Modelling Location-based Services with Temporal Networks to overcome MAC Randomisation

Max Grimmett

21/03/2021

1 Introduction

Location-based services (LBSs) [13, 7, 2, 9] is a field which spans a large array of disciplines and technologies, such services are increasingly used in industry and by people in many differing forms [16].

Recently, they have seen increased use in businesses with high traffic flows of people and customers. Some examples of these include shops, cruise liners, stadiums and airports.

A use-case for LBSs in these scenarios is occupancy sensing [14] and finding customer flow hot-spots. This enables businesses to determine how much value to assign given areas within their confines (and thus how much to charge for leasing and where to place advertisements). Additionally, the services may be used to directly benefit end-customers by providing location-based customer service (for example delivering food or drinks directly to the customer anywhere on the premises).

Another application for LBSs is in assisting with safety and security. Given the modern example of Covid-19, LBSs can detect clusters of people and inform the actions necessary to disperse them.

The data these services are based on is typically collected either through dedicated nodes placed at known locations sending and receiving probe requests or trivially by any WiFi hotspots the devices are already connected to.

2 Specific Topical Issue

2.1 MAC Randomisation

MAC randomisation is a technique employed by phone operating systems to dynamically change the MAC address broadcast by a device either over time or between responding to mobile pings. Since a MAC address is the only trivial means of uniquely identifying an unconnected device, randomly changing it prevents them from being tracked across a series of nodes. It has seen increased use over the past few years as Android, Apple and Windows devices have started shipping with and using it by default [15, 11].

While [6] tries to overcome this using historical probabilistic transition data to predict movement, most other modern techniques, such as those described in [15, 10], try to overcome this by exploiting other information sent by a device (that is derived from the global MAC address). This is achieved by analysing the additional information sent in both 802.11 probe requests and the WPS fields containing connection security information. [12] shows that this information can be used in conjunction with timings of incoming frames (with randomised MAC addresses) to group the frames by the device they were most likely sent from.

2.2 Ethical Considerations

An area of concern relating to LBSs is that of privacy [4, 14]. The reason MAC randomisation is used in the first place is to prevent actors from being able to piece together a map of a users movement over time. While MAC addresses are designed to uniquely identify devices, there is no way to inherently link them to any given person. Additionally, there is no need to collect any personal data traffic being sent over a network as only the control packets sent between devices are required. The scope of the location data should also be confined to individual premises, otherwise additional measures (such as those presented in [5]) should be taken to reduce the resolution of the data. Nevertheless, approval and oversight by an ethical committee would be an important way to ensure any potentially identifying information is ignored and/or removed as early as possible.

3 Research Scenario

Temporal networks [8] are networks in which links appear and disappear over time. They are typically represented as graphs where each edge specifies a set of discrete times at which it is active. Alternatively, at any point in time, a *snapshot* of the temporal network can be represented by a single static network (containing no time information). [8, 1, 3] present the various algorithms, results and metrics used in traditional (static) network and graph theory when converted to be compatible with and applied to temporal networks.

At minimum, research should be performed into how location service data can be modelled using temporal graphs, and how the results of algorithms applied to such models can be interpreted in a real-world context. This would require careful consideration of what the nodes and links are defined to represent, how the links activate and deactivate over time, and whether these representations are compatible with the aforementioned algorithms and metrics.

Ideally, but subject to time constraints, this would be expanded upon to investigate how these algorithms could, in part, be used to track customers movements

throughout their visit, either as a supplement to existing MAC de-randomisation techniques or independently in a novel fingerprinting scheme.

References

[1] Eleni C Akrida et al. "On temporally connected graphs of small cost". In: *International Workshop on Approximation and Online Algorithms*. Springer. 2015, pp. 84–96.

[2] Anind Dey et al. "Location-based services". In: *IEEE Pervasive Computing* 9.1 (2009), pp. 11–12.

[3] Thomas Erlebach, Michael Hoffmann, and Frank Kammer. "On temporal graph exploration". In: *Journal of Computer and System Sciences* 119 (2021), pp. 1–18.

[4] Julien Freudiger. "How talkative is your mobile device? An experimental study of Wi-Fi probe requests". In: *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. 2015, pp. 1–6.

[5] Marco Gruteser and Dirk Grunwald. "Anonymous usage of location-based services through spatial and temporal cloaking". In: *Proceedings of the 1st international conference on Mobile systems, applications and services*. 2003, pp. 31–42.

[6] Hande Hong, Girisha Durrel De Silva, and Mun Choon Chan. "CrowdProbe: Non-invasive crowd monitoring with Wi-Fi probe". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.3 (2018), pp. 1–23.

[7] Iris A Junglas and Richard T Watson. "Location-based services". In: *Communications of the ACM* 51.3 (2008), pp. 65–69.

[8] David Kempe, Jon Kleinberg, and Amit Kumar. "Connectivity and inference problems for temporal networks". In: *Journal of Computer and System Sciences* 64.4 (2002), pp. 820–842.

[9] Axel Küpper. *Location-based services: fundamentals and operation*. John Wiley & Sons, 2005.

[10] Jeremy Martin et al. "A study of MAC address randomization in mobile devices and when it fails". In: *Proceedings on Privacy Enhancing Technologies* 2017.4 (2017), pp. 365–383.

[11] Célestin Matte and Mathieu Cunche. "Spread of MAC address randomization studied using locally administered MAC addresses use historic". PhD thesis. Inria Grenoble Rhône-Alpes, 2018.

[12] Célestin Matte et al. "Defeating MAC address randomization through timing attacks". In: *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. 2016, pp. 15–20.

[13] Jochen Schiller and Agnès Voisard. *Location-based services*. Elsevier, 2004.

[14] Lester S Shen et al. "WI-FI LOCATION-BASED SERVICES (LBS) FOR OCCUPANCY SENSING IN BUILDINGS: A TECHNICAL OVERVIEW". In: (2020).

[15] Mathy Vanhoef et al. "Why MAC address randomization is not enough: An analysis of Wi-Fi network discovery mechanisms". In: *Proceedings of the 11th ACM on Asia conference on computer and communications security*. 2016, pp. 413–424.

[16] Kathryn Zickuhr. "Location-based services". In: *Pew Research* 679 (2013), p. 695.