Estimating real-time highstreet footfall from the Wi-Fi probe requests

Balamurugan Soundararaj[1], James Cheshire[1] and Paul Longley[1]

[1]Department of Geography, University College London, United Kingdom

January 25, 2018

Abstract

Wi-Fi has become an ubiquitous technology used to provide internet access in public and private spaces to people's mobile devices such as smartphones, tablets and laptops. Any point in a dense built environment such as cities has multiple Wi-Fi networks. Anticipating and remembering these networks mobile devices, to be able to switch seamlessly between them, broadcast a special type of management packets known as probe requests. This is a low level package specified in IEEE 802.1 specification which relays information about the source mobile device to any Access Points (AP) listening to them. This often the first steo in establishing a handshake between these devices, These probe requests are constinuous, passive and wireless stream of infromation available at any urban location which can act as a proxy to understanding the number of people present in that area in real-time. In this paper, utilising a set of probe requests collected at a highstreet location, along with manually collected data, we demonstrate that highstreet footfall can be estimated with reasonable accuracy without infringing on people's privacy.

Use of wireless technology to measure movement of people in built environment has been done previously. It is either done at mobile level or network level. At network level it can be done with recording mobile devices coming in and out of network or measuing mobile devices recording aps. These can be done with various technologies such as gps, mobile network where some of them are more accurate than others. In this Wi-Fi provides us a good middle ground - interms of cost, scalability, etc. The major disadvantage is that the unique identifier in wifi based counting is the MAC address which is global and can cause privacy risk. This can be solved by hashing the personal data collected. Recently mobile devices have also started randomising their mac to avoid detection and tracking. Though this can be over come by using various techniques, cite papers, these methods are usually infringing on privacy. We need a method to count people using this data without tracking them which is the major challange in which we address here.

The data collection is done through tshark, data collected were - mac, vendor, length, ssid, tags, signal, duration. - manual count was done in parallel through mobile phone app. The survey was carried out at oxford st. from 12:30 to 13:00. The total numbers. how many people. How many probes? how many unique macs? what are the uncertainities in the

internet access to mobile devices such as smartphones,tablets and laptops. In addition to internet these devices use these networks as a way to quick geo location without waiting for GPS. To produce a list of WiFi networks available, mobile devices constantly broadcast signals called probe requests. These signals have various information regarding the device and its capabilities which are identifiable such as MAC addresses, non identifiable such as signal strength and partly identifiable such as randomised mac addresses. There is a significant use in uniquely identify the number of mobile devices at a specific location wihout actually affecting the privacy of the mobile devices. Such enumeration can be done by looking at identifiable information after anonymising them whenever they are available and from the patterns in the other information when they are not. We collect data on probe requests sent at set of locations in different times and device a methodology for enumerating number of people at these locations from the data. We compare the enumerated counts with the counts collected manually

at these locations to find that we are able to estimate real life with a confidence of XX%.

[4] [3] [2] [5] [1]

References

[1] Constantine E Kontokosta and Nicholas Johnson. "Urban phenology: Toward a real-time census of the city using Wi-Fi data". In: *Computers, Environment and Urban Systems* 64 (2017), pp. 144–153.

[2] Jeremy Martin et al. "A Study of MAC Address Randomization in Mobile Devices and When it Fails". In: *arXiv preprint arXiv:1703.02874* (2017).

[3] 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*. ACM. 2016, pp. 15–20.

[4] 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*. ACM. 2016, pp. 413–424.

[5] Tien Dang Vo-Huu, Triet Dang Vo-Huu, and Guevara Noubir. "Fingerprinting Wi-Fi devices using software defined radios". In: *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. ACM. 2016, pp. 3–14.