Embedded System Security
MAC address anonymization

Mickael Kovel
000396950

December 20, 2024

Abstract

This project examines the privacy risks associated with MAC addresses, focusing on their use in tracking devices. It explores the legal frameworks surrounding data protection, particularly the GDPR, and the European Convention on Human Rights. Additionally, the project discusses various anonymization techniques such as truncation, hashing, and encryption, highlighting their effectiveness in protecting personal data in IoT environments.

Contents

1 Introduction

The increasing reliance on wireless networks and IoT devices has led to a growing concern about the privacy of individuals. Among the various identifiers used in such networks, Media Access Control (MAC) addresses stand out due to their unique, hardware-assigned nature. While MAC addresses play a crucial role in enabling communication within local networks, they also pose potential risks when exploited for tracking and surveillance purposes. The collection and processing of these identifiers raise significant privacy concerns, especially when they are linked to individuals or can be used to track their movements. This project delves into the role of MAC addresses in modern networks, examining the legal frameworks governing their use, the privacy risks associated with their exposure, and the methods employed to anonymize them. By analyzing key legislation, including the General Data Protection Regulation (GDPR) and various international privacy standards, the project aims to provide a comprehensive understanding of the challenges and solutions related to MAC address anonymization, particularly in embedded systems.

2 What is a MAC address?

In the context of networking, we need to identify devices on a network. To exchange information between two devices, we point to an identifier. In the context of the internet, we use IP addresses to identify devices. However, IP addresses are dynamic and can change each time the device connects to the network. This is why we use MAC addresses to uniquely identify devices, in the same way that we use postal addresses to identify houses and send mail.

We can define a MAC address as a unique identifier assigned to a network interface controller (NIC) for use as a network address in communications within a network segment.

2.1 How is a MAC address provided?

MAC addresses are provided by the manufacturer of the network interface controller (NIC). They are typically assigned at the factory and are hardcoded into the hardware. They cannot be changed by the user and are globally unique.

2.1.1 What is the structure of a MAC address?

A MAC address is a 48-bit number that is typically represented as a 12-digit hexadecimal number. The first 24 bits (6 digits) are the Organizationally Unique Identifier (OUI), which identifies the manufacturer of the NIC. The last 24 bits (6 digits) are the device identifier, which is assigned by the manufacturer. Th OUI is assigned by the Institute of Electrical and Electronics Engineers (IEEE) and is unique to each manufacturer. The device identifier is unique to each device and is assigned by the manufacturer.

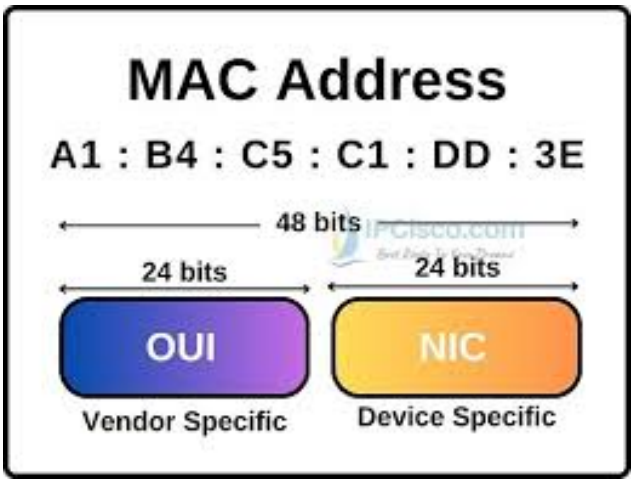

Figure 1: MAC address structure [1]

2.2 What are the privacy implications of a MAC address?

2.3 Where are MAC addresses used?

2.3.1 Wi-Fi - Ethernet

In the context of computer networks, MAC addresses are used to identify devices on a local network. They are linked to the IP address of a device and are used to route data packets to the correct destination. The Address Resolution Protocol (ARP) is used to associate an IP address with a MAC address.

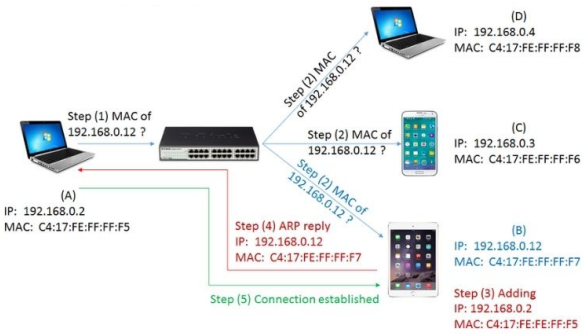

Figure 2: ARP protocol [2]

We can see in the figure 2 that the MAC address is used to identify the device in the network.

2.3.2 Bluetooth

Bluetooth devices do not use MAC addresses in the same way as Wi-Fi devices. Instead, they use a 48-bit Bluetooth device address (BD_ADDR) that is similar to a MAC address. BD_ADDRs are also globally unique and could be used to track devices in the same way as MAC addresses.

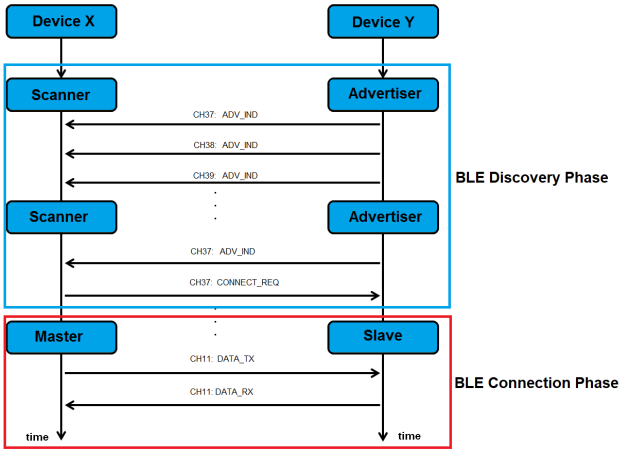

Figure 3: BTLE Link Layer [3]

We can see in the figure 3 that even during the discovery phase, there is exchange in both directions, which means that the BT_ADDR have already been exchanged before even connecting to the device.

3 What does the law say ?

3.1 GDPR

"The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes." [4]

3.2 European Convention on Human Rights

34. "Individual applications The Court may receive applications from any person, non- governmental organisation or group of individuals claiming to be the victim of a violation by one of the High Contracting Parties of the rights set forth in the Convention or the Protocols thereto. The High Contracting Parties undertake not to hinder in any way the effective exercise of this right."

35. "Admissibility criteria

1. [...]
2. The Court shall not deal with any application submitted under Article 34 that
   (a) is anonymous; or [...]" [5]

3.3 Convention 108+

18. "The notion of "identifiable" refers not only to the individual's civil or legal identity as such, but also to what may allow to "individualise" or single out (and thus allow to treat differently) one person from others. This "individualisation" could be done, for instance, by referring to him or her specifically, or to a device or a combination of devices (computer, mobile phone, camera, gaming devices, etc.) on the basis of an iden- tification number, a pseudonym, biometric or genetic data, location data, an IP address, or other identifier. The use of a pseudonym or of any digital identifier/ digital identity does not lead to anonymisation of the data as the data subject can still be identifiable or individualised. Pseudonymous data is thus to be considered as personal data and is covered by the provisions of the Convention. The quality of the pseud- onymisation techniques applied should be duly taken into account when assessing the appropriateness of safeguards implemented to mitigate the risks to data subjects."

19. "Data is to be considered as anonymous only as long as it is impossible to re-identify the data subject or if such re-identification would require unreasonable time, effort or resources, taking into consideration the available technology at the time of the processing and technological developments. Data that appears to be anonymous because it is not accompanied by any obvious identifying element may, nevertheless in particular cases (not requiring unreasonable time, effort or resources), permit the identification of an individual. This is the case, for example, where it is possible for the controller or any person to identify the individual through the combination of different types of data, such as physical, physiological, genetic, economic, or social data (combination of data on the age, sex, occupation, geolocation, family status, etc.). Where this is the case, the data may not be considered anonymous and is covered by the provisions of the Convention."

20. "When data is made anonymous, appropriate means should be put in place to avoid re-identification of data subjects, in particular, all technical means should be implemented in order to guarantee that the individual is not, or is no longer, identifiable. They should be regularly re-evaluated in light of the fast pace of technological development."[6]

3.4 Belgian Law

Articles 101, 134, and 164 of the Belgian Act of 30 July 2018 on the protection of individuals with regard to the processing of personal data stipulate that the personal data referred to in Articles 99, 132, and 162 must be anonymized before they can be accessed. These articles primarily concern the processing of personal data for historical, scientific, or statistical purposes.

99. "outlines conditions under which personal data from intelligence and security services can be consulted for such purposes. The consultation is authorized only if it does not conflict with the service's mandate, obligations under the Act of 30 November 1998, ongoing investigations, or international relations. Any request for further processing of such data for other purposes will be refused unless deemed legitimate by the concerned service."

132. "allows the consultation of personal data held by authorities or appeal boards for historical, scientific, or statistical purposes. However, it is conditional on ensuring that it does not harm any interests protected under the Act of 11 December 1998, particularly regarding the confidentiality and protection of personal data."

162. "similarly authorizes the consultation of personal data from CUTA (the Coordination Unit for Threat Analysis) for historical, scientific, or statistical purposes. Again, this is contingent on ensuring no harm to CUTA's mandate, ongoing investigations, or international relations. Any requests for further processing of such data for different purposes will be denied unless the processing is considered legitimate and does not interfere with the protected interests."

[7]

4 Who can access a MAC address?

In fact, anyone can access a MAC address. It is transmitted in clear text over the network and can be easily intercepted by anyone with the right tools. The privacy risks associated with MAC addresses are not limited to a specific group of people or organizations. However, the real issue arises when a specific location is associated with a MAC address at a particular time, such as when connecting to a Wi-Fi network. The problem becomes even more serious when this data is recorded, as recording real-time location information can enable tracking of an individual over an extended period.

The critical question then becomes: who would have an interest in recording this data, and why?

4.1 Internet Service Provider

Internet Service Providers (ISPs) can access MAC addresses when users connect to their network. By definition, they have access to all the data that passes through their network, including MAC addresses. This data can be used for various purposes, such as network management, troubleshooting, and targeted advertising. However, ISPs are subject to data protection regulations, and the use of MAC addresses for tracking or profiling purposes is generally prohibited.

4.2 Audience Measurement

Because of the probes and probe requests, MAC addresses can be collected by anyone with a Wi-Fi-enabled device. This includes audience measurement companies, which collect data on the behavior of users in public spaces. So audience measurement companies can access MAC addresses when users walk by their Wi-Fi sensors. This data can be used to analyze foot traffic, measure the effectiveness of advertising campaigns, and provide insights to retailers and other businesses. However, the use of MAC addresses for audience measurement is subject to data protection regulations, and the data must be anonymized to protect the privacy of users.

5 What are the risks related to MAC address?

5.1 Tracking

MAC addresses, unique to each device, allow for continuous monitoring when the device interacts with Wi-Fi access points, Bluetooth beacons, or even passive network equipment. These interactions, which can include connection or discovery requests, facilitate the collection of information about users' movements in physical locations such as airports, stores, or streets. For example, crowd management systems use this data to analyze visitor flows, but this information can also be used for less ethical purposes, such as targeted surveillance. This raises major privacy concerns.

5.2 Profiling

Data collected via MAC addresses can be correlated with other information, such as connection times, precise locations, or usage patterns. This information can be used to create detailed user profiles, revealing their personal habits, daily routines, or even consumption preferences. Companies often exploit this data for targeted advertising campaigns, while malicious actors could use it for social engineering attacks. This profiling, although often invisible to users, constitutes a significant intrusion into their privacy.

5.3 Re-identification

Even when MAC addresses are randomized (e.g., through built-in anonymization mechanisms as in iOS or Android), they can often be re-identified by combining different data sources. Researchers have shown that specific usage patterns, such as connection duration or frequency, can link a temporary MAC address to its original address. This seriously compromises anonymization measures and exposes users to risks, especially when stolen or poorly secured databases contain this information.

5.4 Spoofing

MAC address spoofing is a common technique that allows attackers to falsify their network identity. This can be used to bypass MAC address access lists (as in protected WiFi networks) or to steal a device's identity to access restricted resources. This type of attack is particularly problematic in environments where critical systems rely on MAC addresses for their operation, such as Internet of Things (IoT) systems. It illustrates the limitations of using MAC addresses as secure identifiers.

6 How to anonymize a MAC address?

6.1 Truncation

Truncation involves shortening the MAC address by keeping only part of it. For example, you could keep only the first 6 octets (which typically correspond to the manufacturer), or a random portion of the address. This method can keep some of the information while minimizing the link between the MAC address and an individual.

- Advantages:
  - Simple to implement: The method is easy to implement and quick.
  - Less risk of re-identification, as the remaining portion of the MAC address does not contain enough information to be associated with a particular person or device.
- Disadvantages:
  - Loss of information: Relevant data may be lost if a finer identification of the MAC address is needed.
  - May not be sufficient for higher levels of anonymity, as the remaining part of the address can still be used to trace devices over long periods.

6.2 Hashing

Hashing involves applying a mathematical function that takes the MAC address as input and generates a fixed-size output, called a "hash." This process is irreversible, meaning that it is not possible to recover the original MAC address from its hash. One popular technique to enhance the security of hashing is the addition of salt.

Why use salt: Salt is a random value added to the MAC address before hashing. This protects against rainbow table attacks, where an attacker would have pre-computed hashes of different possible MAC addresses. The salt makes each hash unique, even if two users have the same MAC address.

Relevance to your problem: Hashing with salt is useful for anonymizing MAC addresses while providing some security. When collecting data, you can anonymize the MAC address without risking deducing its origin, addressing legal concerns regarding privacy.

- Advantages:
  - Irreversible: Once hashed, the MAC address cannot be recovered.
  - Protection against pre-computation attacks (rainbow table) with the use of salt.
  - Easy to implement in systems where access to the original MAC address is not needed after anonymization.
- Disadvantages:
  - Loss of original integrity: Once hashed, the original MAC address is completely lost.
  - If you need to recover the original MAC address for legitimate reasons (e.g., auditing or re-identification in specific cases), this can be a problem.

6.3 Encryption

Encryption involves encrypting the MAC address using a symmetric or asymmetric encryption algorithm. In cases where you need to recover the original MAC address for legal or administrative purposes, encryption can be used, as it allows data to be recovered using a decryption key.

Recommended encryption scheme: DHIES: The Diffie-Hellman Integrated Encryption Scheme (DHIES) is a hybrid encryption method that combines Diffie-Hellman key exchange and symmetric encryption (e.g., AES). This scheme is used in systems where data confidentiality is crucial while allowing controlled recovery.

Why use DHIES: This scheme offers a very high level of security by using a secure key exchange method (Diffie-Hellman) to establish a secure communication channel, then using a symmetric encryption algorithm to protect the data. It is particularly suitable when access to the original data needs to be limited and controlled.

- Advantages:
  - High security: Data is protected very robustly, with a double layer of security.
  - Allows recovery of the original MAC address with the correct decryption key.
  - Ideal in a context where anonymization needs to be reversible under strict conditions.
- Disadvantages:
  - Complexity: The implementation process is more complex than other methods.
  - Key management: System security relies on proper management of decryption keys, which can be challenging in large-scale systems.
  - Performance: Encryption and decryption can introduce overhead in terms of performance, particularly in large-scale systems.

7 Notable incidents

7.1 Privacy violation

As part of enhancing customer experience and optimizing operations, Nordstrom has implemented technology to track customer movements in its stores through their Wi-Fi connections. This anonymous technology aims to analyze in-store traffic, adjust staffing levels, and rethink department layouts. Since October 2012, sensors in stores collect information on the time customers spend in departments, without following them across sections or identifying personal details. Euclid has experienced rapid growth, increasing by 11,949% in 2012. The service helps retailers analyze customer flow, conversion rates, and optimize workforce management based on traffic patterns. Retailers can adjust staffing levels to maximize sales during peak times and reduce costs during slower periods.

However, after testing the technology in select stores, Nordstrom announced in May 2013 that it would no longer use Euclid, stating that the trial had concluded and that customer feedback would be incorporated into future innovations.[8]

7.2 Google Street View

Since 2007, Google Street View cars, in addition to photographing streets, have also collected information from domestic Wi-Fi networks to improve location services. Although this collection was accidental and involved only fragments of data from open Wi-Fi networks, the situation raised concerns about the security of personal data on these networks. Google admitted to collecting excerpts of personal web activity through its Street View cars. This collection was discovered during a data audit by Germany's data protection authority. Although Google stated it didn't collect sensitive data, the audit revealed that it had inadvertently gathered fragments of data from open Wi-Fi networks, though it was never used in any Google products. Google immediately halted the collection, isolated the data, and plans to delete it under third-party supervision. This highlights the vulnerability of data on unsecured Wi-Fi networks.[9]

7.3 Retail and public space tracking scandals

In London, a pilot project to track smartphones recorded the unique MAC addresses of over half a million phones through a network of recycling bins. Located at 12 spots, these bins captured the MAC addresses of phones with Wi-Fi on, as well as data on the movement, direction, and speed of the devices. The project, run by Renew London, demonstrates the potential for targeted advertising by analyzing users' past behaviors, such as their workplaces, places of interest, and habits. During tests in May and June, over four million events were captured, involving more than 530,000 unique devices. While the data is collected anonymously, Renew emphasized that the devices are not individually tracked and the data is aggregated. However, it remains unclear whether this collection violates the Data Protection Act, as MAC addresses may be considered personal data.[10]

7.4 Data Leaks and Security Breaches Involving MAC Data

7.5 Attacks Exploiting MAC Addresses

7.6 WhatsApp Security Vulnerability

In 2012, a security vulnerability was discovered in WhatsApp that allowed an attacker to impersonate a user if the MAC address of their device could be obtained. At that time, WhatsApp did not use end-to-end encryption and relied on the device's MAC address to authenticate users. By exploiting this vulnerability, an attacker could:

- Log into the victim's WhatsApp account from another device.
- Read and send messages as if they were the victim.

Exploitation of the vulnerability: Attackers could obtain the MAC address through various means, including attacks on public Wi-Fi networks or through malicious applications capable of reading a device's MAC address. They could then use specific tools to log into WhatsApp while posing as the victim.

Consequences and fix: This issue highlighted the risks of using static identifiers like MAC addresses to authenticate users. WhatsApp eventually strengthened its security measures, including adopting end-to-end encryption, and stopped using the MAC address for authentication. This case illustrates the importance of not relying on MAC addresses to authenticate users due to how easily they can be spoofed.

8 Conclusion

8.1 Why should we anonymize a MAC address?

In conclusion, MAC addresses, while essential for network communication, present significant privacy risks when not properly anonymized or protected. The ability to track devices based on their MAC address has raised concerns about surveillance and unauthorized data collection, especially when these addresses are linked to individuals or their activities. The legal landscape, particularly the GDPR and other international standards, plays a critical role in regulating the use of personal data, including MAC addresses, and emphasizes the need for robust privacy protections. Anonymization techniques, such as pseudonymization and encryption, are essential to mitigate these risks, ensuring that MAC addresses can be used without compromising user privacy. As the IoT ecosystem continues to grow, the importance of developing effective anonymization strategies will be pivotal in safeguarding personal data and upholding privacy rights in embedded systems and beyond.

References

[1] IPCisco. What is a mac address? — mac address lookup — mac examples. Accessed: 2024-11-11.

[2] Stanley Arvey. Address resolution protocol (arp): Everything you should know about, 2022. Accessed: 2024-11-11.

[3] EmbeddedCentric. Introduction to bluetooth low energy / bluetooth 5, n.d. Accessed: 2024-11-11.

[4] European Parliament and Council. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), 2016. Official Journal of the European Union, L 119, 4.5.2016.

[5] Council of Europe. Article 34,35 of the convention for the protection of human rights and fundamental freedoms, 1950. Official Journal of the European Union, 4 November 1950, as amended.

[6] Council of Europe. Convention 108+ for the protection of individuals with regard to the processing of personal data, 2018. www.coe.int/dataprotection.

[7] Belgian Government. Act on the protection of natural persons with regard to the processing of personal data. Unofficial translation, 2018. Unofficial translation, Belgian Official Journal, 05 September 2018, Consolidated version (05/09/2018).

[8] Peter Cohan. How nordstrom uses wifi to spy on shoppers, 2013. Accessed: 2024-12-05.

[9] Jemima Kiss. Google admits collecting wi-fi data through street view cars, 2010. Accessed: 2024-12-05.

[10] Kadhim Shubber. Tracking devices hidden in london's recycling bins are stalking your smartphone, 2013. Accessed: 2024-12-05.