Web Tracking and the Arms Race between Digital Advertiser and Adblocker

COMP 531 Paper

Yusong Shen
ys43@rice.edu

Introduction

Internet advertising is becoming more and more important. In 2011, Internet advertising revenues in the United States surpassed those of cable television and nearly exceeded those of broadcast television. And in 2013 Internet advertising revenues in the United States totaled $42.8 billion[1]. Digital Advertiser also focus on displaying ads to target customer by knowing user's preference from other websites, this is called behavioral advertising.[2] Behavioral advertising increase the advertiser's demand for tracking user behaviour, and we will talk about three tracking mechanisms in this paper : Canvas Fingerprinting, Evercookie, Cookie Syncing. We will also talk about the arms race between Adblocker and Digital Advertiser.

How does advertiser track the user?

Canvas Fingerprinting

Introduction

Canvas fingerprinting is a type of browser or device fingerprinting technique that was first presented by Mowery and Shacham in 2012[3]. The authors found that by using the Canvas API of modern browsers, one can exploit the subtle differences in the rendering of the same text to extract a consistent fingerprint that can easily be obtained in a fraction of a second without user's awareness.[4]

---

[1] https://en.wikipedia.org/wiki/Online_advertising
[2] https://en.wikipedia.org/wiki/Behavioral_targeting
[3] Mowery, Keaton, and Hovav Shacham. "Pixel perfect: Fingerprinting canvas in HTML5." *Proceedings of W2SP* (2012).
[4] https://securehomes.esat.kuleuven.be/~gacar/persistent/#canvas-results

How does it work?

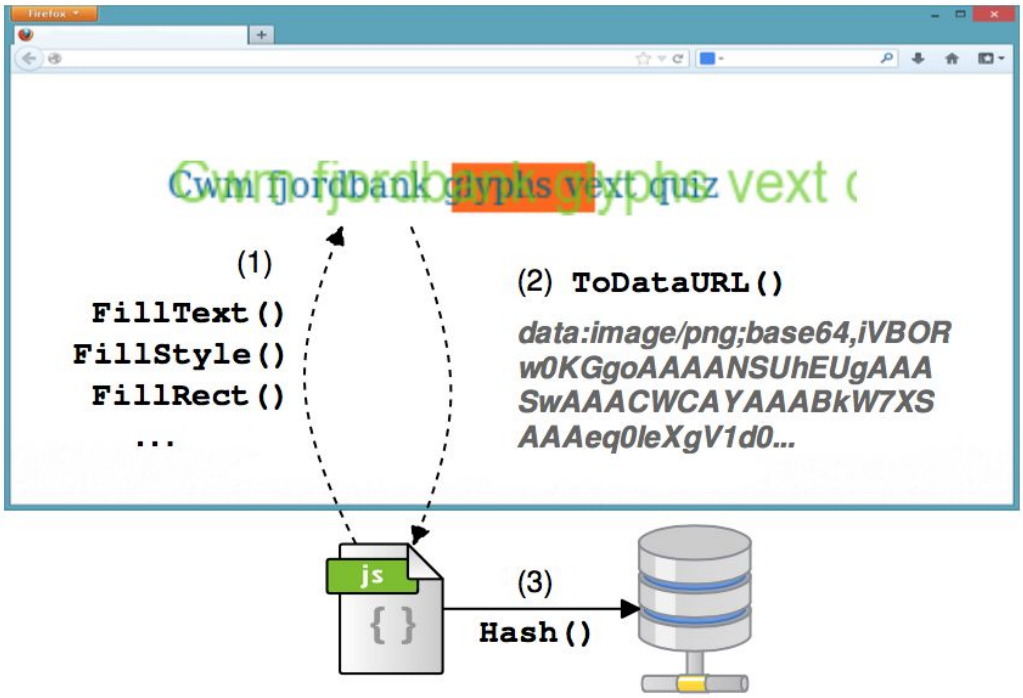

Figure 1: Canvas fingerprinting basic flow of operations[5]

Here is a live demo from BrowserLeaks.com showing how Canvas Fingerprinting works :

[5] Acar, Gunes, et al. "The web never forgets: Persistent tracking mechanisms in the wild." Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014.
APA

Canvas Support in Your Browser :

Canvas (basic support)	✔ True
Text API for Canvas	✔ True
Canvas toDataURL	✔ True

Database Summary :

Unique User-Agents	145237
Unique Fingerprints	5799

Your Fingerprint :

Signature	✔ 7831CE46
Found in DB	✔ True (193 of 145237 unique User-Agents has the same signature as yours)
Image File Details :	
File Size	7079 bytes
Number of Colors	744
SHA256	40DF8E372446F3DAC2A54F9DB3CFB1E8226B9B49E5193149A244A46F0DA7F5AF

PNG Headers

Chunk :	Length :	CRC :	Content :
IHDR	13	477A703E	PNG image header: 220x30, 8 bits/sample, truecolor+alpha, noninterlaced
IDAT	7022	7831CE46	PNG image data
IEND	0	AE426082	end-of-image marker

Browser Detection :

✔ It is very likely that your web-browser is **Chrome** and your operating system is **Mac OS X**.

Operating Systems :		Browsers :		Devices :	
Mac OS X	192/193	Chrome	138/193	Other	193/193
Linux	1/193	Opera (Chromium)	53/193	Platforms :	
OS by Version :		Vivaldi	2/193	MacIntel	193/193
Mac OS X 10.10	125/193	Browsers by Version :			

Unlike the other browser detection tricks, this deals with many OS features related on graphics environment. Potentially it can be used to identify the video adapter, especially if you will use WebGL profiling, not just Canvas 2D Context. By the way different graphics card drivers can also sometimes affect to regular fonts rendering.
This tiny animated GIF shows how canvas image can be variable from 35 different users. The code is not changed, but each frame is different:

Here is the JavaScript code that produce the pixels:

```
// Text with lowercase/uppercase/punctuation symbols
var txt = "BrowserLeaks,com <canvas> 1.0";
ctx.textBaseline = "top";
// The most common type
ctx.font = "14px 'Arial'";
ctx.textBaseline = "alphabetic";
```

```
ctx.fillStyle = "#f60";
ctx.fillRect(125,1,62,20);
// Some tricks for color mixing to increase the difference in rendering
ctx.fillStyle = "#069";
ctx.fillText(txt, 2, 15);
ctx.fillStyle = "rgba(102, 204, 0, 0.7)";
ctx.fillText(txt, 4, 17);
```

To create a signature from the canvas, we must export the pixels from the application's memory using the toDataURL() function, which will return the base64-encoded string of the binary image file. Then we can just create MD5 hash of this string, or even extract CRC checksum from IDAT chunk which is placed from 16 to 12 byte from the end of every PNG file, and this will be our Canvas Fingerprint.[6]

Library

There is a popular open source project called fingerprintjs2 in Github.

Evercookies

Introduction

Evercookie is a technique to make cookie persistent and hard to be cleared. it utilizes multiple options to store cookie, like HTTP cookie, FLASH cookie, IndexDB etc. and respawn cookie even after some of the storages get removed.

[6] https://www.browserleaks.com/canvas#how-does-it-work

How does it works?

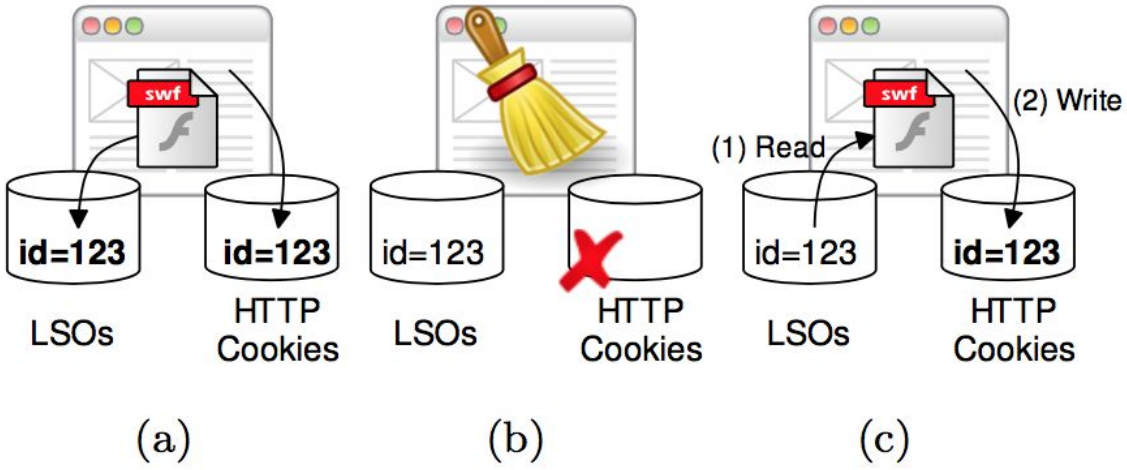

Figure 3 : Respawning HTTP cookies by Flash evercookies[7]

The figure 3 shows how HTTP cookies get respawned by Flash cookies(Local Shared Objects). In (a), the website stores both an HTTP and a Flash cookie. Then in (b), the user clear the HTTP cookie, but even after that, the webpage can respawn the HTTP cookie by copying the value from the Flash cookie as (c) shows.

Cookie Syncing

Introduction

Cookie Syncing or Cookie Synchronization is a practice that domain trackers share information of given user with each other. The identifier associated with Pseudonymous IDs of given user is often stored in Cookie.

---

[7] Acar, Gunes, et al. "The web never forgets: Persistent tracking mechanisms in the wild." Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014.
APA

How does it work?

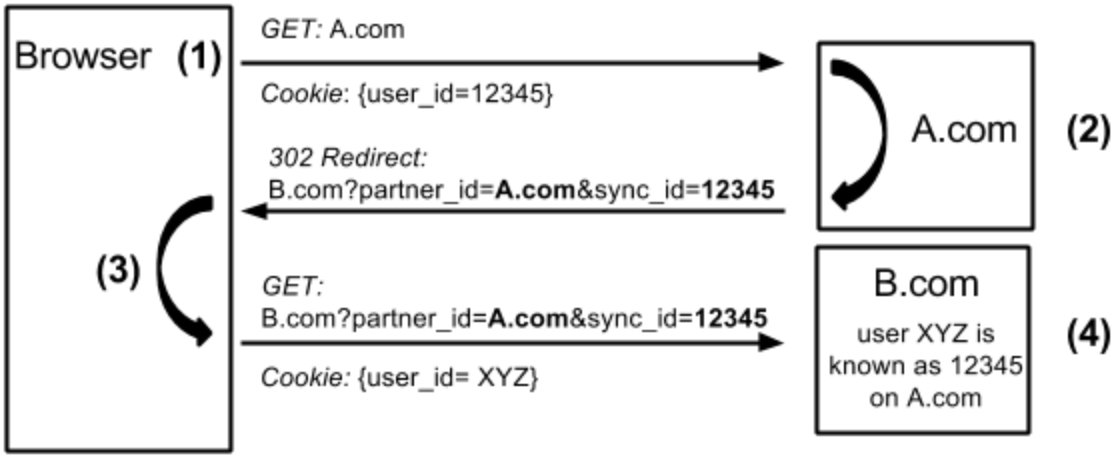

Figure 2: the basic workflow of Cookie Syncing

The process begins when a user visits a site (say example.com, not shown in the figure), which includes A.com as an embedded third-party tracker.
(1) The browser makes a request to A.com, and included in this request is the tracking cookie set by A.com.
(2) A.com retrieves its tracking ID from the cookie, and redirects the browser to B.com, encoding the tracking ID into the URL.
(3) The browser then makes a request to B.com, which includes the full URL A.com redirected to B.com's as well as tracking cookie.
(4) B.com can then link its ID for the user to A.com's ID for the user. All of this is invisible to the user.
(5) Once two trackers sync cookies, they can exchange user data between their servers.[8]

How to prevent tracking?[9]

You could use GHOSTERY plugin, it's a browser plugin that show you which trackers are used in your visiting page.
You could also use Tor browser to enable a more private browsing experience. Tor project is developed to prevent somebody watching your internet connection and hide your actual physical address. Tor browser allows you to use Tor to browsing without extra installation.

[8] https://freedom-to-tinker.com/2014/08/07/the-hidden-perils-of-cookie-syncing/
[9] http://www.digitaltrends.com/computing/how-do-advertisers-track-you-online-we-found-out/

Summary

Canvas Fingerprinting and Evercookie can be used to identified user across the website. Evercookie can be used to bypass the user's privacy setting in browser. Combine Evercookie with Cookie Syncing together can help different websites share user's information with each other, hence they can display ads according to the user's preference in previous website.

Arms Race between Adblock and Anti-Adblock

AdBlock

Introduction

The most popular consumer ad blockers are browser extensions, include Adblock, Adblock Plus etc. They are used to block banners, pop-ups, and other intrusive ads.

How does it works?

The most popular Ad-blockers (AdBlock, ABP) rely on two principal methods:
Communication blocking, in which communication to ad-servers/ad resources is blocked altogether (the client request does not occur). Example: block all requests to URLs that include "google.adsense".
Element hiding, in which certain HTML elements, even if loaded correctly, are still hidden from the page. Example: hide any element with class="Ad".[10]

Block AdBlock

There are quite a few debates[11] about the use of Adblock, as the site owner's viewpoint, ads are one of the most important source of income for the site owners to provide free content. So here comes some anti-adblock methods.

[10] https://www.quora.com/How-do-adblockers-work-technically/answer/Ido-Yablonka?srid=nOve
[11] https://en.wikipedia.org/wiki/Adblock_Plus

How does it works?

AdBlock Detector

Some people[12][13] have used Javascript to detect the effects of the popular Adblock filters. The mechanism is to generate a honeypot-like URL, and verify its delivery. It also verify that the expected advertising DOM elements are present after web page is rendered.

Websocket

Some website also use websocket to get rid of Adblock, this can be done since Chrome previously doesn't allow extensions to block WebSockets. But other adblock extensions like uBlock Origin also figures out a workaround. They use a content script to inject a wrapper for WebSocket into pages. The wrapper performs a dummy web request before WebSocket messages are sent/received. The extension recognises these dummy web requests as representing a WebSocket message. It intercepts and blocks them if the corresponding WebSocket message should be blocked. The WebSocket wrapper then allows / blocks the WebSocket message based on whether the dummy web request was blocked or not.[14]

Summary

Arm races between digital advertiser and adblock are still going on. But there also comes to some agreement. Since the customer complain most about two things of web ads : they are disruptive, and slow down the browsing experience. More and more digital advertisers like Facebook try to provide the lightweight HTML ads with non-interruptive browsing experience. Adblock Plus also provide an acceptable ads whitelist for those non-intrusive ads. We hope to see a better and healthier digital advertising environment in the future.

Conclusion

In this paper, we exam three tracking mechanisms in the wild : Canvas Fingerprinting, Evercookie, Cookie Syncing. This raise some privacy and security concerns among the public. We also exam the arms race between Digital Advertiser and Adblocker, and discuss the self-regulation of ads industrial to display more lightweight and non-intrusive ads.

---

[12] https://web.archive.org/web/20120217010456/http://adblockdetector.com/
[13] https://www.browserleaks.com/proxy
[14] https://issues.adblockplus.org/ticket/1727

Other References

https://blockadblock.com/adblocking/fight-adblock-ask-facebook/
http://www.makeuseof.com/tag/3-tactics-dealing-adblock-users-site/