Defeating MAC address randomisation: A hybrid, multi-layer approach for indoor location-based services

Student Name: Maximilian Grimmett
Supervisor Name: Dr Eleni Akrida

***Abstract* —**

Context/Background

Recently, location-based indoor positioning services using 802.11 (Wi-Fi) have seen a rise in popularity of MAC randomisation to prevent tracking of individual devices throughout their stay.

Aims

Primarily, this paper looks to implement a system that combines previous techniques for MAC de-randomisation at different layers of the OSI model and conduct comparative research evaluating it relative to their individual use. Secondly, this paper looks to generalise these techniques to consider and evaluate groups of devices moving between different 802.11 nodes of the same network. Additionally, to enable such comparisons, this paper aims to produce a novel, modular testing and evaluation framework which uses Monte Carlo simulation to generate realistic, real-time data.

Method

Due to the ethical implications of collecting potentially identifiable real-world data and the impracticality of doing so in a controlled environment, a real-world study is not suitable. While applicable real-world datasets do exist, they do not encompass all the fields required at the different levels of abstraction used by the proposed hybrid system. Instead, this paper proposes the use of Monte Carlo simulation of devices with and without MAC randomisation.

***Keywords* —** Location-based services, MAC randomisation, dynamic networks, privacy-preservation.

I INTRODUCTION

A Background and Motivation

A.1 Location-based Services

Location-based Services (LBSs) is a large field which has seen an explosion in academic and industry interest (Junglas & Watson 2008) since the beginning of the 21st century. Broadly speaking, it covers information-technology services involved in the collection, analysis and application of any data that tracks the location of either objects, devices or people. Much of the interest has stemmed from its potential for interdisciplinary use and insight into geographical social trends and tendencies. It has seen applications in security, health and marketing (mainly for advertising).

Historically, positioning in LBSs has been restricted to use by those who own the infrastructure, namely telecom operators. Bellavista et al. (2008) note the past monopolistic infrastructure-centric nature of both indoor and outdoor localisation and describe this as the shift from operator-centric to user-centric management of LBSs.

As they cover such a wide range of services, LBSs can be grouped by certain characteristics and scopes of usage. Within the field of localisation, an important divide is made between indoor

and outdoor LBSs. Outdoor LBSs primarily make use of GPS and telecom towers for positioning. A large example of an outdoor LBS which uses GPS is satellite navigation and social media geotagging which, as of 2013 saw a large usage uptake (Zickuhr 2013).

On the other hand, GPS and telecom tower triangulation is not suitable for indoor positioning. Even though its possible to receive GPS signal indoors with modern hardware, studies such as (Merry & Bettinger 2019) have shown it to be accurate to between 7 and 13 metres outdoors. This kind of accuracy is only enough to place a device in a building or potentially a large room. Additionally, GPS elevation is only accurate to between 10 and 20 metres, which is not enough to accurately detect what floor a device is on. Hence, for better lateral/longitudinal and vertical accuracy, a different localisation technique is required for indoor LBSs.

A.2 Indoor Position Tracking

Indoor positioning, or indoor localisation, typically involves tracking the location of people, or groups of people as they move around some premises. The hardware and software used to achieve this is called an indoor positioning system (IPS), sometimes also referred to as an indoor localisation system (ILS).

IPSs come in many forms and can be broadly split into two groups: computer vision (CV)-based and device-based. CV-based IPSs have seen increased use recently in line with advancements in state-of-the-art CV object-tracking and facial-recognition methods. Device-based IPSs typically use wireless technologies such as 802.11 (Wi-Fi) (IEEE 1997), Bluetooth, Bluetooth Low Energy (BLE) and RFID (S. Shen & Sui 2020). Despite being an exciting technology, a large problem with CV-based IPSs is that they present a much greater privacy risk than device-based IPSs. This is not only due to the potential for (and incentivised use of) facial recognition, but also the ethical issues with regards to the recording (and storing of recordings) of individuals.

An advantage of 802.11 IPSs is that tracking can be done completely passively. Almost all modern smartphones have Wi-Fi capability and are constantly sending probe requests to scan for Access Points (APs). Because of this, their implementation and use of the protocol is all that is required to detect devices.

Received Signal Strength (RSS) is a measure, in decibels, of how strong a signal between two devices are – these devices are typically a smartphone and a wireless 802.11 AP. Since RSS is inversely proportional to distance, obtaining the RSS from three APs is enough to triangulate a mobile device to a reasonable degree of accuracy, this is demonstrated in Figure 1. Some IPSs additionally make use of Received Response Time (RRT) and some APs contain additional hardware to calculate Angle of Arrival (AOA), both of which can be used to further increase localisation accuracy.

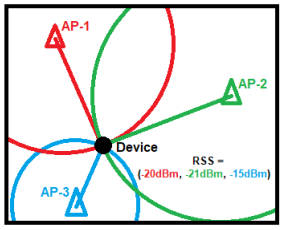

Figure 1: Triangulation of a mobile device from three APs using RSS.

A.3 MAC Address Randomisation

The operating system of a device is ultimately responsible for implementing the protocols it uses. As such, there is no way to enforce a device sending its true physical Media Access Control (MAC) address. This enables the operating system to handle 802.11 communication (including sending probe requests and connecting) with a MAC address of their choice.

Consider the two most widely used operating systems for mobile smartphones: iOS and Android. Both iOS 8, released in 2014, and Android 6, released in 2015, introduced MAC randomisation in probe requests as the default behaviour. Since then, both iOS 14, released in 2020, and Android 9, released in 2018, implement full MAC randomisation when connecting to networks. Whereas the iOS implementation rotates the randomised MAC address every 24 hours, Android retains the same (originally randomised) address per 802.11 SSID. In 2019, Android 10 set this as the default behaviour (as does iOS 14).

While these measures have little practical impact on the user experience when using guest Wi-Fi provided by businesses, it is catastrophic for LBSs which rely on 802.11 probe requests to identify users and their location. In practice, this means that LBSs are, at best, able to pinpoint the location of a device but not detect its movement over a time period (where MAC randomisation for probe requests is on a fixed timer). At worst, they are not able to pinpoint a device at all (when MAC addresses are randomised between probe requests).

B 802.11 fingerprinting

Since before the widespread adoption of MAC randomisation, passive 802.11 fingerprinting has been a hot topic for research (Pang et al. 2007). In the context of the 802.11 protocol, fingerprinting may either refer to the spatio-temporal identification of individual devices or the pre-processing of specific buildings to calibrate or improve localization success rate of an algorithm. For the remainder of this paper, assume fingerprinting refers to the former unless otherwise stated. Additionally, take 'MAC de-randomisation' to mean the process of uniquely identifying a device across space and time.

This paper considers two primary forms of fingerprinting techniques for MAC de-randomisation: implicit identifier fingerprinting proposed by Pang et al. (2007) and timing attacks proposed by Matte et al. (2016) (both of which work on the physical-layer of the OSI model). A third technique that is considered, which does not involve fingerprinting, works on the link-layer of the OSI model and will be referred to as MAC vendor analysis. Broadly-speaking, implicit identifier fingerprinting exploits minor differences in manufacturing within the tolerances of mobile devices to uniquely identify them. Timing attacks use the Inter-frame Arrival Times (IAT) of sets of frames to group them by likelihood of them being sent from the same device. MAC vendor analysis uses knowledge of how manufacturers assign the lower-order bytes of addresses to devices to infer information about the device. These are discussed in greater detail as part of Sections II and III.

C Aims

The primary objective of this paper is to compare a hybrid implementation consisting of implicit-identifier 802.11 fingerprinting, timing attacks and vendor analysis compared to their individual use. To assist with this, the paper also aims to produce a Monte Carlo-based location data simu-

lation method to benchmark and test the systems. Finally, this paper aims to evaluate the hybrid tracking technique when generalised to groups of moving devices.

This research aims to address the following questions:

1. What are the differences and similarities between the performance of 802.11 fingerprinting, timing attacks and MAC vendor analysis applied independently compared to them combined into one hybrid method?
2. How well does this hybrid method generalise to groups of individuals?
3. How well do these different methods work when applied across the different versions of 802.11?
4. How can real-world location data be simulated for a controlled testing and evaluation environment?

II RELATED WORK

A MAC randomisation and 802.11 fingerprinting

There have been many attempts at either overcoming or to point out the flaws with MAC randomisation. The majority of its literature appears post-2015, around the time of its widespread introduction.

Pang et al. (2007) were ahead of their time when they presented a technique which is able to uniquely identify devices without the need for a MAC address. This type of technique will be referred to as *implicit identifier fingerprinting*. Because it yields good results, the method is selected as one of the methods to be used in the hybrid system. However, due the continued maturing of the 802.11 protocol and improved manufacturing techniques, it should be expected that tolerances on the previously discussed fields become more narrow over time. Hence, consideration should be taken into account about how the different 802.11 versions affect both the hybrid system and implicit identifier fingerprinting. This forms the motivation behind the third objective question and is discussed further in Section III.

Vanhoef et al. (2016) perform a technical breakdown of several novel techniques to track mobile devices, bypassing MAC randomisation. It was one of the first papers to specifically aim at bypassing MAC randomisation. One such technique involves looking up one of the WPS fields (the Universally Unique IDentifier-Enrolle (UUID-E)) in a hash table to retrieve the true MAC address. One of the datasets they used was (Barbera et al. 2013), issues have been pointed out in their analysis pertaining to the fact this dataset was anonymised and there was no ground true to validate against, leading to questions being asked about the integrity of the paper and methods presented, especially given MAC address randomisation was not rarely implemented in 2013.

Martin et al. (2017) assessed the effectiveness of MAC randomisation and performed a critical evaluation of Vanhoef et al. (2016). As part of this, they produced their own dataset to better represent probe requests sent by modern devices, and more importantly those including MAC randomisation. They then improved on the fingerprinting technique presented by (Vanhoef et al. 2016), being able to effectively defeat MAC randomisation in about 96% of Android phones. Finally, they presented a new physical-layer flaw which enables, under various circumstances, an active attack able to track any of the devices they tested. Fenske et al. (2021) follows-up on (Martin et al. 2017) by experimentally testing MAC randomisation on 160 mobile phone models and found that many of them still do not implement adequate MAC randomisation –

they still emit implicit identifiers (primarily divided by device make and model). In particular, they conclude that some relatively modern android devices are particularly inconsistent with the MAC randomisation they provide, leaving them strongly susceptible to the implicit identifier fingerprinting attacks. Importantly, they note that the older attacks which are able to recover the true MAC address have been largely phased out.

In parallel to (Vanhoef et al. 2016), Matte et al. (2016) present a novel timing-based attack. This works by statistically analysing and grouping, by device, sets of probe request frames transmitted over time using their IAT. Ultimately, this collects a dictionary of MAC addresses used by the device and has seen success with tracking multiple devices over time. For their PhD thesis, Matte (2017) performed an in-depth review of 802.11 fingerprinting, expanding on their 2016 timing attack and improving on implicit identifier fingerprinting.

In addition to the previously mentioned MAC vendor analysis which works at the data-link layer, Robyns et al. (2017) provides two more data-link layer vulnerabilities. The first technique fingerprints a device transmitter by performing per-bit entropy analysis. The other uses temporal information in a similar way to that of (Matte et al. 2016). With this, they produce a dataset consisting of $\tilde{3}$0,000 unique devices (taken from an outdoor festival) and show the first technique to be capable of successfully fingerprinting 80% of 50 and 68% of 100 observed devices. For 1,000 to 10,000 devices, performance degrades considerably to a 33% to 15% success rate respectively.

MAC addresses consist of six octets, for universal MAC addresses the first three bytes of which uniquely identify the manufacturer (an organisationally unique identifier) and the last three bytes are uniquely assigned by the manufacturer. Here, the second-least significant bit of the third octet, referred to as the locally administered (LA) bit, is always zero. When this bit is set to one, the address becomes an LA MAC address. For 802.11 to work, this must be unique within a local area network (LAN) but otherwise may be freely chosen by the operating system. This is demonstrated in Figure 2. As a continuation of their PhD work, Matte & Cunche (2018) use this to study the spread of MAC randomisation in probe request datasets spanning between 2013 and 2017 by checking if addresses have the LA bit set. This is helpful because it considers data from both before adoption of MAC randomisation, by including the (Barbera et al. 2013) dataset, and after, by using datasets such as those provided by (Martin et al. 2017) and (Robyns et al. 2017). The study comes to an interesting conclusion stating that, where there is a steady increase in usage of MAC randomisation, only 3% of probe requests (as of the 2017 dataset) had the LA bit set. Since it is such a fast-moving field however, with the recent releases of iOS 14 and Android 10, an updated study is required. This lack of modern (post 2020 indoor) data is part of the motivation behind the development of the data simulation framework.

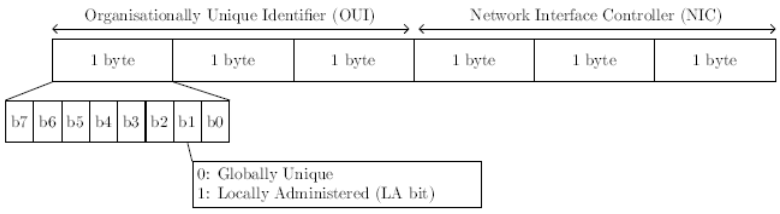

Figure 2: Format of a MAC address from (Matte & Cunche 2018).

B Privacy and Ethical Issues

An inherent problem with LBSs is that of preserving user privacy while still making meaningful use of them. The timestamped location (spatio-temporal information) of a person is generally considered to be very personal information. Because of this, as long as LBSs have existed, there has been research into how privacy can be either guaranteed or personalised by the user. From a technical standpoint, this is also important to consider because the measures employed to provide anonymisation may adversely affect the ability to fingerprint. Within the scope of this paper, it is important to take these into consideration for the design of the testing and evaluation framework.

Most privacy measures focus on LBSs that gain access to the data collected by IPSs. This exchange is typically achieved using third party trusted middleware which anonymises data before handing it off to the end-user LBS. Ghinita et al. (2008) takes a more mathematical approach, employing theoretical Private Information Retrieval to anonymise positioning data returned to LBS queries. An added benefit of this framework is that middleware is no longer required. Alternatively, the IPS may wish to improve privacy by slightly adjusting localisations such that they average out to a true representation, yet provide anonymity on the individual level (Kido et al. 2005). Gedik & Liu (2005) and Gedik & Liu (2008) propose a user-oriented approach to privacy by presenting a framework enabling the user to configure the amount of anonymity it desires, in addition to their desired spatio-temporal resolution. Finally, an alternative to passive user tracking worth briefly mentioning is that proposed in (Furfari et al. 2021). This would mitigate the discussed privacy issues by enabling users to opt-in to discovering location-based services, at which point they give permission to access their location data. From this, the best and simplest privacy measure for simulation of real-world data is the trusted middleware model. Aside from the fact the data is not real, this stops the emulated IPS from having to consider privacy, better enabling it to generalise.

III SOLUTION

The core methodology this paper will implement is a purely quantitative mix between comparative and evaluative research, with the generation of data being slightly exploratory.

A Testing and Evaluation Framework

This section will present the proposed Monte Carlo-based data generation, describe the implementation of the testing and evaluation framework and discuss the measures of effectiveness used for comparison.

A.1 Data Generation with Monte Carlo Simulation and Randomised Algoritms

Due to the previously discussed ethical and privacy issues with real-world data collection and impracticality of setting up a real-world 802.11 study, a seemingly simple solution is to instead generate the data through Monte Carlo simulation. Additionally, while there are plenty of existing datasets, they do not contain all the fields required for the application of each of the independent methods. Because of this, the implementation of a Monte Carlo simulation to generate location data representative of devices with, and without, MAC randomisation would not only be useful for this study, but also future research (potentially in related fields). This is combined into

a testing and evaluation framework that enables the modular insertion or removal of the methods shown in Figure 3 (plus the hybrid model with all of them combined).

Pinpointing location with 802.11 is a more complex technical issue than the implementation of triangulation alone. Much like with GPS, the indoor environment presents localisation difficulty due to walls and other obstructions. A typical existing wireless infrastructure may consist of one access point per floor, or, at best, one per room. If such a system was transformed to include an IPS without additional hardware modification, it would result in high inaccuracy when estimating the location of a device. Because of this, the Monte Carlo simulation will emulate an ideal testing environment, where there are multiple access points in a single room.

The Monte Carlo method is a class of algorithms which leverage random sampling to obtain numerical results (Raychaudhuri 2008). They are well-suited to simulating physical models due to their near-nondeterministic nature. This makes them ideal for simulating physical-layer characteristics such as those used by implicit identifier fingerprinting. Additionally, they can be used to simulate the frame emission times required by the timing attack.

simulations is the employment of random-based algorithms to generate both true and randomised MAC addresses. These can then be used by the timing attack as part of its fingerprinting and be sent for MAC vendor analysis.

Finally, each simulation (whether considering one or more devices) is run a fixed number of times. The exact number of times they are run is determined experimentally trough trial and error until the results converge to within a pre-determined error margin.

A.2 Software Framework

Each de-randomisation technique is independently implemented as a module extending a `DerandomisationTechnique` abstract base class, specifying the location data fields they require as input from the framework. At runtime, through dependency injection, the framework provides the requested data (and continues to update it as the simulation runs). It then calls to an interface method implemented by the technique to return the MAC addresses.

To guarantee more valid, truthful location-data output from the Monte Carlo simulation, its hyperparameters and probability distributions are fine-tuned by running the independent techniques and tweaking them to closely match the results obtained in their respective papers.

The implementation provided by the framework of the 802.11 protocol is designed to enable modularity. For the reasons mentioned earlier, this enables testing of how well the various techniques work on different versions of the 802.11 protocol. An object-oriented approach is proposed whereby an abstract base class, `Simulated80211`, is inherited by concrete classes which implement the respective version of 802.11 (for example, `Simulated80211b` for the 1999 version, or `Simulated80211ax` for the 2021 version). It is within these concrete classes that the Monte Carlo simulation and randomised algorithms are implemented. This is then further abstracted to have `Simulated80211` extend the abstract base class `WirelessProtocol`, allowing any protocol to be used. Because of this, the concrete implementation of the protocol corresponding to the one being tested (or used for evaluation) is instantiated at runtime, and can be freely swapped out as required (even for a real-world source).

With this in mind (and using the separation of concerns principle), the concepts of premises (`Premises`), devices (`Device`), APs (`AccessPoint`) and IPS software (`IpsSoftware`) are defined as different entities. In the 3D model, for simplicity, premises are represented by a rectangular cuboid of arbitrary size, and both APs and devices are represented by a single

point within the bounds of the premises. Implementation-wise, a `Premises` only stores its bounding-box, a `Device` and `AccessPoint` both store a three-dimensional position vector and a `WirelessProtocol` and `IpsSoftware` stores a `DerandomisationTechnique` instance. All of these are externally managed by a `SimulationManager`. The manager maintains a list of `EvaluationMetric` instances which use the results to calculate the metrics listed below. Because the manager has access to both the simulated results and true location assignments, it sends them into each `EvaluationMetric` and outputs the results.

A.3 Measure of Effectiveness

The primary measure of how effective each method is will be based upon the percentage of devices successfully tracked over discrete time frames. The reason for using this is because it closely matches the real-world application of tracking users' locations across space and time. In addition to their use in simulation calibration, the other metrics from the original papers can be used for the hybrid system and compared as a supplement to the primary spatio-temporal measure of success. Finally, as the framework is a controlled environment, the simulation manager is able to undertake performance profiling for speed and memory usage on the simulated techniques. This performance information is also useful knowledge for real-world application because it can be used to determine required hardware specifications.

B Hybrid System

This section will go into more detail on how the three MAC de-randomisation techniques fit together to form the hybrid system and how it fits with the software framework.

B.1 Overview

Figure 3 shows a high-level diagram-description demonstrating the modularity and the layers at which each technique resides in the OSI model. Each layer is directly related to the data it requires: implicit identifier fingerprinting requires low-level parameters such as packet size, (encrypted) packet data and 802.11 options, timing attacks require low-level parameters such as probe request data and IAT, and higher-level parameters like MAC address, and this implementation of MAC vendor analysis requires only the broadcast MAC address.

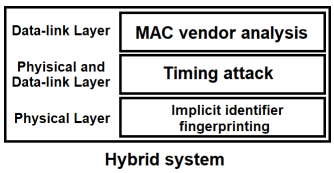

Figure 3: Illustration of the proposed hybrid system, and the OSI layers it encompasses.

B.2 Software Implementation

The modular approach described in Testing and Evaluation Framework continues to apply to the hybrid system. The concrete class `HybridTechnique` extends `Derandomisation-`

`Technique` and contains one field for each independent technique. As it extends the `DerandomisationTechnique` base class, it can be substituted in at runtime as if it were its own technique. All it has to do is implement a way to pass required arguments and the simulated location-data down to the respective techniques.

IV VALIDITY

A Literature Review

Much of the research in the field of LBSs is performed by, or on behalf of, companies which have a vested interest in its success (or apparent success). This means that the cited papers not published in reputable journals such as tech reports need to undergo additional scrutiny to check for conflicts of interest where applicable.

Additionally, given the privacy issues already discussed, it is important to check that cited papers give appropriate consideration to the way in which the methods they implement collect and process potentially identifying information and the actions they take to re-anonymise such data.

To help improve validity, in Section II, most journal articles and conference proceedings are only cited if they are published in a journal of impact factor greater than 4 (2 for newer journals) or a conference of historical impact factor greater than 2. Cited works with lower impact factors are either included for critique or if they are very recent.

A downside to the validity of the literature review is that the majority of it is performed by a few key players, primarily Matte, Célestin. While they appear to produce good work (and some of it is peer reviewed), its hard to validate because the majority is not published in credible journals. This should be partly put down to the fact that its a niche subject area that is still maturing.

B Proposed Methodology

It is important to consider the validity of the proposed methodology. To what degree of certainty does the Monte Carlo simulation of MAC randomisation represent real-world data and how is this calculated.

Whether the testing framework is made before the hybrid system or vice versa, there is knowledge of one leading into the other, meaning there could be unconscious bias which looks to boost results when combining the two.

Further to this, because the evaluation data is simulated, the additional generation and consideration of data for devices not using MAC randomisation provides an important baseline before comparison against the state of the art.

To further increase the validity of the Monte Carlo simulation, where random values between 802.11 tolerances are generated, additional variance in the received value should be simulated to account for physical phenomena such as interference.

One of the most important considerations for validity is how to be confident that the proposed framework is representative of real-world data. One measure put in place to help with this is the use of the previous techniques to calibrate the simulation. As a natural extension to this, the modular ability to swap between 802.11 Monte Carlo simulations provide an abstraction which helps verify the validity of any technique being evaluated using the framework.

V RESULTS

VI EVALUATION

VII CONCLUSION

References

Barbera, M. V., Epasto, A., Mei, A., Kosta, S., Perta, V. C. & Stefa, J. (2013), 'Crawdad dataset sapienza/probe-requests (v. 2013-09-10)'.

Bellavista, P., Küpper, A. & Helal, S. (2008), 'Location-based services: Back to the future', **7**(2), 85–89.

Fenske, E., Brown, D., Martin, J., Mayberry, T., Ryan, P. & Rye, E. (2021), 'Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds', *Proceedings on Privacy Enhancing Technologies* **3**, 164–181.

Furfari, F., Crivello, A., Baronti, P., Barsocchi, P., Girolami, M., Palumbo, F., Quezada-Gaibor, D., Mendoza Silva, G. M. & Torres-Sospedra, J. (2021), 'Discovering location based services: A unified approach for heterogeneous indoor localization systems', **13**, 100334.

Gedik, B. & Liu, L. (2005), Location privacy in mobile systems: A personalized anonymization model, *in* 'Proceedings - International Conference on Distributed Computing Systems'.

Gedik, B. & Liu, L. (2008), 'Protecting location privacy with personalized k-anonymity: Architecture and algorithms', **7**(1), 1–18.

Ghinita, G., Kalnis, P., Khoshgozaran, A., Shahabi, C. & Tan, K. L. (2008), 'Private Queries In Location Based Services Anonymizers Are Not Necessary', **8**(ii), 1–12.

IEEE (1997), 'IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications', *IEEE Std 802.11-1997* pp. 1–445.

Junglas, I. A. & Watson, R. T. (2008), 'Location-based services', **51**(3), 65–69.

Kido, H., Yanagisawa, Y. & Satoh, T. (2005), An anonymous communication technique using dummies for location-based services, *in* 'Proceedings - International Conference on Pervasive Services, ICPS '05', Vol. 2005, pp. 88–97.

Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E. C. & Brown, D. (2017), 'A study of MAC address randomization in mobile devices and when it fails'.

Matte, C. (2017), Wi-Fi tracking: Fingerprinting attacks and counter-measures, PhD thesis, Université de Lyon.

Matte, C. & Cunche, M. (2018), Spread of MAC address randomization studied using locally administered MAC addresses use historic.

Matte, C., Cunche, M., Rousseau, F. & Vanhoef, M. (2016), Defeating MAC address randomization through timing attacks, *in* 'WiSec 2016 - Proceedings of the 9th ACM Conference on Security and Privacy in Wireless and Mobile Networks', Association for Computing Machinery, Inc, pp. 15–20.

Merry, K. & Bettinger, P. (2019), 'Smartphone GPS accuracy study in an urban environment', *PloS one* **14**(7), e0219890.

Pang, J., Greenstein, B., Gummadi, R., Seshan, S. & Wetherall, D. (2007), 802.11 user fingerprinting, *in* 'Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM', pp. 99–110.

Raychaudhuri, S. (2008), Introduction to monte carlo simulation, *in* '2008 Winter simulation conference', IEEE, pp. 91–100.

Robyns, P., Bonné, B., Quax, P. & Lamotte, W. (2017), 'Noncooperative 802.11 mac layer fingerprinting and tracking of mobile devices', *Security and Communication Networks* **2017**.

S. Shen, L. & Sui, D. (2020), Wi-Fi Location-Based Services (Lbs) for Occupancy Sensing in Buildings: A Technical Overview, Technical report, Center for Energy and Environment.

Vanhoef, M., Matte, C., Cunche, M., Cardoso, L. S. & Piessens, F. (2016), Why MAC address randomization is not enough: An analysis of Wi-Fi network discovery mechanisms, *in* 'ASIA CCS 2016 - Proceedings of the 11th ACM Asia Conference on Computer and Communications Security', Association for Computing Machinery, Inc, pp. 413–424.

Zickuhr, K. (2013), 'Location-based services', *Pew Research* **679**, 695.