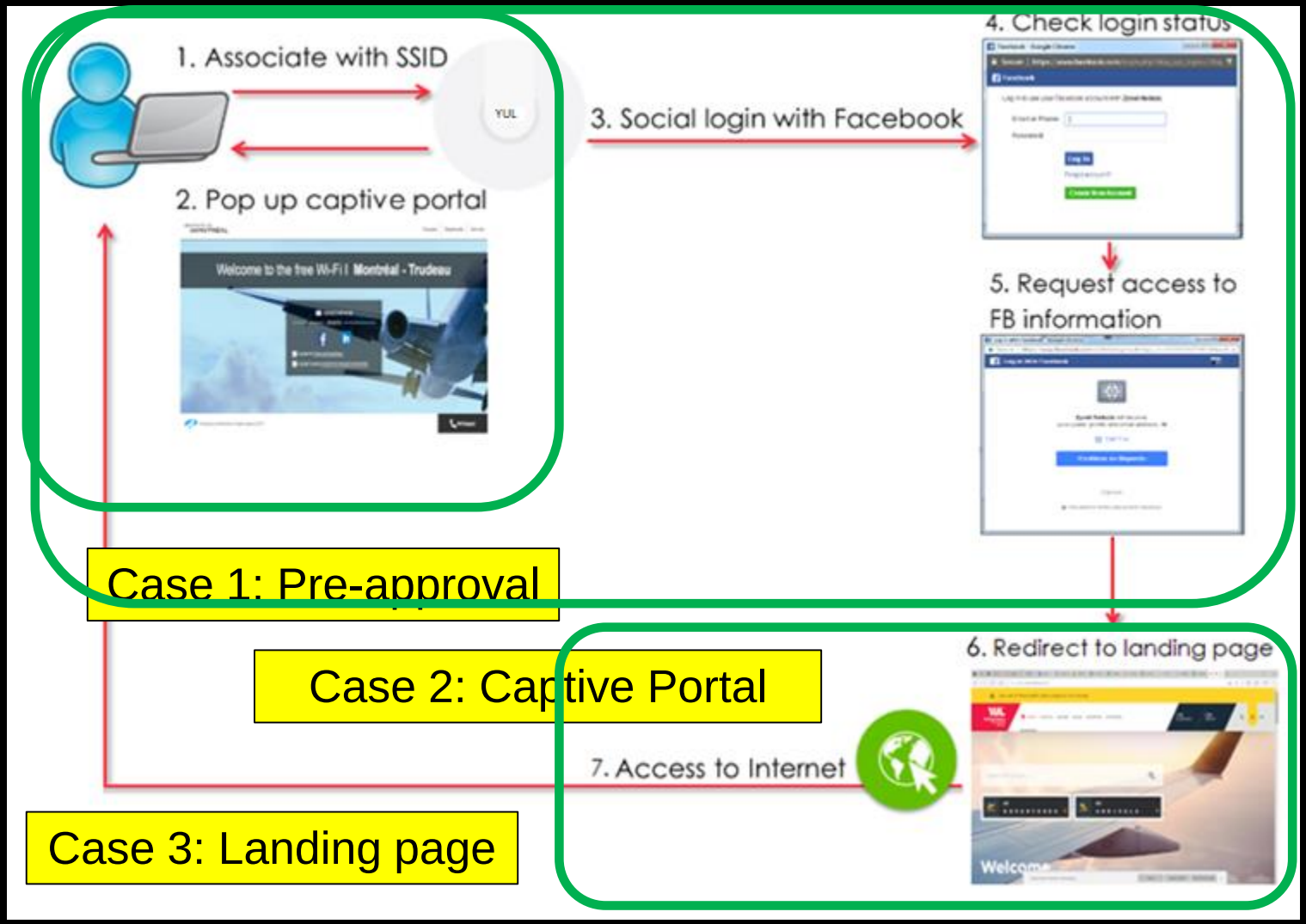
Privacy Risks of Public Wi-Fi Captive Portals

Suzan Ali

Concordia University, Montreal, Canada

Collaborators: Tousif Osman, Mohammad Mannan, and Amr Youssef

Wi-Fi captive portal



What are the risks of Wi-Fi captive portals on user privacy?

Goals

Study privacy leakage

- Collecting sensitive data
- Sharing data with third-parties
- Captive portal lack of HTTPS adoption

Study stateful or stateless tracking behaviors

Hotspots can deduce private information about the user linked to her real identity. e.g., sexual orientation

Why this is important?

1. Hotspots have access to the users' foot traffic and browsing history
2. They can deduce many privacy invasive information from this data
   In a city-WiFi , they can figure out your sex orientation from your foot traffic
3. This could be linked to your real identity
4. One login with your Facebook account is enough to track you for long time
5. If you clear your browser history, they can lookup your real identify using your device MAC address
6. If you try to avoid this by spoofing your MAC address, they could try identify you using your device and browser fingerprint
7. If your device or browser fingerprint has changed, they could link this new fingerprint with your existing profile in those hotspots

Data collection

80 hotspot locations in Montreal, Canada

Local and international brands

576 datasets analyzed for different cases

59.7% use third-party captive portals

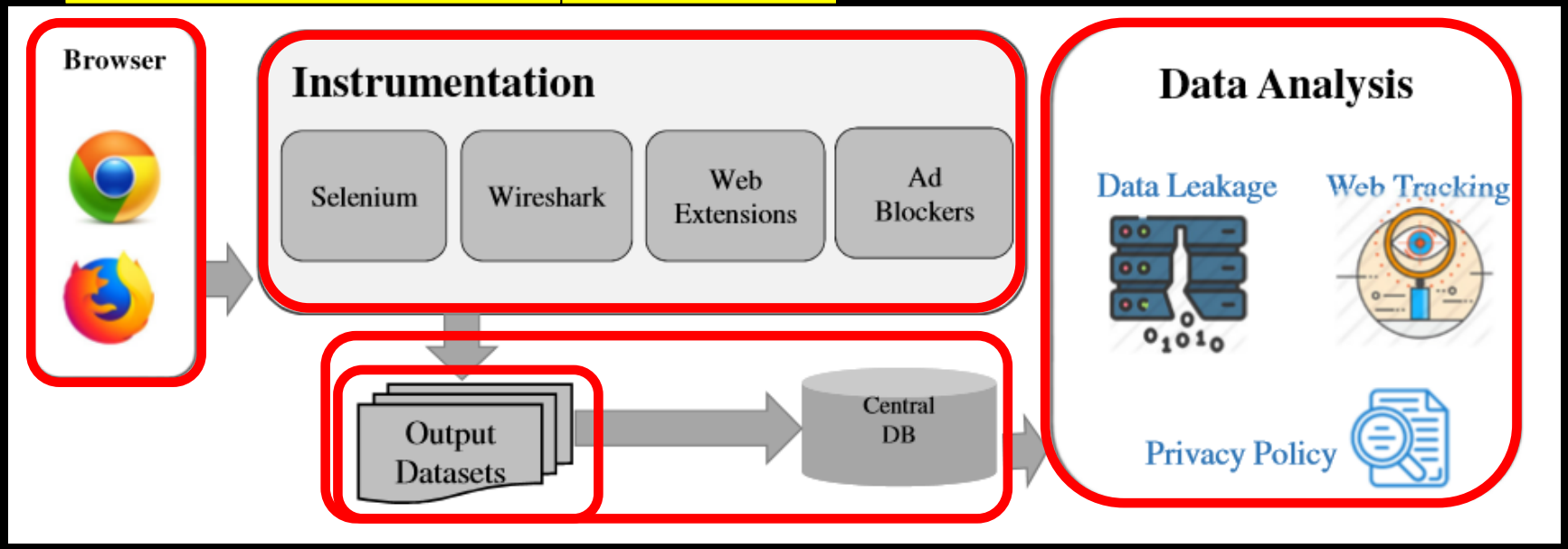
Results might be applicable to a larger geographical scope

Data collection framework

CPInspector collects Web traffic, HTTP cookies, WebStorages, fingerprints, browsing profile, source code, terms of use, privacy policy, and screen shots of the rendered pages

CPInspector on Windows

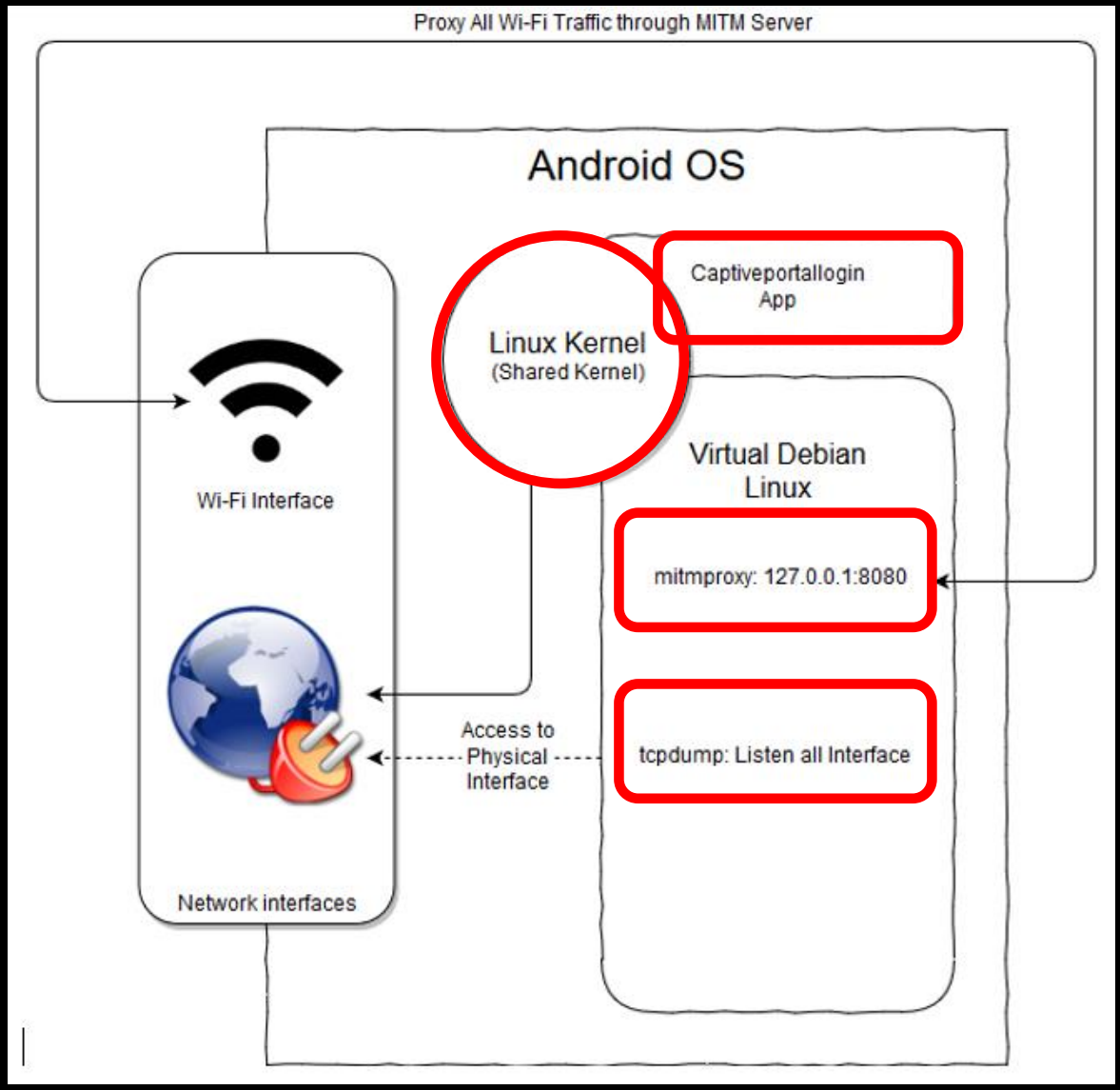
Data analysis tool | framework



The output dataset contains web traffic, HTTP cookies, WebStorages, fingerprints, browsing profile, source code, TOS, privacy policy, and screen shots of the rendered pages

CPInspector on Android

Use special browser to launch captive portals



Results

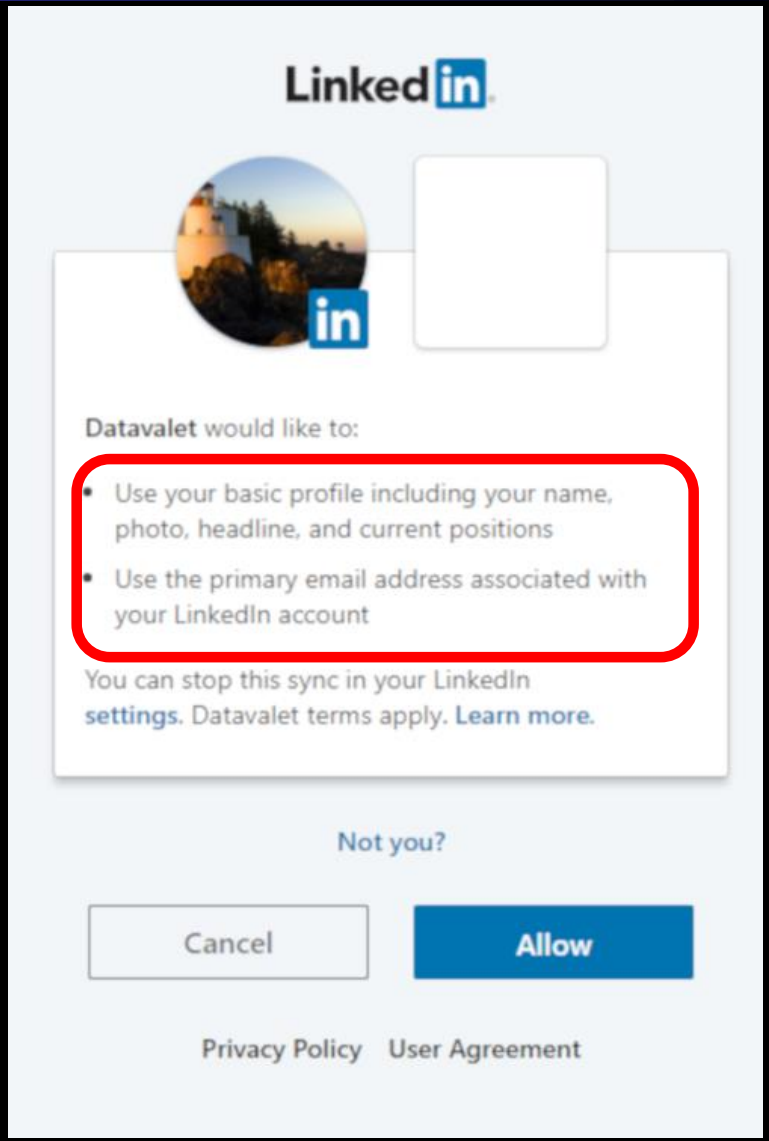
Personal information collection

- LinkedIn
- Facebook
- Twitter
- Instagram
- Google
- Registration form
- User surveys

40% collect personal sensitive information and mandatory collection for 27%

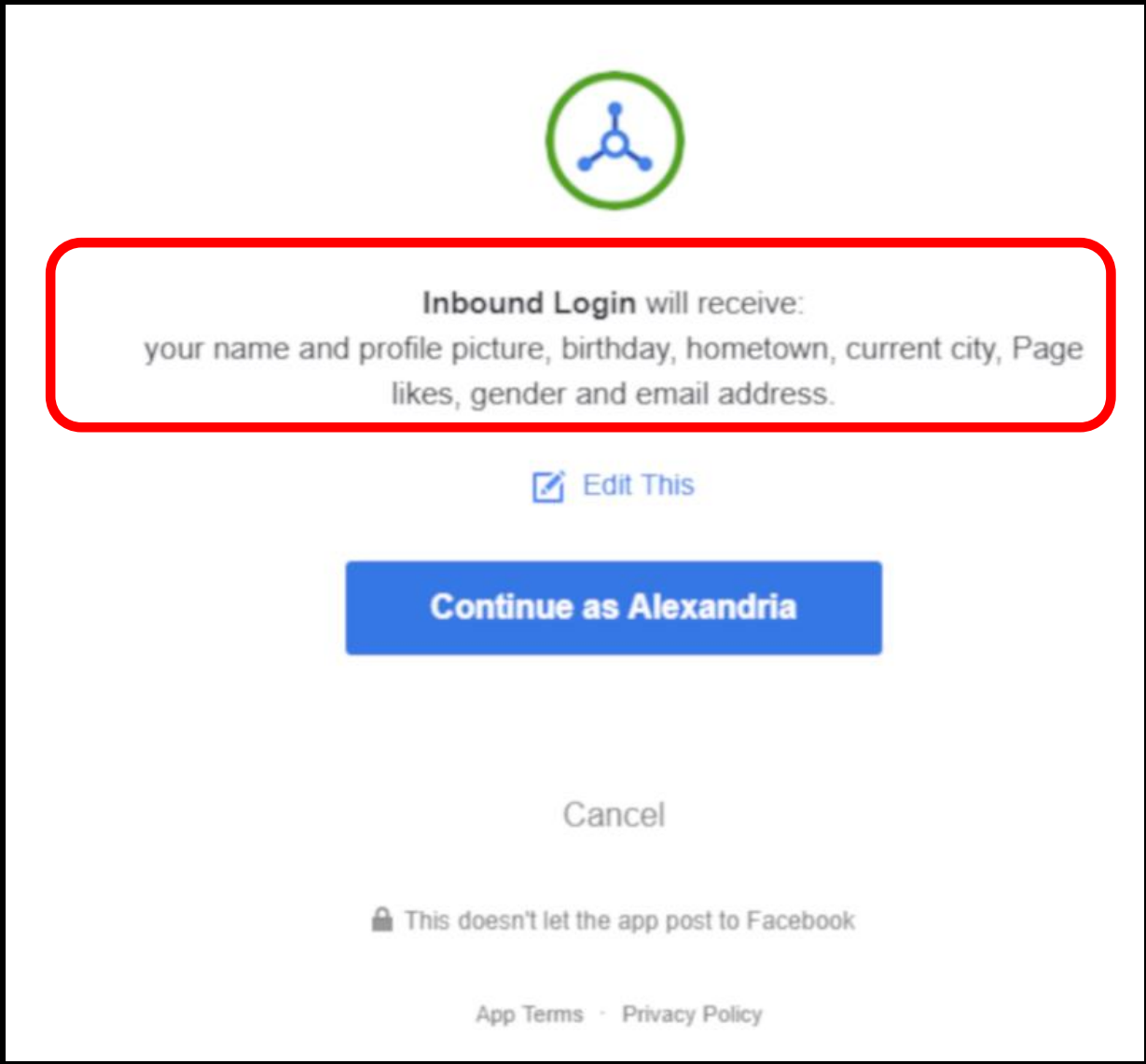
LinkedIn

- Basic profile including:
- Name
- Photo
- Headlines
- Current positions
- Email address



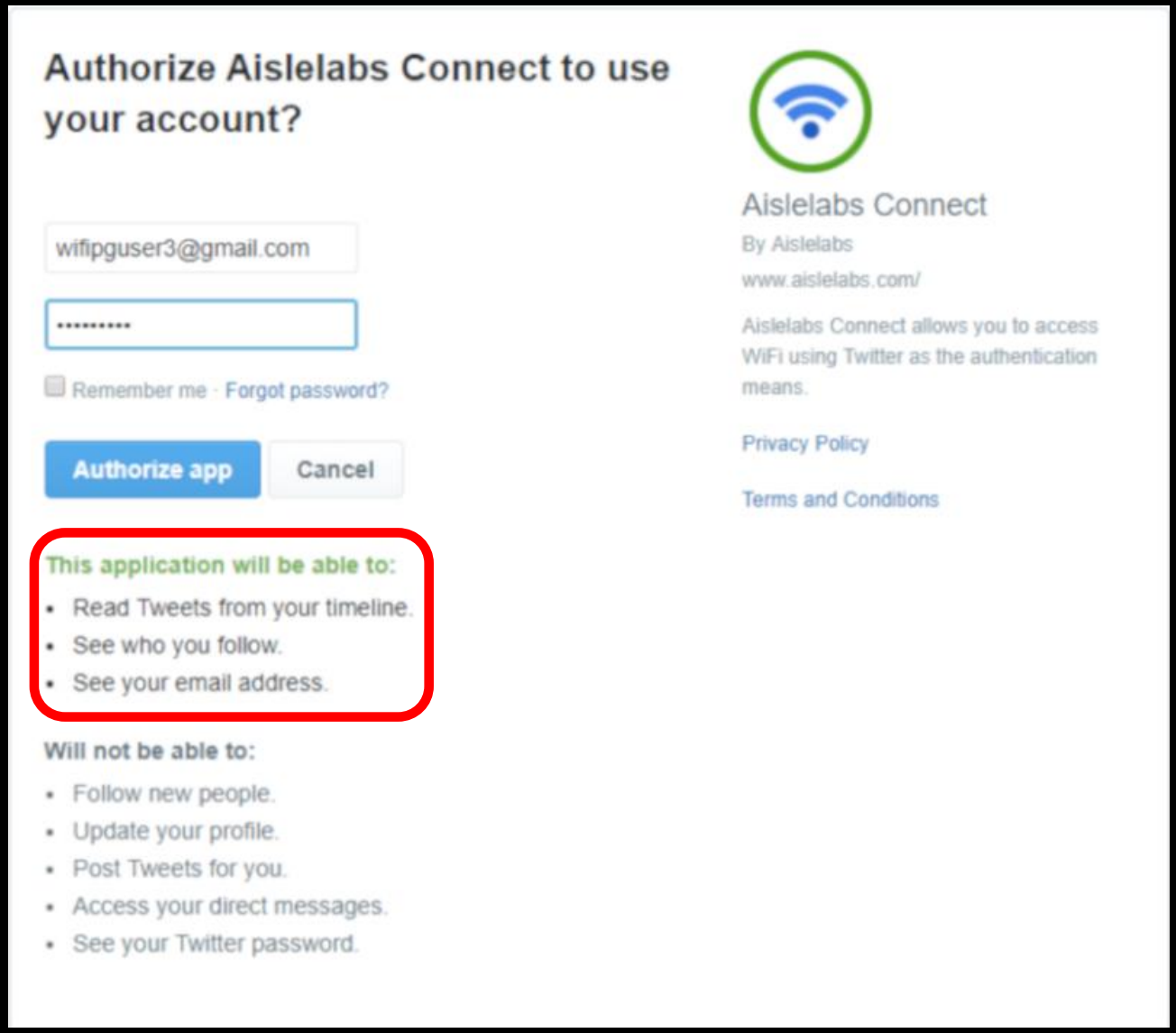
Facebook

- Name
- Profile picture
- Birthday
- Home town
- Current city
- Page likes
- Gender
- Email address

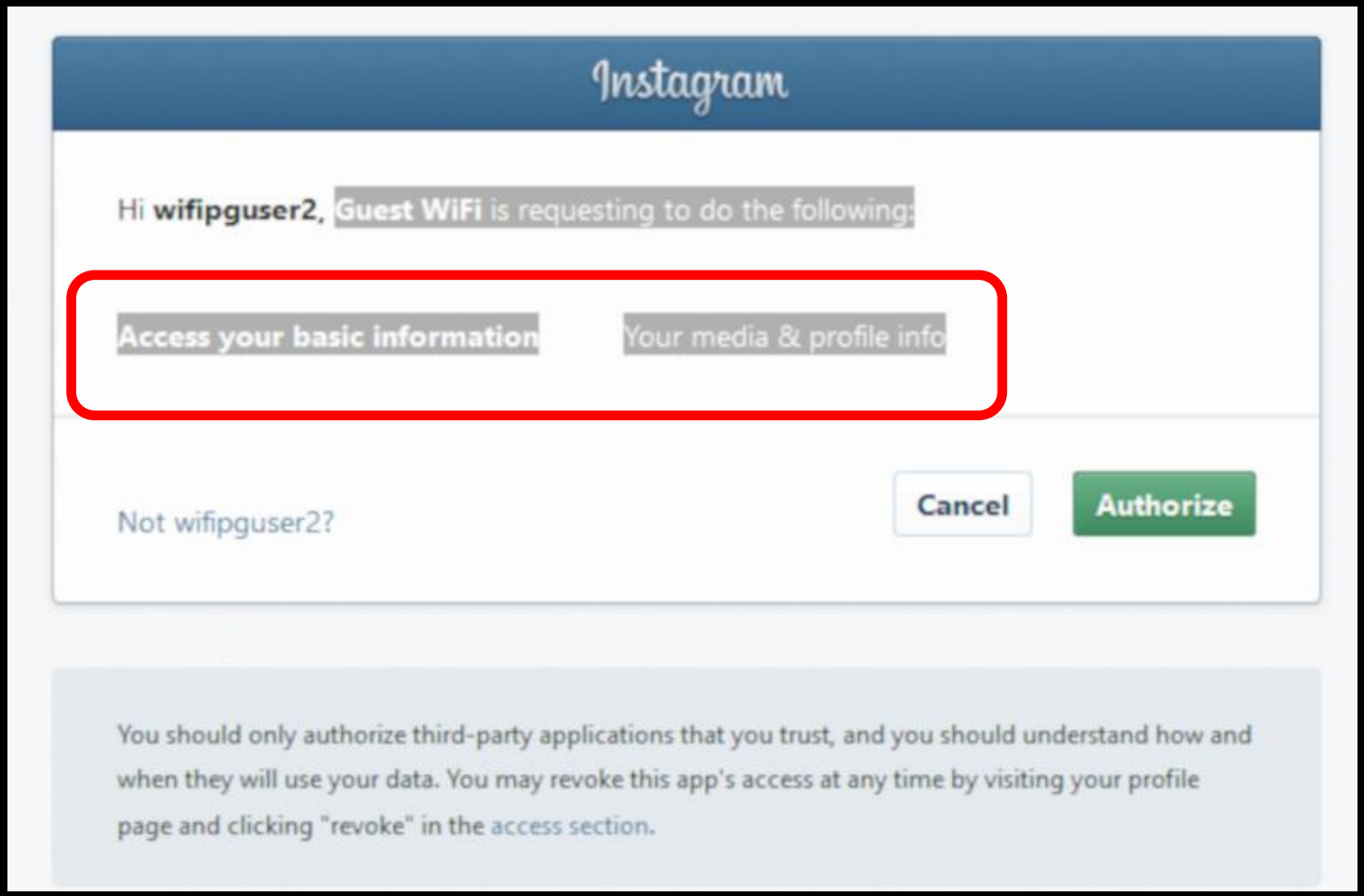


Twitter

- Email address
- Tweets
- The people you follow



Instagram



Media and profile info which includes more information such as name, email, bio, and profile picture

Web Tracking

Except 3, all hotspots use tracking technologies

	Average # of domains	Max # of domains
On captive portals	7.4	34
On landing pages	30.6	186

Google, Facebook, and Datavalet are present on over 10% of the captive portals

Google and Facebook are also present on over 50% of the landing pages

HTTP cookies

1. 38.8% create persistent cookies before user approval to privacy policy
2. Future tracking after leaving the hotspots up to 20 years

	First-party cookies	Third-party cookies
On captive portals	32.8%	59.7%
On landing pages	62.7%	71.6%

3. Track users across hotspots using same origin policy (e.g., a cookie from datavalet.io track user in 8 hotspots)

Device/Browser fingerprinting

"Don't Finger Print Me" extension

35.8% of captive portals fingerprint

Average: 5.9 attribute, Max: 47

76.1% of landing pages fingerprint

Average: 19.4 attribute, Max: 117

14.9% of hotspots fingerprint users before approving the privacy policy

(Eckersley: 83.6% identified by only 8 attributes fingerprint)

Web tracking on Android

1. Store cookies in the captive portal app
2. Cookies are unavailable to the regular browsers
3. 9 out of 22 hotspots store persistent cookies
4. These cookies are not erased when the portal app is closed, or when the user leaves the hotspot

Effective tracking of Android devices despite Android MAC address randomization

Other Risks

1. Five hotspots leak personal information via HTTP
2. 59.7% can uniquely track users for a long period
   Fingerprint the device MAC address
3. 14.9% of which collect personal information and may link it to the device MAC address
   Only 2.9% have declared this in their privacy policies
4. 61.1% leak hotspot's visiting habits to third-parties
5. 50.7% lack a privacy policy, 3 of which lacks terms of use
6. 2.9% use Adobe Device Co-op to track user across devices
   Participant companies share a "hashed login IDs" with Adobe

More results

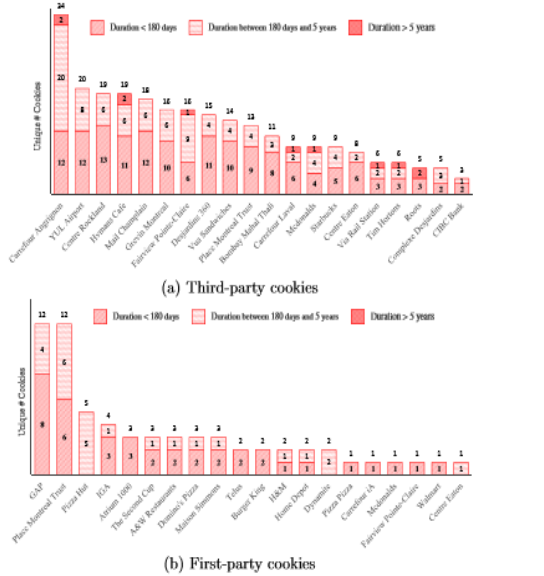


Fig. 3. Number of third-party and first-party cookies on captive portals (top 20). Note that for all reported cookies/domain statistics, we accumulate the distinct cookies as observed in all the datasets collected for a given hotspot.

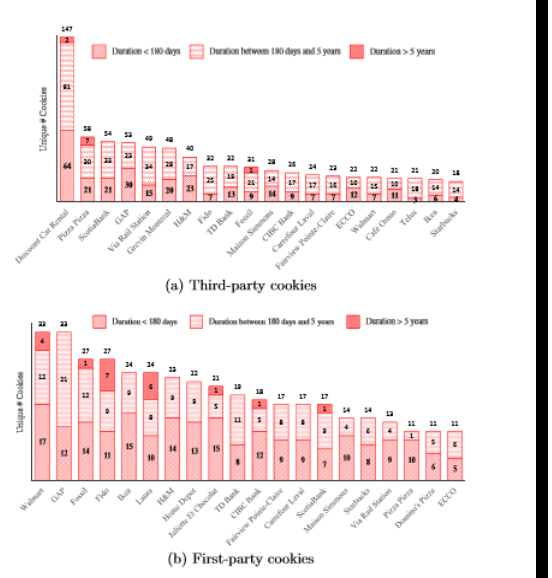


Fig. 4. Number of third-party and first-party cookies on landing pages (top 20). Note that for all reported cookies/domain statistics, we accumulate the distinct cookies as observed in all the datasets collected for a given hotspot.

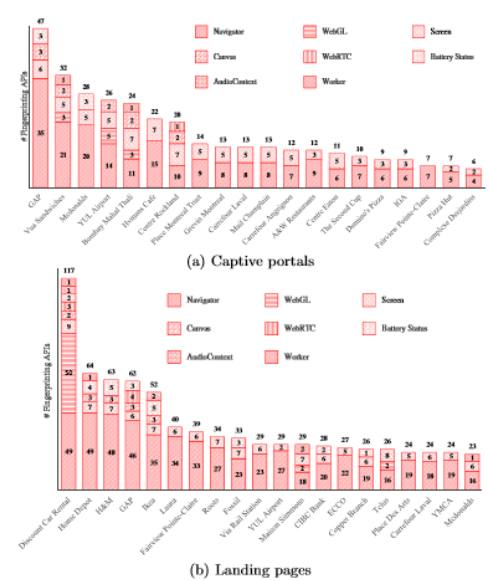


Fig. 5. Number of fingerprinting APIs on captive portals and landing pages (top 20). Note that for all fingerprinting statistics, we accumulate the distinct APIs as observed in all the datasets collected for a given hotspot.

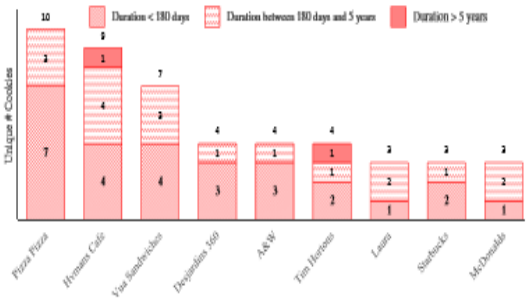


Fig. 6. Number of cookies stored on the Android captive portal app

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

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

Conclusion

1. Privacy sensitive personal data is collected
2. Future tracking is possible due to personal data collection and the use of several web tracking techniques
3. Web tracking starts before approving the privacy policy
4. Effective tracking of Android devices despite Android MAC address randomization
5. 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

Help us collecting data

CPInspectior data collection tool:
https://github.com/MadibaLab/CPInspector

Thank you

Question/Comments?

a_suzan@ciise.concordia.ca

Supporting Slides

Past studies

Characterizing Privacy Leakage of Public WiFi Networks for Users on Travel

Ningning Cheng[1], Xinlei (Oscar) Wang[1], Wei Cheng[1], Prasant Mohapatra[1], Aruna Seneviratne[2]
[1]Department of Computer Science, University of California, Davis, CA, US
[2]National ICT of Australia, ATP, Sydney, Australia
Email: [1]{ninncheng, xlwang}@cs.ucdavis.edu
[1]weicheng@ucdavis.edu [1]prasant@cs.ucdavis.edu [2]Aruna.Seneviratne@nicta.com.au

The continued risks of unsecured public Wi-Fi and why users keep using it: Evidence from Japan

Nissy Sombatruang*, Youki Kadobayashi†, M. Angela Sasse‡, Michelle Baddeley§ and Daisuke Miyamoto†
*Department of Security and Crime Science
University College London, London, UK
E-mail: uctznso@ucl.ac.uk
†Laboratory for Cyber Resilience
Nara Institute of Science and Technology, Nara, Japan
E-mail: {youki-k, daisu-mi}@is.naist.jp
‡Department of Computer Science
University College London, London, UK
E-mail: a.sasse@ucl.ac.uk
§Institute for Choice
University of South Australia, Adelaide, Australia
E-mail: michelle.baddeley@unisa.edu.au

Effectiveness of privacy extensions and private browsing

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

For example, the doubleclick.net cookie is found in 4 captive portals and 30 landing pages, appears 160,508 times in the top 143k Tranco domains (multiple times in some domains)

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

We use OpenWPM between Feb. 28–Mar. 15, 2019 to automatically browse the home pages of the top 143k Tranco domains as of Feb. 27, 2019.

Our recommendations

- Use of VPNs when using public WiFi
- Avoid sharing any personal information with the hotspot (social media or registration forms)
- Use private browsing and possibly some other anti-tracking browser addons
- Clear the browser history after visiting a hotspot if private browsing mode is not used
- Use software programs that may allow to use a fake MAC address on Windows