Privacy in WiFi analytics

whitepaper: best practices

author: privacySIG
date: December 7th 2015

Summary

In this whitepaper we discuss the privacy impact on consumers in WiFi tracking; analyzing visitor crowds by collecting WiFi signals of smartphones. In the second part we talk about solutions how this privacy impact can be mitigated. One approach is our privacySIG code-of-conduct, but a more approachable solution would be that every mobile phone vendor implements an opt-out for WiFi tracking in the phone settings. In this way consumers are informed and have an easy opt-out.

Introduction

Analyzing crowd movements using WiFi signals of smartphones has become very popular in recent years. Popular applications are retail (in store movements), traffic, airport, and events. In fact the technology can be used for every application where people come together. The key technology enabler is that smartphones search from time to time for nearby WLAN networks. In most cases this is multiple times per minute, but depends on the type of device and the software it is running. These mobile devices do this scanning for different purposes such as a low-energy GPS alternative or cellular data-offloading i.e. reduce the network load on cellular networks.

Photo by Ian Muttoo licensed under [Creative Commons Attribution 2.0 Generic.]

When a smartphone scans for nearby networks, an unique hardware identifier is included in the scan, which will be picked up by nearby WiFi equipment. This identifier is called a MAC address. As the MAC address is always the same, it allows to analyze the behavior of an individual smartphone. In this way visit recurrence, visit duration etc. can be determined. On the other hand a MAC address could be considered as indirect personal information because users keep their for smartphone for some time. This means that when the phone is detected, it is quite likely that the user is also present. It should be noted that the MAC in itself does not contain any information except being unique (e.g. no mobile number, IMEI, IMSI etc can be deduced from it).

The purpose of this article is to analyze the privacy impact of WiFi analytics using the MAC address as identifier and give guidelines to vendors to implement measures to protect the privacy of smartphone users. **In our opinion the choice for tracking or not, should be controlled by the end-user, the user of a smartphone.** The outline of this article is as follows, first background information is given about the MAC address. This is followed by a section by measures how privacy of consumers can be protected. In the last part of this white paper best practices are given.

MAC address

What is it?

A MAC address stands for Media Access Control address. It is an unique hardware address that is needed to share a common medium between multiple users. This medium can be a cable, optical cable or in this case a wireless channel. Without an unique hardware address you can't send information from user A to user B. On top of the MAC layer higher protocol are used e.g. TCP/IP protocol. A MAC address should therefore not be confused with an IP address. When your laptop receives an IP address of your home router (DHCP), it uses its MAC address behind the scenes to set it up.

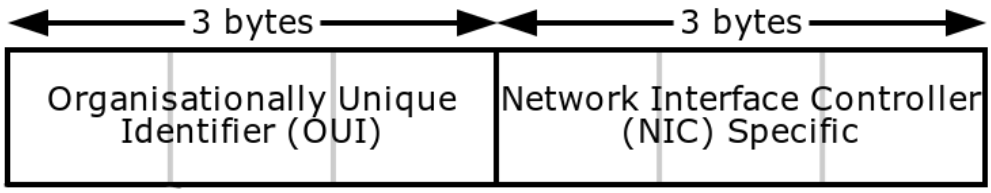
Where is it used?

MAC addresses are used within the IEEE standards. IEEE is an organization that among others manages and develop communication standards. MAC addresses are used within the IEEE 802 family of standards that includes popular standards like Ethernet, Bluetooth and WLAN. As it is the hardware identifier where on top the whole TCP/IP communication is constructed, it is important to avoid collisions. Meaning in a network segment where there are hardware devices with the same MAC address, there is no reliable communication possible. As one does not know beforehand where network equipment will be used, vendors often assign an unique MAC address to each device and network interface.

How is it constructed?

A MAC address is a 48-bit hardware identifier. It consists of two major parts, the OUI field, highest 24-bit. OUI stands for Organizationally unique identifier. This means that every organization, vendor has its own block of MAC addresses. The lower 24-bit in this block can be used by the vendor to assign it to individual devices[1]. In the 24-bit OUI field there are also some special bits, making the OUI field effective 22-bit.

[1] Simplified description, as organisation can apply for smaller blocks. The reader is referred to this link.



MAC address structure from Wikipedia

One of the special bits that is relevant here, is the local bit. If this bit is set, the MAC address is considered as local and not officially registered. One can make the analogy with local IP addresses versus public IP addresses. Public IP addresses are addresses that are managed centrally and companies need to buy a single or block of IP address. The same IP address can't be used twice. On the other hand local IP addresses are addresses used by your home or office router e.g. 192.168.x.x or 10.10.x.x range. These address are local and can't be accessed from outside. Local MAC addresses have a similar function. A local MAC address can be used for instance for setting up additional virtual network adapters.

How can companies apply for a MAC address?

In the above section, the technical details of a MAC address have been described. Here we discuss how companies can apply for a MAC address. The procedure is quite simple and straightforward. The IEEE organization manages a public central database of registered OUI blocks. Companies can ask for a new block when the company states that more than 95% of its current OUI blocks have been used. When an organization applies for a new block, IEEE hands out a new block and adds this to the public database including the company name. Companies do have the option to hide their name in the public listing for an additional fee.

It also means that there is **no central database of individual MAC addresses**, only of OUI blocks. It is up to the vendor to maintain its own database of used MAC addresses.

How should companies use their assigned block of MAC addresses?

As stated before it is up to the vendor how to assign a block of MAC addresses to individual devices. It makes sense for vendors to use one unique MAC address per device (per interface); so you don't run into problems that two devices connect at the same time to the same network

with the same MAC address. In that case you can't set up a reliable link as the communication link can't distinguish between the two devices. Therefore both devices will receive information that is not intended for them, which results in neither being able to communicate reliably.

IEEE does not provide guidelines or requirements to a company that acquired an OUI block. It is up to the company how to assign the block of MAC addresses to individual hardware devices. Large organization tend to use MAC filtering as an authorization tool to whitelist company equipment to their networks. Also the hardware manufacturer does not know where the devices will be used. So from practical point of view, it is easiest to keep using the same MAC address for a particular device and have them unique.

What are legal aspects?

The legal aspects of a MAC address in Europe and therefore WiFi analytics, is that a MAC address is considered as (indirect) personal information, also called pseudonymous data.

pseudonymous data: *any data that could reasonably be associated with a particular consumer or a particular consumer's property, such as a smartphone or other device, or any other unique identifier.*

The MAC address itself is not personal information, as it is just a number. The attached geo-information attached to the MAC address reveals most personal information e.g. when and where is user X seen. For that reason we have set up a Code of Conduct within the privacySIG.

In the new EU privacy law, the General Data Protection Regulation (GDPR), it is allowed to collect pseudonymous data without the prior consent of the consumer. The EU aims that GDPR is adopted in 2017 by member states. Although the contents of this new privacy law are draft, it is very unlikely that the view on pseudonymous data will be changed in the final version.

How to protect consumer privacy

Is privacy implemented in the WiFi standard?

One of the cornerstones in the standard has always been to be fully backwards compatible. This means that equipment which support newer, faster, data rates always should have support for older versions of the standard too. So a consumer does not have to replace its router when he or she buys a new laptop. During the start of the standard, privacy was not a topic of interest and has not been taken into consideration. For that reasons always the same MAC address is used and no temporarily MAC addresses like with 3G/4G.

To protect privacy, several measures have been taken by different stakeholders.

1. The European privacySIG has implemented a Code of Conduct that is compulsory for all members. It includes minimum data collection, storage and pseudomising MAC addresses.
2. Some vendors started to implement measure to randomize the MAC address, the most well known example of this is Apple.
3. There are apps on the market that allow consumers to switch off WiFi automatically when they are outside their homes.

What are the benefits of WiFi analytics?

In every place where people come together, there are benefits from being able to see patterns and trends in the flows of people. Popular examples are retail shops, malls, streets, stations and events. In this analysis individual consumers are not important, **only aggregated numbers** like visit recurrence, duration, number of visitors are relevant KPI (Key Performance Indicators).

The aggregated numbers are needed to make management decisions like having enough staff members during busy hours. Optimize the layout of your shop to have better conversion/sales. Or prevent that train stations become too crowded and steer the crowds. In all these cases it is relevant to have better understanding of the crowd.

Why not use iBeacons?

Apple has developed a technology for location based service called iBeacons. The underlying technology is Bluetooth Low Energy. Similar solutions have been implemented for Android. However, for tracking visitors this solution is not viable for several reasons. First off all the technology is platform dependent. You need to provide a solution for each mobile platform e.g. iOS and Android. Second the consumer needs to install an app as this technology works the other way around compared to WiFi analytics; your smartphone looks for nearby iBeacons, collects the data and then sends this information back to the cloud of analysis. Due to the low penetration of such apps, the capture rate is very low compared to WiFi analytics.

How does Apple randomize MAC addresses in iOS devices?

Starting from iOS8 (September 2014), Apple has implemented a feature to randomize the MAC address in their devices to enhance privacy for consumers against WiFi analytics. With the introduction of iOS9 this feature has been completed. Basically, newer iOS devices that are not connected to any network, will randomize its MAC address when probing for nearby WiFi networks. This temporarily MAC address is changed regularly.

Although the randomization of the MAC address seems to be a good privacy-enhancing feature, it also has several drawbacks. First of all, limited usability. It only randomizes the MAC address in cases when it is not connected to any network. In cases an iOS device is connected to a network, it has to use its real MAC address due to limitations in the WiFi technology. Also the iOS feature can cause degraded WiFi performance. One can compare this with CD copy protection measures. Not all WiFi equipment can cope with a load of extra random MAC addresses.

What are the pros and cons for consumers?

For the average consumer WiFi tracking it is quite difficult to understand and in the media WiFi tracking is often depicted as a big brother tool. This is an incorrect perception by the media. WiFi tracking is similar to cookies on the internet. Just an unique number given by your phone that allow companies to study consumer trends.

So therefore two things are important. First consumers need to be educated on what WiFi tracking is and the impact on their privacy. We, as privacySIG, do our best to contribute to this discussion. Second it should be made easy for consumers to opt out of WiFi analytics. For that one uniform opt-out method should be used. Most obvious way would be to have an setting in the smartphone where a consumer can opt-out of WiFi analytics.

Until an agreement between device manufacturers and analytics companies can be found, there will be several disadvantages.

1. There is an ongoing technology race between vendors and WiFi tracking companies.
2. The end user experience is affected negatively due to unexpected side effects when companies use non standard methods to prevent WiFi analytics from working.
3. Alternative solutions, like iBeacons require a separate dedicated infrastructure which cannot be used for anything else.

4. The ability for users to opt-out is dependent on the choices the device manufacturer has decided on, he/she cannot make their own decision.

What does the privacySIG do to protect your privacy?

All members of the privacySIG have agreed on a compulsory Code of Conduct. Important topics are pseudonymisation of your MAC address, minimum data collection, secure data communication & storage, limited storage of base data and offering an opt-out to consumers.

This means that all members will pseudonymise your MAC address immediately after collecting unique identifiers. Original MAC addresses are discarded and won't be used in further processing. Next member only perform minimum data collection. Typically this is only your (pseudonymised) MAC address, time stamp and signal strength (to know how far you are away from a sensor). We don't collect any other user data. In addition access to sensors, all communication and storage of data is secured. This means that unauthorized persons won't have access to the collected data. Also privacySIG members don't store base data (collected data) longer than necessary for the business case. Finally we offer an opt-out for consumers who don't want to be tracked. This opt-out is used by all members.

What is the best strategy to protect privacy?

In our view the best strategy is to have an option in the settings menu of every smartphone. This would toggle a identifier in the WiFi scan which indicates if the user wishes to take part or not in WiFi analytics, similar to how the Do-Not-Track header is used for browsing. WiFi analytics could be combined with this setting which is already available in several operating systems. If in addition this setting is highlighted when a phone is taken in to use (“get started options”), consumers get informed and can make their preference.