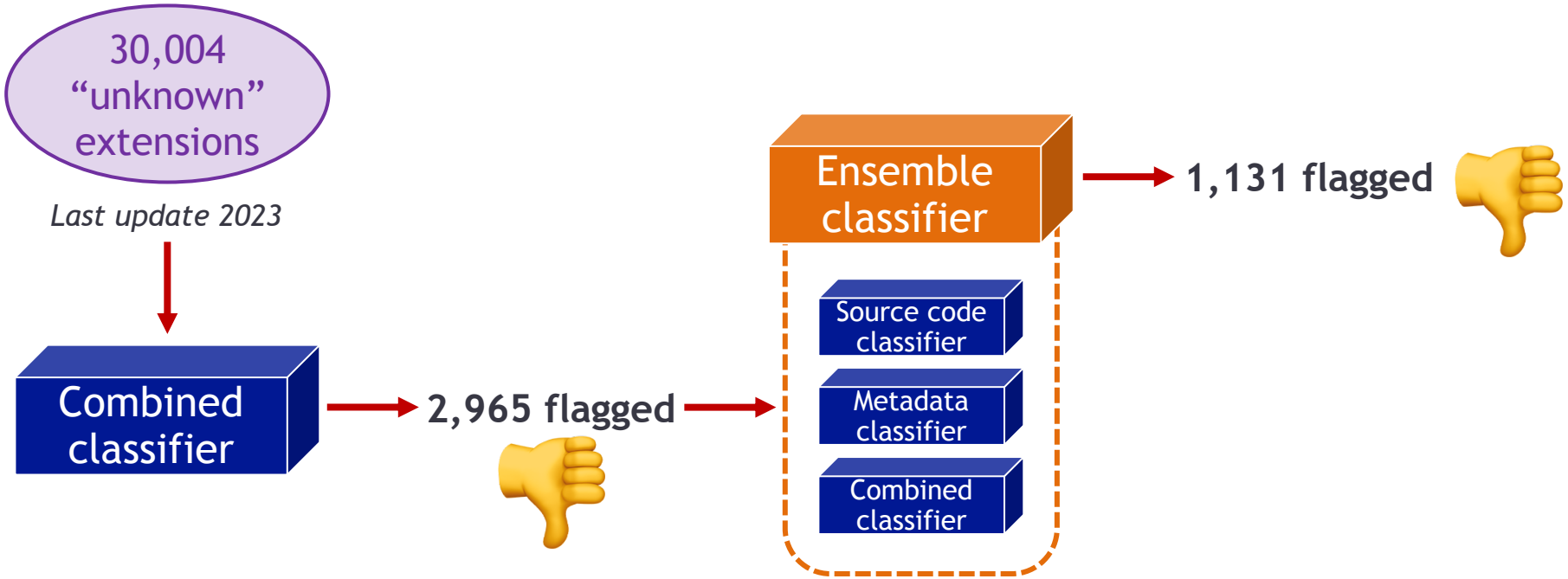
Detecting Malicious Extensions — Real World



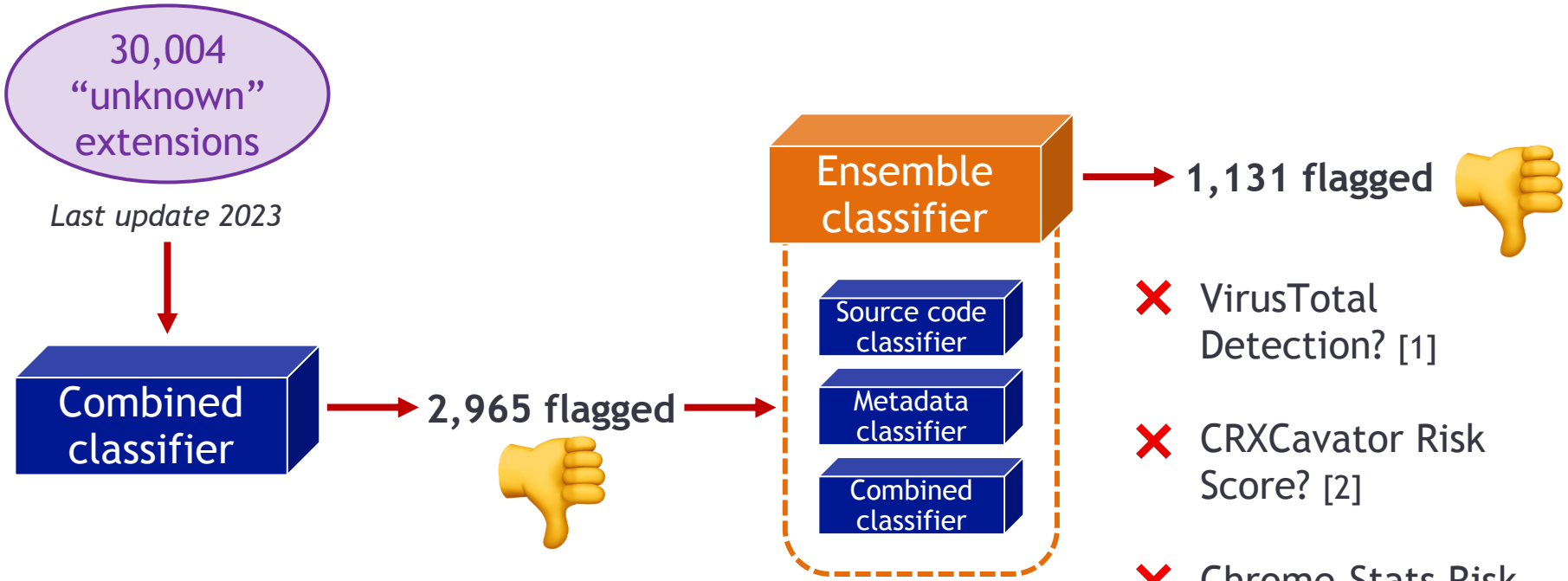
Detecting Malicious Extensions — Real World



Detecting Malicious Extensions — Real World

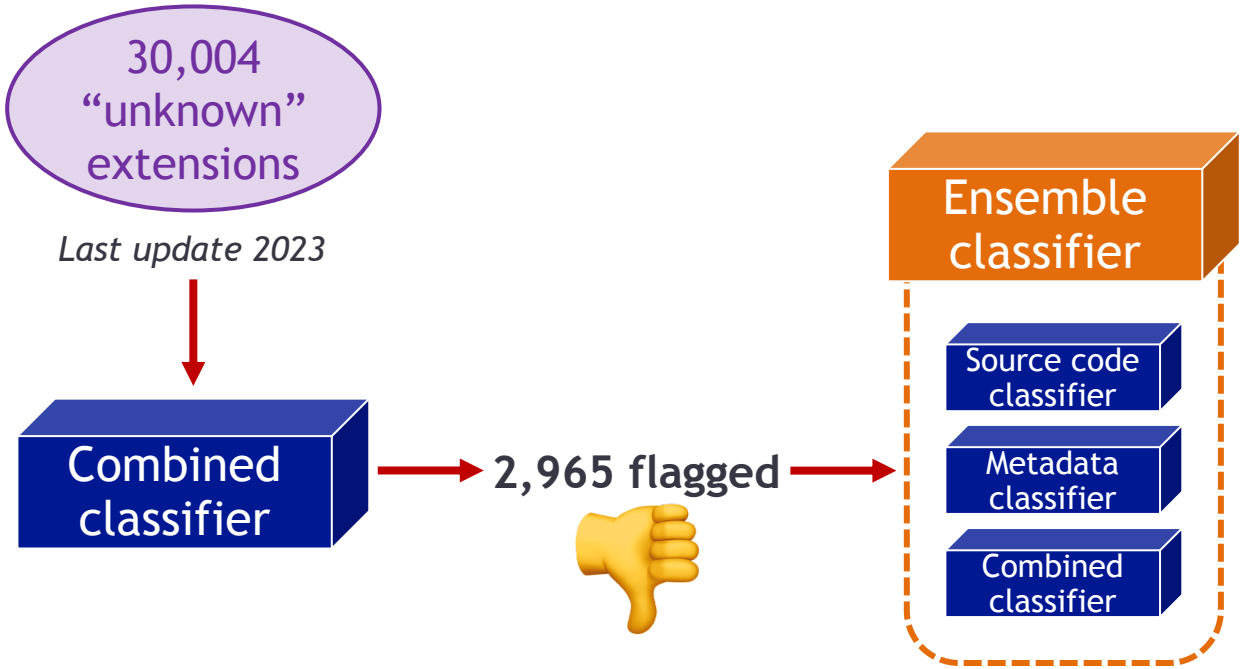


Detecting Malicious Extensions — Real World



[1] https://www.virustotal.com/gui/home/upload
[2] https://crxcavator.io
[3] https://chrome-stats.com

Detecting Malicious Extensions — Real World

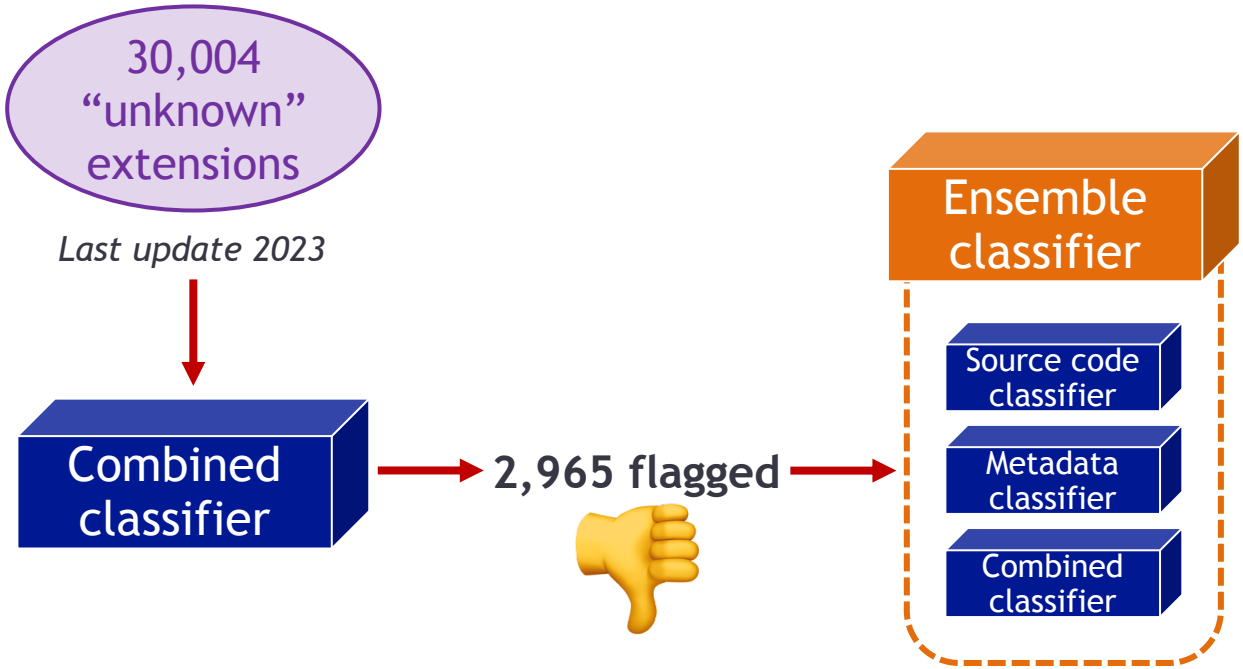


- **28 malicious** extensions removed by Google, **2M+ users**

- **40 malicious** ext. (after clustering + manual analysis on 20% flagged), **11M users**

Total: 68 malicious extensions, 13M users

Detecting Malicious Extensions — Real World



May 2024: **disclosure** of the 40 extensions to Google

As of June 2026: 17 / 40 malicious extensions were taken down

Total: 68 malicious extensions, 13M users

Detecting Malicious Extensions — Lab Setting vs. Real World

- **Lab setting**: 98.37% accuracy

vs.

- **Real world**: 1,131 / 30,004 extensions flagged | 68 verified malicious extensions

???

Detecting Malicious Extensions — Longitudinal Analysis

- We hypothesize that the Browser Extension ecosystem is affected by **concept drift**: extensions (benign and malicious) change over time

- We test this hypothesis…

…and we were right!

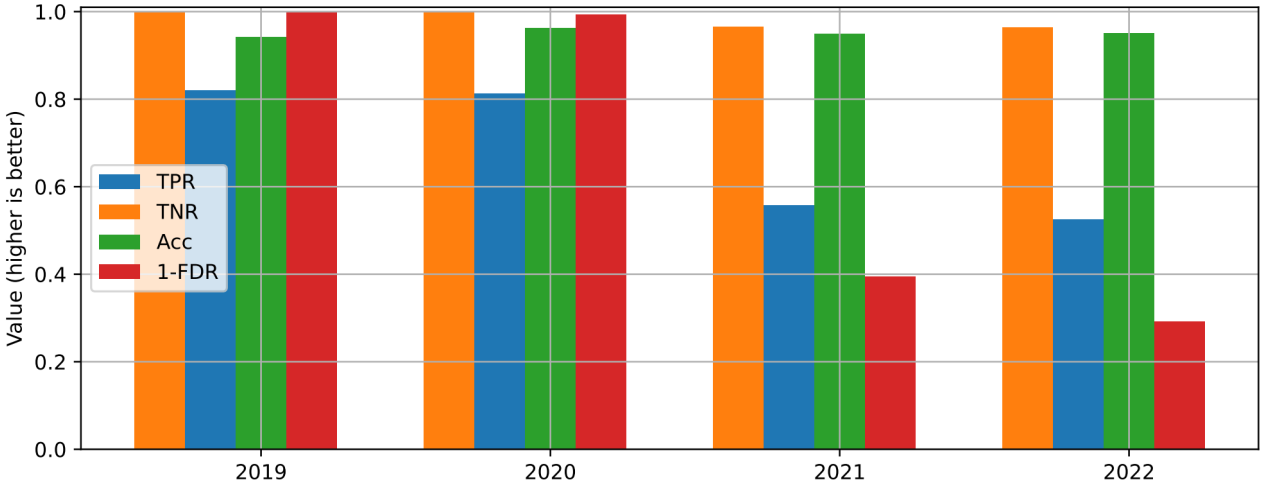


Fig. 7. Longitudinal test. We test the `Combined classifier` on the extensions updated every year from 2019–2022 after training it on the extensions published or updated in the previous years

Detecting Malicious Extensions — Focus on Manifest Version

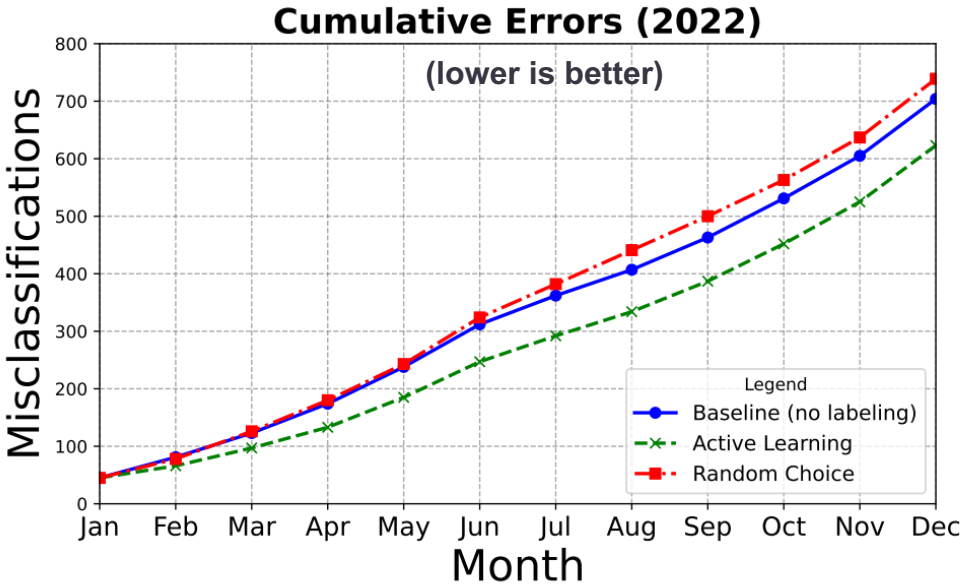
- We further examine the **"concept drift" phenomenon** by focusing on the classification performance of extensions across specific Manifest Versions (MV)

Manifest	Class	2019		2020		2021		2022	
		Benign		Malicious		Benign		Malicious	
		Accuracy	Total	Accuracy	Total	Accuracy	Total	Accuracy	Total
MV2	Benign	99.97%	6,514	99.86%	9,093	95.93%	9,377	94.27%	3,841
	Malicious	82.03%	3,061	81.25%	2,139	61.43%	407	83.13%	83
MV3	Benign	-	-	-	-	99.21%	2,014	97.13%	10,215
	Malicious	-	-	-	-	15.52%	58	44.51%	319

➢ …both benign *and* malicious extensions change substantially over the years!

Detecting Malicious Extensions — Active Learning

- We evaluate if **active learning** can mitigate the impact of concept drift
  - We assume the classifier is retrained on a monthly basis with a labeling budget of 15 extensions
  - (We are the first to assess active learning techniques in this specific context)



➢ There is a **sensible performance improvement**, but blindly picking extensions to label is detrimental

Takeaways: It's not Easy: Applying Supervised ML to Detect Malicious Extensions

In *ACM TWEB* 2026. Ben Rosenzweig, Valentino Dalla Valle, Giovanni Apruzzese, and Aurore Fass

- ML evaluations can be **misleading**: detectors performing well in a lab setting are **no guarantee of practical applicability** in the real world
- ML detectors need to be **retrained regularly**
- **Active learning** could be a **mitigation to concept drift**
- With our classifiers, we **uncovered 68 malicious browser extensions**, affecting **>13M users** (end-to-end, <1s per extension)
- Our code and datasets are **publicly available**:

https://github.com/its-not-easy/tweb25/

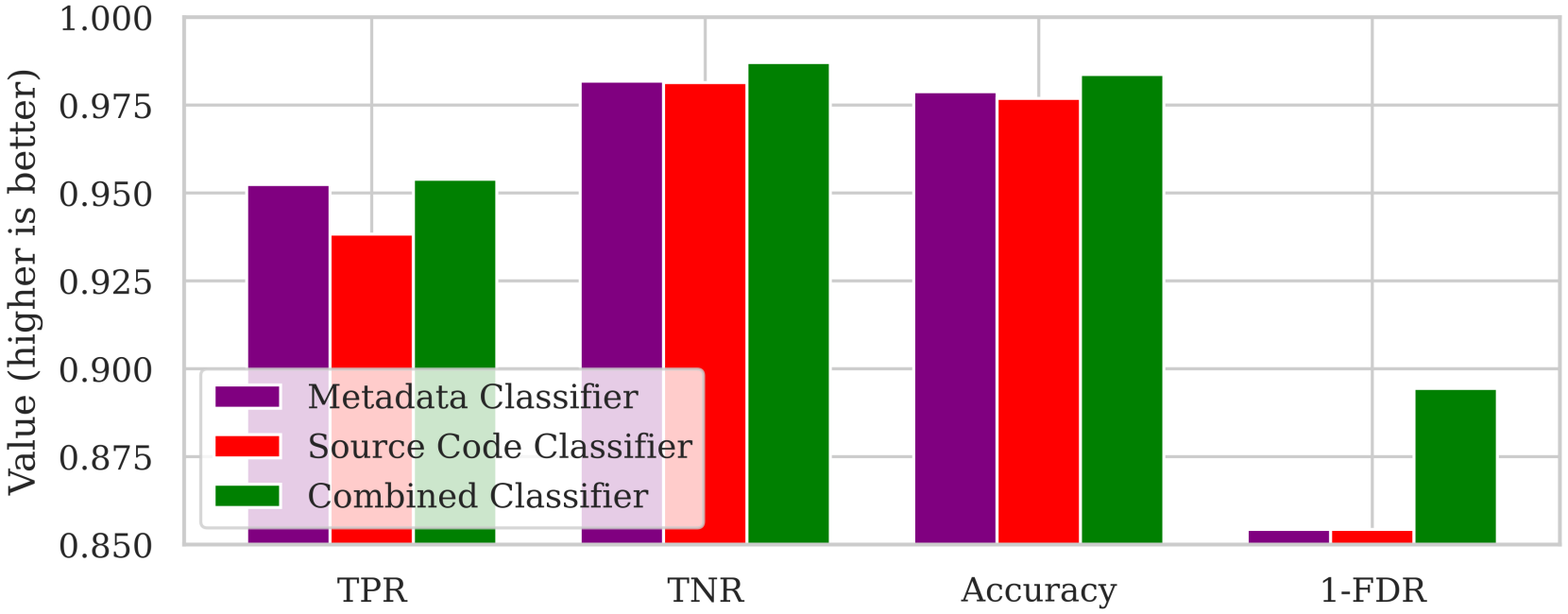
giovannia@ru.is aurore.fass@inria.fr https://www.giovanniapruzzese.com https://aurore54f.github.io

Back up slides

Dataset Summary

Label	Source	Last update	#Extensions
benign	CWS	before 2023-01-01	63,598
malicious	Chrome-Stats	any	7,140
unlabeled	CWS	after 2023-01-01	35,462

Detectors' Performance



Feature Importance Evolution over Time

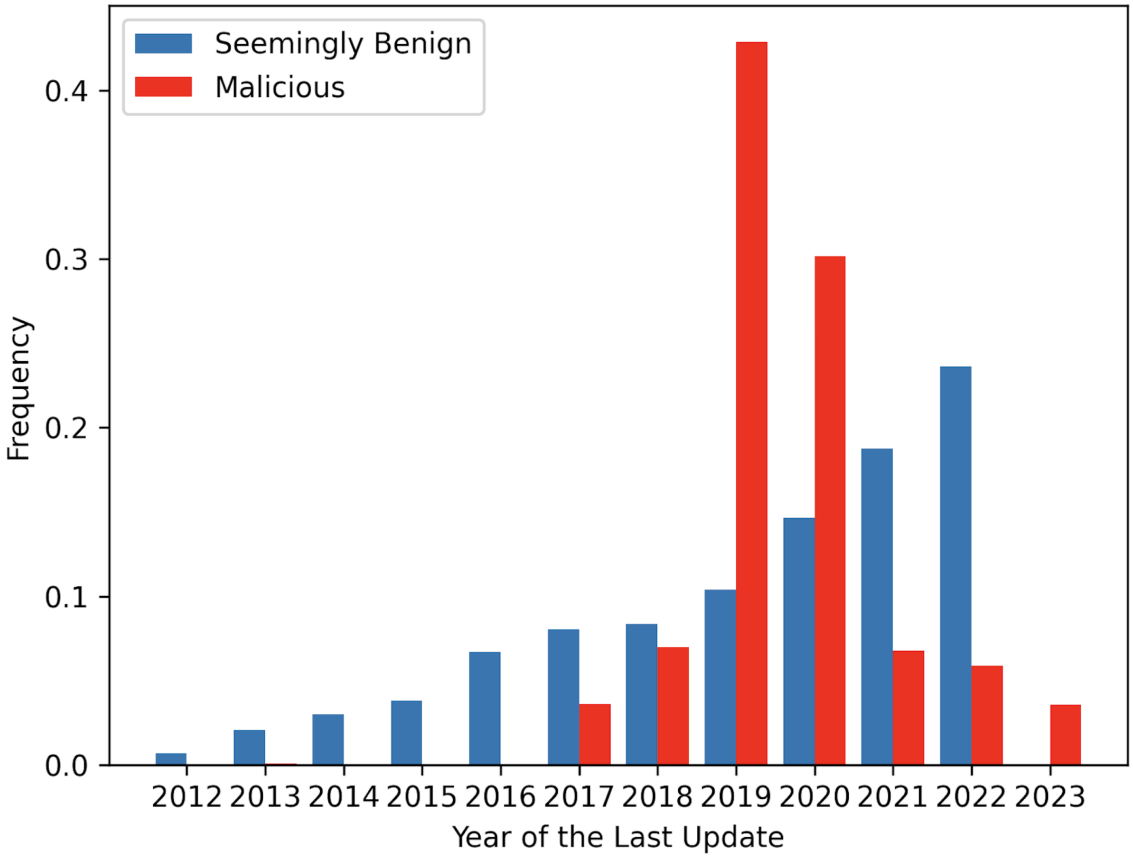
Rank (top-5)	2019		2020		2021	
	Feature Name	**Value**	**Feature Name**	**Value**	**Feature Name**	**Value**
#1	fileCount	0.0272	fileCount	0.0271	fileCount	0.0241
#2	Full-Summary LOVE	0.0212	Permissions topSites	0.0245	Full-Summary WALLPAPER	0.0223
#3	Full-Summary HIGH	0.0211	Full-Summary THEME	0.0196	size	0.0178
#4	(source code #157)	0.0209	Full-Summary WALLPAPER	0.0185	Full-Summary FAVORITE	0.0170
#5	(source code #105)	0.0171	Full-Summary FAVORITE	0.0175	Description TAB	0.0164

Rank (top-5)	2022		2023	
	Feature Name	**Value**	**Feature Name**	**Value**
#1	fileCount	0.0213	fileCount	0.0206
#2	Full-Summary WALLPAPER	0.0196	Full-Summary FAVORITE	0.0154
#3	size	0.0170	Description NEW	0.0151
#4	Description NEW	0.0146	Full-Summary TIME	0.0131
#5	Full-Summary THEME	0.0126	size	0.0117

No Open-Source State of the Art

Paper (1st author)	Year	Considered Browser	Code Public?	Detect Malicious?	Supervised ML used?	Number of Extensions
Kapravelos [58]	2014	C		✓	✗	48k
Jagpal [56]	2015	C		✓	✓	100k
Kurt [91]	2015	C,F,E		✓	✗	50k
Xing [99]	2015	C		✓	✗	18k
Buyukkayhan [35]	2016	F		✓	✗	323
Sanchez-Rola [80]	2017	C,F,S				21k
Starov [90]	2017	C				10k
Starov [89]	2017	C		✓	✗	10k
Weissbacher [97]	2017	C,F		✓	✓	43k
Aggarwal [23]	2018	C		✓	✓	178k
Chen [39]	2018	C,O	✓	✓	✗	19k
Trickel [92]	2019	C	✓			11k
Sjosten [82]	2019	F,C				11k
Some [87]	2019	F,C,O				79k
Starov [88]	2019	C	✓			58k
Karami [60]	2020	C				29k
Pantelaious [74]	2020	C	✓	✓	✗	209k
Borgolte [31]	2020	C,F	✓			8
Laperdrix [65]	2021	C	✓			116k
Fass [48]	2021	C,F	✓			154k
Solomos [84]	2022	C	✓			40k
Karami [61]	2022	C	✓			5.7k
Solomos [85]	2022	C				38k
Picazo-Sanchez [77]	2022	C		✓	✓	159k
Agarwal [21]	2022	C,F	✓			186k
Kim [63]	2023	C,F,S	✗			<100
Bui [34]	2023	C				47k
Moreno [69]	2023	C	✓			–
Yu [102]	2023	C	✓			145k
Xie [98]	2024	C	✓			113k
Nayak [70]	2024	C		✓	✗	160k
Olsson [72]	2024	C				115k
Hsu [55]	2024	C				226k
Solomos [86]	2024	C				60k
Agarwal [22]	2024	C,F	✓			88k
This work	2024	C	[20]	✓	✓	106k

Year of Last Update (Labeled Extensions)



Malicious Behaviors Identified

	#Extensions	#Clusters
Proxying network requests through endpoints	2	1
Navigating the user to sites VirusTotal reports as malicious	3	3
Retrieving the list of friends and account details and publishing a Facebook post that includes the CPU model, CPU, and RAM usage automatically and without permission	2	1
Requests & exfiltrate credentials for popular websites	6	1
Exfiltrating the IP address, operating system, and a generated UUID to various endpoints	22	19
Collecting system information and exfiltrating it to endpoints reported as malicious on VirusTotal	2	1
Ad Blockers that send every visited URLs to an endpoint reported as malicious on VirusTotal	3	3
Total	40	29

Risks Score / Risk Likelihood

Risk score (CRXcavator) [2], in % of extensions:

	<300	300-399	400-499	500-599	600-699	>699	#Ext
Flagged by all 3 classifiers	0.46	23.12	41.71	31.71	1.90	1.10	1,131
Dataset U minus flagged	0.34	44.47	44.52	6.50	2.17	2.00	34,331

Risk likelihood (Chrome-Stats) [3], in % of extensions:

	0	1	2	3	4	#Ext
Flagged by all 3 classifiers	11.09	34.61	34.25	11.98	8.07	1,131
Dataset U minus flagged	12.32	6.18	53.96	26.13	1.41	34,331

[2] https://crxcavator.io

[3] https://chrome-stats.com

Top 10 “Summary Keywords” in the Training Set

Rank	Benign Ext.	Malicious Ext.
1.	EXTENSION	EXTENSION
2.	CLICK	NEW
3.	PAGE	TAB
4.	CHROME	TIME
5.	USE	WALLPAPER
6.	ADD	FAVORITE
7.	NEW	THEME
8.	TIME	HIGH
9.	WEBSITE	BACKGROUND
10.	BROWSER	FEATURE

Top 10 “Host Permissions” in the Training Set

Rank	Benign extensions	Malicious extensions
1.	<all_urls>	*://www.google.com.kh/*
2.	https://*/*	*://mail.google.com/*
3.	http://*/*	*://www.google.com.au/*
4.	*://*/*	*://www.google.us/*
5.	http://*/	*://www.google.ca/*
6.	https://*/	*://www.google.de/*
7.	https://ajax.googleapis.com/	*://www.google.dk/*
8.	chrome://favicon/	*://www.google.fr/*
9.	https://*.1688.com/*	*://www.google.co.jp/*
10.	https://www.youtube.com/*	*://www.google.nl/*

Top 10 “Content Script Matches” in the Training Set

Rank	Benign extensions	Malicious extensions
1.	<all_urls>	<all_urls>
2.	https://*/*	*://duckduckgo.com/*
3.	http://*/*	*://gl-search.com/*
4.	*://*/*	*://redirect.lovelytab.com/*
5.	https://www.youtube.com/*	*://str-search.com/*
6.	https://mail.google.com/*	*://search.yahoo.com/search*
7.	https://twitter.com/*	*://www.google.com/search*
8.	https://github.com/*	*://www.bing.com/search*
9.	*://*.youtube.com/*	https://happyhey.com/*
10.	file://*/*	https://www.happyhey.com/*

Distribution of “File Count” Feature for benign / malicious