Technologie Wi-Fi et vie privée

Mathieu Cunche
mathieu.cunche@inria.fr @Cunchem

INSA-Lyon CITI, Inria Privatics

Ecole d'été Rescom - 26 Juin 2015

Privacy

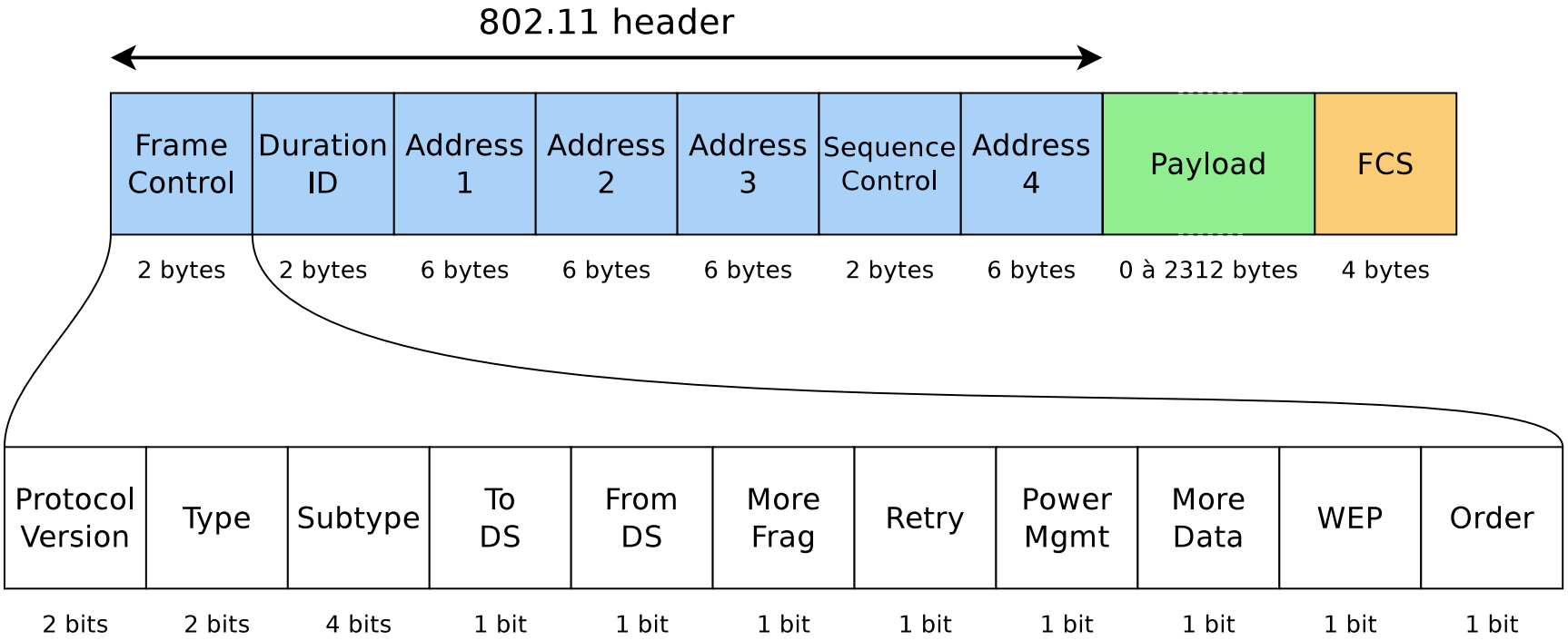
Personally identifiable information (PII)

- Information that can be used on its own or with other information to identify, contact, or locate a single person
- Ex.: Full name, phone number, e-mail address, home address ...

Wi-Fi networking

- IEEE 802.11 standard
  - Specifications for MAC and Physical layers
- Information transmitted by frames
  - **Data**: upper layer datagrams
  - **Management**: beacon, probe request/response, ...
  - **Control**: acknowledgement, ready to send, ...

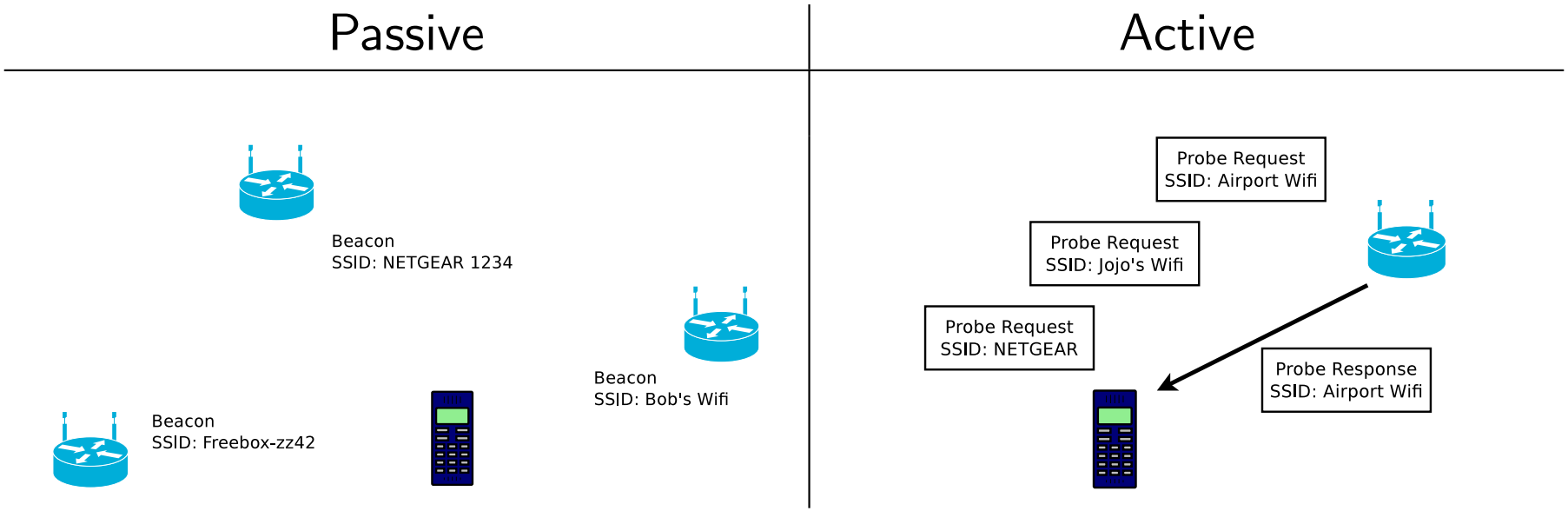
802.11 frame



- Address fields contain MAC addresses (src., dest., ...)
- MAC address: a unique identifier allocated to a network interface

Wi-Fi service discovery I

- Discover surrounding APs and Networks
  - Passive mode: Wi-Fi Beacons
  - Active mode: Probe requests and Probe Responses
  - Probe requests contain an SSID field to specify the searched network
- Active is less costly in energy
  - Preferred mode for mobile devices



Active service discovery

- Probing Frequency: several times per minutes
- Information available in cleartext (headers are not encrypted)
- Broadcast dest. Addr. = `FF:FF:FF:FF:FF:FF`

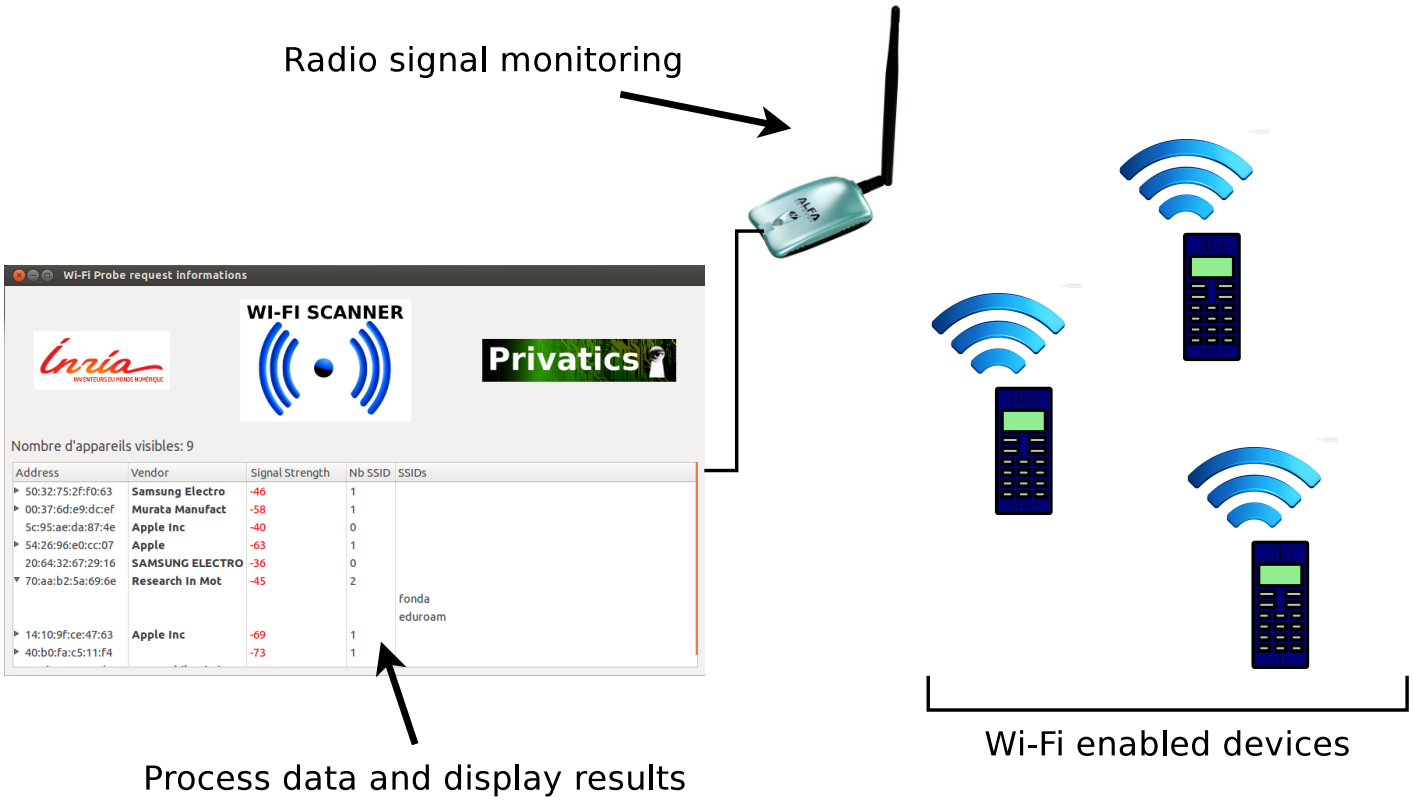
Wi-Fi Fingerprint

Source MAC Address	Destination MAC Address	Signal strength	SSID
00:24:d7:20:4e:45	ff:ff:ff:ff:ff:ff	-70	TECOM-AH4222-561ABC
00:24:d7:20:4e:45	ff:ff:ff:ff:ff:ff	-68	TP-LINK
00:24:d7:20:4e:45	ff:ff:ff:ff:ff:ff	-72	wireless
00:24:d7:20:4e:45	ff:ff:ff:ff:ff:ff	-80	ACCESS-StarHub
00:1f:3b:a2:be:39	ff:ff:ff:ff:ff:ff	-79	A-Company Ltd
00:1f:3b:a2:be:39	ff:ff:ff:ff:ff:ff	-75	Apple Store
00:1f:3b:a2:be:39	ff:ff:ff:ff:ff:ff	-79	dd-wrt
00:19:d2:64:5f:7f	ff:ff:ff:ff:ff:ff	-81	INRIA-guest
00:19:d2:64:5f:7f	ff:ff:ff:ff:ff:ff	-75	INRIA-grenoble
04:46:65:53:8d:ac	ff:ff:ff:ff:ff:ff	-78	A-Company Ltd
04:46:65:53:8d:ac	ff:ff:ff:ff:ff:ff	-77	McDonald's FREE WiFi
04:46:65:53:8d:ac	ff:ff:ff:ff:ff:ff	-74	Cafe_Bello
04:46:65:53:8d:ac	ff:ff:ff:ff:ff:ff	-59	Quality Inn
04:46:65:53:8d:ac	ff:ff:ff:ff:ff:ff	-45	BigPond9568

- Wi-Fi Fingerprint = List of SSIDs broadcast by a device

Monitoring probe requests (Demo.)

- Wi-Fi interface supporting monitoring mode
- Traffic capture and analysis tools

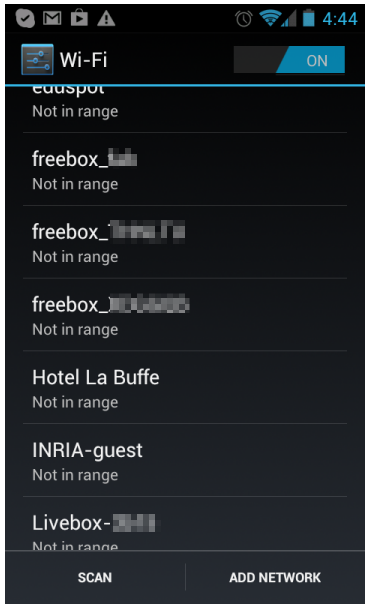


[1]

[1] https://github.com/cunchem/gtk-wifiscanner

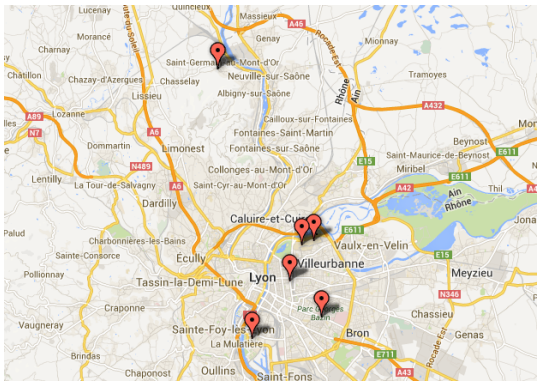
Personal information from SSIDs

- SSIDs: name of the previously connected networks
  - Stored in the Configured Network List (CNL)
  - Observed up to 80 configured networks !

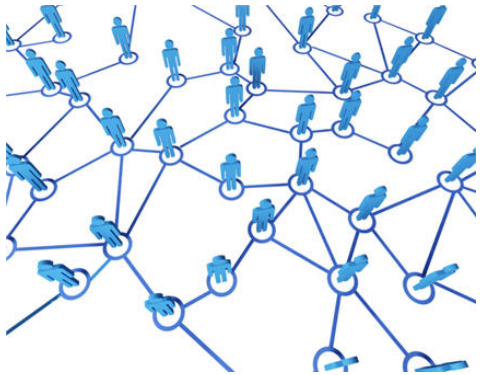


- SSIDs: personal data

Travel history



GPS coordinates

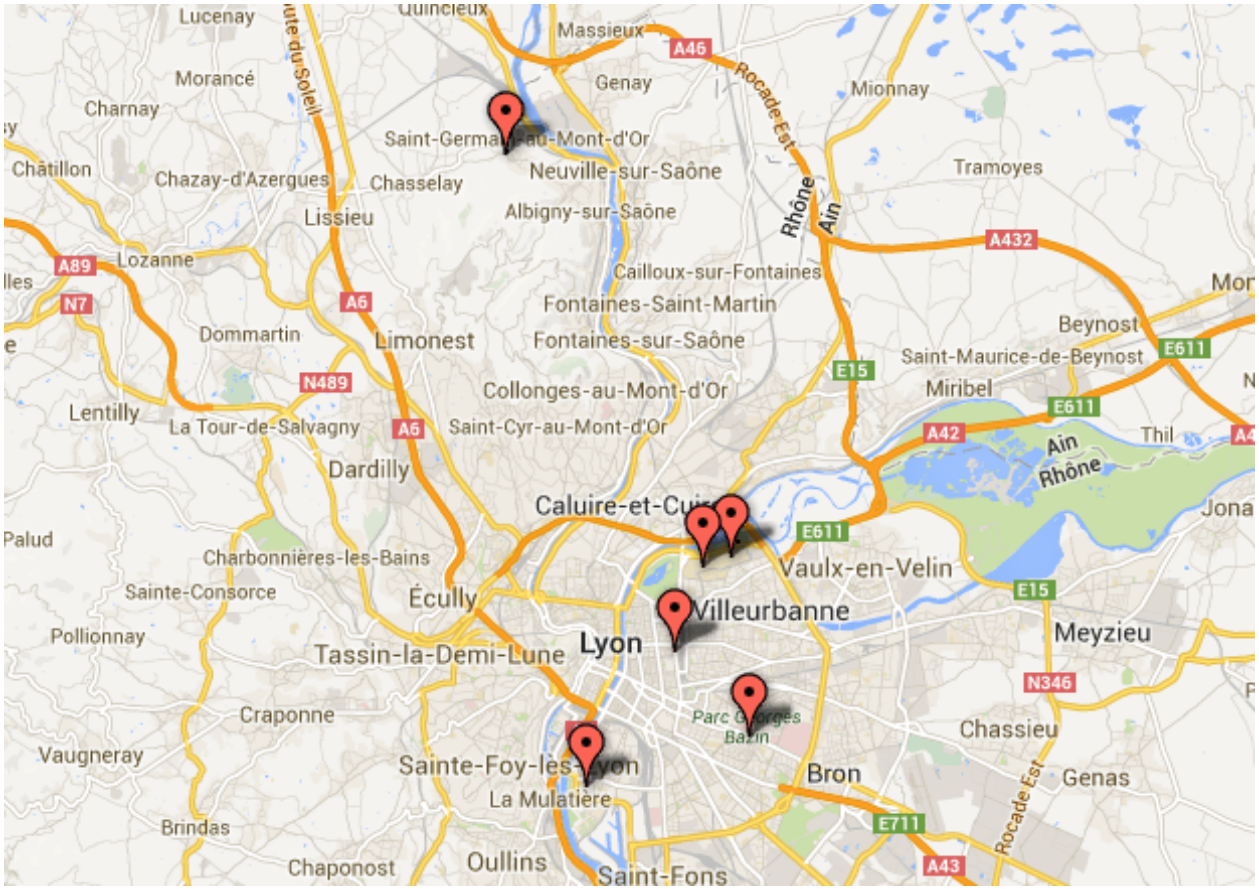


Social links

- Personal information found in SSIDs
  - Company/University/Organization→ INRIA-interne, INSA-INVITE, GlobalCorp Ltd
  - Attended conferences → WiSec14, PETs, CCS
  - Visited places → Hilton-NY WiFi, Aloha Hotel WiFi, Brasserie de l'Est, Sydney-airport-WiFi
  - Individual's identity→ Marc Dupont's iPhone, Bob Fhisher's Network

Precise geolocation information

- From SSIDs to precise geolocation
  - WiGLE database (SSID, BSSID, GPS coord., ...)



Inferring social links I

- Hypothesis: similarity between Wi-Fi fingerprint can betray social links
  - People tends to share their Wi-Fi network with people who are close
- The experiment: "I know who you will meet this evening"[2]
  - A wild dataset: fingerprints of 8000+ devices
  - A control dataset: fingerprint with 30 existing social links

[2] Mathieu Cunche, Mohamed-Ali Kaafar, and Roksana Boreli. "Linking wireless devices using information contained in Wi-Fi probe requests". In: *Pervasive and Mobile Computing* (2013), pp. –.

Inferring social links I

- Quantifying the similarity between fingerprints
  - Metric considering size and rarity of the intersection
- Cosine-IDF and Jaccard index

$$\text{Cosine-idf}(X,Y) = \frac{\sum_{x \in X \cap Y} \text{idf}_x{}^2}{\sqrt{\sum_{x \in X} \text{idf}_x{}^2}\sqrt{\sum_{y \in Y} \text{idf}_y{}^2}} \qquad \text{J}(X,Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

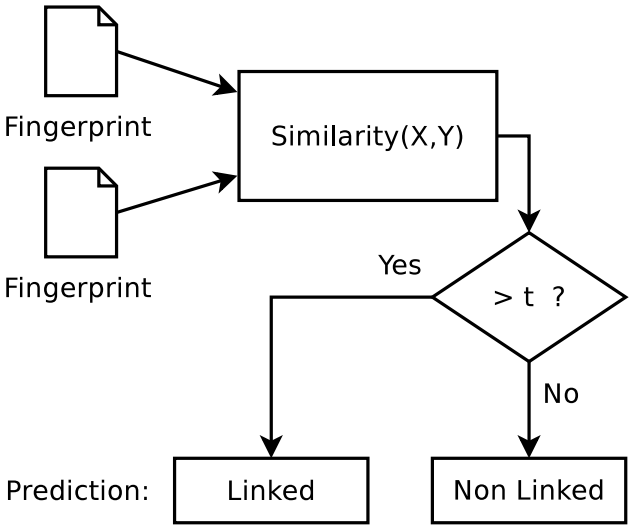
where idf_X : inverse document frequency of x

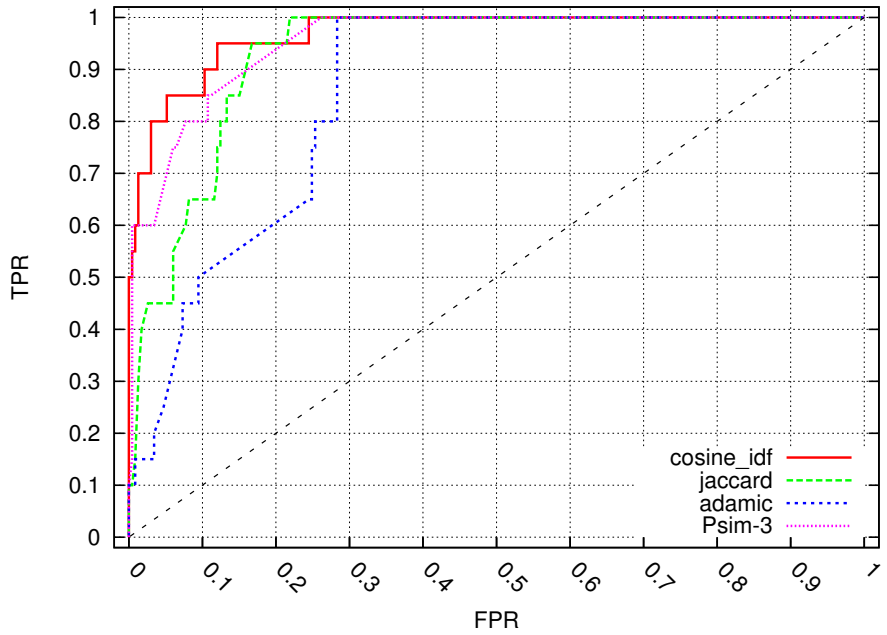
- Adamic, modified Adamic

$$\text{Adamic}(X,Y) = \sum_{x \in X \cap Y} \frac{1}{\log f_x} \qquad \text{Psim-}q(X,Y) = \sum_{x \in X \cap Y} \frac{1}{f_x^q}$$

where f_X : document frequency of x

Inferring social links I





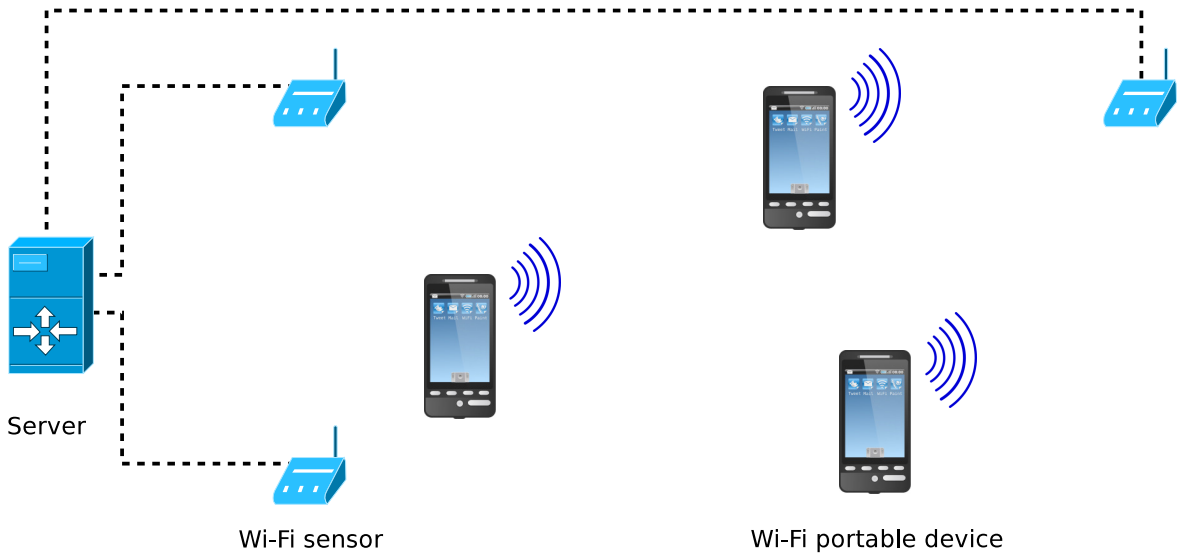
- Performances: detects 80% of social links with less than 8% of error.

The end of broadcast SSIDs

- *NULL* Probe Requests
  - SSID field is left empty
  - AP must responds to all Broadcast Probe Requests
  - Adopted by major vendors to reduce privacy risks
- Hidden Wi-Fi networks
  - Hidden: not broadcasting beacons
  - Probing with SSID is the only way to discover
  - Device continuously broadcast SSID of the network

Wi-Fi tracking

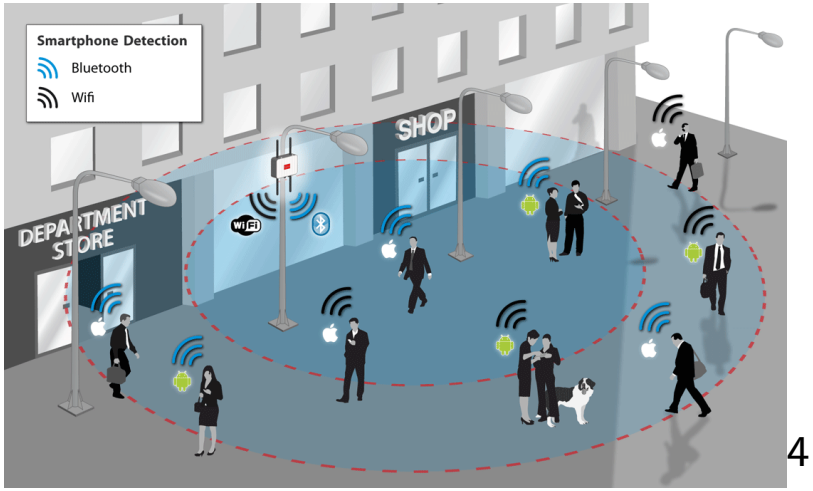
- Wi-Fi enabled smartphone: portable personal beacon
  - Broadcast a unique ID (MAC addr.)
  - Range: several 10s meters
- Wi-Fi tracking system[3]
  - Set of sensors collect Wi-Fi signal
  - Detect and track Wi-Fi devices and their owners



[3] A. B. M. Musa and Jakob Eriksson. "Tracking unmodified smartphones using Wi-Fi monitors". In: *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. 2012.

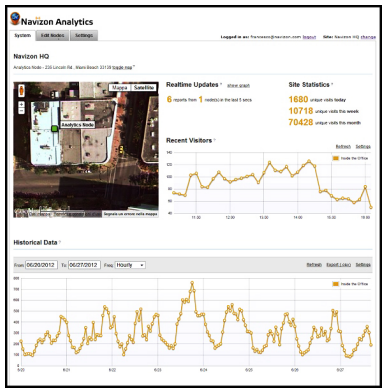
Wi-Fi tracking: applications I

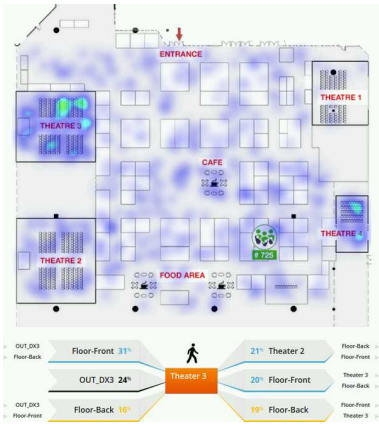
- Shops & shopping center monitoring



4

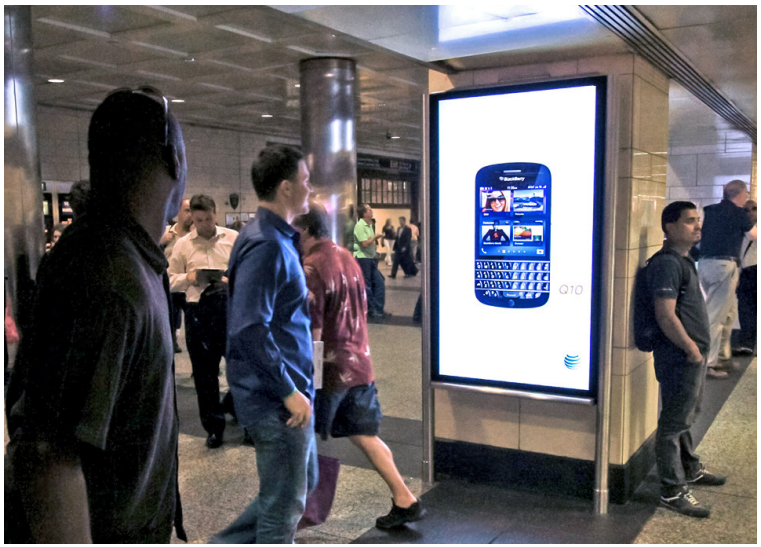
- Physical analytics: Frequency and length of visit, number of visitor,





Wi-Fi tracking: applications II

- Profiling & Targeted advertisement



- Example: London's Wi-Fi bins
  - Detect individuals via Wi-Fi
  - Targeted advertisement displayed on screen
  - Based on a user profile: consuming habits, gender, ...

[4] Source: Euclid Analytics

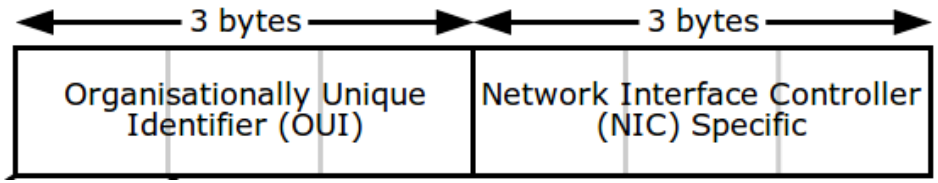
Wi-Fi tracking: privacy

- Privacy concerns

"People have a fundamental right to privacy, and I think neglecting to ask consumers for their permission to track them violates that right" – Senator Al Franken

- Response to privacy concerns
  - User notification & Opt-out mechanisms
  - MAC addr. "does not contain personal information"
  - MAC addr. is "anonymized" (Hash function)

Wi-Fi tracking: privacy



- The MAC address a 48 bits identifier
  - Globally unique identifier allocated to Network Interface
  - Organization Unique Identifier (OUI): 24 left-hand bits
- The MAC address **is** a personal information
  - Unique ID & Personally identifiable information
  - Easy to obtain the MAC addr. of an individual
  - Collected by mobile applications along with other personal information (phone number, email, name, ...)

Wi-Fi tracking: privacy I

- Hash-based anonymization
  - Principle: store the hash of the MAC address instead of the raw value

Time	Location	MAC
12:09	A-4	00:11:11:11:11:11
12:12	B-4	00:11:11:11:11:11
12:13	E-5	00:22:22:22:22:22
12:13	F-4	00:33:33:33:33:33
12:14	B-4	00:11:11:11:11:11

$\rightarrow$

Time	Location	Hash (md5)
12:09	A-4	fb2d5084c0ad1fdf6c29fe2aa323b758
12:12	B-4	fb2d5084c0ad1fdf6c29fe2aa323b758
12:13	E-5	69dc015b56448651561e1a4301ac9b4d
12:13	F-4	07024831442e8b86a06e905fd4d391ce
12:14	B-4	fb2d5084c0ad1fdf6c29fe2aa323b758

  - Motivation: "Hashing is an *Irreversible* operation"
    - Given x, easy to compute $y = H(x)$
    - Given y, hard to find x such as $H(x) = y$

Wi-Fi tracking: privacy II

- Hashed MAC addr. re-identification[5]
  - Test configuration: MD5 + oclhashcatplus + modern GPU (ATI R9 280X)

HASHCAT
Advanced Password Recovery

  - Exhaustive search method
    - Size of the space: 2^{48} values
    - Time: 2.6 days
  - Improved search
    - Only 1% of the space has been allocated
    - Time: 109 seconds

Wi-Fi tracking: privacy III

- Improved search (bis)
  - Wi-Fi devices accounts for a small fraction of OUI
  - Time: 7 seconds to re-identify 99% of Wi-Fi MAC addr.

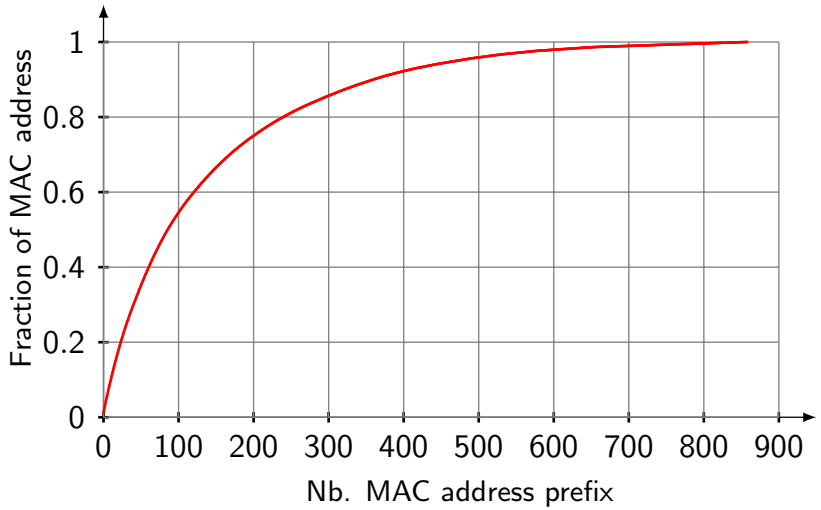
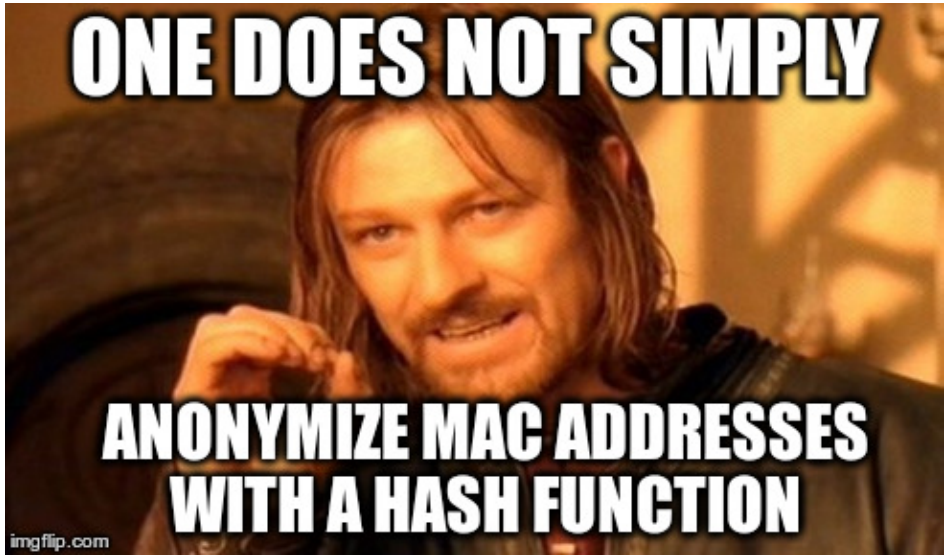


Figure : Cumulative distribution of OUI prefixes in a real world dataset.

Wi-Fi tracking: privacy IV

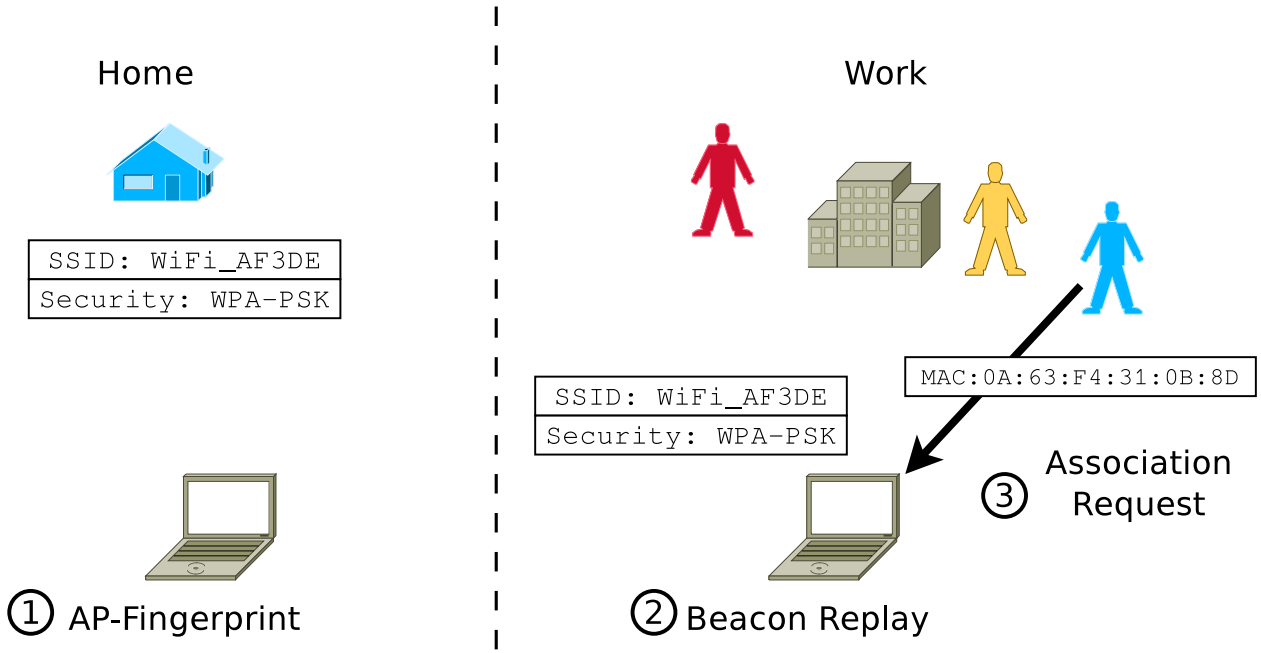


- Simple Hashing does not anonymize MAC addr.
  - Space of origin is too small
  - Exhaustive search is practical
  - Alternate methods are required
    - Loss of information (truncation)
    - Secret salt

[5] Levent Demir, Mathieu Cunche, and Cédric Lauradoux. "Analysing the privacy policies of Wi-Fi trackers". In: *Workshop on Physical Analytics*. Bretton Woods, United States: ACM, June 2014. DOI: 10.1145/2611264.2611266. URL: https://hal.inria.fr/hal-00983363.

Wi-Fi tracking: privacy I

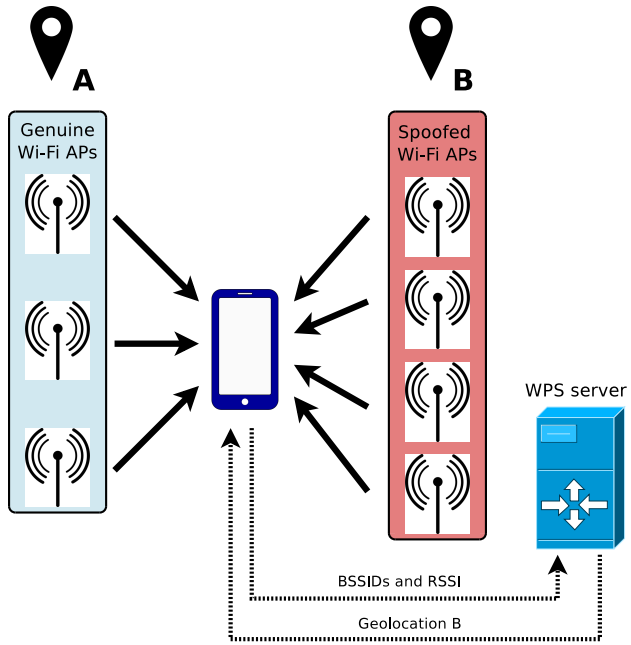
- How to obtain the MAC addr. of an individual ?
  - Without a physical access
- Beacon replay attack[6]
  - Home/work locations uniqueness



[6] Mathieu Cunche. "I know your MAC Address: Targeted tracking of individual using Wi-Fi". In: *International Symposium on Research in Grey-Hat Hacking - GreHack*. Grenoble, France, Nov. 2013.

Wi-Fi tracking: privacy I

- Spoofing Wi-Fi Positioning System (WPS)[7]
  - Spoof WPS location by creating fake Wi-Fi AP
  - Targeted toward a single device (not visible to others)
  - Spoofed geoloc used as sidechannel information for identification on Geotagged platform (Facebook, Twitter, ...)



[7] Célestin Matte, Jagdish Achara, and Mathieu Cunche. "Short: Device-to-Identity Linking Attack Using Targeted Wi-Fi Geolocation Spoofing". In: *Wisec'15*. New York, United States, June 2015.

Wi-Fi tracking

- Surveillance applications
  - MAC addr. used as a selector in NSA's PRISM Framework
  - NSA's ScrapeBear framework
- Hackers' Proof of Concept[8]

CREEPYDOL

[8] Glenn Wilkinson. "Digital Terrestrial Tracking: The Future of Surveillance". In: *Defcon 22* (2014).

Wi-Fi tracking: Botnet of wireless routers I

- Wi-Fi tracking system based on a botnet of wireless routers[9]
  - Suitable features: always powered, connected to the Internet, high quality wireless hardware, ...
  - Simple software modification can turn a Wireless router into a tracking node
    - Proof of Concept with NeufBox V4

  - Wireless routers insecurity: many vulnerabilities, rarely patched, botnets of wireless routers

Wi-Fi tracking: Botnet of wireless routers II

- Simulation of a tracking botnet using a real world dataset
  - Good spatial coverage (especially in urban areas)
  - High tracking potential

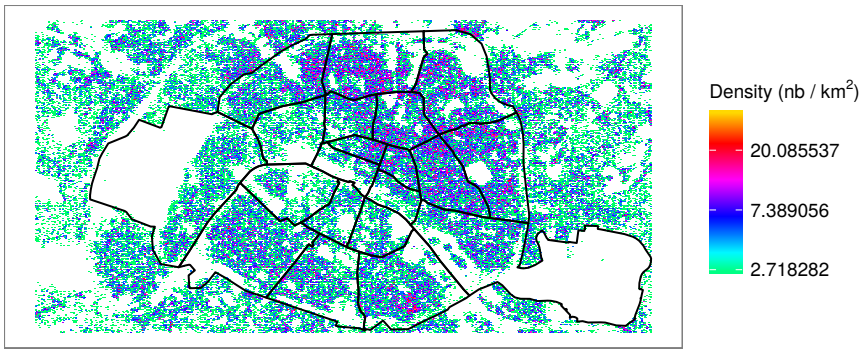


Figure : Density of Freebox in Paris.

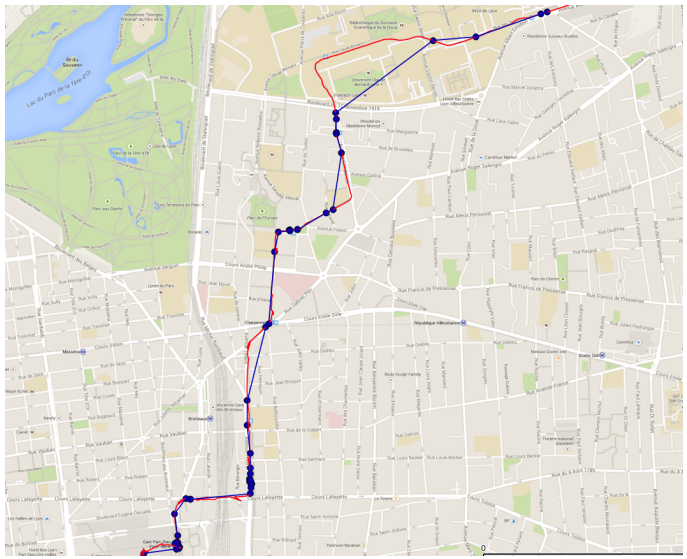


Figure : Trajectory reconstruction with 2% of infected routers.

[9] Pierre Rouveyrol, Patrice Raveneau, and Mathieu Cunche. "Large Scale Wi-Fi tracking using a Botnet of Wireless Routers". In: *Workshop on Surveillance & Technology*. Philadelphie, United States, June 2015. URL: https://hal.inria.fr/hal-01151446.

Countermeasures

- Use Random & Pseudo Random Link Layer identifiers
  - Periodically change MAC address to a random value[10]

- iOS Random MAC address scheme
  - Use new random MAC for each probing burst
  - Only works in very specific configuration (no Data, no Geoloc)
  - Frame sequence number not reseted[11]

[10] Marco Gruteser and Dirk Grunwald. "Enhancing location privacy in wireless LAN through disposable interface identifiers: a quantitative analysis". In: *Mobile Networks and Applications* 10.3 (2005), pp. 315–325.

[11] Julien Freudiger. "Short: How Talkative is your Mobile Device? An Experimental Study of Wi-Fi Probe Requests". In: *Wisec'15*. New York, United States, June 2015.

Countermeasures

- Bluetooth's Resolvable Private Address[12]
  - Requires pairing (shared secret key)
  - Pseudo-random MAC can be resolved iff secret key is known

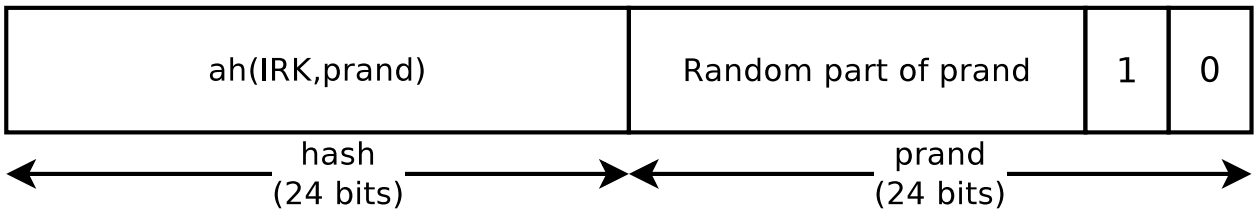


Figure : Resolvable Private Address with shared secret *IK* and hash function *ah*.

[12] *Bluetooth Specification Version 4.2*. Bluetooth SIG. Dec. 2014.

Countermeasures

- Significant modification of the 802.11 protocols[13]
  - Encrypt/obfuscate all identifiers in the 802.11 protocol
  - No backward compatibility
  - Not before several years (decades ?)
- Geofencing
  - Wi-Fi only activated in trusted places (home, office, ...)
  - Apps: Wi-Fi Matic[14] and AVG Privacy Fix[15] (only for Android)

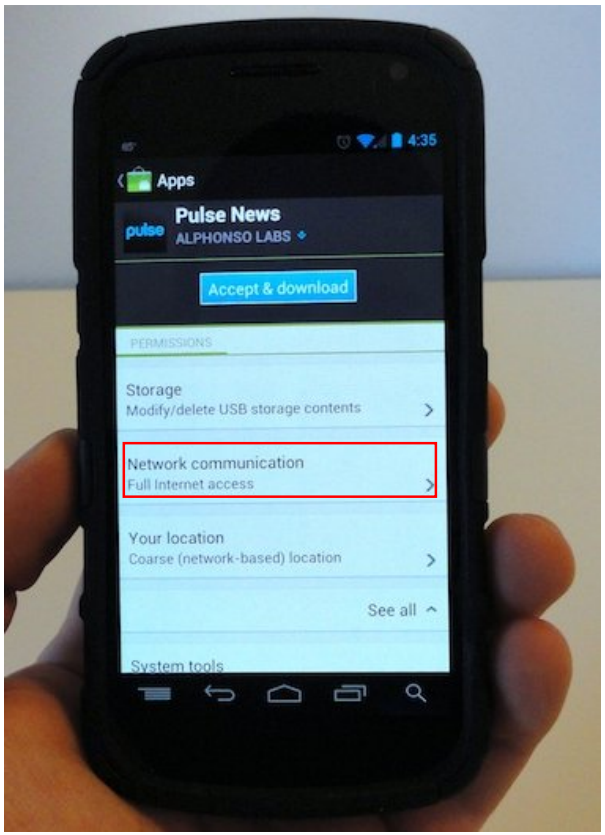
[13] Janne Lindqvist et al. "Privacy-preserving 802.11 access-point discovery". In: WiSec '09. 2009.

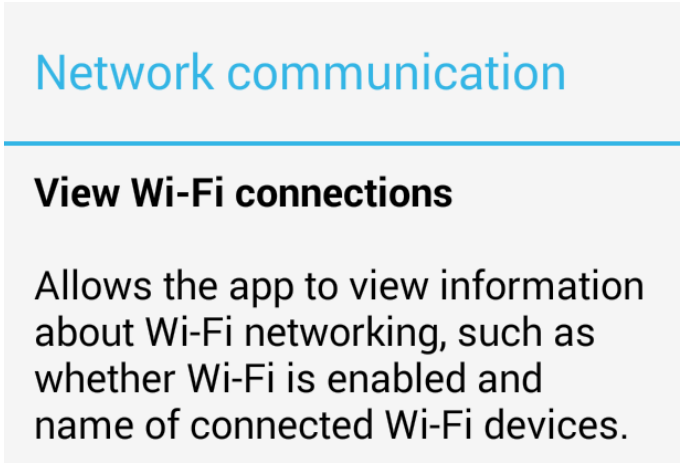
[14] https://play.google.com/store/apps/details?id=org.cprados.wificellmanager

[15] https://play.google.com/store/apps/details?id=com.avg.privacyfix

Mobile applications collecting Wi-Fi data

The ACCESS_WIFI_STATE Android permission





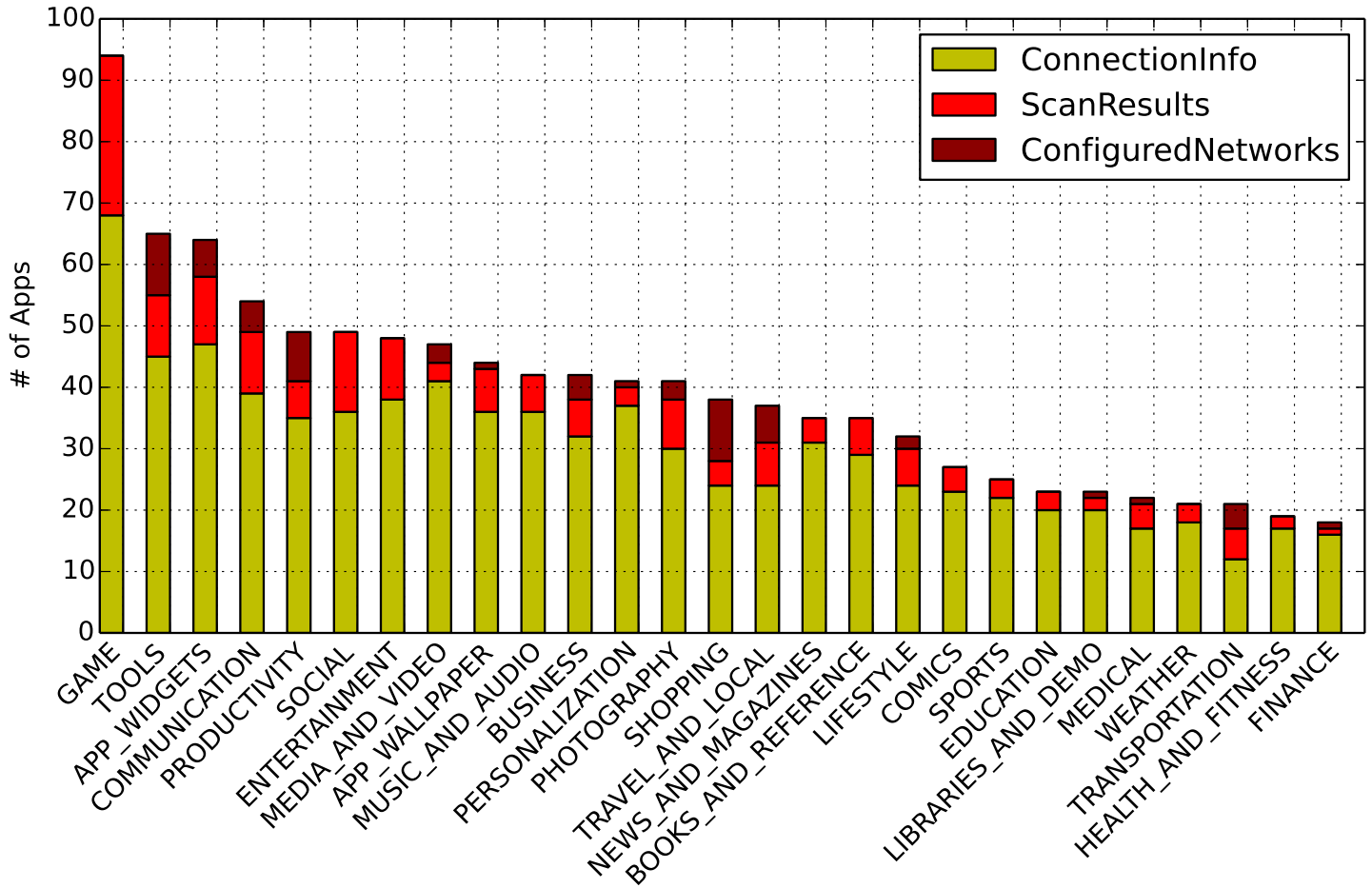
Permission description displayed to a user

- Required to access raw Wi-Fi data
- Protection level : 'Normal'
- Group : 'Network'
- Looks innocuous at first glance!

Mobile applications collecting Wi-Fi data

- Permission analysis through crawling :
  - 2700 Apps (100 * 27 categories)
  - Results: 41% Apps request ACCESS_WIFI_STATE
- Custom tool for static analysis (based on Androguard)
  - Analyses use of various methods of WifiManager class
  - 3 privacy-sensitive methods:
    1. getScanResults(): List of surrounding Wi-Fi APs (Location)
    2. getConnectionInfo(): Connected AP Info + Wi-Fi MAC (Tracking)
    3. getConfiguredNetworks(): SSIDs of previously connected APs (Travel history)

Mobile applications collecting Wi-Fi data



App category wise distribution

Mobile applications collecting Wi-Fi data

- Third-party libraries accessing Wi-Fi data

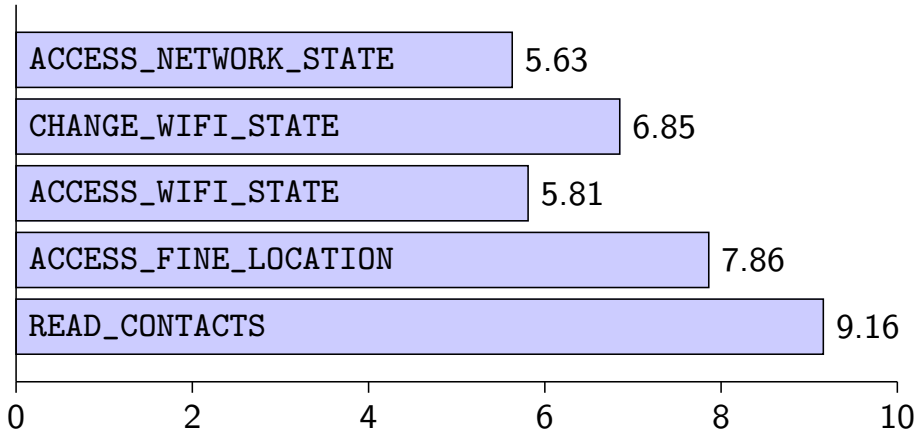
ConnectionInfo		ScanResults		ConfiguredNetworks	
Third-party	**# Apps**	**Third-party**	**# Apps**	**Third-party**	**# Apps**
inmobi.com	74	inmobi.com	9	google.com	10
chartboost.com	55	domob.cn	9	mobiletag.com	4
tapjoy.com	49	mologiq.com	6	lechucksoftware.com	2
vungle.com	47	tencent.com	5	android.com	2
jirbo.com	43	skyhookwireless.com	4	Unibail.com	1

Top 5 third-parties accessing various methods

- Location providers: skyhookwireless.com
- Ads: inmobi.com, tapjoy.com, jirbo.com, mologiq.com, vungle.com
- Game platform: chartboost.com

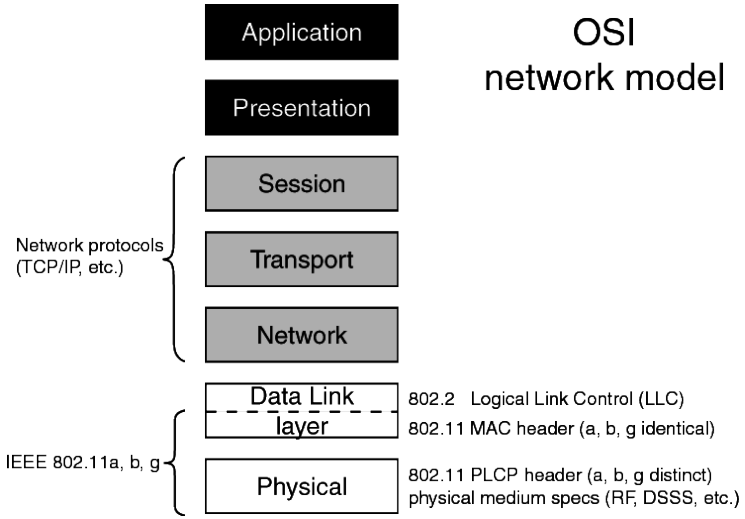
Mobile applications collecting Wi-Fi data

- ACCESS_WIFI_STATE permission: A source of various user PII[16]
- 41% applications request this permission
  - Apps from all categories (including Wallpaper or Comics Apps!)
- Permission exploitation already started:
  - Getting user location without dedicated location permissions
  - Retrieving a unique identifier for tracking purposes
- Privacy implications are not well understood by Android users:



[16] Jagdish Prasad Achara et al. "Short paper: WifiLeaks: underestimated privacy implications of the access_wifi_state android permission". In: *ACM WiSec 2014*.

Conclusion



- Privacy is not restrained to Upper-Layers of the protocol stack
  - Even Link Layer protocols can contain personal information
- Technological legacy
  - Protocol designed in late 1990's
  - Unexpected applications: Wi-Fi in every pocket
  - Security (confidentiality) was considered, but not privacy
  - Difficult to change current standard (backward-compatibility issues)
- Imagination of trackers not to be underestimated
  - Motivated by commercial applications
  - ... or population surveillance & control

Questions ?

Figure : Artist's interpretation[17].

[17] credit P. Treimany