www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

USER BEHAVIOUR MONITORING USING MOBILE PHONES TO IMPROVE 5G SERVICES AND PERFORMANCE

Zhihao Cui[1] and Jize Yan[1,2]

[1]Department of Engineering, University of Cambridge, Cambridge, UK

[2]School of Electronics and Computer Science, University of Southampton, Southampton, UK

Email: J.Yan@soton.ac.uk

Submitted: June 3, 2016 ***Accepted: July 5, 2016*** ***Published: Sep. 1, 2016***

Abstract- 4G has been widely commercialised, and 5G is currently under development. The expected data bandwidth for 5G is 100 times faster than 4G and 500 times faster than 3G; however, the evolution of telecommunication technologies involves both a boost in speed and the enhancement of user experience. The key word used to describe 5G is 'user-centric', rather than 'service-centric' for 4G, and thus user behaviours of mobile data usage should be further investigated. On the other hand, the testing equipment currently being used for base stations is limited to hardware devices, such as spectrum analysers and power meters. These testing methods do not include the considerable potential variations in data demands due to changes in user behaviours, which could be resolved by presuming that all data resources could be dynamically allocated by real-time events.

A complete system has been designed and implemented in this study to investigate current user behaviours regarding mobile data usage. The system consists of three individual parts, including a user iOS application, a web server and an administrative iOS application. Ten devices were tested within the

two-month data collection period. Although the sample size was too small to produce any statistical results, it was found that data usage behaviours differ from user to user, with the exception of using more than 10 times the Wi-Fi over WWAN data at all times. The data also proved that some of the usage case families, which are described in the NGMN 5G white paper, do have strong demands, which could not be fulfilled using current telecommunication technologies due to technological gaps.

This paper shows that the system proposed is a feasible method to investigate user behaviours of mobile data usage. If the sample size of users involved could be increased in the future, it would be possible to develop a model for real-time simulations of mobile users in specific areas so that limited connection resources could be dynamically allocated. Moreover, the basic communication infra-structures, such as base stations, should be well-planned and developed in advance to fulfill the potential 5G demand.

Index terms*: Mobile Phone, 5G, User Behaviour, Cloud Sensing, Wi-Fi, WWAN, Data Analysis.*

I. INTRODUCTION

When mobile communication technology evolves from 3G and 4G to 5G, there will be a boost in data rates as well as an evolution of new concepts (details in Table 1). Most modern lifestyles depend on recent mobile technologies that have been developed, such as video calls and real-time GPS navigation. With the average global connection speed over 1 Mbps in 2014 [1], users now not only use mobile networks for web browsing but also for viewing films and TV shows on mobile devices, which are traditionally fixed-line activities.

Various new 5G technologies are being developed. At the same time, corresponding testing methods should also be developed. During the period of transition from 3G to 4G, the current hardware testing tools had issues involving large integrated equipment and operating speed limitations (more in Section 2), which provides the opportunity for utilising simulation software for tests using modern computers to improve the 5G technologies.

Simulating a mobile usage network would be extremely helpful for different aspects of testing, including base station testing, mobile chip testing and antenna testing. For accurate simulations, real-time data must be collected. Method-like questionnaires can only provide a list of data from each person, which is insufficient in developing a sophisticated simulation model and system.

Due to the daily use of millions of mobile phones globally, relevant data should be collected from these end devices using an installed mobile application.

This paper presents a feasible method that addresses the upcoming 5G development testing issues using mobile phone sensing. Accessing users' mobile data usage behaviours on mobile devices is the primary aim, and the method involves designing a complete system that does not depend on any other services, i.e. collecting data from an iOS app, transmitting data to web services and analysing data from a database.

First, the user application was designed to monitor user behaviours, and it automatically collects data in the background, enables the user to review any data specified in a time period on a map and uploads the data to the web server. The data chosen to be collected are taken from the device data, including time, GPS coordinates, amount of data being used (in both Wi-Fi and WWAN protocols), signal strength and the radio access method, along with any personal data provided by users, i.e. age, gender, occupation and city of residence. These two highly correlated data types were used for further analysis in later stages.

Second, the web server consists of front-end HTML web pages, back-end PHP service and a MySQL database. The front-end web pages introduce project information to encourage users to participate in the project and include a registration form for new users to provide their personal information. The back-end PHP service handles all data requests sent to the database, and all data collected from different devices are centralised from either front-end web pages, user applications or administrative applications, which prevents database security issues.

The third administrative application obtains and stores all data available from the web server for analysis. Any combination of devices and time periods can be chosen to be displayed on the map, which shows the overall geographical pattern using a quad tree clustering technique. Data points that are similar are clustered into one point for simplicity, and statistics within each cluster are given, such as the average speed of data used. The statistics of clusters from different combinations provide a general idea of the way users consume their data throughput.

Small-scale trial runs on multiple iOS devices belonging to different users were carried out, and various data were collected and analysed to show feasibility. Transversal and longitudinal analyses could identify the features required for 5G in large-scale tests in the future.

II. BACKGROUND

Recently, Next Generation Mobile Network (NGMN) Alliance announced the NGMN 5G white paper [2], which is the first public paper published by leading industry members to propose the definition and applications of 5G, and the paper presented various current and future 5G implications. In addition to quantitative measures, such as 1 Gbps bandwidth, 10 ms end-to-end latency and <1 m three-dimensional network based positioning accuracy, one of the key factors mentioned is 'Use Cases' (Table 2).

Technology	Services	Data Bandwidth	Key Factors
2G	Circuit switched	14.4 kbps	Incompatible standards
2.5G	Packet switched	384 kbps	Provide benefits of 3G
3G	Packet switched	2 Mbps	Operator-centric
4G	Wireless broadband	200 Mbps	Service-centric
5G	Emergence of IoT	1 Gbps	User-centric

Table 1: Mobile technology generations comparison [3, 2]

No.	Use Case Families	Example Use Case
1	Broadband access in dense areas	PERVASIVE VIDEO
2	Broadband access everywhere	50+ MBPS EVERYWHERE
3	Higher user mobility	HIGH-SPEED TRAIN
4	Massive Internet of Things	SENSOR NETWORKS
5	Extreme real-time communications	TACTILE INTERNET
6	Lifeline communications	NATURAL DISASTER
7	Ultra-reliable communications	E-HEALTH SERVICES
8	Broadcast-like services	BROADCAST SERVICES

Table 2: 5G use case families and related examples [2]

The features of 5G indicate the need to understand the way that the mobile users allocate their data usage. Different user behaviours would result in a significant difference in traffic generation, which directly influences the performance of a mobile base station.

In terms of technology implementation, current Wi-Fi technology could achieve the speed requirement of 1 Gbps with an IEEE 802.11ac standard, but WWAN technology could not. According to 3GPP released specifications, the average data rate of 100 Mbps can be achieved by current LTE technology, which indicates the need for significant improvement for 5G applications.

2.1 Testing Challenges

Existing base station testing methods highly depend on hardware. For instance, the Anritsu Base Station Analyser combines a number of different measurement functions within a portable package [4], and the size will expand in the future when advanced communication technology needs to be tested. On the other hand, some testing software, such as ROMES4, analyses a large amount of data collected when testing LTE networks [5].

Challenges in testing already existed during the transition from 3G to 4G, especially when data packages needed to be analysed in detail, and both current and future networks will be package switched networks (Table 1). Considering the 5G data rate, which aims to be 100 times faster than 4G, it will be extremely difficult to maintain reliable and durable software simulations during effective testing using existing hardware equipment. A cloud software-based testing method, which is easily altered by number, users or data rates, would be desirable, and it could be expanded by super-computer or cloud computing.

2.2 Mobile Sensing

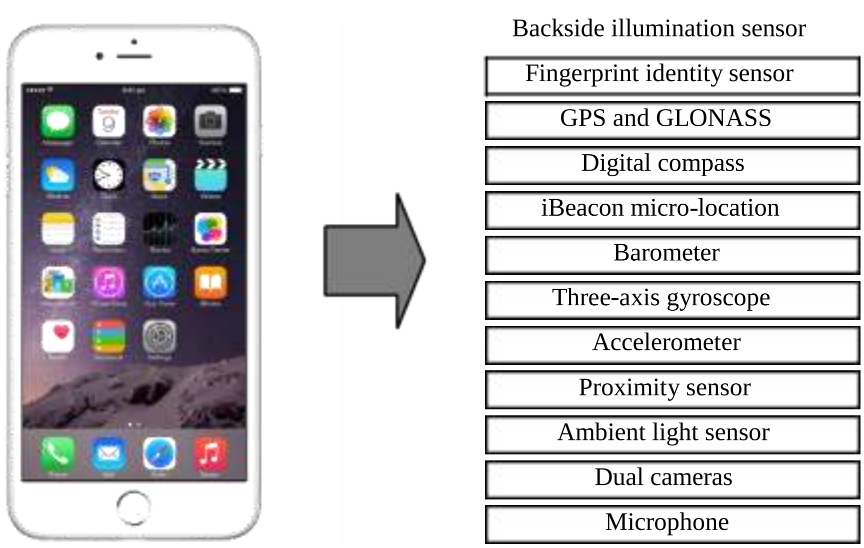


Figure 1: Sensors on an iPhone 6 (potentials related with 5G user cases)

Due to their popularity, smart phones can fulfil the sensing role (instead of dedicated wireless sensor networks) by acting as super sensor nodes and gateways [6]. For example, the recent iPhone model contains more than enough sensors (Figure 1) to achieve the aim of this project, which was collecting data and examining usage behaviours. Mobile applications are often utilised for everyday use. The following subsections include examples.

Traffic sensing traditionally uses external GPS devices to collect data. It is much easier for developers to collect and analyse road data using smart phones. Several existing libraries and APIs are available for both collecting information and post-processing for end users. Speed estimation through mobile sensing [7] focused on using an accelerometer to estimate vehicle speed rather than the power-consuming GPS module (high-speed train in Table 2).

Social networking is a major use of mobile data, which is utilised to share personal moments, photos and videos. Social networking platforms, such as Facebook, are no longer restricted to being used to communicate with friends; they are also used as modern social media platforms. Therefore, collecting data regarding social networking habits is more valuable than before. For example, [8] automatically classified events in people's lives, and then automatically sent them through Facebook or Twitter.

Moreover, after 3G was introduced, it was possible to make video calls without using Wi-Fi. With the cost reduction of cellular data, users are willing to utilise WWAN video calls, allowing their families and friends to view live feeds (Pervasive video in Table 2).

Environmental monitoring is currently used to be conducted by environmental authorities, and all data are collected through special sensors installed in cities, offshore or undersea. Individuals are encouraged to maintain a lower carbon lifestyle, but they usually do not have the access to the data and are not aware of their personal impact on the environment. Now, applications such as participatory sensing systems [9] can track the way that individual actions might affect the environment, such as through carbon emissions, based on data collected from mobile phones. Thus, the impact of individuals can be calculated and simulated by sophisticated models (sensor networks in Table 2).

Health monitoring is used on mobile phones for first responders following health-related data collection. For example, Body Sensor Network [10] uses the body sensor network and mobile phone applications to provide continuous monitoring and analysis of physiological parameters, and activity sensing [11] captures levels of physical activity and related this information to personal health goals by presenting feedback to the user (e-health services in Table 2; health monitoring can be categorised in the ultra-reliable communications use case family).

These methods have been developed for patients with long-term illnesses who live at home, and they utilise environmental measurement nodes and wearable sensing devices. After obtaining and monitoring all necessary raw data wirelessly, typically through Bluetooth, the mobile phone sends data to online servers via USB or WLAN for further medical diagnosis; however, for these methods, mobile phones only serve as transmission media along with other hardware devices, which cannot function alone without wireless connection [12, 13, 14].

Data usage monitoring is dependent on individual use. There are several applications in the Apple App store that can monitor user data usage. For example, DataMan has the ability to calculate how much data has been used by time or location as well as how many applications have used data. The data from these apps cannot be used to discover general trends because they are limited to making predictions and adhering to monthly data plans.

2.2.2 Typical Architecture

The sensible approach would consist of an application installation on individuals' phones, which would upload the data collected and provide a data analysis with feedback to develop a typical mobile sensing architecture, as suggested in [15].

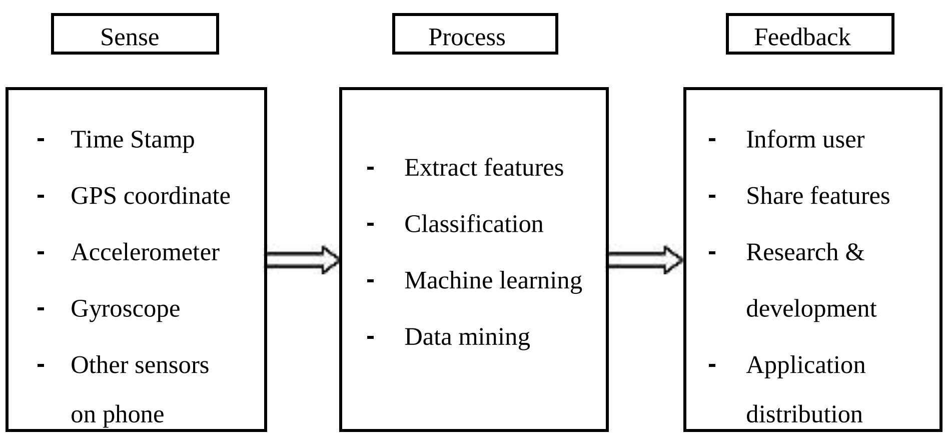


Figure 2: Typical Mobile Sensing Architecture

Sensor data, such as GPS or Gyro, would be easy to access due to the APIs provided; however, other lower level sensors, such as ambient light sensors, are difficult to access on iOS. In addition, cellular data usage is normally not accessible (refer to section 3.1.3).

Processing data: light-weight data processing can be completed on smart phones, but large-scale data analyses must be carried out in the cloud. The actual paritions between data size could be determined by various architectural considerations, such as the amount of raw data stored on a phone, the cost of data transmission and available computing resources.

Provide user feedback: normally, users can access their data though the application provided on their smart phones. There are also other ways, such as web applications or raw data downloading, to access data provided to users for further data mining purposes.

2.3 Mobile Platform

The benefits of mobile sensing compared to traditional testing methods are the time and cost reductions. To collect enough data from various locations, traditional methods involve installing multiple physical sensors or carrying out long-term testing maneuvers, while mobile sensing relies on the movements of mobile users. It saves time and cost to use mobile sensing methods when the target of the amount of the data is large and involves information regarding mobile users.

Various types of app stores, such as the Apple App Store, Android Google Play and Windows Phone Store, make the distribution of the application to massive users possible, e.g. there are more than 200 million active iOS users (the average life of an iOS device is 2 years), and significantly reduce the time and cost in installing physical sensors in multiply locations. For instance, Apple Developer costs $99 per year, and a more powerful enterprise version is $299 per year for in-house distribution. Android developer charges a one-time fee of $25 to display their app on Google Play.

Using all iOS devices as data collecting tools may seem unilateral, but nearly half of the smart phone shares (reported by Nielsen in 2014) in the US proves its feasibility. In terms of data usage, iOS applications are acknowledged as user-friendly and functionality rich, which are key factors when the purpose of the data collection is to analyse the way mobile data is used.

2.4 Mobile User Behaviour Monitoring

Existing methods have only considered data traffic transmitted through one media platform. For example, there was a data analysis based on the traffic data collected from a large Chinese 2G and 3G service provider [16], which focused on the data usage, mobility and application pattern. [17] used mobile search data of queries from a major US mobile service provider to analyse mobile user search behaviours. [18] captured mobile Internet user behaviours through a ‘micro’ web browser. Since the data used in these analyses were collected from service providers, there are limitations in that they could not accurately describe the motion of users (e.g. movement and exact location) and could not systematically analyse all data used in Wi-Fi and WWAN networks.

III. SYSTEM IMPLEMENTATION

A complete system was designed for monitoring user behaviours regarding mobile data usage, which is discussed in this section. The system, shown in Figure 3, consists of three individual parts: the user app, web server and administrative app. The data collected from each user was uploaded to the web server and then centralised to the administrative app for analysis and visualisation. The user app was an iOS application, which is mainly used for monitoring users' daily mobile behaviours and providing users with the option to review any previous data collected. The web server acts as a bridge between user apps and the administrative app, providing public URL website access, back-end data request responses and a database to store data sent from all user apps. The administrative app gathers all available data on one iOS device, typically an iPad with a large screen, where further analysis takes place.

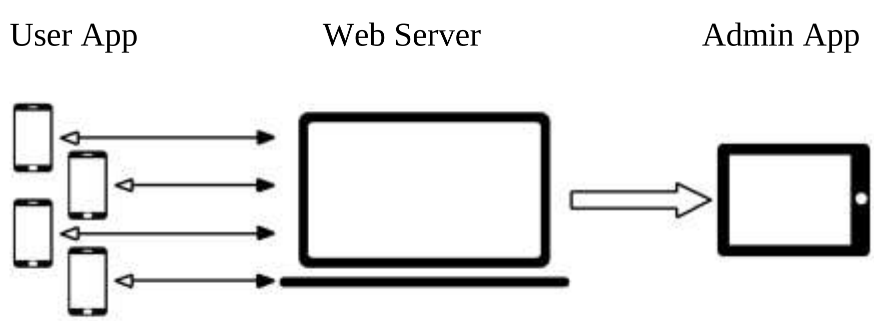


Figure 3: System overview.

3.1 User App

Users were provided with an iOS app to install on their iPhones, providing the ability to automatically collect, examine on a map view and optionally upload data to the web server. Figure 4 shows that the app consists of four individual modules. Data collection module automatically collect data in background. Data storage module then stores the data collected in a local database. The data can be reviewed on a map. The details can be uploaded to the web server. The user will decide to upload the data or not (the dashed arrow).

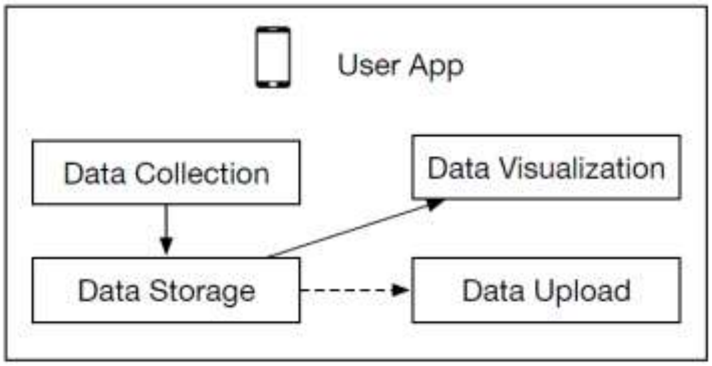


Figure 4: User app work flow.

3.1.1 Application Breakdown

Data types, such as raw data, that need to be collected vary for different types of applications. To provide enough information for further analysis and potential simulations in later stages, the data were separated into two sets, which are phone data and personal data, as shown in Figure 5. On-device and personal data can be highly correlated in mobile user behaviours.

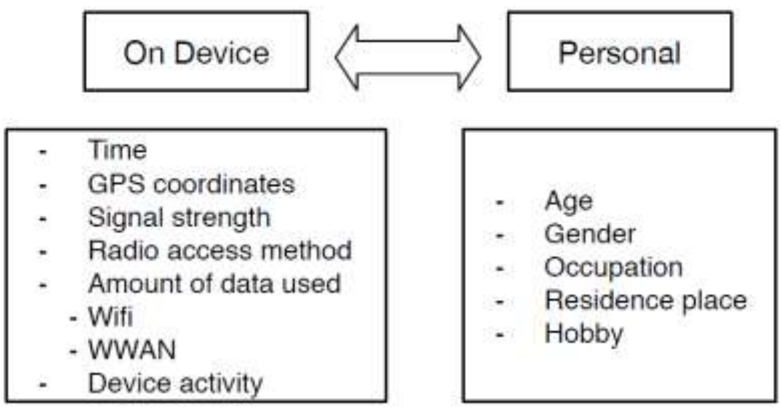


Figure 5: Data Types Used in Sensing.

On-device data, such as time, GPS coordinates and amount of data used, are three essential components for predicting data throughput. Also, to meet the potential of 5G requirements, data usage must be separated between WI-FI and WWAN. Signal strength should also be taken into account, as it significantly affects the performance of cellular networks. For instance, poor signal on a high-speed train usually results in heavy latency of interest usage. Furthermore, radio access methods, such as EDGE, which stands for 2.75G, HSDPA and WCDMA, which stand for 3G, and LTE, which stands for 4G, are used when all smart phones attempt to switch to a lower standard service when the signal strength of the higher standard becomes too low. Here, the term ‘device activity’ refers to the actual physical status of the phone, such as which direction the device is facing or the speed at which the device is travelling, which indirectly reflects the user’s movement and use of the phone. This could provide substantial evidence that the use of data depends on the user’s habits

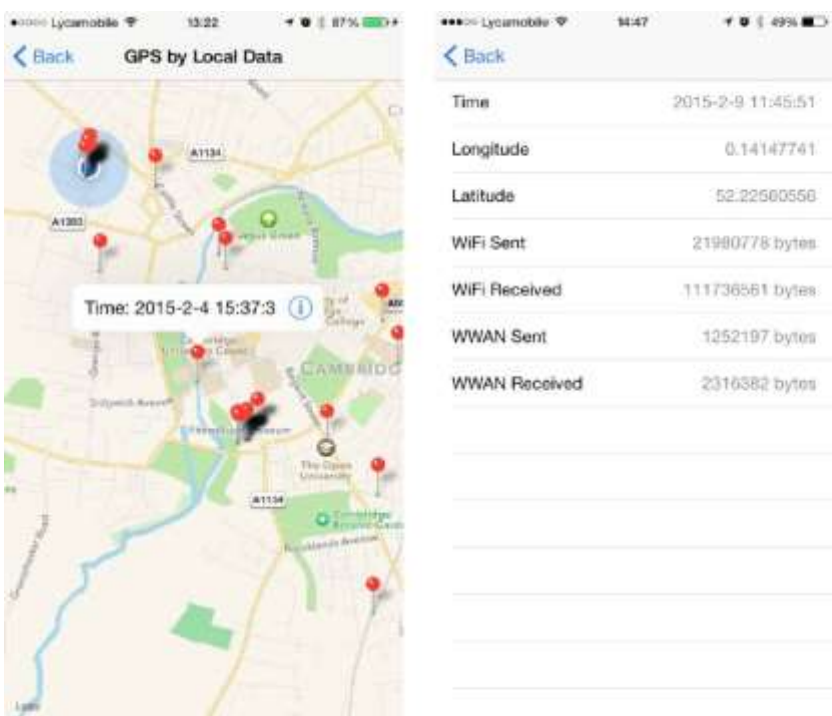
Personal information was collected using a web page questionnaire. This information is equally important because personal routines could affect the geographical position of a phone. For instance, when comparing weekdays and weekends in central London, the decreased population only applies to office buildings on weekends, and there is still a high population on the roads and in shopping centres, which means that office service providers could dynamically adjust the regional bandwidth in the future to meet these predictable requirements. Another good example could be the flow of the population in which a broad range of age or occupation groups determine the movement tendencies. Service providers along heavy traffic routes, especially during peak times, should correspondingly adjust any related resources.

Sampling frequency should be balanced between data accuracy and performance impact. Although the time used for a single data read and store process is brief, it always causes minor delays in the system (explained more in section 3.1.2). Constantly listening to the data exchange port requires a background thread that continuously drives the hardware, which utilises a large amount of system resources, i.e. CPU and RAM, especially for smart phones. Thus, the total amount of data transmitted within a certain period would be a good estimate when the diverse routines of different users are taken into account. More specifically, the update process would only take place when the user changes locations. Therefore, the data transmitted focuses more on a moving device than a still device. In other words, if a user remains at a location for a long period of time, the data traffic rate varies, while the total amount of usage from different internet access methods reflect more information regarding internet usage behaviours because mobile phones are portable devices that are not as powerful as PCs or as enjoyable as entertainment equipment; however, if a user is actually moving or travelling, then the data rate could indicate more behavioural factors, assuming the mobile phone is the only device in possession that can provide an internet connection.

Data storage for this app uses a database rather than simple le based solutions, such as XML, due to the potentially large amount of data stored and the extensibility. On iOS devices, XML manipulation is not recommended by Apple due to the limited RAM availability. Core data is used as the database, as it is the most efficient method that is fully backed-up by the iOS system.

The database design impacts the performance in accessing data. A structure that is too simple limits the extensibility, and a structure that is too complex requires more inter-table transforming. Within the user app, the database has a single table structure (shown in Figure 8(a)). There is no entity relationship involved in this database because all attributes of each entry of the table are accessed together each time the data is requested for map viewing or stored during the collection process. Otherwise, somewhat costly multi-table operations must be carried out each time.

Data visualisation provides a better way for users to review the data collected in the past, which involves two different techniques (Figure 6). The map view (Figure 6(a)) offers a clear overall representation of GPS location histories, as the pure coordinates are not understandable without a map. The table view (Figure 6(b)) shows details of each entry, which is suitable for listing all available information collected at one GPS location.



(a) Map view (b) Table view

Figure 6: On-Device Showing Data Collected

Data upload was optional for all users, although all current participants agreed to provide their information periodically. The advantages of providing information periodically rather than uploading immediately after performing each data collection include decreasing privacy concerns, saving each byte of cellular usage and ensuring the accuracy of the statistics. The upload procedure consists of two steps. The first step is to upload data within the local database, i.e. entries of Figure 8(a), which is performed by packing all data beginning from the last successful upload into a JSON string (also worked around an issue mentioned in Section 3.1.3). The string is then resolved and inserted into a database by the web server PHP service (details in Section 3.2). The second step is to upload device activity data, which is an estimation generated by the Apple M-series motion co-processors chip [19], using the same method.

3.1.2 Programming Implementation

An iOS app was fully developed in Objective-C and was tested using XCode. The key iOS frameworks and services involved are discussed in the subsequent subsections.

General info	
Code written	2600 lines
Compiled size	1.3 Mb
Generated data size	1.4 Mb
Performance measures	
Peak CPU usage	22%
Peak memory usage	61.4 Mb

Table 3: User app software information. Figures shown are based on the debugging tool in XCode when testing on a 64Gb iPhone 6 and loading more than 1000 data points onto the map.

Core Location frameworks use the available hardware to determine the user’s position and direction. The CL Location Manager is initiated when the app is launched to obtain GPS coordinates and a system’s estimated speed. The service desired accuracy is set to the nearest ten meters, which gives an accurate enough location co-ordinate for locating a person within a building but consumes less energy than the best available accuracy. This occurs because the hardware is activated for a shorter period of time, as suggested by Apple. The CL Location Manager listens to system monitoring for significant location change notifications when running in the background (typically after pressing the home button on the iPhone), which can initiate the application or relaunch it if terminated to save battery life by a previous system memory shortage. It switches to listening to continuous location updates when running in the foreground mode to ensure quick responses.

Core Data frameworks provide generalised and automated solutions for common tasks associated with object lifecycles and object graph management, including persistence. The entity structure used is shown in Figure 8(a).

Core Motion frameworks allow the application to receive motion data from device hardware and to process that data. Gyroscope, accelerometer, magnetometer and other related data are retrieved to display in the app but are not saved or uploaded due to the nature of activation by the system. For devices with a built-in gyroscope, the raw gyro data as well as the processed data that reflects the attitude and rotation rates of the device can be received (Apple Framework Reference).

Motion Activity databases are queried by filtering out activities with high-confidence levels, which are then uploaded when available (iOS devices launched after iPhone 5S).

Background refresh services are activated when the location manager requests 'Always Authorization' from users to allow the app to use location services even when the app is running in the background.

3.1.3 Challenges

Most of the data types can be easily accessed by APIs provided by Apple; however, there is no public documentation provided by Apple that explains how to access the total amount of data used, so that usage of private APIs is investigated in which a low-level system header files needs to be included and string manipulation is required.

As a result of using private APIs, the user app cannot be submitted to Apple's official iOS app store. Hence, public iOS users are not able to search for or download it easily. Nevertheless, Ad-hoc provisioning with an online installation method was used to distribute the app globally, which allowed users to install the app directly on their phone rather than connecting to a computer.

The upload process was initiated by sending an HTTP POST request along with the data needed to be transmitted to the server, which is an industry standard le uploading method; however, a system error will occur when a large amount of data is sent, which is a NSNetworkError—'the connection has lost'—after successfully uploading a certain amount of data each time. The error cannot be resolved by either iOS or server side techniques, so it is then reported to Apple as a system bug.

3.2 Web Services

The web server provides services for web pages to introduce the project responses for data uploads or to fetch requests from the database to store data, as shown in Figure 7. It is hosted by an efficient independent server provider with a customised website domain name and HTTPS certificate, providing a stable and secure connection so that any further development in the future would be easier.

3.2.1 Services Breakdown

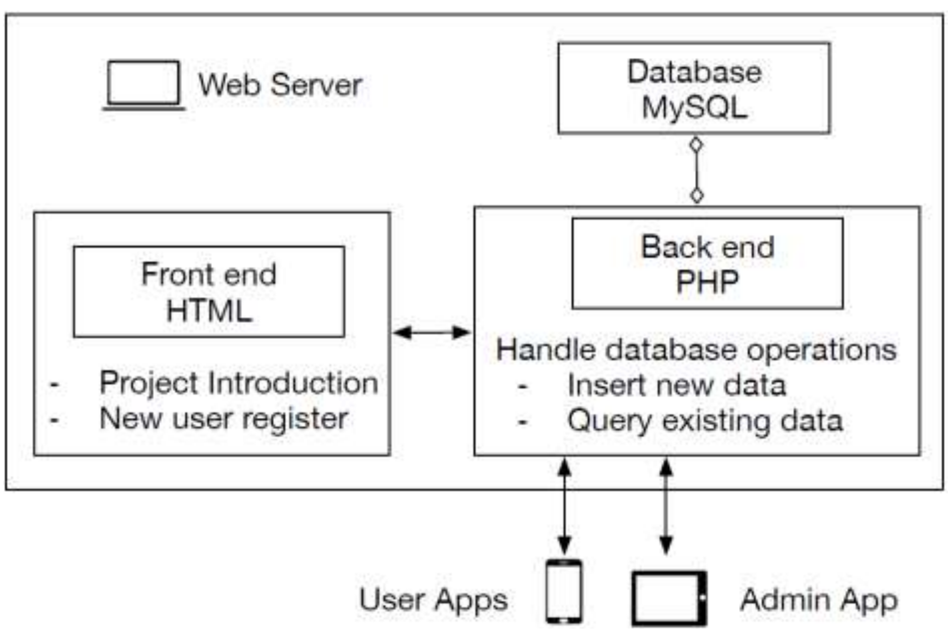
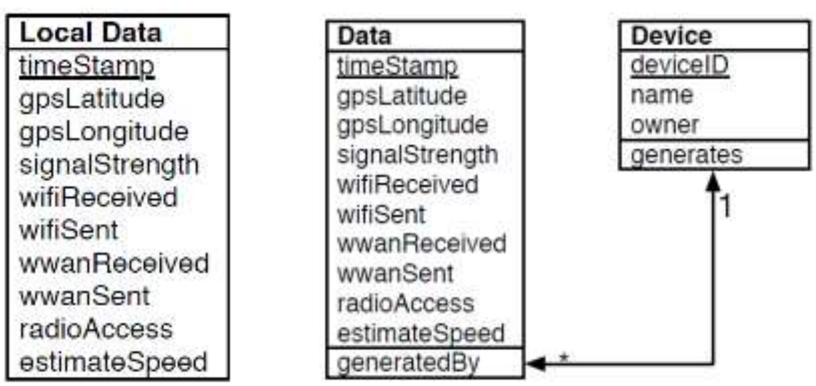


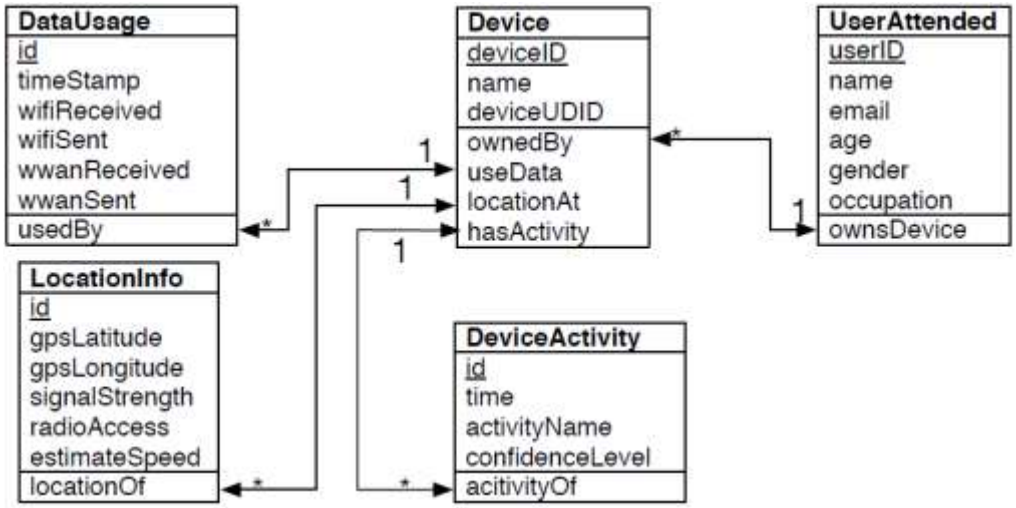
Figure 7: Web Server Work flow. Back-end service handles all requests that need to be queried in the database from the front-end service and the two iOS Apps and performs secured database operations so that data is only available to privileged users.

HTML front-end web pages were used to introduce the project to the public as well as to provide a registration form for new test users to fill in. New users' devices must be registered with Apple before being used under Ad-hoc provisioning distribution. Using this method, testers do not need to be team members or iTunes Connect users to run the app, but their devices need to be registered in the Member Center (Apple Developer App Distribution Guide), which is a manual inputting process conducted by the developer. Therefore, after a new user completes the registration form online, the information is securely stored in the database, and an automatic email is sent to alert the developer to finish the input as soon as possible.

The iOS app online installation service enabled users to automatically download and update the application through a webpage link, providing an alternative way to submit the application rather than through the official app store. This was achieved using a special HTML hyperlink protocol, which is only used by Apple for app distribution. The method is essentially a work-around version of enterprise in-house app distribution provided by Apple, so using HTTPS protocol to communicate and transmit all relevant files is required. PHP back-end services handle all requests from the iOS user app, administrative app and front-end web pages to insert and query the MySQL database. The data structure is shown in Figure 8(c).



(a) User App (b) Administrative App



(c) Web Server

Figure 8: Database Schemes UML Class Diagram

The database on the web server stores more information than the iOS apps, so a more sophisticated database structure was introduced (Figure 8(c)). Multiple one-to-many relationships were used from the top level, e.g. the user could own multiple devices, to the bottom level, e.g. one device generates several data. The table structure of the generated data is a separated version of the on-device structure, except that the device's activity table contains information regarding device activities that is not directly available when each collection action is performed (also discussed in section 4.3).

3.3 Administrative App

The administrative app is another iOS application, which centralises all data from different users and displays different combinations of data needed on the map for analysis. The work flow is shown in Figure 9.

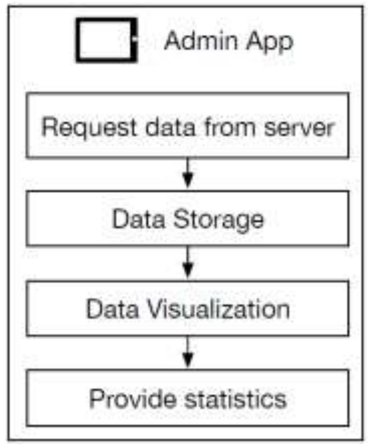


Figure 9: Administrative App Work Flow. Two types of requests, i.e. list of devices and data collected from certain device(s) are sent to a web server, and the returned data is passed on to the storage stage. Then, the visualisation stage provides the demanded data and calculates the corresponding statistics.

3.3.1 Theory Involved

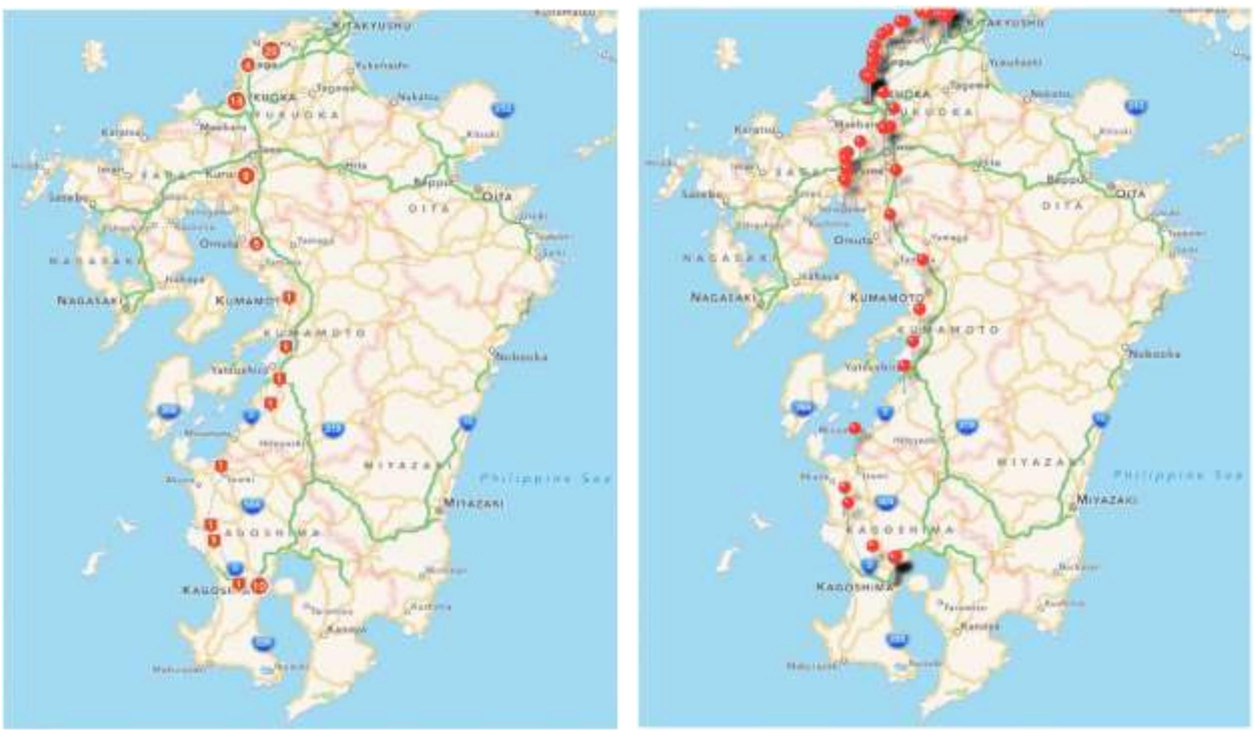
Data visualisation in this app needs an extra method to be displayed on the map because the amount of data is much higher than on a user app for a single device. As all data points are highly correlated with geographical positions, viewing data on a map as a starting point is the most appropriate option. When the number of data selected to be viewed increases to a relatively large amount (typically over 50 points within a single place), it is better for both device performance and for providing a straightforward analysis to cluster data points that share close GPS coordinates.

General Info	
Code written	1800 lines
Compiled size	1.2 Mb
Generated data size	8.5 Mb
Performance Measures	
Peak CPU usage	43%
Peak memory usage	120.5 Mb

Table 4: Administrative app software information. Figures shown are based on the debugging tool in XCode when testing on a 64Gb iPad Air and loading more than 4000 clustered data points onto the map. Lines of code exclude all third-party libraries used.

Furthermore, the data of points within one cluster should represent useful statistical results, i.e. a mean or correlation.

The clustering technique used by the administrative app is called quad tree clustering, which is most often used to partition a two-dimensional space by recursively subdividing it into four quadrants, or regions. It is efficient in comparing two-dimensional, ordered data points, which usually operate in O(n log n) time [20]. A disadvantage of the quad tree structure is that deleting and merging the tree nodes is complex, which is not relevant in this visualisation case because data shown on the map did not change after being loaded onto the map. After implementing the clustered technique into the app, no noticeable delays could be observed with panning, zooming or rotating gestures applied to a map with more than 1000 clustered points. Screenshots of the differences before and after clustering are shown in Figure 10.



(a) Clustered (b) Non-clustered

Figure 10: Map View Options to Display the Same Data Set in the Administrative App

Providing statistics after selecting a certain number of data points within the same cluster will offer a sample mean. Because the user number was small during the only test run (< 10), it was impossible to conduct a proper statistical analysis; however, from the user case perspective, users could be classified into different groups to identify potential similarities or differences. The classification method could be based on several factors.

The geographical area can be as large as countries. For instance, the sample data were collected from the UK, US, China, Korea and Japan. It could also include functionality regions, such as residential and commercial areas, in which mobile behaviours could have a significant impact. Occupations, e.g. student or office worker, provide a general idea of the age groups, for which work, free time and interest trends vary.

Characteristics include users' major social activities or habits, which influence the key information demanded. It could also reveal physical locations that are visited by people with the same interests. Using the administrative app, the proposed classification methods can be performed by selecting the correct combinations of devices and time intervals to be shown on a map view (results given in section 4).

3.3.2 Programming Implementation

Requesting data for both device lists and from detailed data collected by servers is achieved by sending a HTTP Post request, which is returned with a JSON object to be resolved and inserted into a local database.

Data storage is designed based on the scheme of the user app, which contains information collected from different devices, so the extra device table is related to the data table with a one-to-many relationship (Figure 8(b)).

Data visualisation methods can alternate between showing all points, which is used in the user app, or showing a clustered version, as mentioned previously. The clustering method is a modified version of an open source project called ADClusterMapView provided by GitHub.

Statistics are provided in different ways based on whether or not a clustered data point is selected. Average data usage over time is given if a clustered point is selected; otherwise, a list of the original data collected is shown.

IV. RESULTS AND DISCUSSION

4.1 Results

Ten devices with the user app installed were analysed in the current stage, and over five thousand sets of data were collected (one set of data refers to data collected from different sensors at one location). The system functions on a global scale with existing data from the US, Europe and Asia. The processed results of the data are presented in this section.

User	Gender	Age group	Occupation	Residence
A	M	< 30	Student	Cambridge
B	F	< 30	Student	Cambridge
C	M	< 30	Student	Cambridge
D	M	> 30	Employee	Beijing
E	F	> 30	Employee	Shanghai

Table 5: Five typical users were selected to be compared in a later stage, as they all travelled to different cities or countries during the data collection stage. The difference in locations can be observed based on the data collected in different countries due to various travelling purposes.

Figure 11 shows the screenshots of the administrative app with plotted sequences of GPS coordinates on a map. Figure 11(a) shows an eleven-day trip to northern Europe, which began in London and continued through Finland, Sweden and Denmark. Train travel is represented by consecutive points on the map, whereas there are separate points for airplane travel. The points in Figures 11(b) and 11(d) refer to trains, and the points in Figure 11(c) refer to planes.

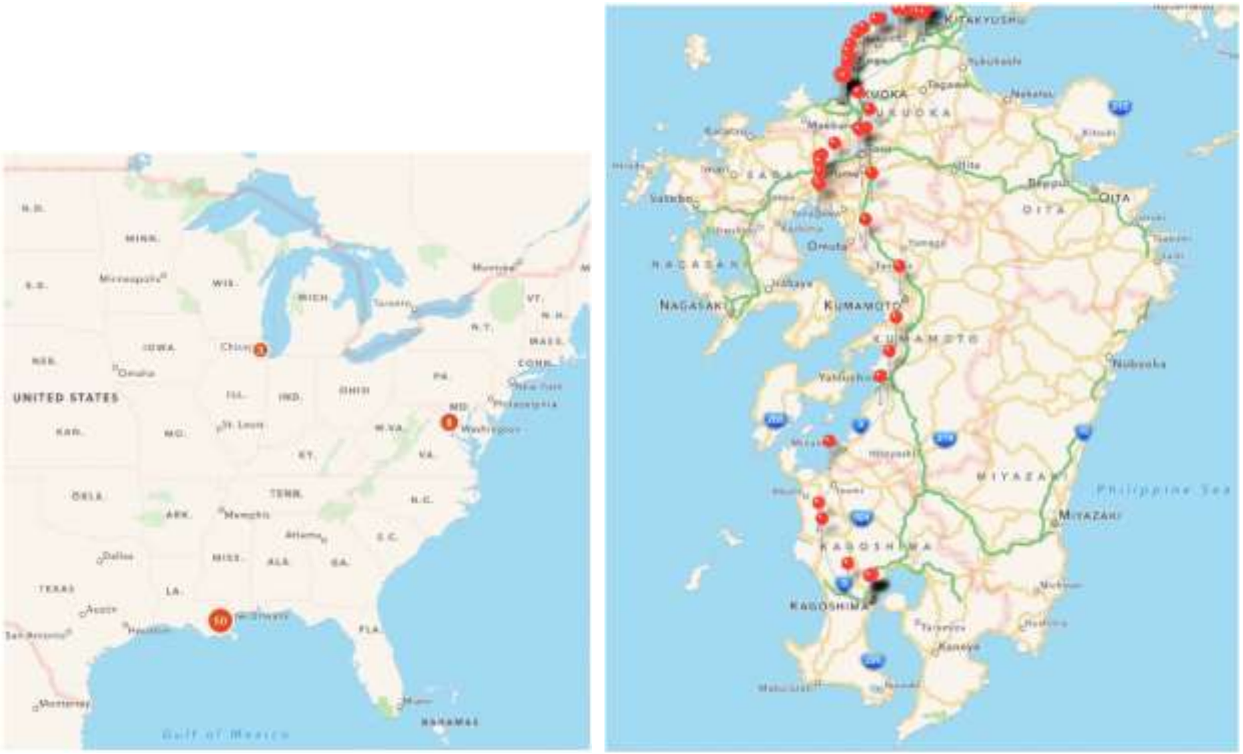
Based on the internet usage, it can be observed that users strongly preferred Wi-Fi to WWAN during travel, as shown in Table 6, even if the cost of WWAN was acceptable; however, Users C and E chose to purchase an external Wi-Fi hotspot so that friends could also have an internet connection. The users may not have enabled the hotspot sharing feature on their phones because they wanted to preserve battery life or because the SIM card provider could remotely switch to this feature on the phone.

User	Destination	Duration	Wi-Fi Usage	WWAN Usage	W2W Ratio
A	Northern Europe	10 days	302 Mb/Day	30 Mb/Day	10
C	South Korean	13 days	292 Mb/Day	0	-
D	US	5 days	535 Mb/Day	50 Mb/Day	11
E	Japan	5 days	1551 Mb/Day	0	-

Table 6: Comparison of Wi-Fi and WWAN usage during trips to foreign countries

Another interesting comparison involved usage at the office and at home, which is shown in Figure 9. For student users, i.e. A, B and C, ‘office’ referred to departments or facilities of the university, and ‘home’ referred to accommodations during the term. Wi-Fi was generally used more than WWAN with one exception in which the duration for User C at work was far less than all other sample durations. Users generally used about the same or more Wi-Fi at home than at the office, though User A was an exception. User A was charged to use Wi-Fi for large traffic at home, but it was free at the office. According to the data, the usage of WWAN was completely random. Figure 12 shows that the clustering function in the administrative app could correctly identify the locations ‘home’ and ‘office’ by filtering out higher visits.

(a) User A travel in Europe (b) User C travel in South Korea



(c) User D business in US (Clustered) (d) User E travel in Japan

Figure 11: Data Points of Different Users Globally

Figure 12: Clustered plot showing the correct home (star) and office (square) locations of different users in Cambridge.

The difference in time periods was also investigated, which was classified into two categories: day and night or weekday and weekend.

Tables 7 and 8 show the differences between office vs home and day vs night, respectively. First, all users still used far more Wi-Fi than WWAN during the day; however, the difference of usage between day and night depended on individual user behaviours. For example, Users D and E used more Wi-Fi at night, but Users A, B and C used more Wi-Fi during the day. The use of WWAN was also varied based on daytime use or nighttime use.

User	Location	Wi-Fi Usage Mb/24hr	WWAN Usage Mb/24hr	W2W Ratio
A	Home Work	239 723	3 5	69 137
B	Home Work	170 174	0.6 58	267 3.0
C	Home Work	159 24	83 33	1.9 0.7
D	Home Work	262 291	20 9	13 31
E	Home Work	107 37	1.1 23	95 1.6

Table 7: Comparison of usage at the office and at home. Note that both usage numbers have been rescaled to 24 hours.

User	Time	Wi-Fi Usage Mb/24hr	WWAN Usage Mb/24hr	W2W Ratio
A	DAY NIGHT	243 154	12 23	20 6.8
B	DAY NIGHT	404 249	11 2	37 107
C	DAY NIGHT	290 95	44 74	7 1.3
D	DAY NIGHT	280 311	16 19	17 17
E	DAY NIGHT	265 380	14 6	19 62

Table 8: Comparison of usage during the day and at night. Note that both usage numbers have been rescaled to 24 hours.

Table 9 shows the differences in Wi-Fi and WWAN use for different users during weekdays and weekends. All users used more Wi-Fi than WWAN during both weekdays and weekends. There was no general preference in data use between these two periods.

User	Time	Wi-Fi Usage Mb/day	WWAN Usage Mb/day	W2W Ratio
A	Weekday	225	13	17
	Weekend	175	25	7.1
B	Weekday	316	5	66
	Weekend	288	7	40
C	Weekday	141	54	2.6
	Weekend	249	90	2.8
D	Weekday	278	15	18
	Weekend	329	23	15
E	Weekday	369	11	32
	Weekend	180	9	21

Table 9: Comparison of usage during weekdays and weekends

The differences in age groups, which is the same as the occupations of users listed in Table 5, is shown in Table 10. WWAN was used three times more often for younger users than senior users, whereas Wi-Fi usage was generally the same. It is worth noting that this could have occurred because the cost of cellular data in the UK, which is where all three younger users lived, is much less expensive than the cost in China.

Age Group	Wi-Fi Usage Mb/day	WWAN Usage Mb/day	W2W Ratio
< 30	234	39	5.0
> 30	278	13	21.2

Table 10: Comparison of usage between age groups

Overall, the data shown in this section revealed that Wi-Fi is used more often than WWAN, which could be because devices are actually designed for Wi-Fi use due to the following reasons. Wi-Fi has about 100 times the bandwidth of WWAN, as discussed in section 2. Wi-Fi is generally less expensive, as most Wi-Fi services are charged by a data rate with an unlimited amount, whereas WWAN is normally charged based on the total amount of usage. The energy consumption of using Wi-Fi is lower than WWAN [21].

4.2 5G Prospect

Based on the results presented in the previous section, the data usage behaviour of each user is unique, though there are some similarities. Of the 5G use case families, which are shown in Table 2, the first four examples are discussed in this section.

Pervasive video is defined as person-to-person or person-to-group video communication with extremely high resolution [2] using suggested hardware, including optical head-mounted displays, collaboration in 3D cyber-real offices or operating rooms with both a physical and virtual presence. Currently, HD video calls from smart phones require at least 1 Mbps upload bandwidth [22, 23], which is achievable for both WWAN and Wi-Fi technologies. In fact, User A placed a long video call on the way from home to work without Wi-Fi access. On the other hand, it would also be helpful for all advanced video encoding technologies, such as H.265, which is used for FaceTime over a mobile network, to achieve twice the compression ratio as the previous H.264 standard.

Another example in the same use case family is HD video sharing in stadiums, which is characterised by a high connection density and potentially temporary use [2]. Currently, a 1080p video is recommended to have a bandwidth of 4.5 Mbps [24], which suggests that a 5G bandwidth would need to target not only an average bandwidth for individual users but also total bandwidth within a certain three-dimensional space. Using current technology, making a traditional phone call would be difficult at a stadium during a popular sport match.

The standard is '50+ Mbps everywhere', and it has been emphasised that the user rate must be delivered consistently across the coverage area even at cell edges. This indicates the importance of signal strength, which is directly proportional to the rate at which a wireless signal can transfer data. An informal explanation is that a better signal strength implies a higher signal to noise (SNR) ratio, and thus less errors. According to the Shannon-Hartley theorem, the maximum achievable data rate depends on SNR. Table 11 shows the average signal strength for different users, and the results of both the lower values and the differences between the two averages indicated that Users A and C normally experienced poor signal, whereas Users D and E had an excellent signal. The results should be correct because both the homes and offices of Users A and C always had poor signal, and Users D and E had a good signal because large cities in China generally have a very good signal. This indicates that significant improvement is required for 5G.

User	Average Signal Strength	
	Exclude 0	Include 0

A	31.26	21.80
C	40.41	38.02
D	75.45	71.32
E	60.55	59.14

Table 11: Average signal strength from different users. Strength of 50 shows full strength on the indicator of the phone, and 0 means no signal at all.

The high-speed train example demonstrates the growing demand for mobile services in vehicles, trains and aircrafts. In the UK, User A travelled from Cambridge to London on a train at speeds around 120 km/hr, and the average signal strength, excluding both end stations, could be as low as around 10, resulting in a mobile webpage that could not been loaded. In contrast, User D travelled on a high-speed train (more than 300 km/hr) in China, and the average signal strength still remained at an acceptable level of 30. Furthermore, user E travelled on a high-speed train (more than 280 km/hr) in Japan, and the average signal strength was above 35. Thus, especially in the UK, mobile services require considerable improvement.

Sensor networks and smart wearables are common applications that have been developed due to the massive growth in the Internet of Things. Using sensors on a smart phone, as shown in Figure 1, smart phones and wearables, such as the iWatch, could grow into a huge mobile network of sensors; however, continuous and reliable internet connections are likely the most important current issue.

Moreover, a very small portion of the GPS coordinates recorded were inaccurate. Since current smart phones mainly use one or many of these locating technologies GPS Network based. For Wi-Fi based positioning systems (WPS) (systems that include indoor positioning systems), the location detected by the system might be slightly off towards any nearby base stations if the GPS signal is weak and Wi-Fi is not connected, which can be observed by comparing base station locations (Figure 13) with any off-location coordinates in Cambridge.

(a) Cambridge Base station location

Figure 13: Location Accuracy Based on Base Station

4.3 Future Work

The current scale of behaviour monitoring is too small to produce useful statistical results, so future studies should increase the size of the user group being monitored. With a greater number of users involved, it would be possible to design a model for real-time simulations on super computers. Potential ways to achieve this goal are discussed in this section.

The system should be promoted to more users globally using either mobile network service providers or major mobile phone manufacturers. With the general influence of these companies, as well as the large amount of data available, additional user behaviour information could be explored. Furthermore, a wider mobile platform of the user app, other than the iOS, would need to be implemented, such as for Android and Windows phones.

Improved existing systems could be designed for the public by providing more useful feedback information after analyses. Improvements could also be made in advanced device activity algorithms, rather than reading results processed by the M-series motion co-processors chip on iOS devices [19]. Websites could also potentially provide web services to access and display personal and overall analysed data.

VI. CONCLUSIONS

A complete system has been implemented to investigate user behaviours for mobile data usage, including an iOS application installed on iPhones, a web server providing data requesting and storage services and another iOS application used for the overall data analysis. An application for data collection purposes was installed on ten users' phones over a two-month period, and the phones were fully functional in eight different countries across three continents. Based on the data collected, it can be concluded that mobile user behaviours for data usage are unique to each individual. Moreover, all users tended to use 10 times more Wi-Fi data than WWAN data.

Although the current results could not offer a full statistical analysis due to the small sample size, it would be possible to achieve this by increasing the number of users involved. If additional data are collected in the future using the system developed, it is possible to design a dynamic simulation system for modelling real-time mobile service usage and providing feasible suggestions to dynamically allocate limited network resources.

According to the data and the feedback from the users, several 5G requirements given in [2] are still quite far from being achieved. Some of the use case families mentioned in [2] have been proven to be in demand even with current technologies. The current quality of mobile services in the UK was found to be inferior to the mobile services in the US and some east Asian countries, such as China and Japan, which indicates that mobile service infrastructures should be planned in advance to accommodate the upcoming generation of technology for 5G.

REFERENCES

[1] GSMA. The mobile economy 2014. http://www.gsmamobileeconomy.com/GSMA_ME_Report_2014_R2_WEB.pdf.

[2] NGMNA. NGMN 5G white paper. https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf.

[3] A.H. Khan, M.A. Qadeer, J.A. Ansari, and S. Waheed. 4G as a next generation wireless network. In Future Computer and Communication, 2009. ICFCC 2009. International Conference on, pages 334-338, April 2009.

[4] Anritsu MT8221B High Performance Base Station Analyzer Manual, 2011.

[5] Rohde-Schwarz ROMES4 Drive Test Software - Product Brochure, Nov 2011.

[6] R. Mizouni and M. El Barachi. Mobile phone sensing as a service: Business model and use cases. In Next Generation Mobile Apps, Services and Technologies (NGMAST), 2013 Seventh International Conference on, pages 116-121, Sept 2013.

[7] P. Leakkaw and S. Panichpapiboon. Speed estimation through mobile sensing. In TENCON 2014 - 2014 IEEE Region 10 Conference, pages 1-5, Oct 2014.

[8] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application. In Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, SenSys '08, pages 337-350, New York, NY, USA, 2008. ACM.

[9] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda. Peir, the personal environmental impact report, as a platform for participatory sensing systems research. In Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, MobiSys '09, pages 55-68, New York, NY, USA, 2009. ACM.

[10] B. P. L. Lo, S. Thiemjarus, R. King, and G. Z. Yang. Body Sensor Network A Wireless Sensor Platform For Pervasive Healthcare Monitoring. In Proceedings of the 3rd International Conference on Pervasive Computing (PerCom), May 2005.

[11] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay. Activity sensing in the wild: A field trial of ubi t garden. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, pages 1797-1806, New York, NY, USA, 2008. ACM.

[12] O. Postolache, P.S. Girao, P. Sinha, A. Anand, and G. Postolache. Health status monitor based on embedded photoplethysmography and smart phone. In Medical Measurements and Applications, 2008. MeMeA 2008. IEEE International Workshop on, pages 39-43, May 2008.

[13] N. Nawka, A.K. Maguliri, D. Sharma, and P. Saluja. SESGARH: A scalable extensible smart-phone based mobile gateway and application for remote health monitoring. In Internet Multimedia Systems Architecture and Application (IMSAA), 2011 IEEE 5th International Conference on, pages 1-6, Dec 2011.

[14] A. Sieber, X. Yong, A. L'Abbate, and R. Bedini. Cardiac sentinel: A smart GSM based embedded solution for continuous remote monitoring of cardiac patients. In Intelligent Solutions in Embedded Systems, 2008 International Workshop on, pages 1-11, July 2008.

[15] N.D. Lane, E. Miluzzo, Hong Lu, D. Peebles, T. Choudhury, and A.T. Campbell. A survey of mobile phone sensing. Communications Magazine, IEEE, 48(9):140-150, Sept 2010.

[16] J. Yang, Y. Qiao, X. Zhang, H. He, F. Liu, and G. Cheng. Characterizing user behavior in mobile internet. Emerging Topics in Computing, IEEE Transactions on, PP(99):1-1, 2014.

[17] M. Vojnovic. On mobile user behaviour patterns. In Communications, 2008 IEEE International Zurich Seminar on, pages 26-29, March 2008.

[18] T. Yamakami. Toward understanding the mobile internet user behavior: a method-ology for user clustering with aging analysis. In Parallel and Distributed Computing, Applications and Technologies, 2003. PDCAT'2003. Proceedings of the Fourth Inter-national Conference on, pages 85-89, Aug 2003.

[19] Apple. iPhone6 technology. https://www.apple.com/iphone-6/technology/.

[20] R.A. Finkel and J.L. Bentley. Quad trees a data structure for retrieval on composite keys. Acta Informatica, 4(1):1-9, 1974.

[21] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani. En-ergy consumption in mobile phones: A measurement study and implications for network applications. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, IMC '09, pages 280-293, New York, NY, USA, 2009. ACM.