Defeating MAC Randomisation: A hybrid, multi-layer approach

Max Grimmett
maximilian.a.grimmett@durham.ac.uk

Durham University

BACKGROUND AND RELATED WORK

Location-based services (LBSs) are software services that include the collection, analysis and presentation / visualisation of any data that includes the location of objects, devices or people. They have seen use in security, cybersecurity, health and advertising. A good example is GPS, which is a user-centric LBS that allows a device to calculate its position anywhere in the world to a high degree of accuracy.

Indoor Positioning Systems (IPSs) are typically business-centric technologies used for the collection of location data from within buildings (typically offering a higher degree of accuracy and reliability compared to GPS). They come in many forms and are built using differing technologies – traditionally these use Wi-Fi (through 802.11 probe requests)[1], Bluetooth or Bluetooth Low Energy, more modern solutions include audio-based and computer vision-assisted positioning[2]. 802.11 IPSs remain popular as most people carry around Wi-Fi-enabled devices and due to the increased privacy concerns caused by the storage of recordings and the greater potential for personal identification.

802.11-based IPSs usually rely on the Media Access Control (MAC) addresses returned as part of a probe-request, it then uses triangulation based on response time, received signal strength (RSS) and angle of arrival (AOA) (if available) to determine the location of the device relative to the access points (APs), this is demonstrated in Figure 1.

The recent introduction of MAC randomisation by smartphone operating systems (since iOS 8 and Android 6) has brought the 802.11 IPSs that rely on MAC address identification to a standstill, at best able to pinpoint the instantaneous location of a device or at worst not being able to locate devices at all.

Since before widespread adoption of MAC randomisation, passive 802.11 fingerprinting has been a hot topic for research[3,4]. Fingerprinting may either refer to the spatio-temporal identification (and localisation) of individual devices or the pre-processing of specific buildings to calibrate or improve localization success rate of an algorithm. More modern research in 802.11 fingerprinting includes the use of convolutional neural networks with a stacked autoencoder which has seen impressive results boasting 100% localization success rate at the building-level and 95% at the floor-level[4] (an example use case of which is sown in Figure 2).

Approaching the problem from another angle, MAC randomisation has been shown to be vulnerable to timing attacks[6] and MAC vendor analysis[7].

The timing attacks lie at the intersection between the physical-layer they work by statistically analysing and grouping – by device – sets of probe request frames transmitted over time using their inter-frame arrival times.

MAC vendor analysis works slightly differently to the aforementioned fingerprinting techniques in that it only relies on the link-layer MAC address. This works by exploiting the known ways in which device manufactures and operating systems assign the low-order bytes in a MAC address. It should be noted that vendor analysis is not yet a fingerprinting technique in itself as it is currently used to infer make and model number from non-randomised MAC addresses.

AIMS/OBJECTIVES

This primary aim of the project is to compare the performance of hybrid method combining multiple methods for MAC de-randomisation to the methods on their own in both individual-device and multi-device (group) settings.

Questions this research aims to address:

- 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?
- How well does this hybrid method generalise to groups of individuals?
- How well do these different methods work when applied across the different versions of 802.11?

Implementation wise, in addition to the hybrid system, it will aim to provide an independently implemented framework for evaluating MAC de-randomisation.

Figure 1: *Triangulation of a device in an open room from three APs using RSS, AOA and response time*

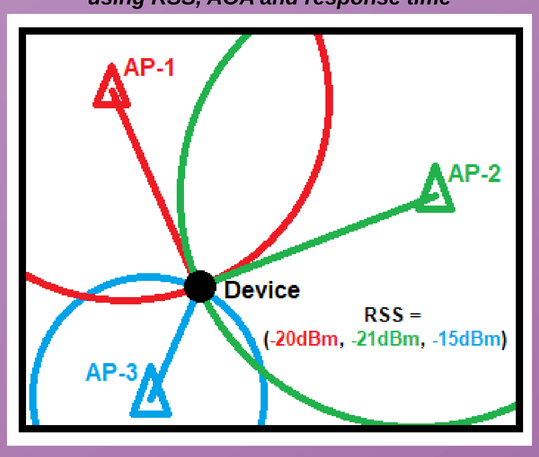
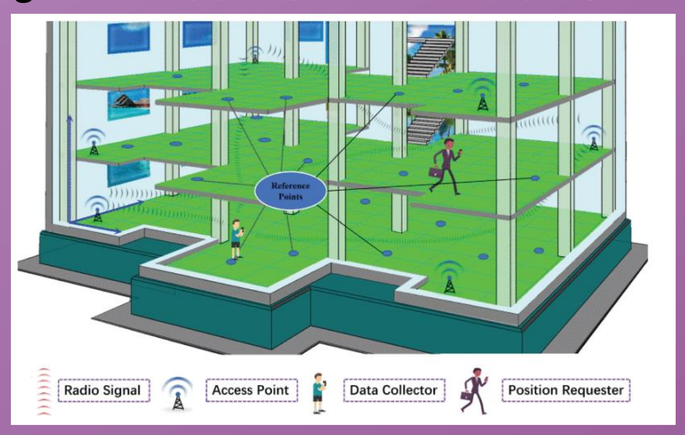


Figure 2: *Fingerprinting-based IPS scenario by Song, et al. (2019)[4]*



PRIVACY AND ETHICAL ISSUES

A recurring topic of great importance in location-based services is that of privacy concerns. While, for this study, no real-world data is collected, it is important to consider the privacy implications if the methods used by this research were applied in the real world and whether it can be personally identifying.

The hybrid system uses the MAC address as broadcast by a device plus various implicit identifiers. These implicit identifiers, while used to uniquely identify a device cannot be used to uncover the true MAC address where randomisation is used and moreover true MAC addresses are only associated with a device, and, while implicitly used to uniquely identify users, they do not give personally identifying information.

An important note is that while de-anonymisation is non-trivial on an individual level, research has shown that 802.11 probe requests can be used for de-anonymisation of large groups[8].

Figure 3: *Proposed multi-layer hybrid system*

Data-link Layer	MAC vendor analysis
Phyisical and Data-link Layer	Timing attack
Physical Layer	802.11 fingerprinting

METHOD

Aside from the previously mentioned privacy risks, there is an impracticality of setting up a real-world 802.11 study . Additionally, while there are plenty of existing datasets, they don't contain all the fields required for the application of each of the independent methods. Because of this, a core element of the methodology is the implementation of a Monte Carlo simulation to generate location data representative of devices with and without MAC randomisation. This is combined into a test and evaluation framework which enables the modular insertion or removal of the methods shown in Figure 3 (plus the hybrid model of with all of them combined).

Once the methods in Figure 3 are implemented, the test and evaluation framework is then fine-tuned against the independent implementations of 802.11 fingerprinting and timing attack to closely match the results obtained in their respective papers.

The primary measure of how effective each method is will be based upon the percentage of successfully tracked over discrete time frames. The reason for using this is because it closely matches the real-world use-case of tracking users' locations across space and time. In addition to 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.

VALIDITY

- Much of the research in the field is performed by or on behalf of companies which have a vested interest. Because of this, it is important to pay particular attention to papers which have not been published in reputable journals and identify conflicts of interest where applicable.
- To try improve validity, articles and conference proceedings are only cited as a source of second-hand information if they are from an academic journal with an impact factor 4 or more or conferences with a historical impact factor of 2 or more, unless where noted as part of critical analysis.
- Either the hybrid system or the test and evaluation framework is implemented first, because of this care must be taken to design the second one without bias and impartiality towards the first.
- 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.

REFERENCES

[1] Vattapparamban, E., Çiftler, B.S., Güvenç, I., Akkaya, K. and Kadri, A., 2016, May. Indoor occupancy tracking in smart buildings using passive sniffing of probe requests. In 2016 IEEE International Conference on Communications Workshops (ICC) (pp. 38-44). IEEE.
[2] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor posi- tioning with mobile phones and visible light," in Proc. ACM MobiCom, 2014, pp. 447–458
[3] Pang, J., Greenstein, B., Gummadi, R., Seshan, S. & Wetherall, D. (2007), 802.11 user fingerprinting, in 'Proceed-ings of the Annual International Conference on Mobile Computing and Networking, MOBICOM', pp. 99–110
[4] Song, X., Fan, X., Xiang, C., Ye, Q., Liu, L., Wang, Z., He, X., Yang, N. and Fang, G., 2019. A novel convolutional neural network based indoor localization framework with WiFi fingerprinting. IEEE Access, 7, pp.110698-110709.
[5] Vo-Huu, T.D., Vo-Huu, T.D. and Noubir, G., 2016, July. Fingerprinting Wi-Fi devices using software defined radios. In Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks (pp. 3-14).
[6] Matte, C., Cunche, M., Rousseau, F. and Vanhoef, M., 2016, July. Defeating MAC address randomization through timing attacks. In Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks (pp. 15-20).
[7] Martin, J., Rye, E. and Beverly, R., 2016, December. Decomposition of MAC address structure for granular device inference. In Proceedings of the 32nd Annual Conference on Computer Security Applications (pp. 78-88).
[8] Di Luzio, A., Mei, A. and Stefa, J., 2016, April. Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests. In IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications (pp. 1-9). IEEE.