

# These Browser Extensions Spy on 8 Million Users

## Extended Abstract

Michael Weissbacher  
Northeastern University, Boston, USA  
<http://mweissbacher.com>

**Abstract**—This work investigates the `upalytics.com` library for Chrome and Firefox extensions, which performs real time tracking of users on all sites they visit. The code is bundled with free extensions in the official extension stores, exfiltrating browsing history as a feature. Within the top 7,000 Chrome extensions, the library is used 42 times with over 8 million installations, the most widely used one has 1.48M installations alone. For Mozilla Firefox we found 400,000 users to be affected. We also look into the relationship of `upalytics` with `similarweb.com`, a third-party web analytics company, which is using that library for their own extension.

We reported the suspicious Chrome extensions in March 2016 and they were deleted from the Google Chrome Web Store within 24 hours. Mozilla deleted three out of five reported extensions. In August 2016 we reviewed the Chrome Web Store and found no evidence of this library in the top 7,000 extensions. While this work focuses on one privacy perpetrator, tracking in browser extensions presents a wider research problem.

**Index Terms**—Web security, Browser extensions, Privacy

### I. INTRODUCTION

Web trackers and analytics are powerful tools that offer owners of websites insights on audience and behavioral information. To track visitors, owners add JavaScript code that triggers a request to the tracking server. The server stores the data for later analysis, the data is accessible to the owner of the website and the tracker itself.

In March 2016 we noticed a website that offered information similar to analytics sites, but had no apparent relation to the sites that were analyzed – `www.similarweb.com`. The data includes links clicked on a site, referrer statistics, the origin of users, and others. While this is interesting, it also raises questions about the sources of data. Based on their website they collect data from millions of devices, but the software they advertise has a number of installations that was orders of magnitude lower than that. Through analysis of their official Chrome extension we noticed a tracking library performing real-time monitoring of all visited websites. In a subsequent analysis of the top Chrome and Firefox extensions we found the library present in both official stores. We found the same library in 42 extensions in the Chrome Web Store and five extensions in the Mozilla store, in total over 8 million installations.

Google removed all extensions within 24 hours and Mozilla removed three out of five.

Compared to tracking on websites, tracking through Chrome extensions is fundamentally different. Extensions have privileged access to the browser and can track all websites indiscriminately. They report data to parties unaffiliated with the monitored website. This type of monitoring allows for a complete view of user browsing behavior. User privacy is violated as these libraries are not always mentioned in terms of service, furthermore, invasive tracking is not expected behavior of extensions.

Monitoring data is often transferred in an extension background script that is not visible to the websites. As no tools exist that would block such trackers or alert users, this presents a novel research problem. The contributions of this work are as follows:

- We found malicious extensions in the official Chrome and Firefox stores that tracked detailed user behavior on unaffiliated websites.
- Our work resulted in 45 extensions being removed from Chrome Web Store and Mozilla Firefox store, with over 8 million installations.
- We highlight privacy invasion through tracking libraries bundled with free browser extensions.

### II. MOTIVATION

SimilarWeb offers insights into third-party web analytics. To the end user the functionality is similar to Google Analytics, except that visitors can see traffic details of websites neither they or SimilarWeb are affiliated with. This is useful for analysis of competitors, or explore new markets for a product. The company was founded 2007 and currently has 300 employees.

Using the free version of their service, the presented information includes information on visitors, search, and advertising. The data is detailed, including number of visitors, average visit duration, search keywords used, countries of origin, referring sites, destination sites that the visitors leave through, and others.

#### A. Origins of Data

As the company does not have direct access to these data sources, the displayed data must be extrapolated from data which is accessible to them. This high resolution

of data without direct access made us curious to further investigate. Their website suggest use of four types of data sources including millions of devices and ISPs, quoted from their website: “A panel of monitored devices, currently the largest in the industry”

### B. SimilarWeb Chrome Extension

As first step we analyzed the extension offered on their website. The offered main functionality consists of suggesting sites similar to the one currently seen. After reviewing their code and analyzing network traffic, we noticed suspicious behavior. The extension intercepts requests for all websites and reports any URL or search queries to SimilarWeb in real time, including metadata such as referrers. We noticed that the JavaScript library used for tracking was developed by another company, Upalytics<sup>1</sup>. The purpose of this library is to track user behavior in Chrome extensions, other platforms are advertised on their website as well, including mobile and desktop. Since this was an external library, we suspected it might be used in other extensions as well for similar purposes.

## III. RESULTS

### A. Finding More Extensions

After crawling the Chrome Web Store we found 42 suspicious extensions by searching for code similarities. To verify malicious behavior we manually analyzed each extension under four aspects: Whether the extension has the capability to exfiltrate private data, whether tracking happens by default, or the user has to opt-in. We also analyzed the terms of service: whether tracking is mentioned directly, and if not, whether it is available through a link in the terms of service.

All suspicious extensions were able to collect history, all but one were tracking as default behavior. Of these 42 extensions 19 explain their data collection practices in the terms of service, while 23 do not. Furthermore, out of these 23 extensions 12 have no URL where this would be explained. One URL that is used across 13 extensions to explain the privacy ramifications is `http://addons-privacy.com`. While the URL is shared between extensions, the developers have no obvious connection. Six of the remaining domains point to the same IP address.

Contents of tracking beacons includes the target URL, referring site, and how the site was navigated to. The IP is automatically included as metadata of the generated request. Tracking beacons can be linked between reboots and location changes through generated session and persistent user IDs.

### B. Network Information

The extensions use nine different hardcoded hostnames to report tracking information, we found relations linking all 42 extensions. All endpoint domains, `addons-privacy`

`.com` and `upalytics.com` were registered through Domains by Proxy<sup>2</sup>, a service used to obfuscate ownership of domain names by hiding WHOIS records. All extensions were reporting to subdomains `http://lb.*`. Some of the names of the domains appear to be misleading, suggesting updates or being a searchhelper. Two of the domains (`connectupdate.com`, `secureweb24.net`) were registered 13 seconds apart. Also, the `robots.txt` file used in all cases is the same.

Furthermore, all these IPs belong to the same hoster, XLHost. 8 out of 9 of these hosts have all addresses in a /18 network, half of the IPs of the `upalytics.com` endpoint are in another XLHost network. All IPs in use are unique, however, this involves consecutive IP addresses and other neighborhood relationships.

All hosts used round robin DNS, using multiple IPs for each domain name. To examine this closer we compared the distance of IP addresses used by these extensions for tracking. In Figure 1b, the nodes are the 9 domain names in use, edges are the grade of distance. By taking into account distances of up to 4, we can link together all hostnames used in all 42 Chrome extensions. For example: IPs `1.1.1.1` and `1.1.1.3` have a distance of 2. As for the labels, the edge between `similarsites.com` and `tetrafficstat.net` reads `6x2`. This means that the domains share 6 IP addresses with a distance of 2. Figure 1a visualizes the distance relationship between `lb.crdui.com` and `lb.datarating.com`.

### C. Reported Extensions–Google Chrome

We reported our findings on March 31<sup>st</sup> 2016, and all extensions were removed from the Chrome store within 24 hours, including the official SimilarWeb and SimilarSites extensions - a partner site. By September 2016, 18 out of 42 deleted extensions have returned without the offending library, 22 remain deleted.

### D. Reported Extensions–Mozilla Firefox

We reported five extensions with over 400,000 total installations to Mozilla which were tracking user behavior outside of extensions. Out of these three were removed from the store because they did not disclose tracking in their privacy statement. However, this type of tracking is generally tolerated for Firefox.

## IV. RELEVANCY TO FTC PRIVACYCON

Trackers are popular on websites and well studied, however, they are fundamentally different from tracking in browser extensions. Websites need to opt-in to use a tracker, and their scope is limited to their own website, unless purposefully shared. Furthermore, visitors can use tracker-blockers to opt-out of tracking with extensions such as Ghostery. Conversely, in browser extensions the scope of tracking is not limited to a single website, but collects information on all websites. This level of tracking

<sup>1</sup><http://www.upalytics.com>

<sup>2</sup><https://www.domainsbyproxy.com>

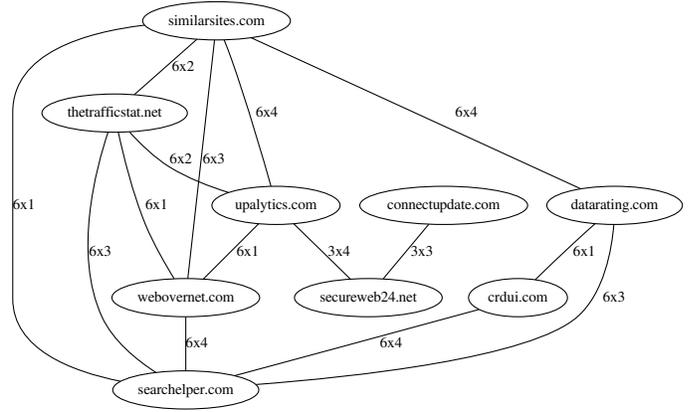
```

## lb.crdui.com
173.45.110.134 <-
173.45.110.153 <-
173.45.110.172 <-
173.45.127.134
173.45.127.153
173.45.127.172

## lb.datarating.com
173.45.110.135 <-
173.45.110.154 <-
173.45.110.173 <-
173.45.127.135
173.45.127.154
173.45.127.173
etc.

```

(a) Neighboring relationships of IPs between seemingly unrelated domains used for monitoring.



(b) Graph linking domain names by IP relationships used in 42 extensions to covertly collect browsing history.

Fig. 1: Domains using `upalytics.com` library reported to a network of domains that can be linked by IP neighborhood.

represents a stronger privacy invasion than third-party tracking on websites. Furthermore, no tools exist to reduce the impact of privacy invasion on the user.

Six of the privacy policies used by extensions reference California Civil Code Section 1798.83<sup>3</sup>. This law allows for inquiry about usage of personal information for direct marketing purposes. We reached out to two of the email addresses, but received no response of time of submission.

## V. RELATED WORK

As any web application, browser extensions are third-party code. However, these programs operate with elevated privilege and have access to powerful APIs that can allow access to all content within the browser. Permission systems allow developers to restrict their programs, but extensions have been shown to over-request permissions and effectively de-sensitizing users. Heule et. al. [1] showed that 71% of the top 500 Chrome extensions use permissions that support leaking private information.

Hulk [2] is a system that was used for the first large scale dynamic analysis of Chrome extensions. The authors introduced the concept of Honeypages. This technique generates web content tailored to an extension to trigger malicious behavior driven by expectations of the extension.

To monetize extensions maliciously inclined authors may add or replace ads in the browser with their own. In 2015 a study has found 249 Chrome extensions in the Chrome web store injecting unwanted ads [3].

Third-party tracking on websites has been studied extensively. Browsing on seemingly unrelated sites can be observed by third-party trackers and combined into a comprehensive browsing history. Mayer et. al. introduced the FourthParty measurement platform [4], discussing privacy implications, technology, and policy perspectives of

third-party tracking. Roesner et. al. [5] developed client-side defenses to classify and prevent third-party tracking. Recent work has analyzed the history of web tracking via the Internet Archive’s Wayback Machine [6]. The authors found that tracking has steadily increased since 1996. Tracking on the web has never been as pervasive as in 2016.

## VI. CONCLUSION

With this work we highlight the scope of privacy invasion through browser extensions available in the official extension stores. We investigate one popular tracking library that affects 8 million users of Chrome and Firefox extensions. Tracking in browser extensions is more widespread than this one library and presents a novel research problem. We suggest that extensions should be both tested more rigorously when admitted to the store, as well as monitored for tracking when updated.

## REFERENCES

- [1] S. Heule, D. Rifkin, A. Russo, and D. Stefan, “The most dangerous code in the browser,” in *USENIX Hot Topics in Operating Systems (HotOS)*, Kartause Ittingen, Switzerland, 2015.
- [2] A. Kapravelos, C. Grier, N. Chachra, C. Kruegel, G. Vigna, and V. Paxson, “Hulk: Eliciting malicious behavior in browser extensions,” in *USENIX Security Symposium*, San Diego, CA, 2014.
- [3] K. Thomas, E. Bursztein, C. Grier, G. Ho, N. Jagpal, A. Kapravelos, D. McCoy, A. Nappa, V. Paxson, P. Pearce, N. Provos, and M. A. Rajab, “Ad injection at scale: Assessing deceptive advertisement modifications,” in *IEEE Symposium on Security and Privacy (Oakland)*, 2015.
- [4] J. R. Mayer and J. C. Mitchell, “Third-party web tracking: Policy and technology,” in *IEEE Symposium on Security and Privacy (Oakland)*, 2012.
- [5] F. Roesner, T. Kohno, and D. Wetherall, “Detecting and defending against third-party tracking on the web,” in *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, San Jose, CA, 2012.
- [6] A. Lerner, A. K. Simpson, T. Kohno, and F. Roesner, “Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016,” in *USENIX Security Symposium*, Austin, TX, 2016.

<sup>3</sup><https://epic.org/privacy/profiling/sb27.html>