WiFi Temporal Coverage: Analysis of Socio-Economics Influences in Malaysia

N. A. Wahab¹, M. R. Ahmad¹, A. A. M. Isa¹, R. F. Malik² and M. R. M. Esa³

¹Broadband and Networking Research Group (BBNET), Centre for Telecommunication Research and Innovation (CeTRI), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
²Faculty of Computer Science, Universitas Sriwijaya (UNSRI), Inderalaya, Sumatera Selatan, Indonesia
³Institute of High Voltage and High Current (IVAT), School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor Bharu, Malaysia

Email: nazdiana.aw@gmail.com; riduan@utem.edu.my; rezafm@unsri.ac.id

Abstract. According to Internet Survey 2017, it was found that 89.4 percent of Internet users in Malaysia used a smartphone as a medium to access the internet. The proliferation of mobile devices and bandwidth hungry application, has resulted in mobile network provider (MNO) struggles to keep up with the ever-increasing demand for higher network capacity and wider network coverage. WiFi offloading seems the most viable solution due to the abundance and readily available infrastructure. Diverted traffic to the WiFi network may possibly reduce cellular network traffic; hence cellular network provider may be able to accommodate traffic growth at a lower cost. By analysing and understanding the user’s offloading behaviour may provide guidance in price and cost restructuring strategies or in network planning. This paper investigates temporal coverage of mobile user’s WiFi utilization and socio-economics influences through an extensive measurement study. The measurement study is carried out for 18 days and collected real data traces of user WiFi utilization on Android smartphone. The observation analysed user’s WiFi temporal coverage where participants are categorized into two different socio economics group. The finding reveals low WiFi temporal coverage for both groups of smartphones users. Group 1 spends more time using WiFi compared to Group 2, while both groups shows higher temporal coverage during active hour. Findings of this research may potentially serve as insights to the relevant stakeholder or regulatory body, as well as a barometer to carry out enhancement measures or any potential joint venture between cellular network provider and ISP.
1. Introduction
From the past 10 years, Malaysia has experienced tremendous growth in mobile broadband subscriptions to 30.6 million in the first quarter of 2017, from only 23.35 million in 2007 [1, 2]. MNO has continuously upgrading and expanding their infrastructure to extend network coverage as well as to improve network capacity. However, mobile broadband coverage is not universal. 4G and 3G coverage usually only available at a strategic location such as metropolitan areas, offices and industrial area or location with high population. Most of the places either still experiencing slow mobile broadband service such as EDGE or does not have mobile broadband coverage at all.

In addition, some of the places with good mobile broadband coverage experiencing problems like a blind spot. For indoor building, there are still issue on coverage even though indoor base station or indoor mobile broadband base station (MBB) has been installed inside the building. Mobile users are also experiencing slow data speed during peak hour due to network congestion, particularly at high traffic areas such as at a crowded airport or football stadium.

[3] highlighted one of the most frequent consumer complaints are related to service speed. The signal fluctuated between 3G signal and EDGE signal even though the area is within 3G coverage. The unstable connection to the end user is caused by the fluctuations of the receiving signals.

Therefore, consumers experienced slow speed during data connection activities. The issues mentioned above pointed out some of the limitations of mobile data service provided by the cellular networks. The issues mainly are caused by the limited network capacity or also known as bandwidth scarcity. The problems are further exacerbated by the proliferation of mobile devices and rich multimedia content, video streaming and cloud services that require reliable and high bandwidth capacity.

According to Internet Survey 2016 [3], it was found that 89.3 percent of internet users in Malaysia used a smartphone to access the internet. In addition, the percentage of smartphone ownership rose to 90.7 percent in 2015 compared to 74.3 percent in 2014.

Table 1. Growth of Internet Subscription and Smartphone Users

|                               | 2015  | 2016  | 2017  |
|-------------------------------|-------|-------|-------|
|                               | (million) |       |       |
| Fixed Broadband               | 2.8   | 2.5   | 2.6   |
| Mobile Cellular               | 27.8  | 28.5  | 35.3  |
| (% of Penetration Rate)       | 99.7  | 99.8  | 117.3 |
| Broadband                     | 143.8 | 139.9 | 131.1 |
| Smartphone User               | 87.3  | 87.9  | 89.4  |

To increase cellular network capacity by deploying more base stations requires expensive capital expenditure (CAPEX) and operation expenditure (OPEX) [4]. Therefore, by leveraging on the existing radio access technology (RAT) is an interesting alternative and offers a promising solution. Data traffic from cellular networks could be offloaded to other existing RAT such as femtocells or WiFi and are referred to as mobile data offloading. Due to the abundance of WiFi infrastructure that is widely deployed by internet service providers (ISP) and residents, Wifi offloading is a viable solution to address the need of additional network capacity and at the same time providing wider network coverage possibly via WiFi features such as roaming.
Furthermore, it was announced by Telekom Malaysia (TM) which is the leading ISP in Malaysia that it has deployed over 10000 WiFi hotspots and is planning to increase its coverage [5]. Therefore, partnership between ISP and cellular network provider might be beneficial to both sides. Such implementation and collaboration need further understanding, feasibility studies and proper planning on consumer demand, as well as user’s WiFi utilization pattern.

However, there is a lack of knowledge and information on the mentioned area, especially the one focusing on smartphone user’s WiFi utilization. Hence, this paper aims to investigate WiFi utilization pattern in terms of temporal coverage, specifically in the context of Malaysian mobile data subscriber. [1] conducted a quantitative study on the performance of WiFi offloading by an experimental analysis of metropolitan area in South Korea. This research intends to simulate the study by focusing on the perspective of Malaysia’s mobile user and understanding the influences of socio-economics group on WiFi offloading.

An Android based mobile application has been designed and developed as a tool to collect real data trace of user’s daily-life pattern. This research extends [1] works by comprehensively investigating the socio-economics group influences on WiFi offloading. On WiFi utilization pattern, the study investigates the temporal coverage or the time portion users would like to spend in WiFi, as well as frequency of connection and properties between different groups.

All observations take into consideration the data behaviour during active hour and all-day connection. A comprehensive analysis of real data trace of smartphone’s WiFi usage for 18 days study, from 100 Android smartphone users has been studied. The main goal of the study is to observe the WiFi utilization details of smartphone users, as well as answering questions such as how much is the frequency and time portion in a day, that a smartphone user stays in WiFi coverage area? What time of the day that user use WiFi the most?

The measurement study is designed and conducted specifically for Android smartphone users which also has been categorized according to different socio-economics group. An Android based application called WiFi Analyzer has been designed and developed to collect real data traces of user’s daily WiFi
The research also considers the correlation of different socio-economics group on the several key parameters which is a method that has never been studied before. WiFi offloading has been highlighted as a promising alternative that could assist cellular network provider in accommodating the ever-increasing demand of network capacity and network coverage. Therefore, it is necessary to further understand the current scenario in the consumer’s market before any proper planning and deployment can be implemented. In this context, this work presents statistical observation, where outcome from the measurement study might be useful in understanding the bigger pictures of Malaysian internet scenario.

Additionally, the end users or consumers also might be interested in WiFi offloading due to economics reason, e.g., potential reduction of subscription fees or some value-added service with the same fees. Although there are some limitations in this study, it is hoped that the results of this study will be significant for future research endeavours. Findings of the study reveals new knowledge and information on the specified research area, which may provide insights and beneficial for cellular network provider, internet service provider and even regulatory body for the purpose of telecommunication network planning and development in Malaysia.

2. Related works

[1] claimed to perform the first trace-driven simulation using the measurement study in South Korea. The study recruited 97 volunteers from iPhone user community in South Korea and asked them to install DTap in their phones for the period of the measurement study. Findings reported the average of temporal coverage for all users are 70% for all day and 63% for active hour only.

From the other side, study by [6] reported only 11% of temporal coverage. Their work has a different approach where measurements were done only when a user is on a vehicle. Typically, users spend most of the time in the office or at home.

To the best of our knowledge, this is the first measurement study that carried out a real data trace of user’s daily WiFi utilization pattern in the context of Malaysian cellular network subscriber. Using an application that has been designed and developed for Android platform to autonomously collect and recorded data, the study is specifically targeting on Android mobile devices such as smartphone and tab users.

Throughput, jitter, latency and QoS are the most common key parameters investigated in improving network performance, whether in fixed broadband environment or mobile networks infrastructure. It is found from the continuous literature search that most of the research carried out in Malaysia focus on mobile data networks and fixed broadband, and less research from wireless LAN, especially from the user’s perspective. However, due to the high broadband demand, as well as to support government policy of NKEA, wireless LAN is an approach that should be given more serious consideration.

To provide internet services at Malaysia rural areas, [7] investigates application of a two-tier wireless network system with integrating wireless access and wireless backhaul as the evolution path of providing rural areas with Internet services. The first tier of the proposed network system is the wireless access network which provides interface to user accessing to Internet and enjoy services such as voice over IP, web browsing, and social network. The second tier is the backhaul network in which comprises of a number of wireless routers that carry traffic from various access networks to a gateway.

[8] conducted study on broadband internet performance in Malaysia, where the performance management is defined by the Quality of Services (QoS). The assessment was carried out among 5 regions in Malaysia with goal to identify the benchmarking in network performance. However, the study
was conducted for fixed line broadband where the assessment mostly was carried out using desktop and laptop users.

[9] conducted experiment to prove that single radio access unit (RAN) not only helps reducing MNO investment but it is also to improve network for coverage and quality compared to conventional radio deployment. The performance of the single RAN was studied in terms of coverage and quality compared to conventional radio. The specific tools which is called Nemo Outdoor is used to measure the coverage, quality and throughput via drive test. Important parameter and data captured are being process using Nemo Analyse and MapInfo Professional.

On the other side, [10] conducted research on the drivers of broadband in Malaysia and focusing on the identification of two relevant research stream covering broadband, which are adoption and continuance. Findings from the paper revealed that ISP, regulatory bodies, cellular network provider and application developer need to co-ordinate to support the adoption growth of broadband internet.

3. Experimental setup

This section briefly explained the three phases which Data Source, Data Understanding and Data Preparation where each phase plays crucial roles in ensuring systematic data collection as well as reliability and validity.

A. Data Source

• Location, Population and Android-based Platform

To further study the WiFi temporal coverage, users from various levels are chosen in order to produce a variation of the sample. The participants are from various areas around Melaka, which has a diverse occupational background such as lecturers, private sectors worker, housewife, teachers, and students.

The accessible population that was selected consisted of public users and undergraduates from University Teknikal Malaysia Melaka with accessibility to WiFi through Android-based mobile devices. The rationale for choosing undergraduates is because they were in the age range of 20 to 24 years and, according to statistical data from the MCMC, the highest internet users are in the age group of 20 to 24 years old.

On the other side, a report by Device Atlas Mobile Report for Q2 2017 [11] reveals that Android as the leading mobile OS in Q1 2017. Figure 2 shows that Android OS dominated Malaysia mobile market share by 83.17% compared to only 14.33% by iOS.
Figure 2. Mobile OS Market Share in Malaysia

- **Sample Size**

For the purpose of this study, Raosoft® software and Cochran's formula were utilised for the research. Raosoft® software is an efficient and user-friendly software that can easily be installed in any technological device including smartphones or tablets.

To determine the sample size using Raosoft® software, 10% of the margin error and a 95% confidence level were input parameters that were included. 100 participants from Android user community has been recruited and asked to install WiFi Analyzer on their phones. As the aims of the research are to investigate temporal coverage according to a different socio-economics group, participants have been categorized into two groups.

Group 1: The sample represents higher institutional students from University Teknikal Malaysia Melaka (UTeM). 25 students are from Faculty of Pengurusan Perniagaan (FPTT) located at Campus Jalan Hang Tuah, while another 25 students are from Faculty of Electronics Engineering and Computer Engineering located at Main Campus Durian Tunggal.

Group 2: The second group of 50 participants represents public users which comes from a diverse occupational background such as government and private sector employees, housewife, and teachers from various locations in Melaka.

- **Instrumentation**

The process of data gathering is done using autonomous Mobile Data Collection (MDC) method. MDC is a method of evaluating quantitative data using mobile phones, tablets or PDAs for programming or data collection. an Android smartphone platform called WiFi Analyzer has been developed. WiFi Analyzer tracks the data traces of WiFi activity in the background, saves the record in a text file and systematically upload the file to an FTP server.

The application captures all WiFi SSID available within its coverage area. The timer starts once the WiFi connection is established, and a log file is created in user’s local storage. Otherwise, the application
will keep checking the WiFi state. When the timer reaches three minutes, the application saves the current date and time stamp. The application also scans location and records the latitude and longitude of the current session of WiFi AP. Note that location only can be obtained if user enables location service on their mobile device.

The Android application has been developed using Android Studio tool with minimum Android SDK 21 (equivalent to Android 5.0 Lollipop) version, or newer is required. Data were sent to the server via file transfer protocol (FTP). Thus, Amazon web server (AWS) has been configured to receive data from users.

- **Procedure of Data Collection**

  The application was ready to be distributed for the actual study after the validity and reliability were conducted and confirmed. In order to conduct the research in the respective faculties, the chronology of events leading to data collection was prepared. The description is as illustrated in Figure 3.

  
  **B. Data Understanding**

  Fraction of time that a user stays connected to a WiFi coverage area could determine offloading performance and is defined as temporal coverage. The key parameters that are crucial to measure temporal coverage are; connection start and end time, date of connection, daily total duration of WiFi connectivity for all day and active hour connection and daily tabulation of users with WiFi access at hourly period.
Figure 3. Chronology of Data Collection Process

C. Data Preparation

• Data Cleaning and Extraction

It is beneficial to understand the accuracy of collected data as well as correcting any errors or outliers in the data. Outliers are a single or small number of data values that are not similar to the rest of data set. The raw log file contained a massive data collection accumulated during the duration of the measurement study. There are a number of information that are not related to the scope of this research. Therefore, a Python script has been developed to automatically extract the related information only. The Python script reads the raw log file line by line and only selects a predetermined field and organized the data into a table systematically. The script processes the wifi-log.txt file and produces an output file in excel .csv format. All necessary information can be extracted after identify the character surround the information. For example, referring to the string as shown in Figure 4 from the “wifi-log.txt”, time is between string “time:” and “(". Identifying these characters is actually the main method implemented in this script. All individual parameter has its own array to save the parameter in sequence order and tabulated into excel table.

Figure 4. Identifying Characters in Python Script
4. Implementation of analysis

In analysis process, summary of tables and graphs can be used to convey information about the data. In this research, a summary table is used, which is a common way of understanding data. Summary tables will often show a count of the number of observation (or percentage) that have that particular value (or range).

Due to the comprehensive data collection, the data log for each participant can reach up to 10MB and obviously is a massive amount of data. The raw data files need to be sorted and organized before analysis can be carried out. Figure 5 shows the steps involved in data analysis procedure.

Figure 5. Data Analysis Procedure

The Python script is responsible to extract all the pre-defined information related to the research scope. However, the information still needs to be processed in order to produce a result and analysis. A MATLAB algorithm has been developed to do autonomous calculation and to analyse the temporal coverage. The program will first organize all the data according to the session of the connection, whether it is during active hour or all-day connection. Then, the algorithm calculates the duration of connection time for each day of the measurement study, which is used to measure the temporal coverage.

Temporal coverage is defined as time portion that user stays in a WiFi coverage areas and is one of the important key performance parameters in determining the performance of offloading. From the data collected during the measurement study, daily average temporal coverage could be determined for each participant.

The difference of time between all day reading and active hour (9.00 – 24.00) has also been highlighted. Temporal coverage can be determined by calculating the duration of connection time, where for each participant is the average percentage of connection duration for each day. Then, the percentage for the whole period of measurement study were averaged.

5. Results and discussion

A. Temporal Coverage

Figure 6 and Figure 7 shows the average of daily temporal coverage recorded by Group 1 and Group 2 participants for the duration of the measurement study. It also shows the plot of coverage recorded during the active hours (0900 to 2400) session. The red and blue bar represents the average of daily
temporal coverage for each participant in the measurement study for all day and active hours respectively. Users are numbered as shown in X-axis and labelled as “User Index”. The horizontal yellow and green line reveals the average reading for both sessions.

![Figure 6. Temporal Coverage for Group 1 (Student)](image1)

![Figure 7. Temporal Coverage for Group 2 (Public)](image2)

As can be seen from the graph, the recorded trace of students in group 1 showed a higher temporal coverage for both categories all day and active hour connection. The average WiFi utilization for all day connection are 17.3% and 14.81% for group 1 and group 2 respectively, with only less than 3% difference. According to the demographics and socio-economics of internet user report by [12], more than 50% of Internet users in Malaysia are from the age group of 20-34 years old, which supported the finding that group 1 generally incline to access longer internet time.
Table 2 also shows that the recorded data trace has a quite significant standard deviation in the data sets from the mean value. Group 1 showed higher standard deviation of 15.6% during all day connection. The data indicates variety of user data pattern with minimum temporal coverage of only 1.7% to the maximum reading of 56.9%. However, the all-day median for both groups are almost similar with only 0.5% difference.

|                | All Day (%) |                      | Active Hour (%) |
|----------------|-------------|----------------------|----------------|
|                | Group 1     | Group 2              | Group 1        | Group 2          |
| Min            | 1.731       | 0.4178               | 0.7            | 0                |
| Max            | 56.9        | 39.29                | 31.12          | 28               |
| Mean           | 17.30       | 14.81                | 10.8           | 10.49            |
| Median         | 11.77       | 11.27                | 6.727          | 8.48             |
| Std            | 15.6        | 10.11                | 9.572          | 7.812            |

For active hour, the data showed similar pattern of higher usage by group 1 compared to group 2 but with only slight difference of 0.31% for mean value. However, the median for group 1 and group 2 are 6.73% and 8.48% respectively.

Note that during active hour, the median for group 2 is higher than group 1. The data suggests that during active hour, public group spends more time using WiFi due to the availability and more reliable WiFi connection in the office or home. On the other side, the pattern showed that students usually spend more time in class or lecture during active hour and therefore has lower temporal coverage. [12] reported in Internet User Survey that place of study is the least popular place to access internet. The reason for lower temporal coverage for group 2 during active hour, also might be caused by the lack of coverage or stability of WiFi in campus during time spent on campus area.

B. User’s Frequency Distribution

The frequency distribution presents information according to the number of observations reported for each ranges of values, for a particular variable. The observation for a range of percentage for both groups during all day temporal coverage is counted and plotted in the following histogram. User’s frequency is plotted as density in Y-axis. The shape of the plot reveals trends, that is the user density for each ranges of temporal coverage percentage.

Figure 8 shows the histogram of all user (group 1 and group 2) temporal coverage during all day connection. The histogram indicates the number of users for each predetermined ranges of temporal coverage percentage. The result reveals that almost 50% from total participants recorded 10% of daily temporal coverage during all day connection.
Figure 8. Combined Histogram for Group 1 and Group 2 (All Day)

C. Tabulation of User (Hourly)

Figure 9 and Figure 10 illustrates the average percentage of users with WiFi by hourly period for group 1 and group 2 respectively. It shows average percentage of users for each hour of the day, for the whole duration of measurement study.

Figure 9. Percentage of users with WiFi access from 0100, Sept.18 to 2400 Oct.6 (Group 1)
The finding reveals that at any given hour of the day, the average percentage of users with WiFi access for group 1 is 21%. Percentage of group 1 user access is higher between 1300 to 0100 and the lowest access is between 0300 to 1200. This is most likely because students usually have class in the morning session and tends to access internet afterwards, whether at home, hostel or at campus. In addition, (Malaysian Communications and Multimedia Commission, 2017) reveals that place of study is the least favorite place to access WiFi.

On the other side, in average only 14% of group 2 users access WiFi by each hour of the day period. The finding reported that the highest WiFi usage for group 2 usually is between 1400 to 2300. Group 2 generally spends less time in accessing WiFi and the result is concurrent with the temporal coverage result as per highlighted in Table 2.

6. Conclusion

This paper presents a real trace of quantitative measurement study on the smartphone user’s WiFi utilization in terms of WiFi temporal coverage in Malaysia. Two comprehensive datasets representing different socio-economics group were collected to further understand its influences in WiFi utilization. An autonomous method of data collection has been conducted using Android application, known as WiFi Analyzer which has been developed specifically for this study.

The finding reveals that the average temporal coverage for all participants are 16% for all day. It means the average daily WiFi usage for each user is only 4 hours. Group 1 spends more time using WiFi compared to Group 2, while both groups shows higher temporal coverage during active hour. The result is consistent with correlation analysis that shows group 1 has higher temporal coverage. The significant correlation is highly related to the difference between all day and active hour. In other words, higher difference between all day and active hour means higher correlation value and resulted in higher
temporal coverage. In addition, histogram of user frequency distribution discovered that 50% from all users in both groups has average of only 10% daily temporal coverage.

Finding on tabulation of users by hourly-period indicates that average of 21% of group 1 access the internet at any given hour of the day, but the highest is between 1300 to 0100. Group 2 only has average of 14% user access at any hour of the day with high WiFi utilization between 1400 to 2300.

The finding indicates low WiFi temporal coverage for both groups of smartphone users. The result is contradicted with [12] data that reported 89.4% of internet users in Malaysia use smartphone as a preferred medium to access internet. But our finding reveals that smartphone users only use average of 4 hours daily, to access internet through WiFi networks. The possible explanation for the low WiFi temporal coverage is, users might access the internet via data networks provided by the cellular network provider instead of using WiFi networks. The theory is supported with MCMC data that reported increase in mobile data subscription to 35 million in 2017. If the situation continues, it will be difficult for cellular network provider to keep up with the high demand of network capacity.

Acknowledgments

The authors would like to acknowledge the support provided by Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer (FKEKK), Universiti Teknikal Malaysia Melaka (UTeM). This project is funded by Short Term Grant (Projek Jangka Pendek PJP) in FKEKK, UTeM (PJP/2018/FKEKK/(3B)/S01615). This project is partially funded by Universiti Teknologi Malaysia (UTM) under Vote No. 04G19 and Ministry of Education (MOE) under 4F966 research grants (FRGS).

References

[1] Lee, K. et al. (2010) ‘Mobile Data Offloading: How Much Can WiFi Deliver?’, ACM CoNEXT, 21(2), pp. 536–550. doi: 10.1145/1921168.1921203.
[2] MCMC (2017b) Pocket Book of Statistics.
[3] MCMC (2017a) ‘Network Performance Report 2016 Measuring broadband and voice performance’, (January).
[4] Suh, D., Ko, H. and Pack, S. (2016) ‘Efficiency Analysis of WiFi Offloading Techniques’, IEEE Transactions on Vehicular Technology, 65(5), pp. 3813–3817. doi: 10.1109/TVT.2015.2437325.
[5] Telekom Malaysia (2018) Unifi Wifi. Available at: https://www.unifi.com.my/unifi-en/ijoin/product-details.page?productId=wifi (Accessed: 7 June 2018).
[6] Balasubramanian, A., Mahajan, R. and Venkataramani, A. (2010) ‘Augmenting mobile 3G using WiFi’, ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 209–222. doi: 10.1145/1814433.1814456.
[7] Kwong, K. H. et al. (2011) ‘WiFied Up Malaysia Villages: A Case Study of Using WiFi Technology to Increase Internet Penetration Rate in Malaysia Rural Areas’, (October), pp. 7–12.
[8] Idris, M. F. B. M. et al. (2014) ‘Broadband Internet performance (QoS measurement) view from home access gateway in Malaysia’, Proceedings - 2014 5th IEEE Control and System Graduate Research Colloquium, ICSGRC 2014, pp. 147–152. doi: 10.1109/ICSGRC.2014.6908712.
[9] Ya, N. et al. (2014) ‘Coverage and Quality Radio Analysis for Single Radio Access Network Technology in GSM’, pp. 158–163.
[10] Lurudusamy, S. N. and Thurasy, R. (2016) ‘The Drivers of Broadband Internet in Malaysia’, Proceedings - 6th International Conference on Computer and Communication Engineering: Innovative Technologies to Serve Humanity, ICCCE 2016, pp. 225–230. doi: 10.1109/ICCCE.2016.57.
[11] DeviceAtlas (2017) ‘DeviceAtlas Web Intelligence Report Q2 2017’, (August).
[12] Malaysian Communications and Multimedia Commission (2017) ‘Internet Users Survey 2017 Statistical Brief Number Twenty-One’, pp. 1–51. Available at: http://www.mcmc.gov.my