IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

A Case Study of WiFi Sniffing Performance Evaluation

YAN LI[1], JOHAN BARTHELEMY[1], SHUAI SUN[2], Pascal Perez[1], and Bill Moran[3] (Member, IEEE)

[1]SMART Infrastructure Facility, University of Wollongong, Wollongong, NSW 2522, Australia
[2]Department of Electrical and Telecommunication Engineering, RMIT University, Melbourne, VIC 3000, Australia
[3]Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia

Corresponding author: Yan Li (e-mail: liyan@uow.edu.au).

ABSTRACT Mobile devices regularly broadcast WiFi probe requests in order to discover available proximal WiFi access points for connection. A probe request, sent automatically in the active scanning mode, consisting of the MAC address of the device expresses an advertisement of its presence. A real-time wireless sniffing system is able to sense WiFi packets and analyse wireless traffic. This provides an opportunity to obtain insights into the interaction between the humans carrying the mobile devices and the environment. Susceptibility to loss of the wireless data transmission is an important limitation on this idea, and this is complicated by the lack of a standard specification for real deployment of WiFi sniffers. In this paper, we present an experimental analysis of sniffing performance under different wireless environments using off-the-shelf products. Our objective is to identify the possible factors including channel settings and access point configurations that affect sniffing behaviours and performances, thereby enabling the design of a protocol for a WiFi sniffing system under the optimal monitoring strategy in a real deployment. Our preliminary results show that four main factors affect the sniffing performance: the number of access points and their corresponding operating channels, the signal strength of the access point and the number of devices in the vicinity. In terms of a real field deployment, we propose assignment of one sniffing device to each specific sub-region based on the local access point signal strength and coverage area and fixing the monitoring channel belongs to the local strongest access point.

INDEX TERMS Channel configuration, probe requests, passive WiFi sniffer, WiFi monitor mode

I. INTRODUCTION

SMART phones have evolved into promising mobile computing platforms enhanced by strong communication capability, network access and multi-function embedded sensors [1]. They are able to provide critical information of human mobility behaviours through the environment. Advances in wireless communication and the consequent ubiquity of WiFi infrastructure provide the ability to extract human-related information, such as location, movement and other activity by analysing wireless connectivity between mobile clients and the Access Points (APs).

In the light of IEEE 802.11 standard, a mobile device can perform two discovery modes to access to the network: passive scanning and active probing [2]. In passive mode, APs are broadcasting beacons to advertise services to clients in range. Mobile clients passively listen on each channel waiting for a periodic beacon from nearby APs. The scan is completely passive and negatively impacted by a high discovery delay, which can result in a latency of 1150 milliseconds for a full scan of 11 channels in the 2.4 GHz band [3], [4]. Alternatively, in active probing mode, the mobile devices continuously send probe requests searching for previously associated networks for auto reconnectivity. Active scan is preferred over passive scan as the time required to scan each channel is 8 ms [3]. These probe request packets carry the unique MAC address of the mobile device in a clear text. In other words, the mobile devices are consistently broadcasting their presence and identification when searching for available WiFi network in range. This offers the opportunity to obtain location information related to mobile users. Acquiring such information is useful in a diverse range of applications such as occupancy estimation [5]–[7], traffic flow monitoring [8], crowd mobility analysis [9]–[13] and building management optimization [14]–[16].

Non-encrypted probe requests can be captured by placing low-cost scan devices in the environment, referred to as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

WiFi sniffers. There are many such small and portable WiFi sniffers available using off-the-shelf hardwares. Such capability is easily achieved by using a linux laptop, enabling the monitor mode on a wireless network interface card (NIC) and installing a linux-based packet capturing software (Wireshark [17]), which allows to capture all traffic between adjacent APs and client devices, without network connection or data transmission, shown in Figure 1.

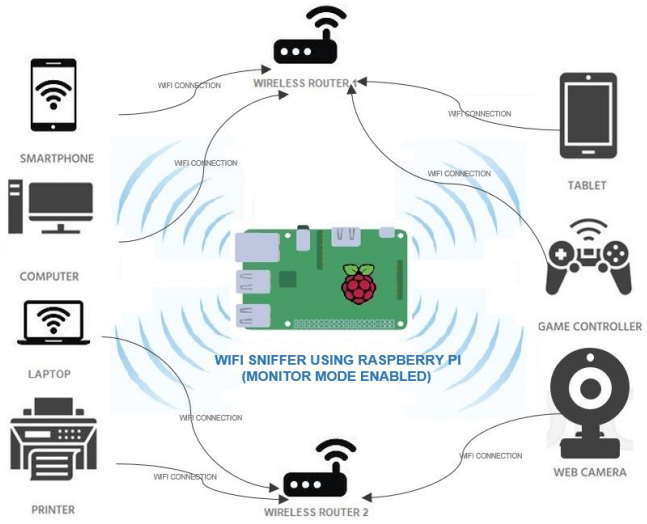


FIGURE 1. A sniffer uses Raspberry Pi to collect Wi-Fi probe requests broadcasted by all nearby wireless devices on 802.11b/g/n channels.

Much past work in the literature has successfully leveraged WiFi sniffers for passive collection of WiFi packets. However, there is no standard for the channel to be used or how to listen for the packets pertaining to implementation details. Probe requests are sent in bursts across multiple channels successively when the device searches for a nearby network. The frequency of probes varies according to multiple factors such as devices manufacturer, operating system and screen status. A device tries to conserve power as much as possible and sets a low frequency of probe request when it is low on battery [18]. Considering all those restrictions, the WiFi sniffer system should functionally cover multiple channels to receive more frames simultaneously.

It is certain that increasing sniffer density would obtain more data [19]. Typically in the 802.11b/g/n (2.4GHz) environment, using a three channel sniffer is suggested [20]. This involves using 3 wireless adapters (antennas) on the sniffing device, with the antennas set to channel 1, 6 and 11; however, this increases cost and design complexity of such devices. Aside from changing the number of sniffers or the coverage area, we are particularly interested in establishing and understanding the principles of such a channel monitoring scheme using only one WiFi module, with a focus on achieving optimal sniffer performance and, accordingly, satisfying appropriate levels of application accuracy.

Two configurations for channel sniffing are most commonly described in the literature: channel hopping, involving rapid switching between channels at a given time interval, and fixed single channel monitoring. An explicit test campaign focused on the use of three non-overlapping channels found that fixed channel monitoring captures more packets than channel hopping [21]. In this work, we further extend this body of research and explore what factors should be taken into consideration in order to maximize the packet collection in a real field deployment.

To this end, we have conducted experiments using different channel monitoring schemes. Our preliminary results showed that: **1.** the duration of the channel hopping interval makes a significant difference to the total number of packets collected and the number of devices detected when the sniffing device is hopping over the three non-overlapping channels. Increasing the time interval spent on each channel benefits the detection of more devices; **2.** fixed channel monitoring captures more packets than hopping channels in most cases; **3.** the connected device is more likely to send direct probe requests on the channel selected by the local strongest AP ; **4.** the total number of packets received from connected devices is highly dependent on the AP configurations, pertaining to the number of supported channels, the corresponding signal strengths of the APs and the number of devices in proximity; **5.** in general, the optimal fixed channel for monitoring should be the channel of the strongest AP in each sub-area.

Our main contributions are listed as follows:

- We have compared the performance between 4 different channel hopping strategies reported in the literature and fixed channel sniffing. We have also evaluated the impact of varying the length of the channel hopping interval in a standard frequency hopping mechanism.
- We have compared the sniffing performance between Raspberry Pi and LoPy4, in terms of numbers of packets, number of devices captured and received signal strength (RSS) levels recorded.
- We have conducted tests in different wireless environments to investigate the possible factors that affect the received number of packets, including the number of APs and their signal strength, the number of devices in proximity and the channel utilization status.

The remainder of the paper is organized as follows: Section 2 briefly presents the background on WiFi active scanning mode and relevant work in the field. Section 3 describes the experiment setup and the experimental results to verify the how different configurations affect the sniffing performance. Additional aspects of the limitation and performance are discussed in Section 4. Finally, conclusions are drawn in Section 5.

II. BACKGROUND AND RELATED WORK

WiFi packets transmitted between mobile devices and the wireless APs carry massive amounts of information, offering new opportunities to learn location information and mobility behaviour related to mobile users using existed WiFi infrastructure.

In conventional wireless local area networks (WLANs), client devices need to discover networks for connection using two scanning methods: passive and active. In passive scanning, client devices iterate over multiple supported channels and listen for beacon frames which are transmitted by APs to advertise their presence. The beacon interval is typically configured to be 100ms [22], which means the device may take a very long time to scan all the channels and hear the beacons broadcast from nearby APs [23], [24]. Discovering the network by scanning all possible channels and listening to beacons passively is not considered to be very efficient.

Alternatively, active scan is the recommended mechanism to enhance the discovery process and efficiently find nearby wireless networks. Client devices locally maintain a list of known networks to which the device has connected before, referred to as the Preferred Network List (PNL). Thus, client devices can perform active scanning constantly to search for a known network to connect to by sending a probe request on each channel, rather than waiting for the network to announce its availability to all the clients. The client device continues to send probe requests automatically, irrespective of an ongoing connection to an AP, in order to discover new and potentially stronger APs in its vicinity to ensure the best network connection quality to the user. By doing so, a client station can maintain and update a list of known APs [25].

There are two types of probe request frames sent by the user devices: directed probes and broadcast probes. Direct probe requests include a specified destination Service Set Identifier (SSID), only APs with a matching SSID will reply with a probe response. A broadcast probe, also referred to as a null probe request, does not target any network in particular. It triggers a probe-response from all APs for each SSID they support. Both types of probe request frames are transmitted without encryption and can be easily captured with cheap off-the-shelves sniffers. Moreover they contain unique device identifiers (MAC addresses), thereby enabling the possibility of detecting distinct devices around the sniffer and ultimately providing a measure of the occupancy status and the movement traces of the mobile user across multiple places.

Capturing probe requests frames can be simply achieved with any IEEE 802.11 compliant wireless adapter set in monitor mode while listening on specific Wi-Fi channels. Each received probe request is allowed to access typical information from the client device, namely, the source device MAC address and the RSS. The source MAC address is a globally unique 48-bit string which identifies the device and whose first 3 bytes contain the Organizationally Unique Identifier (OUI) which identifies the radio chip manufacturer. RSS measures the average signal power at the receiver in decibel-milliwatts (dBm) and is primarily related to the transmission power and distance between the device and the receiver.

Collecting all wireless communications from a specific device is difficult for various reasons. First, mobile devices send the probe frames on different channels, but packet capture software (e.g., TCPDump or Wireshark) must be configured to listen on specific channels. Second, some packets may be lost due to the noisy nature of the wireless medium [21]. In order to cover as much of the spectrum as possible, the sniffer can choose to perform channel hopping, in which the wireless card is configured to listen on a channel with a designated switching time interval and then hop to another channel based on a specific hopping sequence. However, as already stated, many studies have shown that more probe requests are captured when channel hopping is not used [9], [21]. The reason lies in the fact that the wireless adapter can only capture on a single channel at any given time. It may be desirable to sniff on a single channel from among the non-overlapping channels; channels 1, 6 and 11 are non-overlapping channels in the 2.4 GHz band and most frequently used. Other studies have claimed that no intensive knowledge about the statistics on which channel is the most used, though channel 1 is commonly selected for sniffing [26]. The choice of channel is assumed not to have significant impact on the tests, but no exhaustive studies seem to have been done.

Freudiger [21] has done a comprehensive experimental study of how different factors influence the WiFi probe requests, including monitor channel configurations, number of SSIDs stored in the PNL and device configurations. It has demonstrated that three antennas with each set to a fixed non-overlapping channel collects the largest number of probes. The probing behaviour is subject to device manufacturers, where the number of probes is linearly dependent on the number of known SSIDs in general. Device with unlocked screen exhibits more probes and a forged WiFi beacon in proximity will push an increasing burst of probe requests.

In this paper, we extend the previous body of research by investigating the relative performance of different channel hopping schemes with fixed single channel monitoring. We further explore other possible factors that affect the number of received probe packets in different scenarios. It is noted that a number of factors including signal strength of the AP, channel utilization frequency and the number of devices in the area all have an effect on the number of received probes. We conclude that in a real deployment, multiple sniffers should be placed at each sub-area where the area is tessellated according to signal strength. In order to maximize the collected probe data, the optimal monitor channel should be the one associated with the strongest AP in each sub-area, rather than a choice of the three non-overlapping channels, though most APs are configured to operate on non-overlapping channels to avoid interference.

III. WIFI SNIFFING SYSTEM ARCHITECTURE

The IEEE 802.11 standard defines a set of specifications at the physical layer (PHY) and the MAC layer of the Open Systems Interconnection (OSI) model [27], enabling functionality for WLAN communications. The hardware components can all be mapped to the physical layer, in the form of electronic circuitry, media and connectors. Whereas any software required to enable 802.11 functionality maps to the data link layer.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

The MAC layer is a sublayer of the data link layer directly on top of the physical layer, which specifies the behaviours of the wireless communications [28]. The 802.11 MAC layer implements three main functions: data delivery, access control and security. In general, the IEEE MAC specification defines MAC addresses, which enables unique identification of multiple devices at the data link layer. The MAC layer manages and maintains communications between 802.11 stations (client devices and APs) by scheduling access to a shared radio channel and utilizing protocols to facilitate information transfer over a wireless medium. To support the exchanging functions between stations and APs, the 802.11 protocol defines three broad categories of MAC layer frames, which are management, control, and data frames [29].

WiFi sniffer consists of hardware and software application to demodulate the frame and display the payload conveying the WLAN PHY and MAC layer information [30], shown in Figure 2. Details of MAC layer frame format can be found in [28]. Before the basic WLAN communication is yet established between the station and the AP, the client implements an active scan for available network by broadcasting management frames (known as probe request) on every channel its physical layer supports, to which surrounding APs reply with a probe response. This is accomplished by a MAC layer management operation [31]. Figure 3 shows the MAC layer handoff scheme [32]. With a NIC placed into monitor mode, the sniffer will capture the wireless traffic in the network. In this paper, we are focusing on the analysis of the probe requests at which 802.11 MAC layer management frames are transmitted by a wireless device [33].

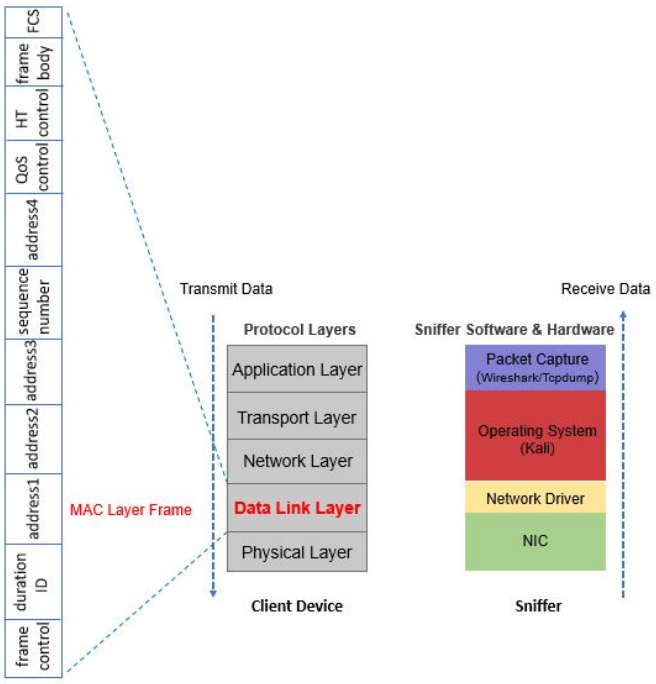


FIGURE 2. WiFi sniffer system architecture

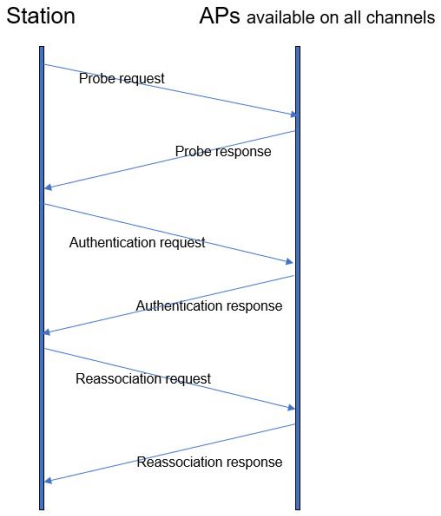


FIGURE 3. MAC layer handoff process

IV. EXPERIMENTAL EVALUATION

In order to evaluate the performance of different channel monitoring strategies and investigate the impact on the number of received probe request packets, we carried out tests in a range of spaces in a university campus, as well as in a home living environment used for working, studying, living and recreational activities.

The first indoor test took place at Level 2 of the Smart Infrastruture Facility Building, University of Wollongong. It is worth pointing out that in a standard enterprise Wi-Fi networks, only the three non-overlapping channels 1, 6 and 11 are used and APs are deployed with overlapping coverage cells in a manner that avoids Adjacent Channel Interference (ACI) and Co-Channel Interference (CCI) [34].

The second indoor test was conducted in a home-living environment with 2 manually set up APs, one in the living room and the other one in the study room located at the other end of the house. In order to distinguish the channels from the other APs in neighbourhood, the two APs are set to operate on channel 3 and 7 respectively.

A. TEST SETUP

1) Hardware and Software

We used the Raspberry Pi (RPi) 3 Model B V1.2 which offers features including 1.2GHz Quad-Core Broadcom BCM2837, 2.4GHz 802.11n wireless LAN, Bluetooth 4.1. We utilized cheap 16 GiB microSD cards as mass storage. We installed the off-the-shelf Kali Linux operating system on the RPi. The in-built WiFi module supports monitor mode, so an external USB WiFi adapter is not needed. The received data frames were captured using *tcpdump*, only probe requests were logged to persistent memory. The resulting dumps were transferred to an external computer and converted to *.pcap* files that can be opened by *Wireshark*. Since the RPi 3 model B only supports 2.4 GHz frequency ranges, we only consider

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

the 11 channels in the 2.4 GHz bands in the following discussion.

We also used Pycom LoPy4 development board for a performance comparison with RPi. This board features an Espressif ESP32 chipset which interfaces with a Xtensa dual-core 32-bit LX6 microprocessor, Bluetooth, LoRa, Sigfox and 802.11b/g/n Wi-Fi radios [35].

2) Probes Collection

Each probe request message includes the following fields:

- the source MAC address
- the OUI which identifies the radio chip vendor
- the SSID of the probe request which can be either *"Broadcast"* or *"Direct"* with a string containing the SSID of a known Wi-Fi network
- the ID of the sniffing device
- the RSS of the received probe packet
- the timestamp of the probe frame.

3) Data Pre-Processing

Most operating systems for mobile devices have now implemented MAC randomization to protect user privacy before associating to the wireless APs [36]. However, in our cases, we assume most of the devices are connected to the university wireless network, which will reveal their true MAC addresses. On the other hand, our dataset indicates that out of the 2855 MAC addresses being detected, only approximately 21% of the MAC addresses have valid, globally unique OUIs. We deduce that the remaining MAC addresses without a valid OUI, are therefore locally randomized by the operating systems. We discard the probe request packets sent from MAC address with invalid OUIs, resulting in approximate 87% of the packets for further analysis.

B. TEST SCENARIO I: UNIVERSITY ENTERPRISE WIRELESS NETWORK

1) Experiment 1: Channel Hopping Schemes

In this section, we present the comparison results of different channel configurations using the in-built WiFi module on board. We chose 4 different channel hopping strategies that can be found in the literature; these are: **1).ch1**: hopping across the 11 802.11b/g/n channels sequentially; **2).ch2**: hopping across the three non-overlapping channels ***(1, 6 and 11)***; **3).ch3**: hopping across the channels from 1 to 13 by jumping to the next non-overlapping channel ***(1,7,13,2,8,3,9,4,10,5,11,6,12)*** [37]; **4).ch4**: hopping across the specific sequence of channels ***(1,6,11,2,7,3,8,4,9,5,10)***, also refereed to as the default Kismet hopping schedule [38].

For fixed channel monitoring, we configured three sensors with each set to a non-overlapping 802.11b/g/n channel (***1,6,11***).

In general, it is expected that more packets will be received when the channel is fixed. Hopping channels over the 3 non-interval channels (**ch2**) outperforms other channel hopping schemes, and **ch4** has the worst performance in terms of probe requests collection, shown in Figure4.

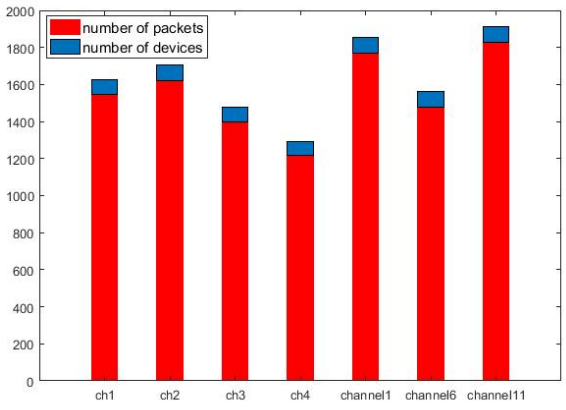


FIGURE 4. Number of devices and probes under different channel configurations

2) Experiment 2: Channel Hopping Interval Impact

In this section, we present the comparison results of channel hopping intervals under two hopping schemes; hopping across the 11 802.11b/g/n channels (**ch1**) and hopping across the three non-overlapping channels (**ch2**). The two RPis are set to perform channel hopping with a hopping interval of 0.5 seconds and 1.5 seconds respectively. The tests were conducted at the same place on different days.

As shown in Figures 5–8, the channel hopping intervals affect both the number of total packets and the number of detected devices. However, the effect of the duration of the channel hopping intervals seems to show contradictory performance under the two different hopping schemes. When hopping all the channels from *1* to *11*, it is desirable to shorten the hopping intervals to receive more packets and devices. While in the case of hopping across the three non-overlapping channels (*1, 6, 11*), increasing the channel hopping intervals is beneficial for maximization of the number of collected probes.

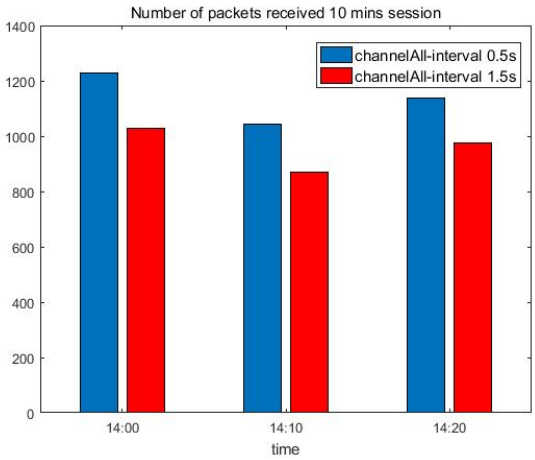


FIGURE 5. Number of packets received using channel hopping **ch1**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

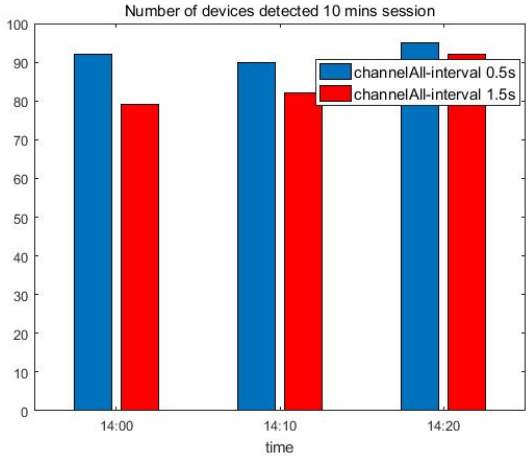


FIGURE 6. Number of devices detected using channel hopping **ch1**

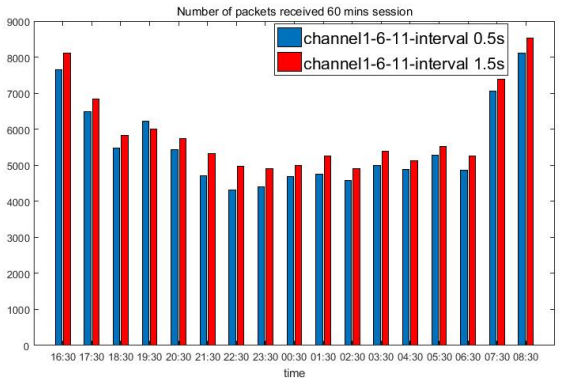


FIGURE 7. Number of packets received using channel hopping **ch2**

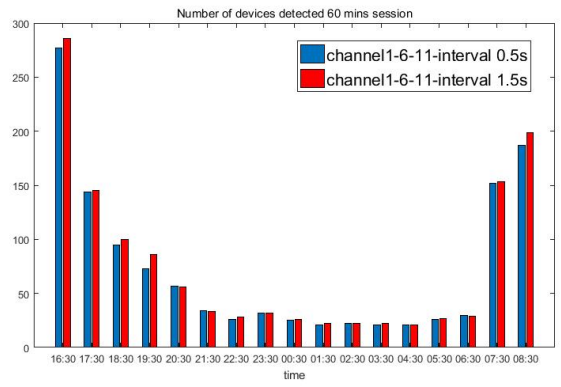


FIGURE 8. Number of devices detected using channel hopping **ch2**

3) Experiment 3: Performance Comparison between Raspberry Pi and LoPy4

In this section, we compared the performance between RPi and LoPy4 in terms of total number of probes, signal strengths and number of packets captured from each user device.

The RPi and LoPy4 were both set to hop across the 11 channels (**ch1**) with an hopping interval of 0.5 seconds. We have recorded the MAC addresses from 9 user devices in order to compare the the RSS levels recorded by the two different sensors. It is noted that the MAC addresses and the brands of the 9 user devices are the only prior information we obtain, their locations and phone status are preserved to user privacy.

During a 60 minutes test, the RPi has captured 7137 packets from 215 MAC addresses with valid OUIs, whereas the LoPy4 detected 221 valid devices emitting 4255 probes, which is about 60% of the number of packets received by RPi. Figure 9, 10 show the number of probes and devices detected in each 10 minutes time period. We believe the antenna of the WiFi module has similar range for both types of sensors.

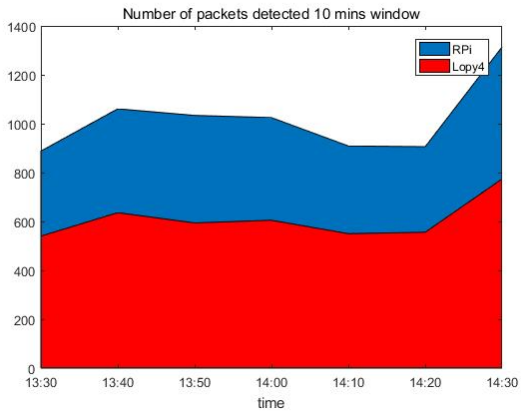


FIGURE 9. Number of packets captured by RPi VS. LoPy4 (**ch1**)

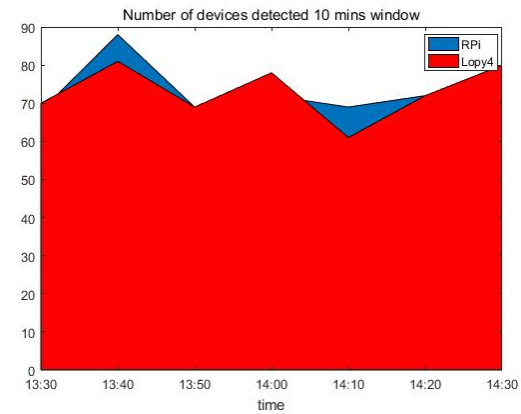


FIGURE 10. Number of devices detected by RPi VS. LoPy4 (**ch1**)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

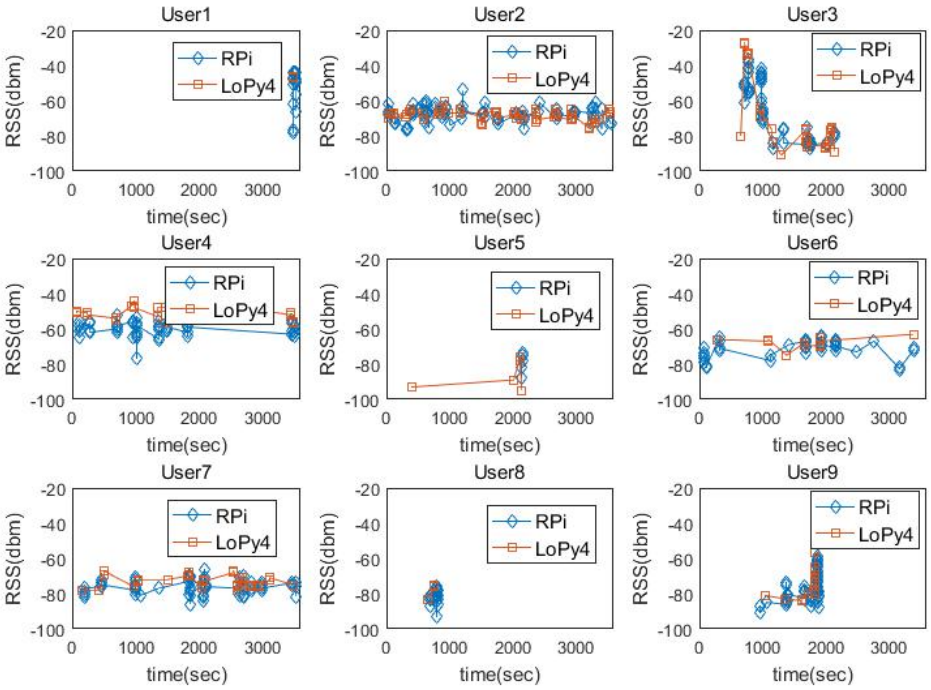


FIGURE 11. RSS recorded by RPi VS. LoPy4 for each User Device.

TABLE 1. Number of Probes from User Devices (RPi VS. LoPy4)

User Device	Brand	RPi	LoPy4
1	Apple	36	8
2	Samsung	115	96
3	Motorola	68	38
4	Samsung	154	61
5	Apple	8	6
6	Apple	102	32
7	Apple	94	44
8	Motorola	18	3
9	Huawei	82	21

In terms of number of probe requests from each user device, LoPy4 is more susceptible to loss data compared to RPi, see Table 1. Though the RSS values recorded by both sensors are at similar level, shown in Figure 11.

C. TEST SCENARIO II: LIVING HOME WIRELESS NETWORK

We set up two routers (Linksys EA6900 AC1900 Dual-Band Wi-Fi Router and ASUS RT-AC68U Wi-Fi Router) in a simple home environment with floor plan as shown in Figure 12. It is noted that the two APs have emerged as the strongest signal strength transmitters covering the whole area but the sniffer also captures the wireless packets from the neighbouring wireless networks within antenna range. AP1 is set to channel 3 at living room and AP2 is set to channel 7 at study room respectively in order to be distinguishable from the other sensed APs in the neighbourhood, which normally operate on channel 1, 6 and 11. Accordingly, the 3 Raspberry Pis are configured to monitor each of the fixed single channel 1, 3 and 7 respectively.

FIGURE 12. Floor plan of a living home environment and AP deployment

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

There are about 20 wireless devices connected to AP1, including wireless adapters, laptops, smart cameras, smart light bulbs, Google Home Mini, tablets and mobile devices. Most of the wireless devices are in the study room near AP2 while only 3 devices (one smart camera, one smart light bulb and one Google Home Mini) are in the living room close to AP1. For simplicity, we focus on one user device, an Apple IPhone XR and connect it to AP1.

1) Test in the Study Room

The sniffers and the IPhone are both placed in the study room. AP2 operates on channel 7 presenting the strongest signal strength of -31 dBm while AP1 works as the associated AP with a signal strength of -59 dBm. There is another AP in the neighbourhood operating on channel 1 with a signal strength of -87 dBm.

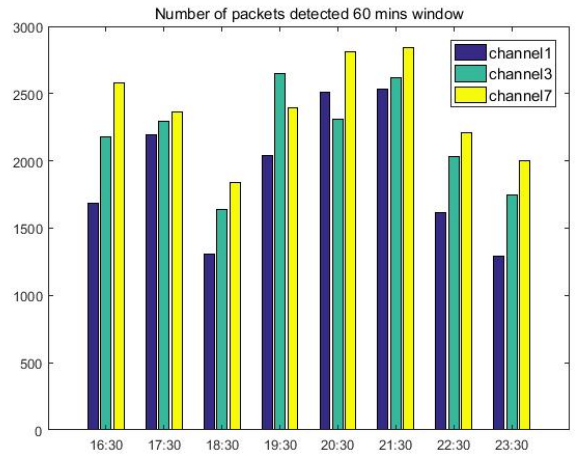


FIGURE 13. Number of packets received in the study room

In most cases, channel 7 received most of the packets that the locally strongest AP (AP2) operates on. For the user device, we notice that the phone tends to send more probe packets over the channel to its locally strongest AP when its associated AP signal is weaker. Accordingly Channel 1 corresponds to the least number of packets in most cases, shown in Table 2.

TABLE 2. Number of Probes from User Devices (Study)

Iphone	Channel 1	Channel 3	Channel 7
Direct	408	504	548
Broadcast	163	120	188
Total	571	624	736

We also perform the analysis of variance (ANOVA) to statistically determine whether the impacts are significantly different between the channels. Specifically, we define the null hypothesis as: $H_0 : \mu_1 = \mu_3 = \mu_7$, where μ_i is the average number of captured packets on channel $i, i \in \{1, 3, 7\}$. At this point, three main assumptions are made: *1.* the number of packets captured on each channel is normally distributed; *2.* the homogeneity of variances; *3.* independence of observations. In this paper, we use *Lilliefors* test for normality and the *Bartlett* test for homogeneity of variance. The *Lilliefors* test shows that the captured packet number comes from normal distribution at the default 5% significance level. The p value of the *Bartlett* test is 0.5291 indicating variances are equal across the captured packet number on different channels. The ANOVA report a statistic result of $(F(2, 45) = 4.71, P = 0.0139)$, so we reject the null hypothesis at the 5% significance level and conclude that the channels have significantly different behaviours. The notched box plot is shown in Figure 14 indicating the confidence interval of the median.

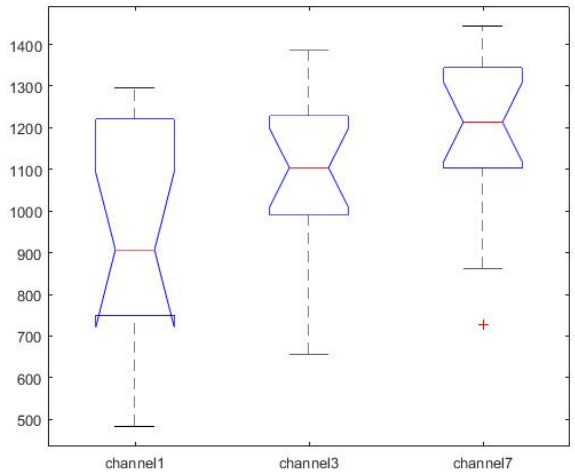


FIGURE 14. Box plot of number of packets (study room)

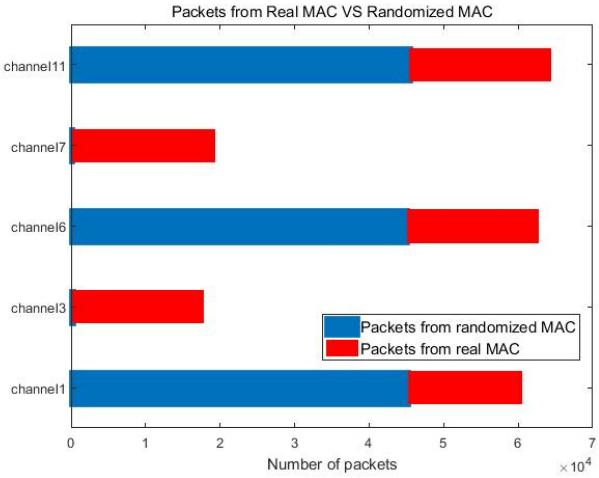


FIGURE 15. Number of packets from real MAC VS randomized MAC (study)

Figure 13 shows the total number of packets received from devices with globally unique OUIs. Figure 15 further

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

presents the comparison between the number of packets received from devices with real MAC addresses and devices with randomized MAC addresses on each channel. As mentioned in section IV-A3, we manually discarded packets emitted from locally assigned MAC addresses for further analysis. However, the dramatically different characteristics in terms of the number of packets from randomized MAC addresses on the *not-in-proximity* channels is noteworthy.

Channel 1, 6 and 11 have collected 45263, 45149 and 45481 packets from 39, 27 and 23 devices with invalid OUIs respectively. While channel 3 has observed 22 devices with locally assigned MAC addresses but only received 246 such packets. Similarly channel 7 has collected 166 packets from 17 devices with invalid OUIs, as shown in Table 3.

TABLE 3. Number of Devices & Packets with Invalid OUIs (Study)

Channel	Devices (locally assigned MAC)	Packets from invalid OUIs
1	39	45263
3	22	246
6	27	45149
7	17	166
11	23	45481

Figure 16 presents the histogram probability distribution of the RSS values of packets with invalid MAC addresses, with a strong average signal strength of -47dBm, which demonstrates that most devices sending fake MAC addresses over channel 1, 6 and 11 are located within the house. Recall that the APs operating on channel 1, 6 and 11 are all from neighbourhood which are *not-in-proximity*, when the devices actively scan for available APs around, they tend to send fake MAC addresses before they get associated with an AP. This also explains channel 3 and 7 receive much less packets with invalid MAC addresses.

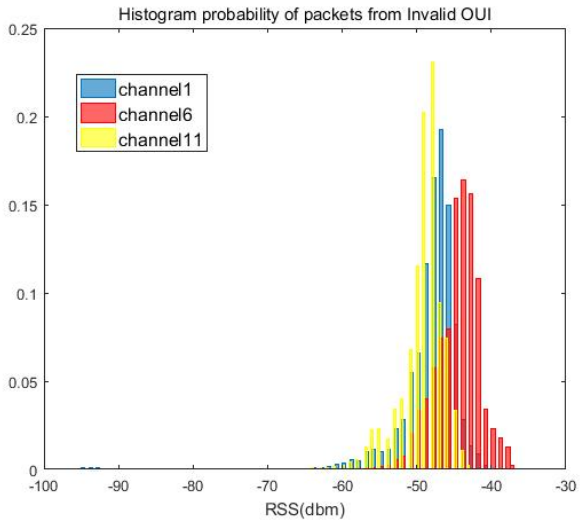


FIGURE 16. Histogram plot of packets from randomized MAC (study)

2) Test in the Living Room

The sniffers and the Iphone are placed in the living room close to AP1.

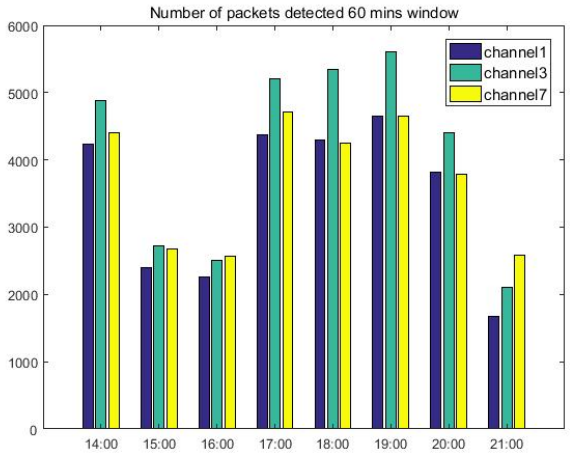


FIGURE 17. Number of packets received in the living room

TABLE 4. Number of Probes from User Devices (Living)

Iphone	Channel 1	Channel 3	Channel 7
Direct	250	361	289
Broadcast	196	164	206
Total	446	525	495

The Iphone is connected to AP1 which also acts as the local strongest AP. Channel 3 outperforms the other two channels in terms of total number of probing packets and data loss for individual device, shown in Figure 17 and Table 4, which matches the test results of the study room, namely, that monitoring the channel which belongs to the local strongest AP achieves the best sniffing performance.

3) Test in the Front Yard

The sniffers are placed outside the house in the front yard and most of the wireless devices are in the study room near AP2. Although AP1 is physically closer to the sniffer with an average signal strength of -64 dBm, due to the signal attenuation and the multipath effects, the signal strength from other APs in the neighbourhood are also at a comparable level (-74 dBm). AP2 is also in the visible range of the sniffer with a weak signal strength of around -85 dBm.

An interesting observation is that in the first hour of the experiment, channel 3 captures most packets as expected, surprisingly channel 3 exhibits a significant probing drop in the next one and half hours of the test regardless of the signal strength, shown in Figure 18. The major reason is that channel 3 suffers a significant drop of around 40% of the number of devices captured in the last one and half hour, bringing the detected number of devices from 22 down to 14. The is due to the neighbouring devices leave the test area,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

shown in Figure 19. Although the number of packets on each channel are all decreased due to the drop of detected devices, channel 7 maintains the sniffing performance. Actually the devices close to AP2 (operating on channel 7) were configured to either watch Youtube videos or play online games during the last hour of the test, which contributes to the wireless traffic on channel 7. This is reasonable in the light of the active discovery mechanism, whereby the device is always searching for local stronger APs in order to ensure the connection quality. It has demonstrated that both the signal strength of the AP and the number of devices in proximity impact the sniffing performance. In particular, the signal strength of APs plays the primary role in preserving the number of captured packets, whereas the number of devices in proximity also contributes to affect the sniffing performance when the environment around each AP dramatically changed, as presented in the last one and half hour of the test.

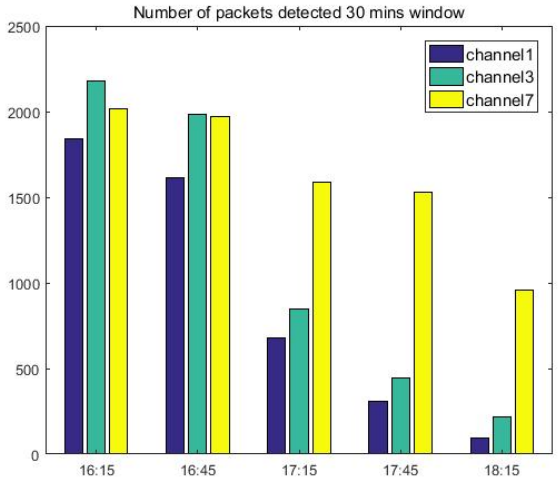


FIGURE 18. Number of packets received in the front yard

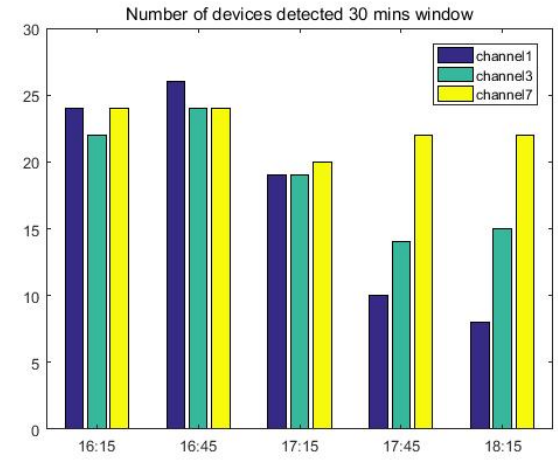


FIGURE 19. Number of devices received in the front yard

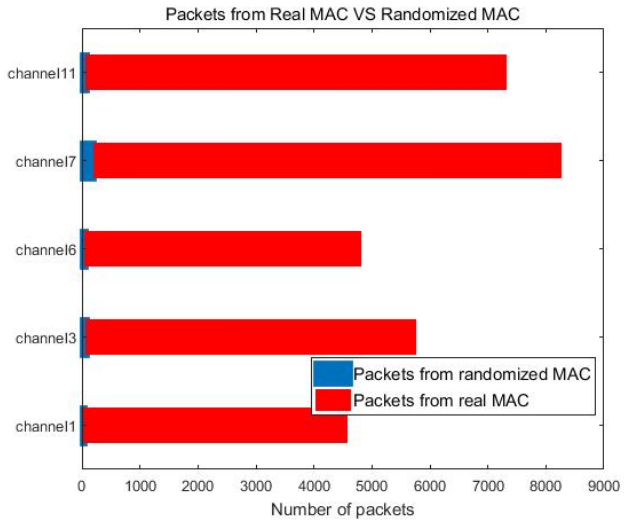


FIGURE 20. Number of packets from real MAC VS. randomized MAC (Front yard)

In terms of the number of packets from randomized MAC addresses, Figure 20 presents different characteristics comparing to Figure 15. All the monitoring channels have shown a significant decrease in the number of packets from devices with randomized MAC addresses. The possible reason might be the location of the sniffed devices. Figure 21 provides further evidence for this assumption on the basis of RSS measurements. Recall that in active discovery mode, the wireless device sends probe requests on each channel to search for available proximal APs, but only reveal its true MAC address when it becomes associated with an AP. Regarding the tests carried out at the study and living room, most of the client devices detected by the sniffer are the devices within the house with an average RSS of -49 dBm, with a preference of sending probe requests to either the associated AP or the strongest AP with their true MAC addresses. In other words, when sniffing the probe requests sent over the channels other than 3 and 7 in this case, a large number of packets are received without the globally unique MAC address. This also explains the test results in the outdoor area where the most detected devices are neighbouring devices which connect to their local APs (normally operate on channel 1, 6 and 11), thus most of the probe request packets sniffed on channel 1, 6 and 11 contain true MAC addresses with valid OUIs.

From the tests conducted in a rather simple home wireless environment, we may conclude that, when the sniffer is sufficiently near the AP so that its signal strength is significantly stronger than those from all other APs sensed within the range ($\geq$ $20dBm$), it will typically receive the largest number of packets on the channel that the local strongest AP operates on.

However, if there are no significant strong APs close to the sniffer, the number of packets is also highly related to the wireless communication activity regardless of signal strength, including the number of APs within the range, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3008533, IEEE Access

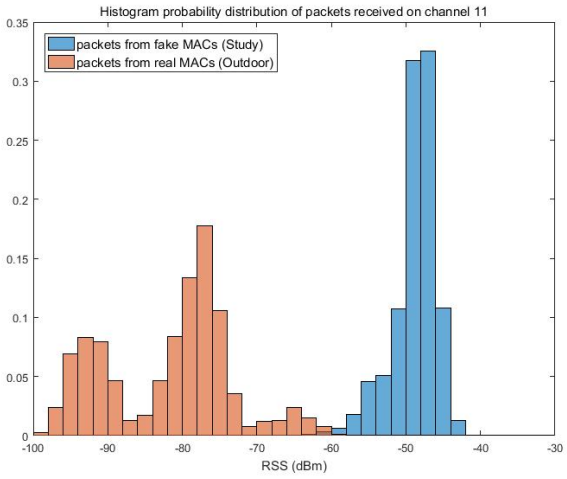


FIGURE 21. Histogram probability distribution of the packets collected at the study room and the front yard

number of devices around each visible AP and the device status.

V. DISCUSSION

- MAC randomization.
  Most of the mobile devices perform MAC randomization as a privacy-preserving feature in active discovery mode; discussion of this is beyond the scope of this paper. In this work, we make reasonable assumption that most of the devices are connected to the WiFi network in each of the working and living environments (i.e. university and home) and this will always result in revelation of the real MAC address in the probe request packets. Therefore MAC randomization is expected to have very little impact on the analytic results in this paper.
- During the tests, it has been observed that some devices do not send direct probes under default factory settings. Moreover, some advanced home routers will automatically adjust their transmission channels based on current channel occupancy and interference, rather than remain on a fixed channel. As for the tests conducted in a house, we manually set the channels to avoid interference from other APs in the neighbourhood on the same channels. Nonetheless, under the channel allocation regime at 2.4GHz, there are only three non-overlapping channels, so interference is likely regardless in our results.
- We have demonstrated the possible factors that affect the number of received probe data based on the data collected in a relatively simple home wireless environment where the AP configurations are simple. Because of lack of knowledge of the wireless environment in the neighbourhood, we can only investigate the data based on what is known about the environment, for example, number of devices, connection status and phone activity. As for the dynamic environment in an enterprise WLAN deployment (such as university network), those factors also impact the sniffing performance. However, the situation is more complicated, we believe there are other factors such as channel capacity and link quality that affect the overall sniffing performance. For example, commercial WiFi APs normally support automatic detection of surrounding interference and apply a radio calibration algorithm that allows dynamic channel selection and power adjustment to minimize such kinds of interference [39]. In addition, most of the client devices are mobile, typically being carried around by humans walking around, so that seamless roaming between APs should be taken into consideration. Moreover, client roaming decision is subject to vendor-specific configurations, including the signal strength, communication quality, error rate and missing probes etc. Therefore, it is suggested to automatically adjust the sniffing channels according to the WiFi AP configurations.
- Limitation of WiFi sniffer.
  1) some people may not carry devices with a wireless interface;
  2) some devices may not have their WiFi enabled;
  3) some people may have more than one wireless device;
  4) some devices might have multiple WiFi adapters;
  5) some transmissions may not be detected, as the mobile device passes through different areas quickly while the probe request frequency is relatively low.

VI. CONCLUSION

In this paper, we have investigated the performance of WiFi sniffers under different channel configurations using off-the-shelf products in different wireless scenarios. We conduct the ANOVA to statistically analyse the sniffing impacts between channels. We also further investigate the probing behaviours over **not-in-proximity** channels which exhibits a large number of probes with randomized MAC addresses. This research proposes a WiFi sniffer protocol using the optimal monitoring channel. We have demonstrated that the number of received probe packets are affected by a range of factors, among which the number of APs and their corresponding operating channels, the signal strength of the AP and the number of devices in the vicinity play significant roles. In a real deployment, it is suggested to assign one sniffer as close as possible to the AP in each sub-area and fix the monitor channel to be the one that the local strongest AP operates on.

ACKNOWLEDGMENT

This research is supported by the SMART Infrastructure Facility, University of Wollongong, Australia.

REFERENCES

[1] P. Hafner, T. Moder, M. Wieser, and T. Bernoulli, “Evaluation of smartphone-based indoor positioning using different bayes filters,” in

International Conference on Indoor Positioning and Indoor Navigation. IEEE, 2013, pp. 1–10.

[2] G. Bartlett, J. Heidemann, and C. Papadopoulos, "Understanding passive and active service discovery," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, pp. 57–70.

[3] I. Purushothaman and S. Roy, "Fastscan: a handoff scheme for voice over ieee 802.11 wlans," *Wireless Networks*, vol. 16, no. 7, pp. 2049–2063, 2010.

[4] Y. A. Powar and V. Apte, "Improving the ieee 802.11 mac layer handoff latency to support multimedia traffic," in *2009 IEEE wireless communications and networking conference*. IEEE, 2009, pp. 1–6.

[5] E. Vattapparamban, B. S. Çiftler, I. Güvenç, K. Akkaya, and A. Kadri, "Indoor occupancy tracking in smart buildings using passive sniffing of probe requests," in *2016 IEEE International Conference on Communications Workshops (ICC)*. IEEE, 2016, pp. 38–44.

[6] L. Mikkelsen, R. Buchakchiev, T. Madsen, and H. P. Schwefel, "Public transport occupancy estimation using wlan probing," in *2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*. IEEE, 2016, pp. 302–308.

[7] B. S. Ciftler, S. Dikmese, I. Güvenç, K. Akkaya, and A. Kadri, "Occupancy counting with burst and intermittent signals in smart buildings," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 724–735, 2017.

[8] P. Fuxjaeger, S. Ruehrup, H. Weisgrab, and B. Rainer, "Highway traffic flow measurement by passive monitoring of wi-fi signals," in *2014 International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, 2014, pp. 396–401.

[9] A. Basalamah, "Crowd mobility analysis using wifi sniffers," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 7, no. 12, pp. 374–378, 2016.

[10] H. Hong, C. Luo, and M. C. Chan, "Socialprobe: Understanding social interaction through passive wifi monitoring," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2016, pp. 94–103.

[11] J. Scheuner, G. Mazlami, D. Schöni, S. Stephan, A. De Carli, T. Bocek, and B. Stiller, "Probr-a generic and passive wifi tracking system," in *2016 IEEE 41st Conference on Local Computer Networks (LCN)*. IEEE, 2016, pp. 495–502.

[12] M. W. Traunmueller, N. Johnson, A. Malik, and C. E. Kontokosta, "Digital footprints: Using wifi probe and locational data to analyze human mobility trajectories in cities," *Computers, Environment and Urban Systems*, vol. 72, pp. 4–12, 2018.

[13] Y. Chon, S. Kim, S. Lee, D. Kim, Y. Kim, and H. Cha, "Sensing wifi packets in the air: practicality and implications in urban mobility monitoring," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014, pp. 189–200.

[14] X. Li, X. Liu, and Z. Qian, "Towards an occupancy-enhanced building hvac control strategy using wi-fi probe request information," in *Computing in Civil Engineering 2017*, 2017, pp. 17–24.

[15] H. Zou, Y. Zhou, H. Jiang, S.-C. Chien, L. Xie, and C. J. Spanos, "Winlight: A wifi-based occupancy-driven lighting control system for smart building," *Energy and Buildings*, vol. 158, pp. 924–938, 2018.

[16] S. G. Shinde and B. G. Jaind, "Iot framework for energy efficient smart building," *International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume*, vol. 5, 2016.

[17] L. Chappell, *Wireshark network analysis*. PODBOOKS. COM, LLC, 2012.

[18] M. Gast, *802.11 wireless networks: the definitive guide*. " O'Reilly Media, Inc.", 2005.

[19] L. Oliveira, D. Schneider, J. De Souza, and W. Shen, "Mobile device detection through wifi probe request analysis," *IEEE Access*, vol. 7, pp. 98 579–98 588, 2019.

[20] CISCO, "Fundamentals of 802.11 wireless sniffing." [Online]. Available: https://www.cisco.com/c/en/us/support/docs/wireless-mobility/80211/200527-Fundamentals-of-802-11-Wireless-Sniffing.html

[21] J. 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.

[22] D. Murray, M. Dixon, and T. Koziniec, "Scanning delays in 802.11 networks," in *The 2007 International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2007)*. IEEE, 2007, pp. 255–260.

[23] C. Pei, Z. Wang, Y. Zhao, Z. Wang, Y. Meng, D. Pei, Y. Peng, W. Tang, and X. Qu, "Why it takes so long to connect to a wifi access point," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[24] K. V. Singh and M. Pandey, "Sdn-based fast handover approach to improve the qos of video streaming over wi-fi networks," in *Computing and Network Sustainability*. Springer, 2019, pp. 137–146.

[25] X. Hu, L. Song, D. Van Bruggen, and A. Striegel, "Is there wifi yet? how aggressive probe requests deteriorate energy and throughput," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 317–323.

[26] R. Buchakchiev, "People density estimation using wi-fi infrastructure."

[27] L. S. Committee *et al.*, "Ieee standard for local and metropolitan area networks: Overview and architecture (en línea)," *New York-NY-USA. The Institute of Electrical and Electronics Engineers Inc*, 2002.

[28] I. . W. Group *et al.*, "Part 11: wireless lan medium access control (mac) and physical layer (phy) specifications: higher-speed physical layer extension in the 2.4 ghz band," *ANSI/IEEE Std 802.11*, 1999.

[29] P. Brenner, "A technical tutorial on the ieee 802.11 protocol," *BreezeCom Wireless Communications*, vol. 1, 1997.

[30] C. Matte, "Wi-fi tracking: Fingerprinting attacks and counter-measures," Ph.D. dissertation, 2017.

[31] W.-F. Primer, "Overview of the 802.11 physical layer and transmitter measurements," *Beaverton: Tektronix Inc*, pp. 4–7, 2013.

[32] B. O'hara and A. Petrick, *IEEE 802.11 handbook: a designer's companion*. IEEE Standards Association, 2005.

[33] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. V. Randwyk, and D. Sicker, "Passive data link layer 802.11 wireless device driver fingerprinting." in *USENIX Security Symposium*, vol. 3, 2006, pp. 16–89.

[34] H. Gebre-Amlak, M. T. Islam, D. Cummins, M. Al Mansoori, and B.-Y. Choi, "Protocol heterogeneity issues of incremental high-density wi-fi deployment," in *International Conference on Wired/Wireless Internet Communication*. Springer, 2018, pp. 159–170.

[35] Pycom. [Online]. Available: https://pycom.io/product/lopy4/

[36] A. E. Redondi and M. Cesana, "Building up knowledge through passive wifi probes," *Computer Communications*, vol. 117, pp. 1–12, 2018.

[37] K. Friess, "Multichannel-sniffing-system for real-world analysing of wi-fi-packets," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2018, pp. 358–364.

[38] C. Hurley, R. Rogers, F. Thornton, and B. Baker, *Wardriving and wireless penetration testing*. Syngress, 2007.

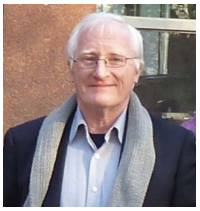
[39] Y. Li, "What are the differences between enterprise wi-fi and home wi-fi?" 2018. [Online]. Available: https://e.huawei.com/se/eblog/enterprise-networking/What-the-difference-between-corporate-Wi-Fi-and-home-Wi-Fi

YAN LI is currently working as an associate research fellow at SMART Infrastructure Facility, University of Wollongong. She obtained her PhD degree in Department of Electrical, Electronic Engineering, University of Melbourne in 2020. Her research interests include wireless sensor networks, mobile sensing, inertial sensors, Internet of Things (IoT) technology, navigation and positioning applications.

JOHAN BARTHELEMY is a lecturer at the SMART Infrastructure Facility of the University of Wollongong (Australia). He is the leader of the SMART IoT Hub and the Digital Living Lab developing sensors and edge computing devices for IoT applications using LPWAN networks including connected beer kegs, smart cameras and water level monitoring and low cost gas sensing. He is currently focusing on the development of innovative applications of AI and Intelligent Video Analytics for smart cities and environmental monitoring.

SHUAI SUN received the B.E. (Hons.) and master's degrees from Northwestern Polytechnical University, China, in 2014 and 2017, respectively. He is currently pursuing the PhD degree with the Department of Electrical and Telecommunication Engineering, RMIT University. His research interests include indoor localization, variational Bayesian, and information fusion.

PASCAL PEREZ is currently the Director of the SMART Infrastructure Facility at the University of Wollongong, overseeing research in infrastructure-related fields such as water and energy efficiency, future transport and mobility, smart cities and communities, and infrastructure system engineering and logistics. As Director he is responsible for SMART's academic governance and for establishing strategic scientific partnerships in Australia and beyond. He is a Fellow of the Royal Society of NSW and of the Modelling and Simulation Society of Australia and New Zealand (MSSANZ). He is also a member of the national Scientific Committee of the Australian Urban Research Infrastructure Network (AURIN). In 2002, he received an ARC-International Linkage Fellowship to develop social modelling research at the Australian National University.



BILL MORAN currently serves, since 2017, as Professor of Defence Technology in the University of Melbourne. From 2014 to 2017, he was Director of the Signal Processing and Sensor Control Group in the School of Engineering at RMIT University, from 2001 to 2014, a Professor in the Department of Electrical Engineering, University of Melbourne, Director of Defence Science Institute in University of Melbourne (2011-14), Professor of Mathematics (1976–1991), Head of the Department of Pure Mathematics (1977–79, 1984–86), Dean of Mathematical and Computer Sciences (1981, 1982, 1989) at the University of Adelaide, and Head of the Mathematics Discipline at the Flinders University of South Australia (1991–95). He was Head of the Medical Signal Processing Program (1995–99) in the Cooperative Research Centre for Sensor Signal and information Processing. He was a member of the Australian Research Council College of Experts from 2007 to 2009. He was elected to the Fellowship of the Australian Academy of Science in 1984. He holds a Ph.D. in Pure Mathematics from the University of Sheffield, UK (1968), and a First Class Honours B.Sc. in Mathematics from the University of Birmingham (1965). He has been a Principal Investigator on numerous research grants and contracts, in areas spanning pure mathematics to radar development, from both Australian and US Research Funding Agencies, including DARPA, AFOSR, AFRL, Australian Research Council (ARC), Australian Department of Education, Science and Training, and Defence Science and Technology, Australia. His main areas of research interest are in signal processing both theoretically and in applications to radar, waveform design and radar theory, sensor networks, and sensor management. He also works in various areas of mathematics including harmonic analysis, representation theory, and number theory.