



# Arcanum: Detecting and Evaluating the Privacy Risks of Browser Extensions on Web Pages and Web Content

Qinge Xie, Manoj Vignesh K M, Paul Pearce and Frank Li



Georgia Tech College of Computing  
School of Cybersecurity  
and Privacy



**BEES** Lab

# Motivation: Browser Extension Scraping

Home / Tech / Security

## Facebook sues two Chrome makers for scraping user data

Facebook has sued today the makers of the **UpVoice** and **Ads Feed** Chrome extensions.



Written by **Catalin Cimpanu**, Contributor

Oct. 1, 2020 at 2:34 p.m. PT

Are there a lot more cases?

2

Meta

## Combating Scraping by Malicious Browser Extensions

January 14, 2021

By Jessica Romero, Director of Platform Enforcement and Litigation

Facebook Inc. and Facebook Ireland have filed a legal action in Portugal against two people for scraping user-profiles and other data from Facebook’s website, in violation of our Terms of Service and Portugal’s Database Protection Law.

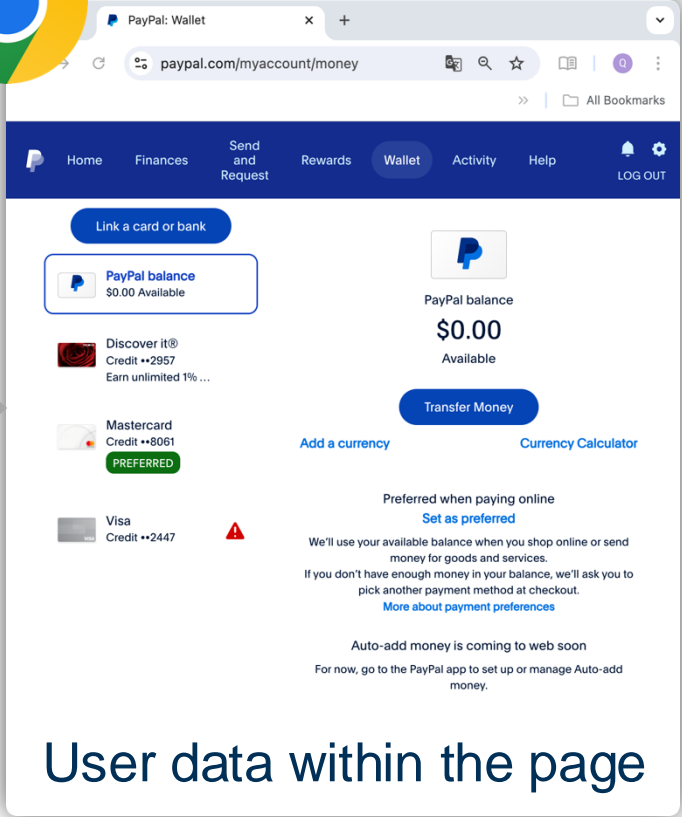
Using the business name “Oink and Stuff,” the defendants developed browser extensions and made them available on the Chrome store. They misled users into installing the extensions with a privacy policy that claimed they did not collect any personal information. **Four of their extensions — Web for Instagram plus DM, Blue Messenger, Emoji keyboard and Green Messenger** were malicious and contained hidden computer code that functioned like spyware.

# Privacy Issue: Browser Extension Access

**Access**  
Browsing Data  
(e.g., chrome.history)

  
**Extension**

**Access**  
DOM



The screenshot shows the PayPal Wallet interface in a Chrome browser. The address bar displays 'paypal.com/myaccount/money'. The page content includes a navigation menu, a 'PayPal balance' section showing '\$0.00 Available', and a list of credit cards: Discover it@, Mastercard (marked 'PREFERRED'), and Visa. A 'Transfer Money' button is visible. At the bottom of the page, there is a notice about 'Auto-add money'.

User data within the page

# Prior Work and Motivation

Several prior works have looked at this privacy issue, but only in terms of browser **APIs** and a **limited set of DOM properties**.

2009 Annual Computer Security Applications Conference

## **Analyzing Information Flow in JavaScript-based Browser Extensions**

Mohan Dhawan and Vinod Ganapathy  
Department of Computer Science, Rutgers University

Session 9A: Web 2

CCS'18, October 15-19, 2018, Toronto, ON, Canada

## **Mystique: Uncovering Information Leakage from Browser Extensions**

Quan Chen  
North Carolina State University  
qchen10@ncsu.edu

Alexandros Kapravelos  
North Carolina State University  
akaprav@ncsu.edu

# Prior Work and Motivation

Several prior works have looked at this privacy issue, but only in terms of browser **APIs** and a **limited set of DOM properties**.

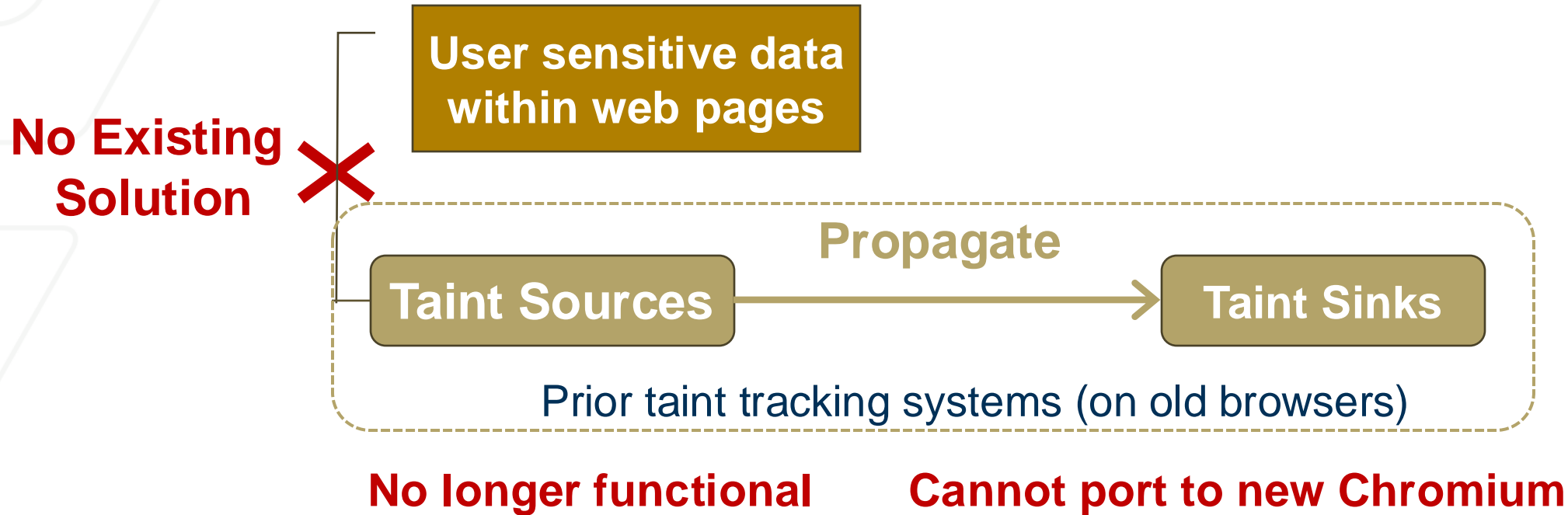
*What about all the other **sensitive data** within webpages?*



***Our Goal:***

*To understand how extensions automatically collect web page content.*

# Dynamic Taint Tracking Approach



## 1. Modern Browser Architecture:

- New JS execution pipeline in V8 engine
- Migrated JS implementation to native C++

## 2. Modern Extensions: Do not support Manifest Version 3 (MV3) extensions

## 3. Modern Websites: New JS expressions/operators (e.g., LinkedIn page)

# Our System

**Arcanum:** A *dynamic taint tracking* system for Chromium designed to track sensitive user *content* on modern web pages and extensions.

**Key distinction from prior systems:** Arcanum can

- **Main:** Track user sensitive data from within web pages,
- **Secondary:** Operate on modern browser architecture, support taint propagation across a broader set of browser, web, and JavaScript APIs

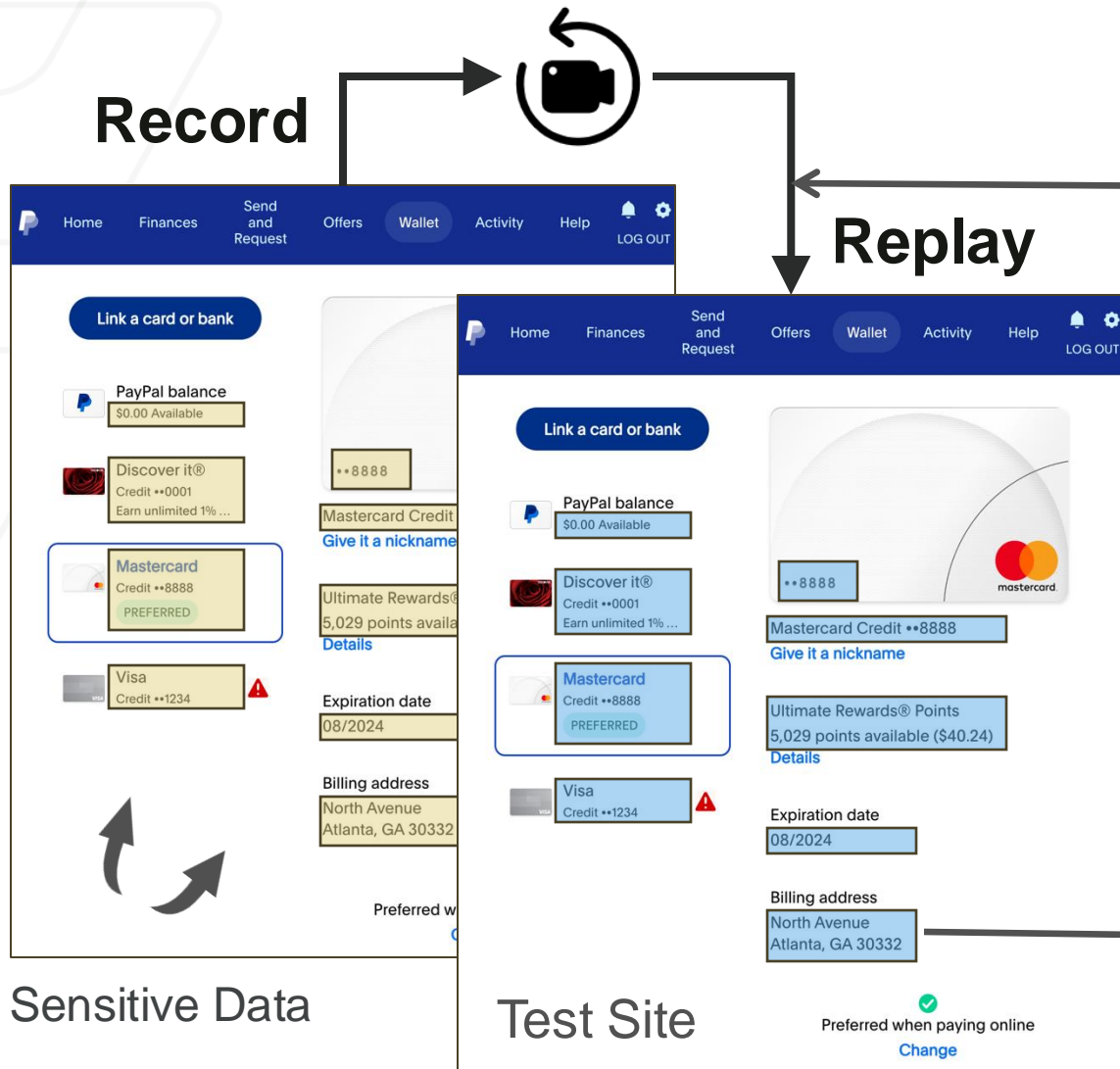
# Overview of Arcanum

The screenshot shows the PayPal Arcanum interface. At the top is a dark blue navigation bar with the PayPal logo and links for Home, Finances, Send and Request, Offers, Wallet (highlighted), Activity, Help, and a LOG OUT button. Below the navigation bar, there is a 'Link a card or bank' button. A list of cards is displayed: PayPal balance (\$0.00 Available), Discover it® Credit (••0001), Mastercard Credit (••8888, highlighted with a blue box and labeled 'PREFERRED'), and Visa Credit (••1234). To the right, a detailed view of the Mastercard Credit card is shown, including the card number (••8888), the name 'Mastercard Credit ••8888', and a 'Give it a nickname' link. Below this, the 'Ultimate Rewards® Points' section shows 5,029 points available (\$40.24) with a 'Details' link. The 'Expiration date' is 08/2024, and the 'Billing address' is North Avenue, Atlanta, GA 30332. At the bottom, there is a green checkmark icon and the text 'Preferred when paying online' with a 'Change' link. A large grey arrow icon is positioned to the left of the bottom right section.

Researchers identify sensitive information



# Overview of Arcanum



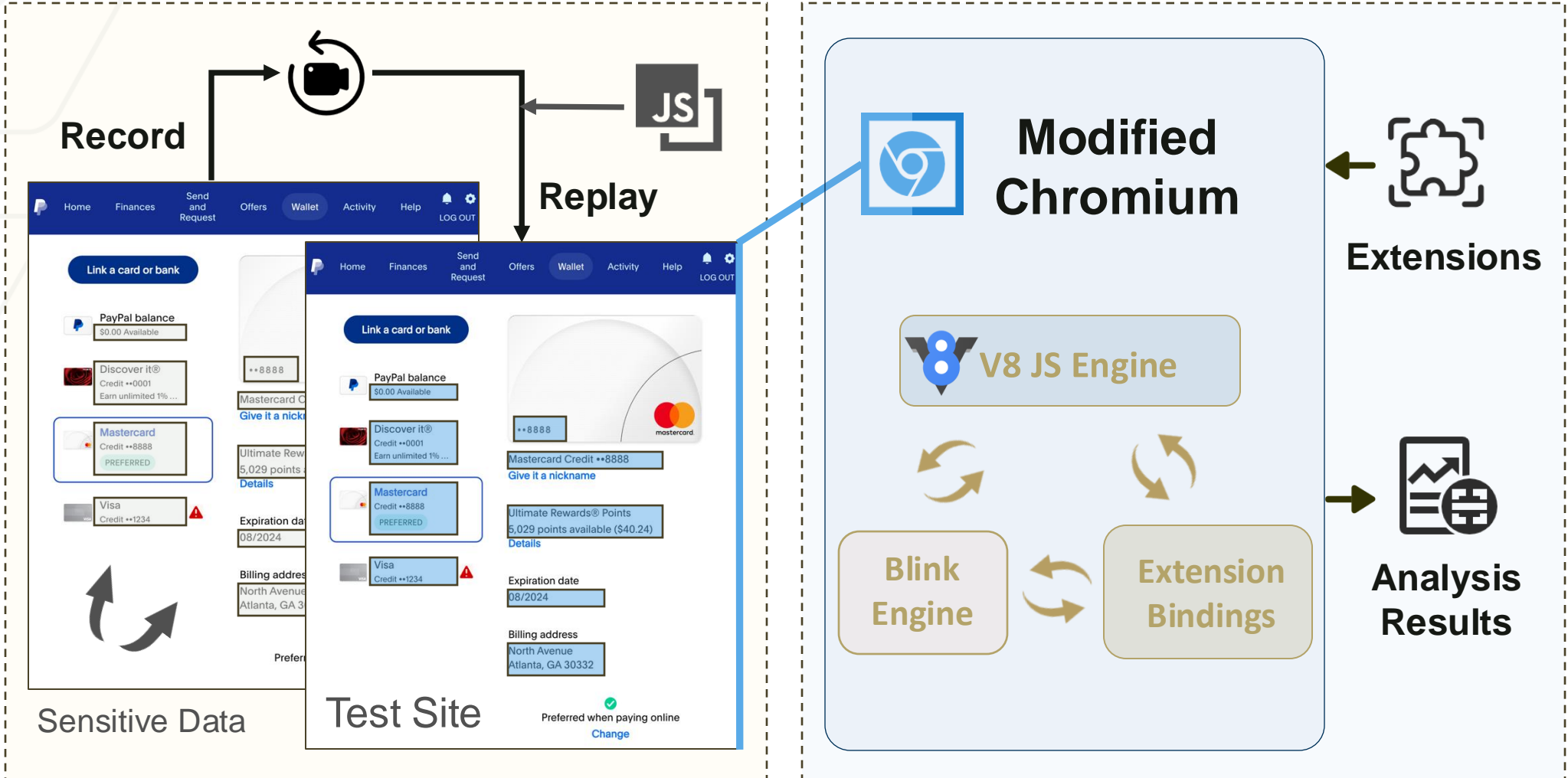
**JS** Inline Script for Annotations

```
...  
let sensitive_elm1 =  
  document.getElementsByClassName("{ $name }")[0];  
sensitive_elm1.setAttribute("data-taint","1");  
...
```

Recursive internally

```
<div class="ppv_x_text--body"  
  data-taint="1">  
<div class="fiDetails-content">  
  North Avenue</div>  
<div class="fiDetails-content">  
  Atlanta, GA 30332</div></div>
```

# Overview of Arcanum



# Method: Taint Sources/Sinks

## Taint Sources

DOM	<ul style="list-style-type: none"><li>• <b>DOM custom elements</b> <small>NEW ↻</small></li><li>• DOM location</li><li>• DOM property</li><li>• DOM Input Element</li></ul>
Extension API	<ul style="list-style-type: none"><li>• Chrome.history API</li><li>• Chrome.tabs API</li><li>• <b>Chrome.cookies API</b></li><li>• Chrome.webNavigation API</li><li>• Chrome.webRequest API</li></ul>
Web API	<ul style="list-style-type: none"><li>• <b>History Web API</b> <small>NEW ↻</small></li><li>• <b>Geolocation Web API</b> <small>NEW ↻</small></li><li>• <b>User-Agent Client Hints</b> <small>NEW ↻</small></li></ul>

Taint Propagation










## Taint Sinks

Web Request	<ul style="list-style-type: none"><li>• <b>Fetch</b> <small>NEW ↻</small></li><li>• XMLHttpRequest</li><li>• WebSocket</li><li>• <b>Beacon</b> <small>NEW ↻</small></li></ul>
DOM	<ul style="list-style-type: none"><li>• DOM elements injection</li></ul>
Storage	<ul style="list-style-type: none"><li>• Chrome.storage API</li><li>• Web Storage API</li><li>• <b>IndexedDB</b> <small>NEW ↻</small></li></ul>

# Large-Scale Experiments

- **All extensions** (both MV2 and MV3) on Chrome Web Store
- **7 Target Sites: Amazon** (address), **Facebook** (profile), **Gmail** (inbox), **Instagram** (profile), **LinkedIn** (profile), **Outlook** (inbox), **Paypal** (credit card)

	Target page	URL	Title	Tainted information on the page
	Amazon-Address	-	-	Name, Physical address (including address and phone number)
	Facebook-Profile	User ID	-	Name, Profile, Friend, Post (including Post content, Location, Comments)
	Gmail-Inbox	-	Email address	Name, Email address, Last account activity timestamp, Email content (including Email content, Title, Sender, Timestamp)
	Instagram-Profile	User ID	User ID, User Name	Name, Profile, Image sources and captions (in alt attributes)
	LinkedIn-Profile	User ID	User Name	Name, Profile, Friend (“People you may know”), Message
	Outlook-Inbox	-	User Name	Name, Email address, Email content (including Email content, Title, Sender, Timestamp)
	PayPal-Card	Payment ID	-	PayPal balance, Last 4 digits of the card number, Card issuance institution, Card expiration date, Physical (Billing) address

# Facebook Profile/Post Page Annotations

**Name**  
Amy Lee  
2 friends

**#Friends**

**Profile** my bio  
Edit bio

- Works at MyComputerCareer
- Studied at Scream Queens
- Went to High School Reunion
- Lives in Atlanta, Georgia
- From Koski TI
- In a relationship with Karen Kim (Pending)
- Pronounces name A-mee LEE

anu  
google.com

Intro

What's on your mind?

Live video Photo/video Life event

**Post** Filters Manage posts  
List view Grid view

Amy Lee is 😊 feeling blessed.  
September 16 · 🌐

What's on my mind? IDK

Amy Lee 1 comment

Like Comment Share

Amy Lee  
ahhhhh awwwww  
7w Like Reply

Write a comment...

Amy Lee is 😊 feeling happy in Atlanta, GA.  
December 3, 2022 · 🌐

Add featured

**Location**  
Amy Lee is 😊 feeling happy in Atlanta, GA.  
December 3, 2022 · 🌐

Photos See all photos

**Post Content**  
This is a rainy day!

**Friends** See all friends  
2 friends  
谢秦歌 Karen Kim

**Life Event** See all

- Other Life Event December 3, 2022
- In a Relationship with Karen Kim January 1, 2020

Amy Lee 2 comments 1 share

Love Comment Share

Amy Lee  
wowo 🍷🍷🍷🍷🍷  
48w Like Reply

Amy Lee  
Thank you! 😊  
48w Like Reply

Write a comment...

Amy Lee

# Results: 1. Overview

- Extension pose a significant privacy risk for users.



	Total	Amazon	Facebook	Gmail	Instagram	LinkedIn	Outlook	Paypal
#Flagged Extensions	3,028 (2.68%)	2,048 (1.81%)	1,730 (1.53%)	2,198 (1.94%)	2,067 (1.83%)	2,088 (1.85%)	1,964 (1.74%)	1,943 (1.70%)
#Total Users	144.0M	89.6M	66.3M	86.1M	91.6M	95.7M	85.7M	83.4M

- Sum of each extension's users
- An upper bound on distinct users

Flagged extensions are more popular!

# Results: 2. Automated Web Page Content Collection

- 202 extensions exfiltrated sensitive page content types, impacting 300k+ users.
- User's names (130) and profile information (124) are the most common content collected.
- Many extensions collecting other types of user data from page content.

Content Type	Extensions	Max Extension # Users
Name	130	80k+
Profile	124	300k+
Email Address	73	10k+
Location	63	30k+
Friend	56	30k+
Credit Card	49	10k+
Post	49	3k+
Email Content	46	10k+
Physical Address	46	10k+
Comments	39	3k+
Whole HTML	30	1k+
Image alt Attribute	1	205
Total	202	300k+

# Results: 3. Text Encrypting/Encoding

- **Tracking of encryption/encoding is needed.**

**159 extensions** transmitting tainted data after using some form of encoding, encryption, or obfuscation.

- `TextEncoder.encode[Into]()` (85 extensions)
- `base64` encoding (78 extensions)
- `SubtleCrypto.encrypt()` (31 extensions)





# Results: 4. Privacy Impact Case Studies

Whether the **automated data collection** we observed is specified in two places.

## 1) Extension's **privacy policy**

- **No Policy**
- **Not in Policy**
- **In Policy**

## 2) Extension's **Chrome Web Store description**

- **Clear**
- **Vague**
- **Violative**

Here's how it works:

- Upload your resume (only needs to be done once)
- Go to any job posting on the internet
- **Click the button**
- Get a custom cover letter that you can download as pdf or copy to clipboard
- Add it to your application, submit it having saved significant time, and start getting interviews!

An extension for creating cover letters

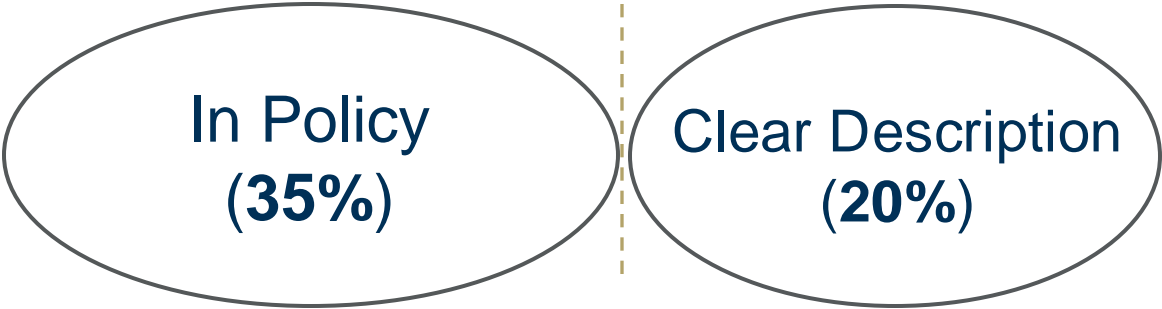
### **Violative Example**

The extensions automatically collect all texts on the 7 target pages

# Results: 4. Privacy Impact Case Studies

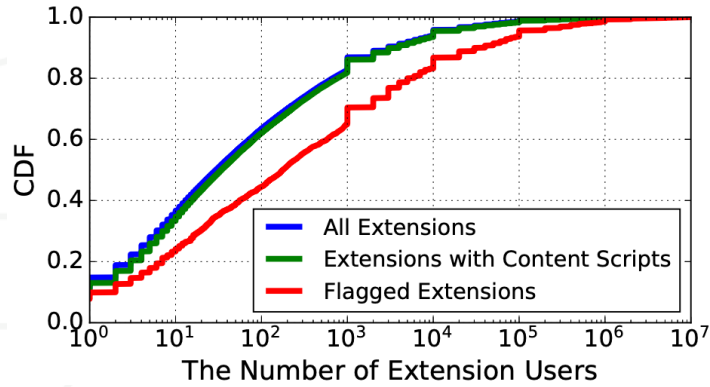
Random Sample Group	Privacy Policy			Web Store Description		
	#In Policy	#Not in Policy	#No Policy	#Clear	#Vague	#Violative
Web Content (20)	8	7	5	3	10	7
All Flagged extensions (20)	6	11	3	5	5	10
<b>Total (40)</b>	<b>14 (35%)</b>	18 (45%)	8 (20%)	<b>8 (20%)</b>	15 (37.5%)	17 (42.5)

No sampled extension provides both



*Users reasonably would not expect the automated data exfiltration*

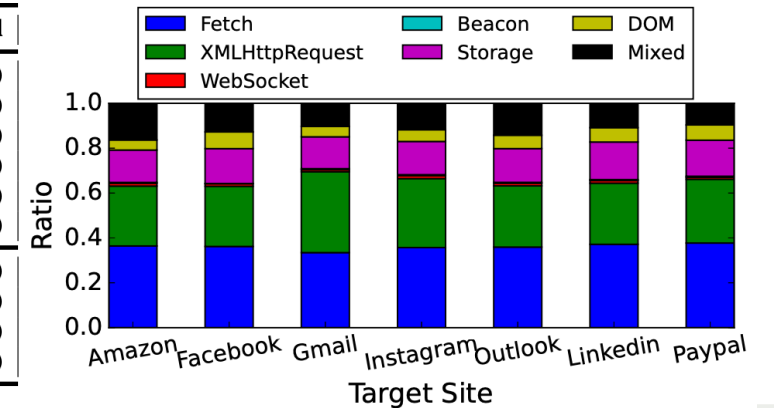
# More Results



Rank	Extension Name	#Users	Taint Sink(s)	Details	Encoded?
1	Honey: Automatic Coupons & Rewards [47]	10M+	fetch, storage	URL, Timestamp	No
2	Online Security [52]	10M+	fetch	URL	No
3	Avast SafePrice [34]	10M+	XMLHttpRequest	URL	Yes
4	Capital One Shopping [38]	8M+	fetch, XMLHttpRequest, storage, DOM	URL, Title, Device	Partial
5	Touch VPN - Secure and Unlimited VPN Proxy [57]	8M+	storage	URL, Country	Partial
6	Avira Browser Safety [35]	6M+	XMLHttpRequest	URL	No
7	Hola VPN - The Website Unblocker [46]	6M+	XMLHttpRequest, storage	URL	No
8	Avira Safe Shopping [36]	5M+	XMLHttpRequest	URL	No
9	NordVPN - VPN Proxy for Privacy and Security [51]	3M+	fetch, storage	Domain, Timestamp	No
10	QuillBot: AI Grammar and Writing Tool [54]	3M+	fetch, storage	Device	Yes

More results in the paper

	Category	Total	Amazon	Facebook	Gmail	Instagram	LinkedIn	Outlook	PayPal
Sources	Domain	463 (15.3%)	543 (26.5%)	308 (17.8%)	575 (26.2%)	395 (19.1%)	304 (14.6%)	435 (22.1%)	435 (22.4%)
	URL	1,551 (51.2%)	947 (46.2%)	902 (52.1%)	1,014 (46.1%)	1,112 (53.8%)	1,175 (56.2%)	971 (49.4%)	984 (50.6%)
	Identification	375 (12.4%)	248 (12.1%)	177 (10.2%)	223 (10.1%)	217 (10.5%)	215 (10.3%)	235 (12.0%)	206 (10.6%)
	Title	251 ( 8.3%)	149 (7.3%)	184 (10.6%)	193 (8.8%)	161 (7.8%)	184 (8.8%)	186 (9.5%)	164 (8.4%)
	Page Content	202 ( 6.7%)	109 (5.3%)	124 (7.2%)	127 (5.8%)	133 (6.4%)	154 (7.4%)	105 (5.3%)	122 (6.3%)
	Uncategorized	186 ( 6.1%)	52 (2.5%)	35 (2.0%)	66 (3.0%)	49 (2.4%)	56 (2.7%)	34 (1.7%)	32 (1.6%)
Sinks	Web Requests	2064 (68.1%)	1,405 (68.6%)	1,198 (69.2%)	1,613 (73.4%)	1,478 (71.5%)	1,448 (69.3%)	1,361 (69.3%)	1,353 (69.7%)
	Storage	362 (12.0%)	318 (15.6%)	249 (14.4%)	312 (14.2%)	306 (14.8%)	349 (16.7%)	296 (15.0%)	312 (16.0%)
	DOM	133 ( 4.4%)	132 ( 6.4%)	60 ( 3.5%)	80 ( 3.6%)	87 ( 4.2%)	113 ( 5.5%)	97 ( 5.0%)	113 ( 5.8%)
	Mixed	469 (15.5%)	193 ( 9.4%)	223 (12.9%)	193 ( 8.8%)	196 ( 9.5%)	178 ( 8.5%)	211 (10.7%)	165 ( 8.5%)



# Use Arcanum in Practice



<https://github.com/BEESLab/Arcanum/>



We released:

- **Chromium patches** (20k+ LOC) of the Arcanum implementation
- **Test cases** for using Arcanum
  - Custom extensions
  - Real-world extensions
- Our **experimental taint logs**

# Conclusion

Privacy risks discovered by Arcanum point to the need for significant changes in extensions, policies, and systems.

- **Web Content Matters**
- **Researcher-Driven Annotations Helps**
- **Extension Permissions are Coarse and Opaque**
- **Taint Tracking for Extension Vetting**
- **Future Work: Build on Arcanum**

*Thank you!*  
**Q&A**

*Qinge Xie*  
*qxie47@gatech.edu*