On Privacy Risks of Public WiFi Captive Portals

Suzan Ali, Tousif Osman, Mohammad Mannan, and Amr Youssef

Concordia University, Montreal, Canada
{a_suzan,t_osma,mmannan,youssef}@ciise.concordia.ca

Abstract. Open access WiFi hotspots are widely deployed in many public places, including restaurants, parks, coffee shops, shopping malls, trains, airports, hotels, and libraries. While these hotspots provide an attractive option to stay connected, they may also track user activities and share user/device information with third-parties, through the use of trackers in their captive portal and landing websites. In this paper, we present a comprehensive privacy analysis of 67 unique public WiFi hotspots located in Montreal, Canada, and shed light on the web tracking and data collection behaviors of these hotspots. Our study reveals the collection of a significant amount of privacy-sensitive personal data through the use of social login (e.g., Facebook and Google) and registration forms, and many instances of tracking activities, sometimes even before the user accepts the hotspot's privacy and terms of service policies. Most hotspots use persistent third-party tracking cookies within their captive portal site; these cookies can be used to follow the user's browsing behavior long after the user leaves the hotspots, e.g., up to 20 years. Additionally, several hotspots explicitly share (sometimes via HTTP) the collected personal and unique device information with many third-party tracking domains.

1 Introduction

Public WiFi hotspots are growing in popularity across the globe. Most users frequently connect to hotspots due to their free-of-cost service, (as opposed to mobile data connections) and ubiquity. According to a Symantec study [24] conducted among 15,532 users across 15 global markets, 46% of participants do not wait more than a few minutes before connecting to a WiFi network after arriving at an airport, restaurant, shopping mall, hotel or similar locations. Furthermore, 60% of the participants are unaware of any risks associated with using an untrusted network, and feel their personal information is safe.

A hotspot may have a captive portal, which is usually used to communicate the hotspot's privacy and terms-of-service (TOS) policies, and collect personal identification information such as name and email for future communications, and authentication if needed (e.g., by asking the user to login to their social media sites). Upon acceptance of the hotspot's policy, the user is connected to the internet and her web browser is often automatically directed to load a landing page (usually the hotspot brand's webpage).

Several past studies (e.g., [6, 23]) focus on privacy leakage from browsing the internet or using mobile apps in an open hotspot, due to the lack of encryption, e.g., no WPA/WPA2 support at the hotspot, and the use of HTTP, as opposed to HTTPS for connections between the user device and the web service. However, in recent years, HTTPS adoption across web servers has increased dramatically, mitigating privacy exposure through plain network traffic. For example, according to the Google Transparency Report [11], as of Apr. 6, 2019, 82% of web pages are served via HTTPS for Chrome users on Windows. On the other hand, in the recent years, there have also been several comprehensive studies on web tracking on regular web services and mobile apps with an emphasis on most popular domains/services (see e.g., [9, 4, 3]).

In contrast to past hotspot and web privacy measurement studies, we analyze tracking behaviors and privacy leakage in WiFi captive portals and landing pages. We design a data collection framework (`CPInspector`)[1] for both Windows and Android, and capture raw traffic traces from several public WiFi (in Montreal, Canada) that require users to go through a captive portal before allowing internet access. Challenges here include: manual collection of captive portal data by physically visiting each hotspot; making our test environment separate from the regular user environment so that we do not affect the user's browsing profiles; ensuring that our tests remain unaffected by the user's past browsing behaviors (e.g., saved tracking cookies); and creating and monitoring several test accounts in popular social media or email services as some hotspots mandate such authentication. CPInspector does not include any real user information in the collected dataset, or leak such information to the hotspots (e.g., by using fake MAC addresses).

From each hotspot, we collect traffic using both Chrome and Firefox on Windows. In addition to the default browsing mode, we also use private browsing, and deploy two ad-blockers to check if such privacy-friendly environments help against captive portal trackers—leading to a total of eight datasets for each hotspot. We also use social logins if required by the captive portal, or provided as an option; we again use both browsers for social login tests (two to six additional datasets as we have observed at most three social login options per hotspot). Some hotspots also require the user to complete a registration form that collects the user's PII—in such cases, we collect two more datasets (from both browsers).

On Android, we collect traffic only from the custom captive portal app (as opposed to Chrome/Firefox on Windows) as the cookie store of this app is separate from browsers. Consequently, tracking cookies from the Android captive portal app cannot be used by websites loaded in a browser. Recent Android OSes also use dynamic MAC addresses, limiting MAC address based tracking. However, we found that cookies in the captive portal app may remain valid for up to 20 years, allowing effective tracking by hotspot providers.

We also design our framework to detect ad/content injection by hotspots; however, we observed no content modification attempts by the hotspots. Furthermore, we manually evaluate various privacy aspects of some hotspots, as

[1] `https://github.com/MadibaLab/CPInspector`

documented in their privacy/terms-of-service policies, and then compare the stated policies against what happens in practice.
Note: *By default all our statistics refer to the measurements on Windows; we explicitly mention when results are for Android (mostly in Sec.5).*

Contributions and summary of findings.

1. We collected a total of 679 datasets from the captive portal and landing page of 80 hotspot locations between Sept. 2018 to Apr. 2019. 103 datasets were discarded due to some errors (e.g., network failure). We analyzed over 18.5GB of collected traffic for privacy exposure and tracking, and report the results from 67 unique hotspots (576 datasets), making this the largest such study to characterize hotspots in terms of their privacy risks.
2. Our hotspots include cafes and restaurants, shopping malls, retail businesses, banks, and transportation companies (bus, train and airport), some of which are local to Montreal, but many are national and international brands. 40 hotspots (59.7%) use third-party captive portals that appear to have many other business customers across Canada and elsewhere. Thus our results might be applicable to a larger geographical scope.
3. 27 hotspots (40.3%) use social login or a registration page to collect personal information (19 hotspots make this process mandatory for internet access). Social login providers may share several privacy-sensitive PII items—e.g., we found that LinkedIn shares the user's full name, email address, profile picture, full employment history, and the current location.
4. Except three, all hotspots employ varying levels of user tracking technologies on their captive portals and landing pages. On average, we found 7.4 third-party tracking domains per captive portal (max: 34 domains).

   40 hotspots (59.7%) create persistent third-party tracking HTTP cookies (validity up to 20 years); 4.2 cookies on average on each captive portal (max: 34 cookies). Surprisingly, 26 hotspots (38.8%) create persistent cookies even *before* getting user consent on their privacy/TOS document.
5. Several hotspots explicitly share (sometimes even without HTTPS) personal and unique device information with many third-party domains. 40 hotspots (59.7%) expose the user's device MAC address; five hotspots leak PII via HTTP, including the user's full name, email address, phone number, address, postal code, date of birth, and age (despite some of them claiming to use TLS for communicating such information). Two hotspots appear to perform cross-device tracking via Device Co-op [2].
6. Two hotspots (3.0%) state in their privacy policies that they explicitly link the user's MAC address to the collected PII, allowing long-term user tracking, especially for desktop OSes with fixed MAC.
7. From our Android experiments, we reveal that 9 out of 22 hotspots can effectively track Android devices even though Android uses a separate captive portal app and randomizes MAC address as visible to the hotspot.

2 Related Work

In this section, we briefly review related previous studies on hotspots, web tracking, and ad injection.

Several prior studies have demonstrated the possibility of eavesdropping WiFi traffic to identify personal sensitive information in public hotspots. Cheng et al. [6] collected WiFi traffic from 20 airports in four countries, and found that two thirds of the travelers leak private information while using airport hotspots for web browsing and smartphone app usage. Sombatruang et al. [23] conducted a similar study in Japan by setting up 11 experimental open public WiFi networks. The 150 hour experiment confirmed the exposure of private information, including photos, and users' credentials— transmitted via HTTP. In contrast, we analyze web tracking and privacy leakage within WiFi captive portals and landing pages.

Web tracking, a widespread phenomenon on the internet, is used for varying purposes, including: targeted advertisements, identity checking, website analytics, and personalization. Eckersley [7] showed that 83.6% of the `panopticlick.eff.org` visitors could be uniquely identified from a fingerprint composed of only 8 attributes. Laperdrix et al. [14] showed that `AmIUnique.org` can uniquely identify 89.4% of fingerprints composed of 17 attributes, including the HTML5 canvas element and the WebGL API. In a more recent large-scale study, Gómez-Boix et al. [10] collected over 2 million real-world device fingerprints (composed of 17 attributes) from a top French website; they found that only 33.6% device fingerprints are unique, raising questions on the effectiveness of fingerprinting in the wild. Note that developing advanced fingerprinting techniques to detect the so-called *golden image* (the same software and hardware as often deployed in large enterprises), is an active research area—see e.g., [13, 22]. Several automated frameworks have also been designed for large-scale measurement of web tracking in the wild; see e.g., FPDetective [1] and OpenWPM [9]. In this work, we measure tracking techniques in captive portals and landing pages, and use OpenWPM to verify the prevalence of the found trackers on popular websites.

Previous work has also looked into ad injection in web content, see e.g., [21, 25]. We use similar methods for detecting potential similar content injection in hotspots since such incidents have been reported in the past (e.g., [16, 20]).

3 CPInspector on Windows: Design and Data Collection

In this section, we describe CPInspector, the platform we develop for measuring captive portal web-tracking and privacy leakages; see Fig. 1 for the Windows variant. As Android uses a special app for captive portal, we modify CPInspector accordingly; see Sec. 5. The main components of CPInspector include: a browser automation framework, a data migration tool and an analysis module. Selenium is used to visit the hotspot captive portal and perform a wide range of measurements. It collects web traffic, HTTP cookies, WebStorage, fingerprints, browsing profiles, page source code, privacy policy, and screen shots of rendered pages (used to verify the data collection process). CPInspector utilizes Wireshark to capture traffic between the instrumented browser and the hotspot access point. CPInspector uses WebExtensions APIs[2] to collect relevant data (e.g., HTTP cookies, JavaScript calls) from the instrumented browser. Selenium is also used to isolate the test environment from the regular user environment, ensuring that our tests remain unaffected by the user's past browsing behaviors. We also save a copy of the privacy policy, if available. The datasets collected from the hotspots are parsed and committed to a central SQLite database. CPInspector's analysis module then examines the recorded data for tracking behaviors or privacy leaks.

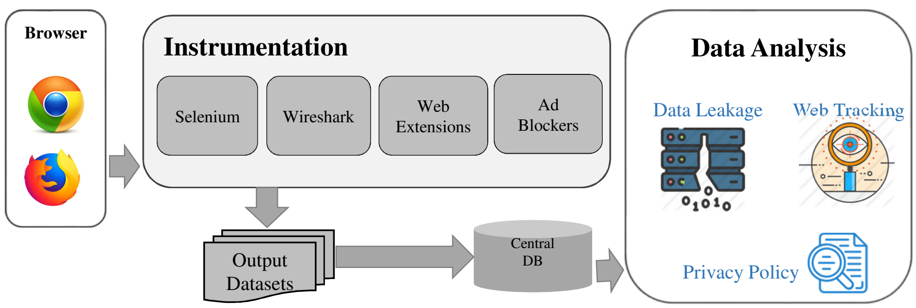


Fig. 1. CPInspector components.

Identifying third-parties. We identified the corporate websites for each hotspot. Then, we use the WHOIS registration records to identify third-party domains by comparing the domain owner name to the hotspot corporate website owner. In cases where the domain information is protected by the WHOIS privacy policy, we visit the domain to detect any redirect to a parent site; we then lookup the parent site's registration information. If this fails, we manually review the domain's `Organization` in its TLS certificate, if available. Otherwise, we try to identify the domain owner based on its WHOIS registration email. We also use `Crunchbase.com` and `Hoovers.com` to determine if the organizations are subsidiaries or acquisitions of larger companies.

Identifying third-party trackers. We use EasyList, EasyPrivacy, and Fanboy to identify known third-party trackers. These lists rely on blacklisted script names, URLs, or domains, which may fail to detect new trackers or variations of known trackers. For this reason, we classify third-party trackers as follows: (a) A *known tracker* is a third-party that has already been identified in the above blacklists. (b) A *possible tracker* is any third-party that can potentially track the user's browsing activities but not included in a blacklist. We observed variations of well-known trackers such as Google Analytics, were missed by the blacklists.

[2] `https://wiki.mozilla.org/WebExtensions`

Table 1. List of evaluated hotspots

Category	Count	Hotspot Name
Cafe and Restaurant	19	A&W, Bombay Mahal Thali, Burger King, Cafe Osmo, Copper Branch, Domino's Pizza, Harvey's, Hvmans Cafe, Juliette Et Chocolat, Mcdonalds, Moose BAWR, Nespresso, Pizza Hut, Pizz izza, Starbucks, Sushi STE-Catherine, The Second Cup, Tim Hortons, Vua Sandwiches
Retail business	17	Canadian Tire, Dynamite, ECCO, Fossil, GAP, Garage, H&M, Home Depot, IGA, Ikea, Laura, Maison Simmons, Michael Kors, Roots, SAQ, Sephora, Walmart
Shopping Mall	12	Atrium 1000, Carrefour Angrignon, Carrefour Laval, Carrefour iA, Centre Eaton, Centre Rockland, Complexe Desjardins, Fairview Pointe-Claire, Mail Champlain, Place Montreal Trust, Place Vertu, Place Ville Marie
Bank	5	CIBC Bank, Desjardins 360, RBC Bank, ScotiaBank, TD Bank
Art and Entertainment	4	Grevin Montreal, YMCA, Montreal Science Centre, Place Des Arts
Transportation	3	Gare d'Autocars de Montreal, Via Rail Station, YUL Airport
Telecom Kiosk	2	Fido, Telus
Car Rental	1	Discount Car Rental
Gymnasium	1	Nautilus Plus
Hospital	1	CHU Sainte-Justine
Hotel	1	Fairmont Hotel
Library	1	Westmount Public Library

Ad injection detection. Our framework also includes a module to detect modifications to user traffic, e.g., for ad injections. We visit two decoy websites (i.e., honeysites in our control and hosted on AmazonAWS) and `BBC.com`, via a home network and a public hotspot, and then compare the differences in the retrieved content (i.e., DOM trees [8]). The use of honeysites allows us to avoid any false positive issues due to the website's dynamic content (e.g., dynamic ads). The first honysite is a static web page while the second is comprised of dynamic content that has four fake ads. The fake ads were created based on source code snippets from Google Adsense, Google TagManager, `Taboola.com`, and `BuySellAds.com`.

Data collection. We collected a total of 679 datasets from the captive portal and landing page of 80 hotspots (12 hotspots are measured at multiple physical locations) between Sept. 2018 to Apr. 2019. We discarded 103 datasets due to some errors (e.g., network failures). We analyzed over 18.5GB of collected traffic for privacy exposure and tracking measurements, and report the results from 67 unique hotspots (576 datasets). We discuss the results in Sec. 4. For the ad injection experiments, we collected a total of 368 datasets from crawling the two honey websites and the `BBC.com` website at 98 hotspot locations. We analyzed over 8.7GB of collected traffic for ad injection, and report the results from 87 unique hotspots (368 datasets). We did not observe any content modification attempts.

4 Analysis and Results for Windows

In this section, we present the results of our analysis on collected personal information, privacy leaks, web trackers, HTTP cookies, and fingerprinting , and the effectiveness of two anti-tracking extensions and private browsing mode.

4.1 Personal Information Collection, Sharing, and Leaking

Personal identifiable information (PII) collection. Most hotspots (40; 59.7%) allow internet access without seeking any explicit personal data. The remaining 27 (40.3%) hotspots use social login , or a registration page to collect significant amount of personal information; 19 (28.4%) of these hotspots mandate social login or user registration, see Table 2.

Sharing with third-parties. Most hotspots share personal information and browser/device information with third-parties via the referrer header, the request-URL, HTTP cookie or WebStorage. We identified 40 hotspots (59.7%) that use third-party captive portals where they share personal information, including 18 (26.9%) share email address; 15 (22.4%) share user's full name; 12 (17.9%) share profile picture; 5 (7.5%) share birthday, current city, current employment and LinkedIn headline; see Table 2. We also found some captive portals leak device/browser information to third-parties, including 40 (59.7%) leak MAC address and last visited site; 18 (26.9%) leak screen resolution; 26 (38.8%) leak user agent; 24 (35.8%) leak browser Information and language; and 15 (22.4%) leak plugins. Moreover, some hotspots leak the MAC address to third-parties, e.g., Pizza Hut to 11 domains, and H&M, Place Montreal Trust and Discount Car Rental to six domains each. Top organizations that receive the MAC addresses include: `Network-auth.com` from 21 hotspots, Alphabet 18, `Openh264.org` 12, Facebook 10, Datavalet 8, and Amazon 6.

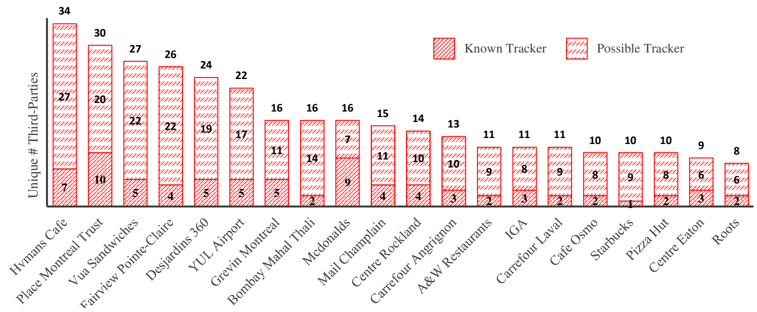
PII leaks via HTTP. We searched for personal information of our used accounts in the collected HTTP traffic, and record the leaked information, including the HTTP request URL, and source (captive portal vs. landing page). Three hotspots transmit the user's full name via HTTP (Place Montreal Trust, Nautilus Plus and Roots). In Place Montreal Trust, the user's full name is saved in a cookie (valid for five years), and each time the user connects to the captive portal, the cookie is automatically transmitted via HTTP. Moreover, three hotspots leak the user's email address via HTTP (Dynamite, Roots, and Garage). In Nautilus Plus, a user must enter her membership number in the captive portal. For partially entered membership numbers, the captive portal verifies the identity by displaying personal information of five people in a scrambled way (first and last names, postal codes, ages, dates of birth, and phone numbers), over HTTP. The user then chooses the right combination corresponding to her personal information. We also confirmed that some of this data belongs to real people by authenticating to this hotspot using ten randomly generated partial membership numbers. Then, we used the reverse lookup in `canada411.ca` to confirm the correlation between the returned phone numbers, names, and addresses.

Table 2. Personal information collected via social login, registration, or optional surveys. The "Powered By" column refers to third-parties that provide hotspot services (when used/identified). F refers to Facebook, L: LinkedIn, I: Instagram, G: Google, T: Twitter, R: registration form, and S: survey; *: personal information is mandatory to access the service.

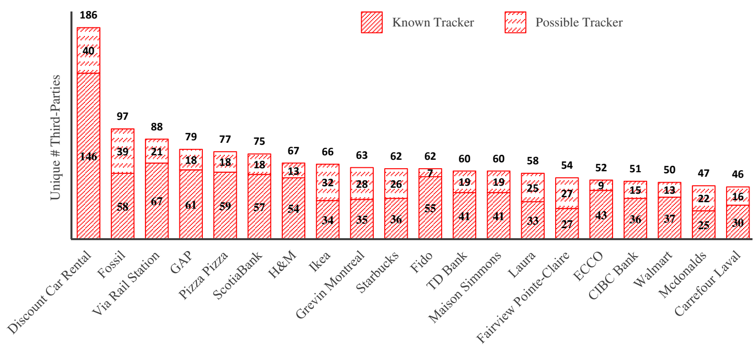
Hotspot	Powered By	Name	Email	Gender	Birthday	Phone Number	Current City	Profile Picture	Home Town	Country	Facebook Likes	Facebook Friends	LinkedIn Headline	Current Employment	Postal Code	# of Children	Basic Profile	Instagram Media	Tweets	People You Follow
Bombay Mahal Thali*	Sy5	FR	FR		F		F													
Carrefour Laval*	Aislelabs	FR	FR	FR	F		F	F	F		F									
Fairview Pointe-Claire*	Aislelabs	FR	FRT	FR	F		F	F	F		F								T	T
Carrefour Angrignon	Eye-In	FGL	FGL					FGL					L	L			L			
Centre Eaton	Eye-In	F	F					F												
Centre Rockland	Eye-In	FL	FL					FL					L	L			L			
Desjardins 360*	JoGoGo	F	F	F	F	R		F				F								
Domino's Pizza			R																	
Dynamite*			R																	
GAP			R																	
Garage*			R																	
Grevin Montreal	Eye-In	FL	FL				F	FL					L	L	S	S	L			
Harvey's*	Colony Networks	F	FR					F												
Hvmans Cafe*	Purple	FR	FR		F		F	F			F						I	I		
Mail Champlain	Eye-In	FL	FL					FL					L	L			L			
Maison Simmon*			R																	
Michael Kors*	Purple	R	R	R	R	R									R					
Montreal Science Centre*	Telus		R																	
Moose BAWR*	Sticky WiFi		R																	
Nautilus Plus*			R																	
Nespresso*	Orange					R														
Place Montreal Trust			R							R					R					
Roots*	Yelp WiFi	R	R																	
Telus*			R																	
Sushi STE-Catherine*	MyWiFi		R																	
Vua Sandwiches*	Coolblue	FR	FR			R		F												
YUL Airport*	Datavalet	FL	FRL					FL					L	L			L			

4.2 Presence of Third-Party Tracking Domains and HTTP Cookies

Tracking domains. We detect third-party tracking domains using: EasyList, EasyPrivacy, and Fanboy's List. On average, each captive portal hosts 7.4 third-party tracking domains (max: 34 domains, including 10 known trackers); see Fig. 2(a). We noticed that the hotspots that use the same third-party captive portal still have a different number of third-parties. For example, for the Datavalet hotspots (YUL Airport, McDonald's, Starbucks, Via Rail Station, Tim Hortons, CIBC Bank, Place Vertu), the number of third-parties are 22, 16, 10, 8, 5, 5, and



(a) Captive portals

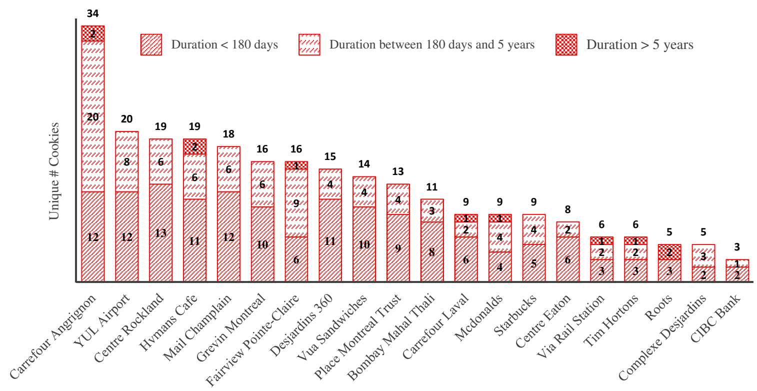


(b) Landing pages

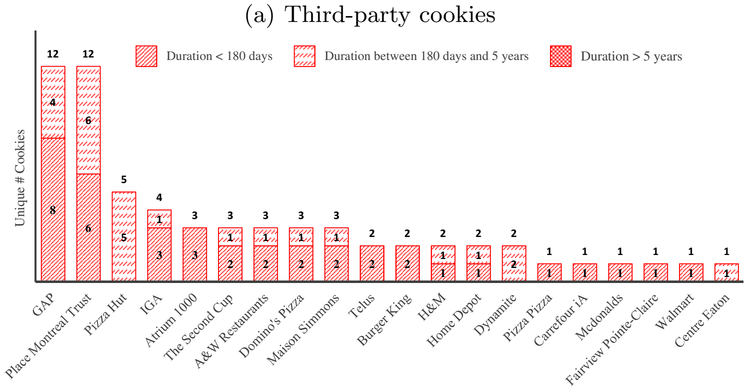
Fig. 2. Number of third-party domains on captive portals and landing pages (top 20). For example, Hvmans Cafe captive portal hosts a total of 34 tracking domains, including 7 known trackers. Note that for all reported tracking/domain statistics, we accumulate the distinct tracking domains as observed in all the datasets collected for a given hotspot (e.g., from both browsers and for different social logins, if required). For list of evaluated hotspots see Table 1.

2 respectively. The hotspots (46; 68.7%) that redirect users to their corporate websites, host more known third-party tracking domains—on average, 30.6 domains per landing page; see Fig. 2(b). We also analyzed the organizations with the highest known-tracker representations. We group domains by the larger parent company that owns these domains. Alphabet, Facebook, and Datavalet are present on over 10% of the captive portals. Alphabet and Facebook are also present on over 50% of the landing pages.

HTTP tracking cookies on captive portals. We found 40 (59.7%) hotspots create third-party cookies valid for various duration—e.g., over 5 years from 10 (14.9%) hotspots, six months to five years from 23 (34.3%) hotspots, and under six months from 38 (56.7%) hotspots; see Fig. 3(a). Via Rail Station, Fairview Pointe-Claire, Carrefour Laval, Roots, McDonald's, Tim Hortons, and Harvey's have a third-party cookie from `network-auth.com`, valid for 20 years. Moreover,



(a) Third-party cookies

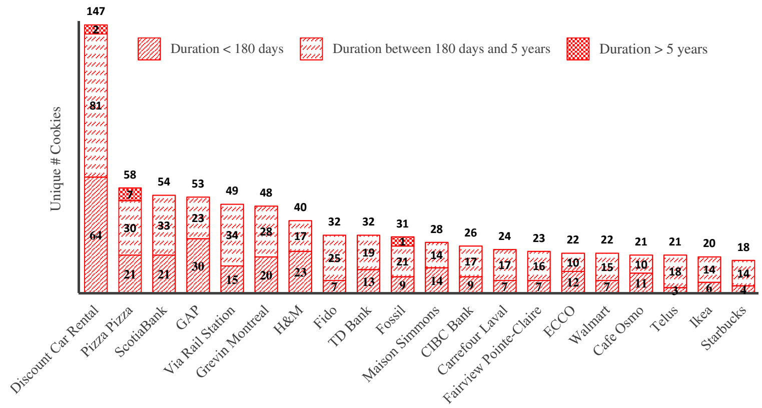


(b) First-party cookies

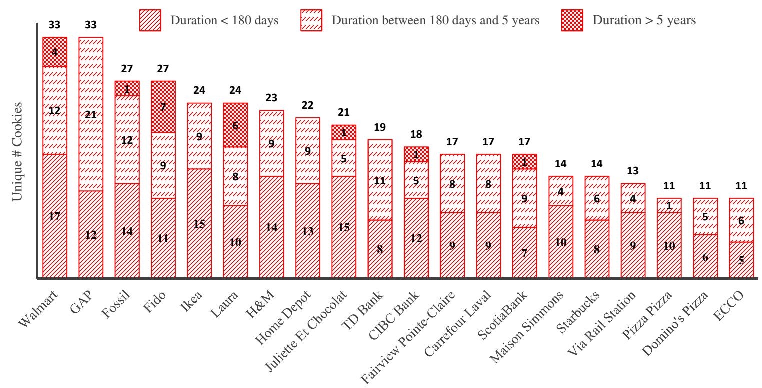
Fig. 3. Number of third-party and first-party cookies on captive portals (top 20). Note that for all reported cookies/domain statistics, we accumulate the distinct cookies as observed in all the datasets collected for a given hotspot.

YUL Airport, Via Rail Station, Complexe Desjardins, McDonald's, Starbucks, Tim Hortons, CIBC Bank have a common 1-year valid cookie from Datavalet, except for CIBC (17 days). This cookie uniquely identifies a device based on the MAC address (set to the same value unless the MAC address is spoofed). Some hotspots save the MAC address in HTTP cookies, including CHU Sainte-Justine, Moose BAWR, and Centre Rockland.

We also analyze first-party cookies on captive portals; see Fig. 3(b). 22 (32.8%) hotspots create first-party cookies valid for various durations; 14 (20.9%) hotspots include cookies valid for periods ranging from six months to five years, and 17 (25.4%) hotspots for less than 6 months. Place Montreal Trust saves the user's full name in a first-party cookie valid for five years; this cookie is transmitted via HTTP. Finally, we analyzed hotspots that create persistent cookies before explicit consent from the user, we found 26 (38.8%) hotspots create cookies that are valid for periods varying from 30 minutes to a year, including Domino's Pizza, Fido, GAP, H&M, McDonald's, Roots, Starbucks, and Tim Hortons.



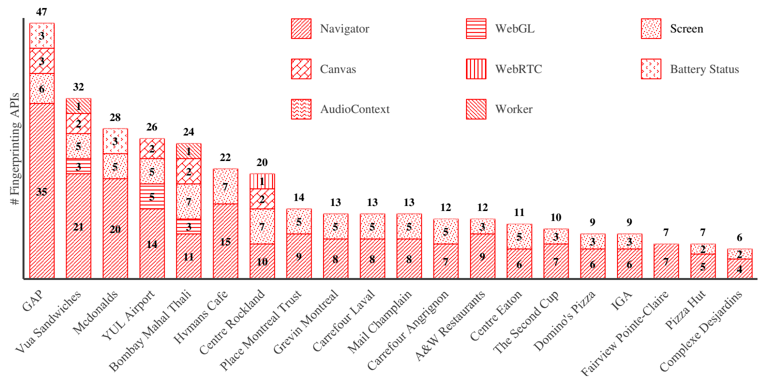
(a) Third-party cookies



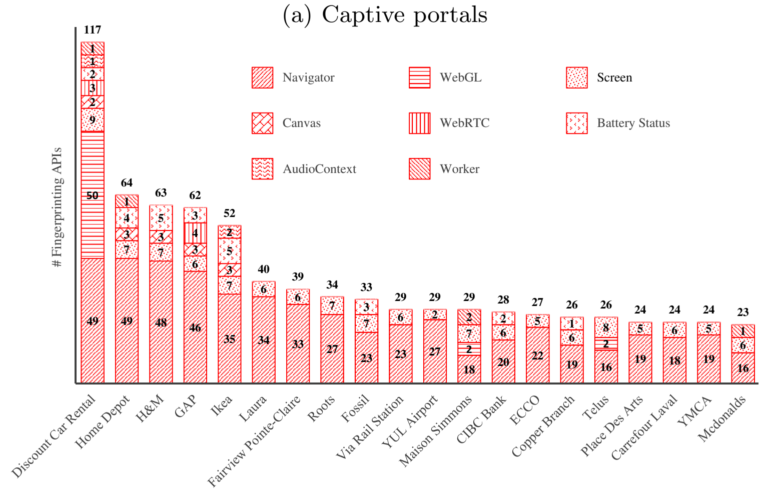
(b) First-party cookies

Fig. 4. Number of third-party and first-party cookies on landing pages (top 20). Note that for all reported cookies/domain statistics, we accumulate the distinct cookies as observed in all the datasets collected for a given hotspot.

HTTP tracking cookies on landing pages. We found 48 (71.6%) hotspots create third-party cookies valid for various durations—e.g., over 5 years from 4 (6.0%) hotspots, six months to five years from 47 (70.1%) hotspots, and under six months from 42 (62.7%) hotspots, see Fig. 4(a). Prominent examples include the following. Fossil has a 25-year valid cookie from `pbbl.com`; CIBC Bank has two 5-year valid cookies from `stackadapt.com`, a known tracker. We also analyzed the first-party cookies on landing pages; see Fig. 4(b). 41 (62.7%) hotspots create first-party cookies valid for various durations—e.g., over 5 years from 10 (14.9%) hotspots, six months to five years from 42 (62.7%) hotspots, and under six months from 41 (61.2%) hotspots. Notable examples: Fossil has a 99-year valid cookie, Fido has three cookies valid for 68–81 years, CHU Sainte-Justine has a 20-year valid cookie, CIBC Bank has a 19-year cookie, and Walmart has four cookies valid for 9–20 years.



(a) Captive portals



(b) Landing pages

Fig. 5. Number of fingerprinting APIs on captive portals and landing pages (top 20). Note that for all fingerprinting statistics, we accumulate the distinct APIs as observed in all the datasets collected for a given hotspot.

4.3 Device and Browser Fingerprinting

We analyzed fingerprinting attempts in captive portals and landing pages. We use Don't FingerPrint Me (DFPM [12]) for detecting known fingerprinting techniques, including the screen object, navigator object, WebRTC, Font, WebGL, Canvas, AudioContext, and Battery Status [9, 18, 17, 19]. We use attribute and API interchangeably, when referring to fingerprinting JavaScript APIs.

Captive portal. 24 (35.8%) hotspots perform some form of fingerprinting. On average, each captive portal uses 5.9 attributes (max: 47 attributes, including 35 Navigator, 6 Screen, 3 Canvas, and 3 Battery Status); see Fig. 5(a). We also found 10 (14.9%) hotspots fingerprint user device/browser before explicit consent from the user, including GAP, McDonald's, and Place Montreal Trust, using 6–46 attributes. Moreover, 46 (68.7%) hotspots fingerprint MAC addresses.
Landing pages. 51 (76.1%) hotspots perform fingerprinting on their landing pages. On average, each landing page fingerprints 19.4 attributes (max: 117 attributes, including 49 Navigator, 9 Screen, 2 Canvas, 3 WebRTC, 50 WebGL, 1 AudioContext, 1 Worker and 2 Battery Status); see Fig. 5(b). Prominent examples include the following. Discount Car Rental includes script from Sizmek Technologies Inc., which uses a total of 67 APIs (48 WebGL, 12 Navigator, five Screen, and two Canvas APIs). Manual analysis also reveals Font fingerprinting via side-channel inference [18]; this script is also highly similar to FingerprintJS [26]. Discount Car Rental also uses script from Integral Ad Science, which uses 41 attributes, including: 31 Navigator, seven Screen APIs, two WebRTC, and one AudioContext (cf. [9]). The navigator APIs are used to collect attributes such as the USB gamepad controllers, and list MIDI input and output devices. H&M and Home Depot host the same JavaScript that collects 42 attributes, including 34 Navigator, six Screen, and two Canvas APIs. Laura has a script from PerimeterX that collects 27 attributes, including 21 Navigator and 6 Screen APIs; code manual analysis reveals WebGL and Canvas fingerprinting.

5 CPInspector for Android

In contrast to Windows, Android OS handles captive portals with a dedicated application. The Android Developers documentation and Android Source documentation omit details of how Android handles captive portals. Here we briefly document the inner working of Android captive portals, and discuss our preliminary findings, specifically on tracking cookies on Android devices.
Android captive portal login app. Using Android `ps` (Process Status), we observe that a new process named `com.android.captiveportallogin` appears whenever the captive portal is launched. The Manifest file for CaptivePortalLogin explicitly defines that its activity class will receive all captive portal broadcasts by any application installed on the OS and handle the captive portal. We observe that files in the data folder of this application are populated and altered during a captive portal session; we collect these files from our tests.
Capturing network traffic. To capture traffic from Android apps, several readily-available VPN apps from Google Play can be used (e.g., Packet Capture, NetCapture, NetKeeper). However, Android does not use VPN for captive portals. On the other hand, using an MITM Proxy server such as mitmproxy (`https://mitmproxy.org/`) requires the server to run on a desktop environment, which would make the internet traffic come out of the desktop OS, i.e., the mobile device would not be visible to the hotspot. To overcome this, we set up a virtual Linux environment within the Android OS by using Linux Deploy

(https://github.com/meefik/linuxdeploy), enabling us to run Linux desktop applications within Android with access to the core component of Android OS, e.g., Android OS processes, network interfaces, etc. We use Debian and mitmproxy on the virtual environment, and configure Android's network settings to proxy all the traffic going through the WiFi adapter to the mitmproxy server. The proxy provides us the shared session keys established with a destination server, enabling us to decrypt HTTPS traffic. We use tcpdump to capture the network traffic.

Data collection and analysis. We visited 22 hotspots and collected network traffic from their captive portals. First, we clear the data and cache of the CaptivePortalLogin app and collect data from a given hotspot. Next, we change the MAC address of our test devices (Google Pixel 3 with Android 9 and Nexus 4 with Android 5.1.1) and collect data again without clearing the data and cache. From the proxy's request packets, we confirm that the browser agent correctly reflects our test devices, and the traffic is being originated from the CaptivePortalLogin app. Next, we analyze the data extracted from the app. The structure of the data directory is similar to Google Chrome on Android. We locate the `.\app_webview\Cookies` SQLite file in the data directory, storing the CaptivePortalLogin app's cookies.

We observe that 9 out of 22 hotspots store persistent cookies in the captive portal app; see Fig. 6. These cookies are not erased when the portal app is closed, or when the user leaves the hotspot. Instead, the cookies remain active as set in their validity periods, although they are unavailable to the regular browser apps. Prominent examples include: Tim Hortons inserts a 20-year valid cookie from `network-auth.com`, and Hvmans Cafe stores a 10-year valid cookie from Instagram. In the captive portal traffic, we confirm that these cookies are indeed present and shared in subsequent visits, and follow the Same-Origin Policy. Hotspots can use these cookies to uniquely identify and authenticate user devices even when the device MAC address is dynamically changed; Tim Hortons hotspot uses its cookies for authentication. However, McDonald's did not authenticate the device even though the cookies were present but the MAC was new.

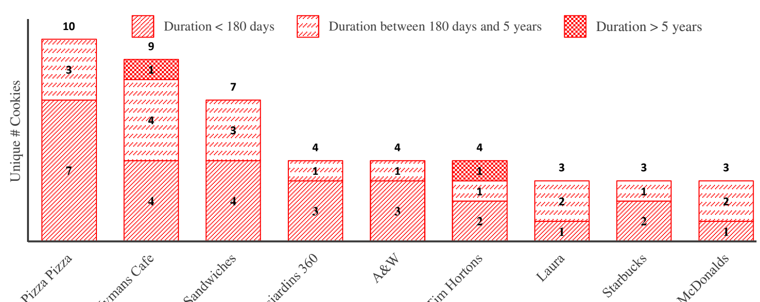


Fig. 6. Number of cookies stored on the Android captive portal app

6 Privacy Policy and Anti-Tracking

We performed a preliminary manual analysis of privacy policy and TOS documents from hotspots that appear to be most risky. Roots states clearly in their privacy policy that they use SSL to protect PII, but their captive portal transmits a user's full name and email address via HTTP. Place Montreal Trust transmits the user's full name via HTTP, and they explicitly state that transmission of information over the public networks cannot be guaranteed to be 100% secure. Nautilus Plus has a very basic TOS that omits important information such as the laws they comply with and privacy implications of using their hotspot. They state clearly that the assurance of confidentiality of the user's information is of great concern to Nautilus Plus, but they use HTTP for all communications, leaking personal information while they attempt to verify the customer's identity; see Sec. 4.1. Their privacy policy is also inaccessible from the captive portal and omits any reference to WiFi. Dynamite and Garage transmit the user's email address via HTTP despite claiming to use SSL. Their privacy policy is inaccessible from the captive portal and omits any reference to the WiFi. GAP explicitly mentions their collection of browser/device information, and they indeed collect 46 such attributes, *before* the user accepts the hotspot's policies.

Although McDonald's tracks users in their captive portal (9 known trackers, 28 fingerprinting attributes), the captive portal itself lacks a privacy policy stating their use of web tracking. Carrefour Laval and Fairview Pointe-Claire perform cross device tracking by participating in the Device Co-op [2], where they may collect and share information about devices linked to the user. Two hotspots link the users MAC address to the collected personal information, including Roots, and Bombay Mahal Thali. Sharing the harvested personal data with subsidiaries and third-party affiliates is also the norm. Eight hotspots (including Hvmans Cafe, Fairview Pointe-Claire, and Carrefour Laval) state that PII may be stored outside Canada. Ten hotspots omit any information about the PII storage location, including Dominos's Pizza and Roots. However, five hotspots have their captive portal domain in the US, including Bombay Mahal Thali, Carrefour Angrignon, Domino's Pizza, Grevin Montreal, and Roots. We found 34 (50.7%) hotspots have a TOS document but lack a privacy policy on their captive portal, including TD Bank, and Burger King. Three hotspots lack both the privacy policy and TOS document on their captive portals, including Laura, ECCO, and Maison Simmons.

The same hotspot captive portal in different locations. 12 hotspots are measured at multiple physical locations. We stopped collecting datasets from different locations of the same chain-business as the collected datasets were largely the same. We provide an example where some minor differences occur: Starbucks' captive portal domain varies in the two evaluated locations (`am.datavalet.io` vs. `sbux-j2.datavalet.io`). However, the number of known trackers remained the same, while the number of third-parties increased by one domain. Moreover, the `--sf-device` cookie validity increased from 17 days to 1 year, and the `--sf-landing` cookie was not created in the second location.

Effectiveness of privacy extensions and private browsing. To evaluate the effectiveness anti-tracking solutions against hotspot trackers, we collected traffic from both Chrome and Firefox in private browsing modes, and by enabling Adblock Plus, and Privacy Badger extensions—leading to a total of six datasets for each hotspot. Then, we use the EasyList, EasyPrivacy, and Fanboy's lists to determine whether known trackers remain in the collected datasets; see Table 3.

Table 3. The number of unique known trackers not blocked by our anti-tracking solutions.

	W/O Ad Blockers	AdBlock Plus	Privacy Badger	Private Browsing
Firefox	382	33	180	315
Chrome	488	117	212	356

Hotspot trackers in the wild. We measured the prevalence of trackers found in captive portals and landing pages, in popular websites—to understand the reach and consequences of hotspot trackers. We use OpenWPM [9] between Feb. 28–Mar. 15, 2019 to automatically browse the home pages of the top 143k Tranco domains [15] as of Feb. 27, 2019. We extract the tracking persistent cookie domains from captive portals or landing pages; we define such cookies to have validity ≥ 1 day and the sum of the value lengths from all the cookies from the same third-party website longer than 35 characters—cf. [5]. Then, we counted those tracking domains in the OpenWPM database; see Table 6. For example, the `doubleclick.net` cookie as found in 4 captive portals and 30 landing pages, appears 160,508 times in the top 143k Tranco domains (mutiple times in some domains). Overall, hotspot users can be tracked across websites, even long time after the user has left a hotspot.

Table 4. Count of tracking domains from captive portals and landing pages in Alexa 143k home pages (top 10).

Captive Portal		Landing Page	
Tracker	**Count**	**Tracker**	**Count**
doubleclick.net	160508	pubmatic.com	326991
linkedin.com	48726	rubiconproject.com	257643
facebook.com	37107	doubleclick.net	160508
twitter.com	14874	casalemedia.com	131626
google.com	13676	adsrvr.org	116438
atdmt.com	5198	addthis.com	83221
instagram.com	3466	demdex.net	83160
gap.com	295	contextweb.com	82965
maxmind.com	294	rlcdn.com	75295
gapcanada.ca	64	livechatinc.com	69919

7 Conclusion and Future Work

Many people across the world use public WiFi offered by an increasing number of businesses and public/government services. The use of VPNs, and the adoption of HTTPS in most websites and mobile apps largely secure users' personal/financial data from a malicious hotspot provider and other users of the same hotspot. However, device/user tracking as enabled by hotspots due to their access to MAC address and PII, remains as a significant privacy threat, which has not been explored thus far. Our analysis shows clear evidence of privacy risks and calls for more thorough scrutiny of these public hotspots by e.g., privacy advocates and government regulators.

Our study covers hotspots in Montreal, Canada, and we are currently working on collecting data from other parts of the world. Our recommendations for hotspots users include the following: avoid sharing any personal information with the hotspot (social media or registration forms); use private browsing and possibly some other anti-tracking browser addons, and software programs that may allow to use a fake MAC address on Windows; and clear the browser history after visiting a hotspot if private browsing mode is not used. Additional suggestions are available at: `https://madiba.encs.concordia.ca/reports/OPC-2018/`.

8 Acknowledgement

This work was partly supported by a grant from the Office of the Privacy Commissioner of Canada (OPC) Contributions Program. We thank the anonymous DPM 2019 reviewers for their insightful suggestions and comments, and all the volunteers for their hotspot data collection. We also thank the members of Concordia's Madiba Security Research Group, especially Nayanamana Samarasinghe, for his help in running OpenWPM to automatically browse the home pages of the top 143k Tranco domains.

References

1. Acar, G., Juarez, M., Nikiforakis, N., Diaz, C., Gürses, S., Piessens, F., Preneel, B.: FPDetective: Dusting the web for fingerprinters. In: ACM CCS'13. Berlin, Germany (Nov 2013)
2. Adobe.com: Adobe experiance cloud: Device Co-op privacy control, `https://cross-device-privacy.adobe.com`
3. Binns, R., Zhao, J., Kleek, M.V., Shadbolt, N.: Measuring third-party tracker power across web and mobile. ACM Transaction on Internet Technology **18**(4), 52:1–52:22 (Aug 2018)
4. Brookman, J., Rouge, P., Alva, A., Yeung, C.: Cross-device tracking: Measurement and disclosures. In: Proceedings on Privacy Enhancing Technologies (PETS). Minneapolis, MN, USA (Jul 2017)
5. Bujlow, T., Carela-Español, V., Sole-Pareta, J., Barlet-Ros, P.: A survey on web tracking: Mechanisms, implications, and defenses. Proceedings of the IEEE **105**(8), 1476–1510 (2017)

6. Cheng, N., Wang, X.O., Cheng, W., Mohapatra, P., Seneviratne, A.: Characterizing privacy leakage of public wifi networks for users on travel. In: 2013 Proceedings IEEE INFOCOM. Turin, Italy (Apr 2013)
7. Eckersley, P.: How unique is your web browser? In: International Symposium on Privacy Enhancing Technologies Symposium (2010)
8. Elifantiev, O.: NodeJS module to compare two DOM-trees, https://github.com/Olegas/dom-compare
9. Englehardt, S., Narayanan, A.: Online tracking: A 1-million-site measurement and analysis. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna, Austria (Oct 2016)
10. Gómez-Boix, A., Laperdrix, P., Baudry, B.: Hiding in the crowd: an analysis of the effectiveness of browser fingerprinting at large scale. In: TheWebConf (WWW'18). Lyon, France (Apr 2018)
11. Google: HTTPS encryption on the web, https://transparencyreport.google.com/https/overview?hl=en
12. Klafter, R.: Don't FingerPrint Me, https://github.com/freethenation/DFPM
13. Klein, A., Pinkas, B.: DNS cache-based user tracking. In: Network and Distributed System Security Symposium (NDSS'19). San Diego, CA, USA (Feb 2019)
14. Laperdrix, P., Rudametkin, W., Baudry, B.: Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In: IEEE Symposium on Security and Privacy (SP). San Jose, CA, USA (2016)
15. Le Pochat, V., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., Joosen, W.: Tranco: A research-oriented top sites ranking hardened against manipulation. In: NDSS'19. San Diego, CA, USA (Feb 2019)
16. Medium.com: My hotel WiFi injects ads. does yours?, news article (Mar. 25, 2016). https://medium.com/@nicklum/my-hotel-WiFi-injects-ads-does-yours-6356710fa180
17. Mowery, K., Shacham, H.: Pixel perfect: Fingerprinting canvas in HTML5. Proceedings of W2SP pp. 1–12 (2012)
18. Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., Vigna, G.: Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In: 2013 IEEE Symposium on Security and Privacy. Berkeley, CA, USA (May 2013)
19. Olejnik, Ł., Acar, G., Castelluccia, C., Diaz, C.: The leaking battery. In: Data Privacy Management, and Security Assurance, pp. 254–263. Springer (2015)
20. PCWorld.com: Comcast's open WiFi hotspots inject ads into your browser, news article (Sep. 09, 2014). https://www.pcworld.com/article/2604422/comcasts-open-wi-fi-hotspots-inject-ads-into-your-browser.html
21. Reis, C., Gribble, S.D., Kohno, T., Weaver, N.C.: Detecting in-flight page changes with web tripwires. In: NSDI'08. San Francisco, CA, USA (2008)
22. Sanchez-Rola, I., Santos, I., Balzarotti, D.: Clock around the clock: Time-based device fingerprinting. In: ACM CCS'18. Toronto, Canada (Oct 2018)
23. Sombatruang, N., Kadobayashi, Y., Sasse, M.A., Baddeley, M., Miyamoto, D.: The continued risks of unsecured public WiFi and why users keep using it: Evidence from Japan. In: Privacy, Security and Trust (PST'18). Belfast, UK (Aug 2018)
24. Symantec: Norton WiFi risk report: Summary of global results, tech report (May 5, 2017). https://www.symantec.com/content/dam/symantec/docs/reports/2017-norton-wifi-risk-report-global-results-summary-en.pdf
25. Tsirantonakis, G., Ilia, P., Ioannidis, S., Athanasopoulos, E., Polychronakis, M.: A large-scale analysis of content modification by open HTTP proxies. In: Network and Distributed System Security Symposium (NDSS'18) (2018)
26. Valve: Fingerprintjs by Valve, https://valve.github.io/fingerprintjs/