SwiftScan: Efficient Wi-Fi Scanning for Background Location-Based Services

Ramsey Faragher, Andrew Rice
Email: firstname.surname@cl.cam.ac.uk
Computer Laboratory
University of Cambridge
Cambridge
United Kingdom

***Abstract*—The provision of location-based services on consumer devices has moved from on-demand navigation capabilities to always-on ubiquitous location-aware tools such as weather updates, travel information, location-based reminders and many more. Background localisation is generally provided by Wi-Fi fingerprinting, since GPS does not provide service in indoor environments where we spend 80% of our time. However the power consumption of a Wi-Fi scan is proportional to the number of channels scanned, and so naive full-channel scans are inefficient. Here we describe and validate SwiftScan, an intelligent, self-training Wi-Fi fingerprinting scheme that reduces the energy consumption of periodic background Wi-Fi scanning for localisation. SwiftScan is tested with data from more than a thousand Android users over a six month time period and we show that energy savings of over 90% are possible, and that the majority of users benefit from more than a 70% reduction in the energy consumption associated with a Wi-Fi scan for localisation purposes.**

I. Introduction

Wi-Fi fingerprinting is currently the most common method for providing consumer indoor positioning on smartphones and similar body-worn devices. Fingerprinting builds upon the assumption that the list of Wi-Fi MAC addresses that are detected during a scan provide a unique label for the current global location. Since the range of a Wi-Fi Access Point (AP) is typically around 30 metres, this sets the approximate upper-limit on accuracy of Wi-Fi fingerprinting, assuming a good survey. If the signal strength measurements of each AP are also logged during surveys, the positioning accuracy of Wi-Fi fingerprinting in cluttered indoor environments can be better than 10 metres [2].

Location-based services on smartphones were originally on-demand functions; the user enabled the GPS chip and directly accessed a mapping or navigation application. Today there are a wide range of background location-based services that operate without direct user intervention, such as Google Now, opportunistic retail advertising, and weather notifications. These background services need to provide moderate positioning performance ($\sim$50 m) with minimal impact on battery life. The focus of this paper is the investigation of an intelligent positioning scheme that aims to minimise the number of Wi-Fi channels scanned per location update, in order to minimise the impact on battery life of background location-based services.

Over the last decade, the number of Wi-Fi channels available to users has grown due to the deployment of access points using the 5 GHz radio band. The traditional 2.4 GHz band makes use of only 14 Wi-Fi channels, whereas the 5 GHz band contains over 40 channels. Brouwers et al. demonstrated that there is a linear relationship between number of channels scanned and the energy consumed [1]. The energy consumption of a background location scan has therefore increased as more radio channels have become available. Long battery life is not only a priority for uses of smart devices, but will be a key requirement for future location-aware Internet of Things devices that will employ Wi-Fi and BLE fingerprinting for indoor positioning.

A recent attempt to reduce the battery consumption of Wi-Fi scanning on Android devices was a move from passive to active scans. Active scans consume less power than passive scans due to the computational load associated with processing the many seconds of captured radio data in a passive scan searching for AP information. However, the author noted in a previous study [2] that this approach can lead to interference issues with BLE scanning on certain Android handsets (those with Broadcom chipsets), reducing the performance of BLE fingerprinting. Active scanning also raises security concerns by broadcasting requests for APs to respond; it does not carry the privacy benefits of GPS localisation or passive Wi-Fi fingerprinting schemes.

Passive location monitoring can also be performed by noting the current *Cell-ID* from the cellular radio. This is the identifier of the cellular mast currently providing telephony services. The coverage area of a cellular mast can range from a few hundred metres to more than a dozen kilometres, and so the positioning accuracy of a Cell-ID fix covers the same range. For some background services (such as weather information) Cell-ID fixes alone can be useful but for many (personal location-based reminders, transport time information, retailer advertising, etc.) the positioning error needs to be around 100 metres or better.

An intelligent approach to Wi-Fi scanning is proposed here in order to reduce battery consumption while maintaining desirable positioning performance, privacy and BLE interoperability for background location-based services. This approach relies on the ability to command single-channel Wi-Fi scans, which can be performed on current Android smartphones by replacing the Wi-Fi driver, as verified by Brouwers et. al. Previous work by other authors has considered the use of inertial sensing for reducing Wi-Fi scan frequency [1] [5], or simply determined the order of channel scanning in order

to scan popular channels first [1]. In this work we provide a tailored optimal scanning procedure for each user based on their common fingerprints and Cell-IDs. This new method and its validation using data from over a thousand Android smartphones users are the major contributions of the research.

II. Device Analyzer dataset

A large dataset is required to test the performance of SwiftScan. Device Analyzer [7] is an usage-monitoring Android application built and maintained by the University of Cambridge Computer Laboratory with over 23,000 users. The app gathers a wide range of usage data, including performing background Wi-Fi and Cell-ID scanning. This provides a very rich dataset for testing SwiftScan.

III. Intelligent Wi-Fi channel scanning

The intelligent approach to passive scanning used in SwiftScan builds on the following assumptions:

- *Most users spend most of their days and nights following the same behavioural patterns.* For example, sleeping in the same house, working in the same office building, etc. Location can therefore be generally predicted [4], and only a small amount of information (and so power consumption) should actually be required to confirm whether the user is in a common location.
- *Smartphone Cell-ID data is effectively a free measurement in terms of power consumption*. The cellular radio monitors the Cell-ID regularly in order to maintain cellular connectivity. The smartphone has access to the current status of Cell-ID automatically whenever the smartphone has cellular reception, and so using Cell-ID as the initial confirmation step of a time-and-day-based location prediction has no impact on battery life beyond the algorithm's processing overhead [6].
- *The majority of Wi-Fi Access Points are broadcasting on a small minority of channels*. This is caused by the method employed by *Enterprise* deployments of Wi-Fi networks to use a cellular layout employing channel numbers 1, 6 and 11 on the 2.4 GHz band to deploy their entire network, since these three radio channels have minimal simultaneous overlap. This suggests that when no prior information is available, a Wi-Fi scan should concentrate first on just these three channels. The Wi-Fi AP channel distribution determined using the Device Analyzer dataset is given below in Figure 2.
- *Only a coarse location fingerprint is required for most background location services*. Recent work by the author has demonstrated that a signal strength fingerprint measurement in the 2.4GHz band increases in location accuracy with more APs within the fingerprint, but with diminished returns above around 6-8 APs [2]. A fingerprint scan of dozens of access points across dozens of channels will provide little improvement in positioning accuracy over a fingerprint of half a dozen access points from a scan of just one or two channels. Wi-Fi scans can therefore be stopped after just a few populated channels have been detected, saving power and time.

A. Efficient location estimation

The new intelligent fingerprinting scheme described below undergoes a training stage for the first month of use in order to create a personalised lookup table. Training also continues dynamically as the positioning system is used, as discussed below.

Training consists of infrequent full-band passive Wi-Fi scanning to log Wi-Fi fingerprints with the corresponding Cell-ID at the revealed locations provided by the Android location API. A fingerprint is defined as the set of Wi-Fi MAC addresses observed within a single scan of all Wi-Fi channels. Crucially, the Wi-Fi radio channel number for each MAC address is also recorded. The Cell-ID is only considered to be valid if it has been updated within 5 seconds of the Wi-Fi scan. When comparing fingerprints, they are judged to be the same if there is more than 50% correspondence between the collections of MAC addresses recorded within them.

The training data is searched for two types of links: the set of Wi-Fi fingerprints that are all observed within the same Cell-ID, and the set of Cell-IDs that are all observed with the same Wi-Fi fingerprint. Multiple Cell-IDs can be recorded at the same Wi-Fi fingerprint for many reasons, such as handover to overlapping cells and a device using a mixture of 2G, 3G and 4G technologies. Telecommunications providers may also regularly remap their Cell-IDs, which is discussed later. It is also possible that many distinct regular user locations (all with different Wi-Fi fingerprints) can all be contained in the same Cell-ID (the homes of the user and their local friends, shops near the user's home and workplace, etc.).

Therefore, for each Cell-ID, the Wi-Fi fingerprints associated with it are studied for the most-populated common Wi-Fi channel. This provides a simple scheme to check first the Cell-ID, then the most common Wi-Fi channel associated with that Cell-ID, then the APs available on that channel in order to determine if the user is at one of their most common locations. It is likely, based on the population density of Wi-Fi channels 1, 6 and 11, that one radio channel can be common to all fingerprints, but if not, the minimum required number of Wi-Fi channels to scan within that particular Cell-ID is also logged accordingly in the lookup table, with the channel associated with the most likely fingerprint scanned first.

The lookup table can be stored on the device rather than by the location provider. This enhances both user privacy and further benefits battery performance by not requiring network access to provide the location solution during times when the user is at a typical location. Efficient background location monitoring then proceeds as follows:

1) The current Cell-ID is checked against the list of most common locations. If there is no match then proceed to step 2, else proceed to step 3.
2) As we are not at a typical location, a Wi-Fi scan is initiated, with order based on the channel population given in Figure 2 before moving on to the 5 GHz band and scanning again in order determined by the typical population statistics. The scan continues channel-by-channel until seven APs have been detected. This number provides a good quality position fix and ensures that there is some redundancy in the measurement set in the case of missing,

new, or relocated APs within the scan. This strategy reduces the number of Wi-Fi channels scanned (and so the battery consumption of the position fix) in order to reach a fingerprint of a reasonable size. The fingerprint and Cell-ID could be added to the lookup table if desired, providing dynamic retraining of the lookup table.

3) If the test in step 1 resulted in a Cell-ID match, then the Wi-Fi radio channel recorded in the lookup table as the most common channel across the most likely fingerprints is scanned first. The presence of any of the expected APs on this channel for common locations in conjunction with the corresponding Cell-ID confirms the user is in a typical known location but at much lower cost than a full Wi-Fi scan. If the scan does not reveal expected APs then the scan continues using the order provided in the lookup table followed by the default strategy given for step 2 and the location is requested from the usual network provider once a minimal number of APs, such as seven, is detected.
4) The success rate of this scheme can be monitored by noting how often step 2 of this sequence is needed. The scheme could also be extended to generating most common locations by time of the day and day of the week to provide a richer positioning scheme for users with particularly variable days.

IV. Validation

The proposed method and its underlying assumptions were tested against the Device Analyzer dataset. In order to provide an up-to-date set of findings for the approach proposed here, only the most recent six months of data for each user were tested, and any user that could not provide six months of contiguous data between the beginning of 2014 and the time of the study in 2015 was ignored. Users were also disqualified from the trials if their dataset contained fewer than 5,000 Wi-Fi scans, or fewer than 1,000 cellular scans. In all 1,317 users met the necessary criteria, their geographical distribution is shown in Figure 1. Since different users may have contributed very different numbers of scans, the normalised AP channel distribution for each user was determined and the normalised distributions were then averaged. This accounted for differences in individual sample sizes. The distribution of APs on each Wi-Fi channel is shown in Figure 2. It is clear from the Device Analyser data that the majority of Wi-Fi APs are on channels 1, 6 and 11 in the 2.4 GHz band. These data suggest that when having to perform a Wi-Fi scan with no prior knowledge, the optimal channel scan sequence is {6, 1, 11, 9, 3, 5, ...}, following the distributions given in Figure 2. The total number of APs seen across all users in the 5 GHz band is 100 times smaller than the total number seen in the 2.4 GHz bands.

A. Fingerprint Statistics

The distribution of fingerprint sizes across all users was also tested using the most recent month of data. For each fingerprint, the number of BSSIDs on an "Enterprise channel" (channel number 1, 6 or 11) was counted, and the total across all channels was also counted.

Both the distribution across all time (i.e. counting all fingerprints as many time as they are seen) and the distribution of unique fingerprints (each fingerprint is only counted once)

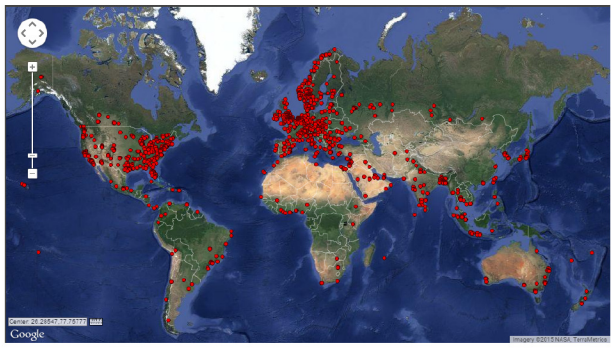


Fig. 1. The distribution of users from the Device Analyser dataset used in this study. Imagery is from Google Maps, Copyright 2015 NASA and TerraMetrics)

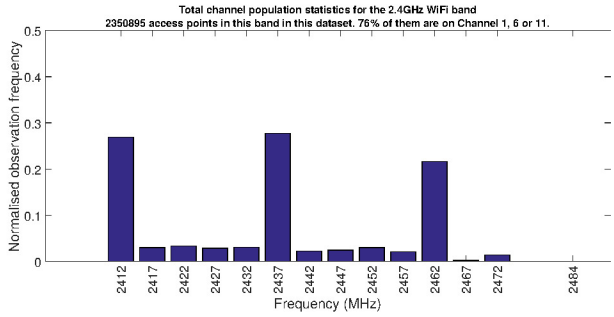


Fig. 2. The distribution of access points by radio channel in the 2.4 GHz band generated the most recent month of Device Analyser data for 1317 users. The 5 GHz access points form less than 1% of the total population.

were considered, as shown in Figure 3 and Figure 4. The distributions demonstrate again the strong bias towards the enterprise channels. Any given fingerprint is more likely to contain four enterprise channel Access Points than to contain no enterprise channel access points at all.

B. Most Common Fingerprints

Some typical fingerprint distributions for individual users are shown in Figure 5. Across the full six months, the three most common fingerprints for a given user typically accounts for at least 50% of all location fixes, and for some users this can be over 90%.

C. Intelligent Positioning

To test the proposed new positioning scheme, the most recent six months of data from each user were selected for analysis. Two sets of analyses were performed. In the first, the system was trained for three months, then tested for three months. In the second, the system was trained for one month, then the following five months were each tested separately. In neither case was dynamic retraining during the testing phase employed. This allowed an analysis of the effect of the age of the training data to be performed.

The results of the long training test (three months) are shown in Figure 6. In this case the majority of users experienced a reduction in the number of channels scanned for localisation of over 70%, with 10% of users experiencing a reduction over of 90%.

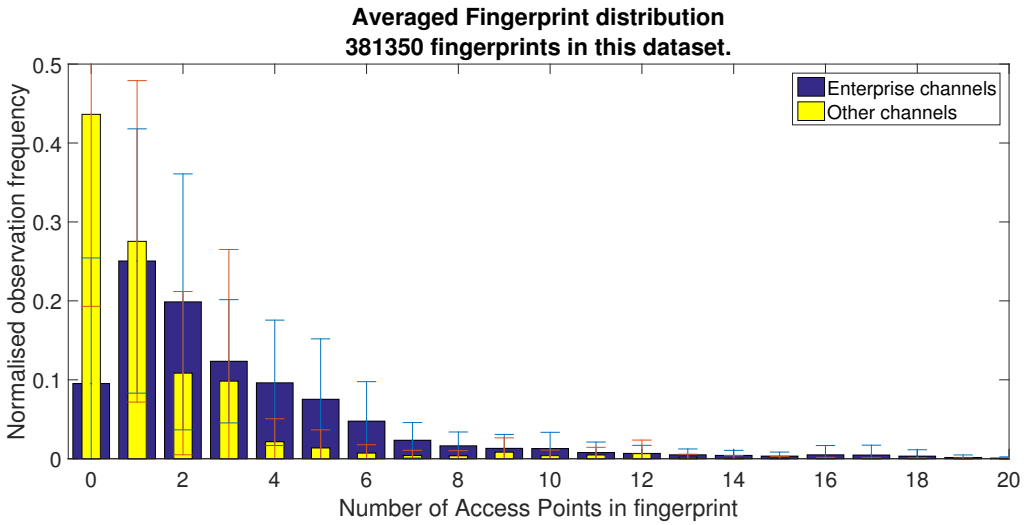


Fig. 3. The distribution of fingerprint sizes, averaged across all uses, and separated into Enterprise and non-Enterprise channels. This distribution is across all time (i.e. any given access point is counted as many times as it is seen)

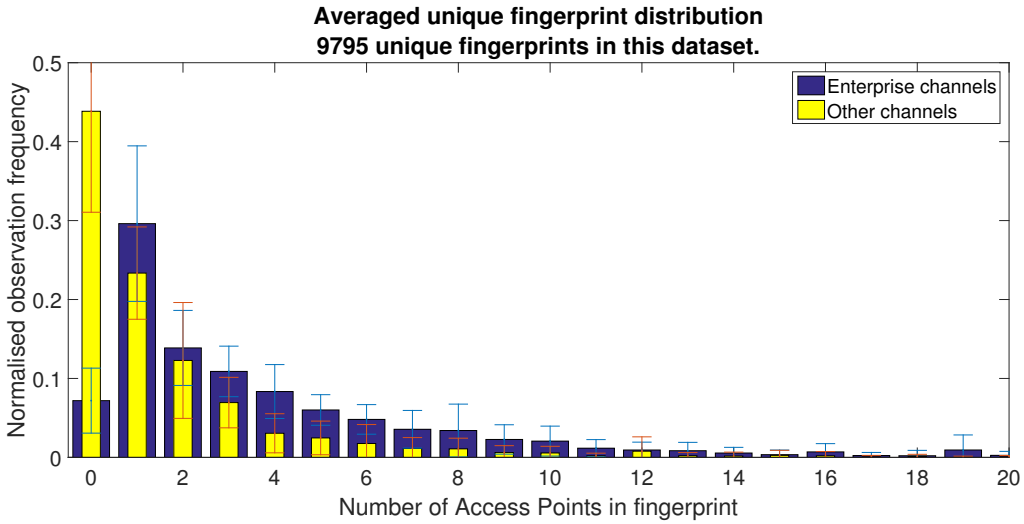


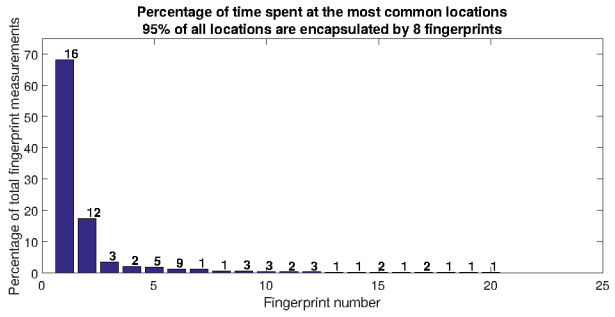
Fig. 4. The distribution of fingerprint sizes, averaged across all uses, and separated into Enterprise and non-Enterprise channels. This distribution is based on unique views; any given access point is only counted once)

The ageing test, shown in Figure 7, demonstrated a slight decline in system performance over time. Dynamic retraining is recommended to ensure good performance throughout the year (i.e. accounting for holidays and time spent away from home), in regions where Cell-ID reallocation and BSSID reallocation is common, and to account for infrequent changes in common locations such as changing home or employment.

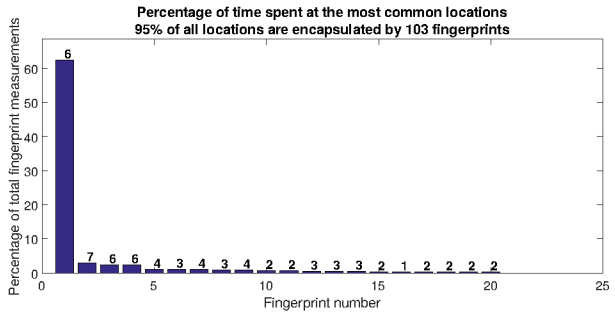
Both of these tests demonstrated that for the majority of users, this new method can reduce the number of Wi-Fi channels needing to be scanned for background positioning to at least half of the current level. Brouwers et al. demonstrated that a full WiFi scan across all 2.4 GHz and 5 GHz channels draws around 100 mA of current for around 3.5 seconds [1], and that there is a linear relationship between number of channels scanned and the energy consumed. The age of devices tested by Brouwers suggests that his tests were for passive Wi-Fi scans. His findings suggest that the new intelligent channel scanning method proposed here can provide energy savings of around 70% to 90% for the majority of users.

D. Cell-ID and BSSID variations

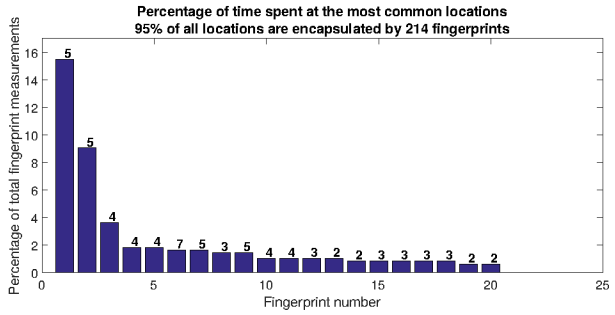
A serious issue associated with opportunistic radio positioning using Cell-ID and Wi-Fi BSSIDs is that of identity reallocation. The ability to perform fingerprinting relies strongly on the ability to correctly identify base stations by their broadcast identity, and if these identities are regularly changed then databases must be refreshed at a higher rate in order to be usable. In the past it has been known for cellular operators to infrequently remap their network in a region and to change all of the LAC and Cell-ID fields for each base station on a timescale of around once per year. However, there is evidence within the Device Analyser datasets that some operators now dither their Cell-IDs over timescales of hours (or minutes) not months.



(a)



(b)



(c)

Fig. 5. Example fingerprint distribution plots, ordered by frequency, from three randomly-selected users. The number above each bar shows the number of Cell-IDs that have been recorded at that particular Wi-Fi fingerprint location. Note the different y-axis scalings for each plot

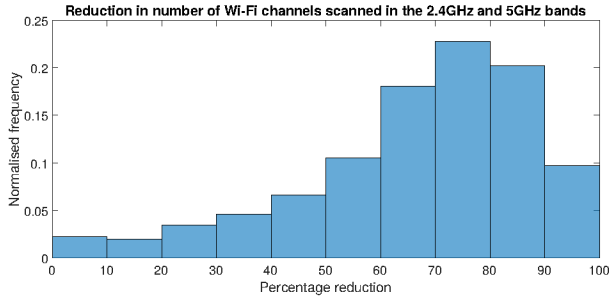
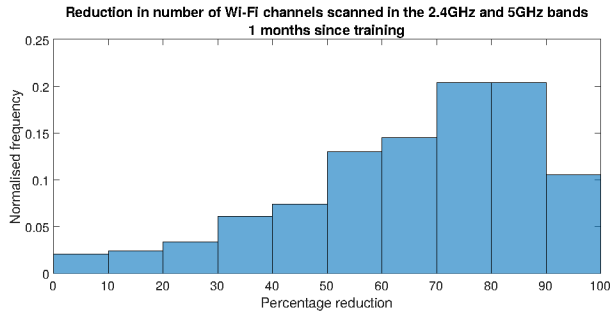
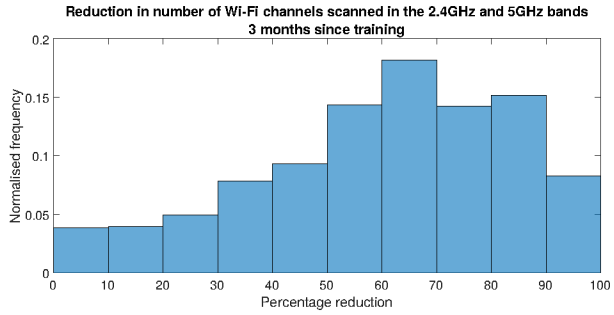


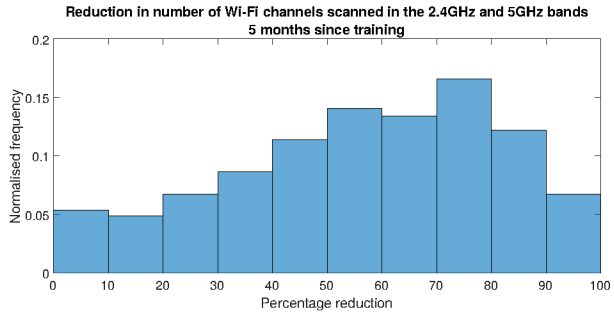
Fig. 6. The reduction in number of channels scanned to determine location for the three month training and three month trialling test dataset of 1,317 users.



(a)



(b)



(c)

Fig. 7. The performance of the low power fingerprinting scheme reduces slightly as the time since training increases, with very similar performance evident after one month (a), three months (b), and after five months (c). In these tests only one month of training was employed.

Here we consider Cell-IDs for an example device in our dataset. We firstly partition the data into periods in which it is likely the device is stationary. Our criteria for a phone to be stationary are that it must be plugged into an A/C charger, the screen must be off and any Wi-Fi fingerprints must not change. We only consider stationary periods of longer than 5 minutes. For 34% of these stationary periods the device sees more than 5 Cell-IDs. Figure 8 shows the of rate Cell-ID switching for this device during stationary periods. We can see that a switching rate of 1 Cell-ID per minute is not particularly uncommon.

Changing the identity of cellular masts so frequently would seem to have little purpose other than to hinder the crowd-sourced use of the ID for location [3]. Large enterprise deployments of Wi-Fi access points could also be regularly remapped by the system provider for the same reason. These changes will impact all fingerprint-based indoor positioning schemes based on crowd sourcing and opportunistic surveying.

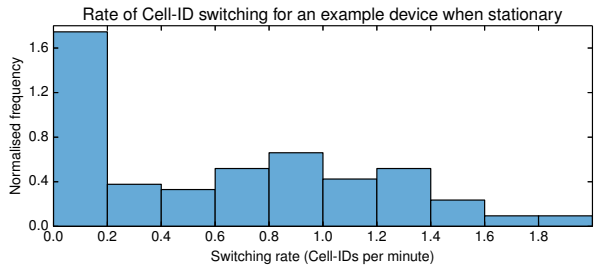


Fig. 8. The rate of Cell-ID switching for an example device when stationary

V. Conclusion

A new Wi-Fi fingerprinting scheme has been described that can reduce the number of channels scanned, and therefore the energy consumed, by background Wi-Fi position fixes by over 70% for half of those tested, and over 90% for 10% of the test users. The new method was tested using over 1,000 six-month datasets provided by Android contributors to the Device Analyser project. The new method involves a period of learning where a lookup table is generated for each user in order to provide future efficient location estimates. The lookup tables combine Cell-ID, Wi-Fi radio channel and Wi-Fi MAC address details to allow a device to scan the most populated channel observed for a given Cell-ID in order to efficiently search for known MAC addresses in order to confirm user location with minimal radio channels scanned.

The study also revealed evidence of telecommunication operators re-allocating Cell IDs over short timescales of minutes and hours. This is likely to be an attempt to prevent crowdsourcing methods of opportunistic fingerprinting such that the telecommunication operators are the sole providers of their location information. If this trend is also adopted by companies that provide large scale domestic and commercial Wi-Fi deployments then the accuracy and integrity of crowdsourced signal fingerprint databases will be adversely affected.

VI. Acknowledgements

We gratefully acknowledge the help of the Device Analyzer team. In particular, we thank Daniel Thomas for his assistance working with the analysis code and processing infrastructure.

References

[1] N. Brouwers, M. Zuniga, and K. Langendoen. Incremental wi-fi scanning for energy-efficient localization. In *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*, pages 156–162, March 2014.

[2] R. Faragher and R. Harle. An analysis of the accuracy of bluetooth low energy for indoor positioning applications. In *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014), Tampa, Florida, September 2014*, pages 201–210, 2014.

[3] Michal Ficek, Nathaniel Clark, and Lukáš Kencl. Can crowdsensing beat dynamic cell-id? In *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, PhoneSense '12, pages 10:1–10:5, New York, NY, USA, 2012. ACM.

[4] Marta C González, César A Hidalgo, Albert-László Barabási, M C Gonzalez, and A L Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):479–482, June 2008.

[5] Kyu-Han Kim, Alexander W Min, Dhruv Gupta, Prasant Mohapatra, and Jatinder Pal Singh. Improving energy efficiency of wi-fi sensing on smartphones. In *INFOCOM, 2011 Proceedings IEEE*, pages 2930–2938. IEEE, 2011.

[6] Arvind Thiagarajan, Lenin Ravindranath, Hari Balakrishnan, Samuel Madden, Lewis Girod, and Others. Accurate, low-energy trajectory mapping for mobile devices. *Proceedings of the 8th USENIX conference on Networked systems design and implementation*, pages 20–20, 2011.

[7] Daniel T. Wagner, Andrew Rice, and Alastair R. Beresford. Device analyzer: Large-scale mobile data collection. *SIGMETRICS Perform. Eval. Rev.*, 41(4):53–56, April 2014.