The following publication P. Wepulanon, A. Sumalee and W. H. K. Lam, "Temporal Signatures of Passive Wi-Fi Data for Estimating Bus Passenger Waiting Time at a Single Bus Stop," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 8, pp. 3366-3376, Aug. 2020 is available at https://doi.org/10.1109/TITS.2019.2926577.

Temporal Signatures of Passive Wi-Fi Data for Estimating Bus Passenger Waiting Time at a Single Bus Stop

Piyanit Wepulanon, Agachai Sumalee, and William H.K. Lam

***Abstract*— This paper proposes an alternative method for bus passenger waiting time estimation using passive Wi-Fi data. With the underlying mechanism for Wi-Fi communication, the presence of Wi-Fi enabled devices in a particular area can be passively discoverable. The mobile devices carried by bus passengers can be exploited to estimate passenger waiting time at a bus stop without the passengers' direct participation. Passenger waiting time estimation using such opportunistic data is challenging due to the particular characteristics of the Wi-Fi data collected from bus stop environments. This paper proposes a methodology to handle massive noise in Wi-Fi data and identify the potential Wi-Fi records which are derived from passengers' devices. The filtered data can then be used to estimate passenger waiting time. Wi-Fi data collected from a bus stop in Hong Kong are used as a case study for evaluating the proposed system. The practicality is investigated in terms of estimation accuracy and insightful analysis.**

***Index Terms*— Average waiting time, Bus stop based waiting time, Opportunistic data, Passenger waiting time estimation**

I. Introduction

Passenger waiting time constitutes the part of the total transit trip time which could be the most challenging factor to measure. For bus transport, the value of waiting time is crucial as it could represent the duration of the passengers' exposure to traffic emissions. Since passengers have direct experiences of waiting for buses, the waiting time could affect their perceived performance of bus services. Average passenger waiting time (AWT) has been suggested as one of the performance measures to evaluate the quality of bus services. Furthermore, AWT has been considered in transit assignment models for operational improvement and planning [1], [2].

A direct method used to derive AWT at a bus stop involves calculating the average length of waiting times by observing individual waiting passengers. However, such direct measurement is costly and impractical to perform over a long time period. In previous literature on this topic, indirect methods have been developed to estimate AWT at a bus stop as a function of bus headway. This could be a viable solution for AWT estimation as observing bus headway is simpler. However, the estimation relies on two major assumptions regarding passenger arrival patterns and bus headway distribution at the bus stop.

Two major assumptions were made in the traditional models namely uniform passenger arrival and independent bus headway [3]. From these, AWT was assumed as half of the bus headway. Nevertheless, several studies demonstrated that these major assumptions could be invalid in actual bus operations [4] and that the half-headway approach could overestimate the actual AWT. More accurate estimation models have been developed based on other assumptions regarding passenger arrival patterns and bus headway distribution for an observed bus stop [5].

The fact remains that estimating AWT at a bus stop is challenging in practice. Unlike other modes of public transport, the AWT of bus passengers could be affected by a wider range of factors. First, the reliability of bus services can vary due to the effects of several factors on actual bus operations e.g. road traffic conditions. Second, individual passengers may have their own criteria for boarding a bus, especially when common bus routes are operating at the same bus stops. In addition, some passengers may choose to wait for the next bus if the arriving bus is overcrowded.

In densely populated cities such as Hong Kong, making assumptions on passenger arrival patterns and bus services is an onerous task. Firstly, passengers may arrive at a bus stop in a batch [6]. The batch arrival phenomenon could be encountered at multimodal transit stops, and/or in the peak time period of routine activities such as after office hours. Next, high passenger loading could affect passenger behaviour. Passengers serving patterns may not be based on a first-in, first-out order as some passengers may prefer to wait for the next bus. Although high-frequency bus services are operated in Hong Kong, passengers waiting time could be longer than expected. Another challenge for Hong Kong is the accessibility of bus arrival time information. The bus system is competitive and bus agencies are responsible for their own investment and operating costs. Therefore, bus agencies have the right to withhold any information (i.e. bus arrival time) which could affect their revenue.

To tackle these challenges, this paper proposes an alternative solution for estimating bus passenger waiting time. To be more specific, it is proposed that passive Wi-Fi data be considered for AWT estimation owing to the ubiquitous nature of the mobile devices carried by bus passengers (e.g. smartphones) as well as the general use of Wi-Fi technologies. Moreover, Wi-Fi data can be collected without the direct participation of the bus passengers in the data provision. Since the underlying mechanism for Wi-Fi communication is exploited by the proposed method for

This work was supported by the Research Grants Council of the Hong Kong Special Administrative Region (No. PolyU 152101/17E and R5029-18), and the Research Committee of The Hong Kong Polytechnic University (No. 1-BBAR, 1-ZVFJ, and 4-ZZFY).

Piyanit Wepulanon, Agachai Sumalee and William H.K. Lam are with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China. (e-mail:piyanit.wepulanon@connect.polyu.hk; ceasumal@polyu.edu.hk; william.lam@polyu.edu.hk)

AWT estimation, the strong assumptions regarding passenger arrivals and bus headway could be compromised.

In this study, AWT is initially estimated for each bus stop regardless of individual bus lines. With the continuous availability of stop-based AWT, bus transit performance can be evaluated in the spatial aspect over the transit network and in the temporal dimension over time.

Despite the advantages of passive Wi-Fi data, a number of challenges pose extra difficulties in AWT estimation. First, massive noise is included in Wi-Fi data collected from bus stop environments. Second, the temporal resolution of Wi-Fi data is uncertain among different devices. Finally, the Wi-Fi data would only be partial as it is unlikely to capture the data from all waiting passengers. The primary contributions of this paper are summarized as follows:

- An alternative method for estimating AWT at individual bus stops using passive Wi-Fi data,
- A methodological framework and data mining techniques for handling the challenges in passive Wi-Fi data collected from bus stop environments, and
- Practicality analysis in terms of AWT estimation accuracy and system’s limitations based on the actual Wi-Fi data collected from bus stop environments.

The paper is organized as follows. Section II provides the essential background on acquiring passive Wi-Fi data, and thereafter reviews the related works. Section III presents an overview of the proposed system. The details of significant processes including data preparation, and AWT estimation are described in Section IV and V, respectively. Section VI presents the results with insightful analysis performed on a case study. Finally, the paper is concluded in Section VII with a discussion on the direction of future studies.

II. Background

This section provides preliminary knowledge of passive Wi-Fi data, as well as the previous works on deriving human mobility information from passive data.

A. Motivation

Recently, the potential of smart mobile devices as sensing devices for human mobility tracking has been considered. As smart mobile devices become increasingly ubiquitous due to the greater market penetration, the embedded sensors within them create opportunities for deriving digital footprints while the users are conducting activities. Data mining techniques have been already applied to extract the mobility information from various types of sensors. The essential mobility information consists of meaningful stay locations identifying where people conducted activities, and the time when the activities were conducted.

The potential of tracking Wi-Fi enabled devices has been gaining attention for several reasons. The conventional technologies e.g. Global System for Mobile Communications (GSM) and Global Positioning System (GPS) are impractical for indoor tracking. Also, the underlying mechanisms for monitoring Wi-Fi devices offer a non-participatory means of tracking individuals with less privacy invasion.

The method of Wi-Fi data acquisition can be applied to Bluetooth devices since the communication standards of the technologies are similar [7]. However, the use of Wi-Fi is more general. Wi-Fi tracking has been suggested for indoor mobility analysis, particularly in small-scale study areas such as in a hospital and or on a campus.

B. Passive Tracking for Wi-Fi Devices

Wi-Fi devices can be tracked passively using a Wi-Fi monitor. The monitor is responsible for capturing the data in Wi-Fi communication channels. According to the communication standard, a Wi-Fi device is able to communicate with other Wi-Fi devices by transmitting a set of packets which consist of the information needed for the communication. One of the most signification information is the Media Access Control address (MAC-ID) which is a unique identifier assigned to each Wi-Fi device.

A number of studies have considered the potential of a captured MAC-ID with its timestamp for human mobility analysis under an assumption that a Wi-Fi device can be captured as long as the Wi-Fi function is enabled. As a Wi-Fi monitor is capable of sensing Wi-Fi signals within its detection range, human activity in each study location can be analyzed in terms of occupancy time [8]. The analysis can be extended in the spatial dimension using multiple monitors installed at different points of interest [9].

C. Related Works

In related literature, one study using methodology similar to the proposed system adopted passive Wi-Fi data to estimate people dwell time for indoor activities. The most challenging task is identifying the precise entry and exit time at a study location since the resolution of passive Wi-Fi data can be sparse over the temporal dimension. A device can broadcast a Wi-Fi packet anywhere between few seconds and at periods of several minutes. Therefore, the estimation accuracy relies on the frequency with which Wi-Fi data can be detected from the device.

Due to the variability in detection frequency, three major approaches have been implemented for dwell time estimation. Table I summarizes the combination of approaches which have been adopted in previous studies.

To begin with participatory-based systems, participants can be asked to install a smartphone application which can improve the detection frequency by increasing Wi-Fi data broadcasting. The participatory-based systems are capable of estimating short occupancy durations with high accuracy [10], [11], whereas non-participatory based systems are suitable for capturing the activities on which people tend to spend a considerable amount of time in the study locations.

TABLE I
SUMMARY OF THE PREVIOUS STUDIES ON PEOPLE DWELL TIME/WAITING TIME ESTIMATION USING WI-FI DATA

Authors	Participation of Smartphone Users	No. of Wi-Fi Monitors	Waiting Time/Dwell Time Resolution	Study Location	Occupancy Duration
Wang et al., 2014 [10]	Requisite	Single	Continuous	Indoor (a coffee shop, an airport)	Short-Intermediate
Manweiler et al., 2013 [11]	Requisite	Single	Discrete	Indoor (a café, a library)	Short-Intermediate
Shu et al., 2016 [12]	Non-requisite	Multiple	Continuous	Indoor (an airport terminal)	Intermediate-Long
Yan et al., 2017 [13]	Non-requisite	Multiple	Discrete	Indoor (a shopping mall)	Short-Long
Le et al. [14]	Non-requisite	Single	Discrete	Indoor (an office)	Short-Long
Wepulanon et al. (in press)	**Non-requisite**	**Single**	**Continuous**	**Outdoor (a bus stop with common bus routes)**	**Short-Intermediate**

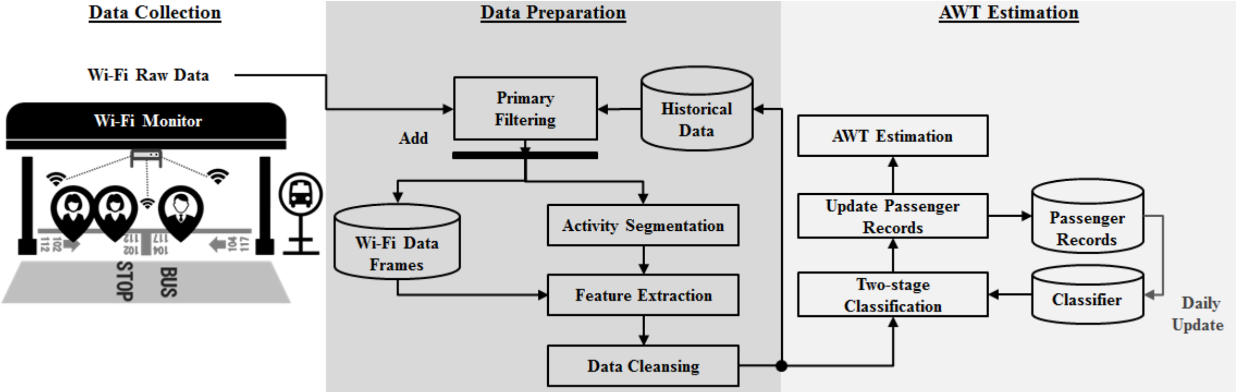

Fig.1 Overviews of the proposed system for AWT estimation at a single bus stop

Next, multiple-location based systems were introduced. Wi-Fi monitors can be installed at a point of interest as well as in its surrounding areas. In this way, people mobility can be tracked over space. Localization techniques were implemented to enhance the accuracy of spatial information on a human trace [12], [13]. Dwell time at the point of interest was estimated by considering the entry and exit time at the location. This method was implemented in large buildings with more open space. In contrast, single-location based systems were employed for smaller study areas. Using a single monitor substantially reduces the monitoring area and results in the lower accuracy of dwell time estimation.

Finally, customer behaviors were classified based on the time spent at the study locations [11], [13], [14]. With this method, the time resolution is reduced into a discrete scale since the major objective is not estimating dwell time. As this paper aims to estimate AWT at a bus stop, none of the three major approaches can be directly adopted. Firstly, the system should be passive. Direct participation in data collection should not be a requirement. Moreover, the multiple-location based system is not suitable since bus stop areas are limited. Lastly, the waiting time needs to be estimated as a continuous variable.

The proposed system involves two additional challenges. First, massive noise can be included in Wi-Fi data collected from bus stop environments. Second, there is a time limitation for detecting waiting passengers. Waiting time could be short due to the regularity of bus services at a bus stop. Consequently, the waiting time might be insufficient for detecting Wi-Fi data from passengers' devices.

III. System Overviews

The proposed system is comprised of three major processes: data collection, data preparation and AWT estimation. The operational overviews are shown in Fig.1.

A. *Data Collection*

1) *Wi-Fi Monitor Set-up*

The first step is to justify the monitor detection range. In general, the range is not a deterministic value since it can be varied based on the transmission power of the monitor and the Wi-Fi devices. A regular method for estimating the detection range is conducting empirical experiments using various smartphone models. The monitor is capable of detecting Wi-Fi data while the devices are in the detection range. By making gradual changes in the devices' distance from the monitor, the detection range can be determined at a distance where most of the devices are no longer detected.

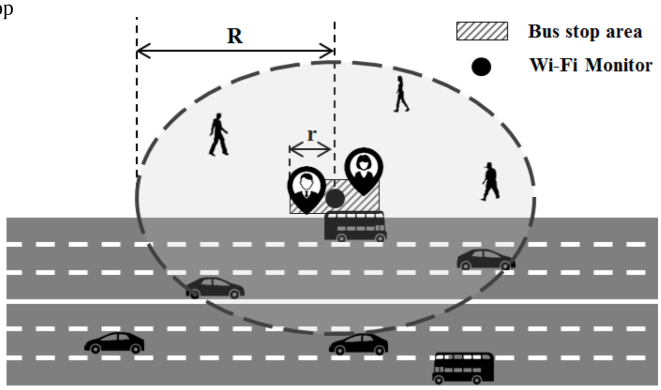

Fig.2. Detection range of a Wi-Fi monitor installed at a bus stop area

Suppose that a Wi-Fi monitor is installed in the middle of the passenger waiting area of a bus stop, and an omnidirectional antenna is applied to the monitor. The detection area becomes a circular shape with a radius R. Then, the distance from the monitor to the end of the passenger waiting area, r, can be measured. Fig.2 shows the detection range of a monitor installed at a bus stop.

2) *Data Description*

A Wi-Fi packet consists of numerous data required for communication. To minimize database storage and lessen privacy concerns, only four data fields are retained as a basic packet for AWT estimation:

a) MAC-ID: An identifier for each Wi-Fi device.
b) Timestamp: The timestamp of packet detection.
c) RSSI: Received Signal Strength Indicator.
d) Frame subtype: Identifying communication purpose.

B. *Data Preparation*

Meaningful information from basic packets is extracted in this process. The information is used to facilitate data filtering which aims to remove noise from the Wi-Fi data.

C. *AWT Estimation*

The main objective is to distinguish the set of Wi-Fi data which belong to waiting passengers from the remaining noise. This is the most challenging task since the proportion of Wi-Fi data from waiting passengers is remarkably low. To overcome this difficulty, classification features are introduced and adopted with a machine learning technique.

IV. Data Preparation

This section describes the processes of preparing the raw data for AWT estimation: primary filtering, activity segmentation, feature extraction, and data cleansing. Most of the noise data will be removed through each process. Also, the raw data will be transformed in order to facilitate AWT estimation.

A. Primary Filtering

The primary sources of noise in Wi-Fi data are stationary devices nearby the bus stop, i.e. access points. The Wi-Fi data captured from access points can be promptly identified using frame subtype information in the basic packets. Particular frame subtypes which only identify the action of the access points (e.g. probe response and beacon) can be recognized. Moreover, a dynamic lookup database recording a list of stationary devices nearby the bus stop can be established. The method for establishing the lookup database has been proposed [15]. With the lookup database, frame subtype information may not be necessary for primary filtering. However, the data field is beneficial for instantaneous filtering which could minimize the computational time of the primary filtering process.

B. Activity Segmentation

Since a passenger may appear in a bus stop area several times in a day, the Wi-Fi data of a MAC-ID can be captured from multiple appearances. It is essential that the sequence of appearance for each basic Wi-Fi packet be identified. The sequence can be identified by considering the continuity of data detection over time. The basic idea is that consecutive packets of the same appearance are supposed to be captured within a short time period and an appearance should be ceased when a MAC-ID is no longer detected within a pre-defined time window. The time window can be determined using the detection period between consecutive packets of the same MAC-ID.

In this paper, a 10-minute interval is identified. To study the detection period, various smartphone models were used for conducting experiments. The state of the smartphones was configured in order to study the detection period at various states. The results are consistent with the previous study [16] which showed that the average detection period of some smartphone models exceeds five minutes for an idle state but never reaches ten minutes.

It is noteworthy that the amount of Wi-Fi data is not reduced during this process. The outcome only assigns an appearance sequence to the existing Wi-Fi packets of individual MAC-IDs. Once the basic packets are segmented into individual appearances, the next step is to extract essential information for each appearance.

C. Feature Extraction

A Wi-Fi record is defined as representing the meaningful information of an appearance. The information can be constructed from the basic Wi-Fi packets captured during the appearance duration. The Wi-Fi record of a MAC-ID, MAC, during an appearance sequence, seq, is described by a vector $\overrightarrow{fv}_{MAC}^{seq}$. Table II summarizes the information attributes of a Wi-Fi record categorized into 3 groups to describe (i) time information, (ii) statistics of detection period, and (iii) statistics of RSSI.

TABLE II
THE INFORMATION OF A WI-FI RECORD

Attribute	Description
entry_time	The timestamp of the first observation
exit_time	The timestamp of the last observation
observation_count	Total observations
observation_period_mean	Average time period between consecutive observations
observation_period_sd	Standard deviation of the period between observations
observation_period_max	Maximum period between consecutive observations
RSSI_mean	Average RSSI
RSSI_sd	Standard deviation of RSSI
RSSI_frequency [range_1]	Frequency of RSSI values in range 1
…	…
RSSI_frequency [range_L]	Frequency of RSSI values in range L

During Wi-Fi communication, a Wi-Fi device could broadcast Wi-Fi packets as a burst. A set of packets from a MAC-ID can be detected several times in just a few seconds. As a result, some features including the statistics of the detection period and RSSI need a low-pass filter to reduce potential bias in the statistical values. Herein, for each MAC-ID, a basic Wi-Fi packet is considered if the MAC-ID has not been detected within the previous 3-second time window. Otherwise, the packet will be omitted. The low-pass filter is based on an empirical finding on the detection period of regular mobile devices [15].

One attribute set that should be highlighted is the frequency of RSSI values in various ranges. Given a set of Wi-Fi observations and supposing that RSSI values are discretized into L ranges, there will be L attributes which summarize the RSSI frequency for each range. Empirical experiments should be conducted to justify the number of ranges as well as the minimum and maximum RSSI values that could be detected by the Wi-Fi monitor at the study location. In general, RSSI observations from a Wi-Fi device can fluctuate even though the device remains in the same state (e.g. the same location). The range of values should be based on the general standard deviation of RSSI observations during an appearance. In this study, RSSI values are discretized into ten ranges from -80 dBm (weak) to -40 dBm (strong). In cases where RSSI values exceed the defined range, the closest range is considered.

D. Data Cleansing

The final step of data preparation is to filter the Wi-Fi records which do not correspond to waiting passengers. Two basic characteristics are considered for filtering: (i) the appearance duration of the device, and (ii) the number of Wi-Fi observations.

1) Appearance duration

The appearance duration of a Wi-Fi record is derived from entry_time and exit_time attributes. Minimum and maximum duration thresholds are determined for filtering the devices which are unlikely to belong to waiting passengers. First, the minimum duration can be determined from the statistical information of bus dwell time at the bus stop whereas the maximum duration can be based on the manual observation of passenger waiting time. As a result, Wi-Fi records from on-board bus passengers and/or passing vehicles can be removed by the minimum duration threshold, while the records from other long-duration activities conducted nearby the bus stop can be filtered by the maximum duration threshold.

2) Unreliable observation

It is assumed that Wi-Fi packets will be detected regularly during an appearance. However, a Wi-Fi record may lack detection continuity during the appearance and the record can be considerably insufficient for passenger waiting time estimation. In this study, the records which were observed only twice are removed from the dataset.

V. AWT Estimation

The majority of noise data are filtered out during data preparation. The remaining noise is assumed to be derived from mobile data sources e.g. passing vehicles, and passers-by. A promising way to distinguish between the data from waiting passengers and the data from other sources is identifying the device's position. If the device is in the bus stop area, it can be assumed to be a waiting passenger. Since the proposed system is non-participatory and based on a single Wi-Fi monitor, RSSI attributes could be a dominant indicator for such data classification.

This section firstly introduces the generalized classification features. Then, the methodologies for training a classifier are proposed based on the sparse nature of Wi-Fi detection in the temporal dimension, followed by a two-stage classification which could reduce computational time.

A. Classification Features

In order to perform data classification, the difference between the characteristics of the Wi-Fi records from waiting passengers and the noise data needs to be recognized. In spatial-temporal dimensions, waiting passengers tend to spend a considerable period at the bus stop area. Consequently, it can be assumed that most Wi-Fi data will be detected at the waiting area. On the other hand, passers-by who walk through the detection range of the Wi-Fi monitor with a constant walking speed should not spend a considerable amount of time at a particular location. Most of the walking time would be spent outside the bus stop.

Suppose that the relationship between the RSSI and the distance from the Wi-Fi monitor is a direct variation. The RSSI can represent the device location at the detection time. To this end, the statistics of RSSI observations are used for establishing the classification features for each Wi-Fi record. First, the frequency distribution of the RSSI values is generalized to simplify the classification. The generalized features consist of four data fields which describe RSSI distribution in terms of the distribution peak and range:

a) dist_min: The range of minimum RSSI.
b) dist_max: The range of maximum RSSI.
c) dist_width: The number of ranges from dist_min to dist_max.
d) dist_mean: The range of average RSSI based on the value of RSSI_mean in the Wi-Fi record.

Fig.3 demonstrates the feature generalization of two Wi-Fi records. It can be seen that most of the RSSI observations from a waiting passenger (MAC A) are in the high range of values, whereas the observations from a passing vehicle (MAC B) lie in the lower ranges. This could imply that most observations from a waiting passenger were detected in the bus stop area near the location of the Wi-Fi monitor, while the observations from a passing vehicle were detected at a location away from the monitor.

Range Index	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**RSSI_mean**
Upper Bound RSSI	**-81**	**-76**	**-71**	**-66**	**-61**	**-56**	**-51**	**-46**	**-41**	**∞**	
Lower Bound RSSI	**-∞**	**-80**	**-75**	**-70**	**-65**	**-60**	**-55**	**-50**	**-45**	**-∞**	
MAC A	0	0	0	0	1	1	2	2	1	0	52.13
MAC B	0	0	1	1	4	0	0	0	0	0	65.33

Generalized value	**dist_min**	**dist_max**	**dist_width**	**dist_mean**
MAC A	5	9	5	7
MAC B	3	5	3	5

Fig.3. Generalized classification feature of the RSSI observations from a waiting passenger (MAC A), and a passing vehicle (MAC B)

B. Classifier

The classifier requires a learning phase for adjusting the relationships between the input features and classification results. In this study, labelling the features is challenging due to the lack of known MAC-IDs from bus passengers. A practical way to identify noise data is based on manual observation of individual waiting passengers at the bus stop.

1) Training Assumptions

With the manual observation of passenger waiting times, the duration of time when there is no waiting passenger at the bus stop can be identified. Then, the Wi-Fi records captured during such duration can be assumed as noise data. In addition, repeated MAC-IDs can be discovered in multi-day observations. The MAC-IDs could be assumed as waiting passengers in the cases where the MAC-IDs are detected at a similar period for several days.

2) Data Matching Problem

Also, the Wi-Fi records from waiting passengers can be assumed by data matching. Let M , and N be the number of Wi-Fi records and total waiting passengers respectively. The objective is to match an observed waiting time N to a potential Wi-Fi record M . This becomes a matching problem and an assignment function μ can be defined:

$$\mu : \begin{cases} \{1, 2, \ldots, N\} \rightarrow \{1, 2, \ldots, M\} \\ a \mapsto b, a = 1, 2, \ldots, N, b = 1, 2, \ldots, M \end{cases} \quad (1)$$

where $\mu(a) = b$ indicates that the Wi-Fi record b is assumed to be detected from the observed passenger a .

The challenge is that a Wi-Fi record cannot be matched to an observed waiting time directly due to the sparseness of Wi-Fi data in the temporal dimension. To solve the matching problem, time window constraints are introduced for identifying potential Wi-Fi records.

3) Identifying Potential Candidates

The time window constraints can be formulated based on the observed passenger arrival time at the bus stop, and the departure time of the boarded bus. An assumption is that the Wi-Fi record from an observed passenger should be detected during the wait or a similar time period. Fig.4 illustrates the time window constraints of an observed passenger in space and time dimensions, while Table III summarizes the case of constraints with the factors for determining time windows.

For the case of A1, the first Wi-Fi observation wf_A^j can be detected before the actual passenger arrival at the bus stop obs_A^i since the monitor detection range covers a distance further than the bus stop area. In contrast, the observation can also be detected after the passenger arrival due to the sparseness of Wi-Fi detection (case A2). Hence, the first time window is specified based on the observed passenger arrival time with a lower bound $obs_A^i - e_A^{A1}$ and an upper bound $obs_A^i + e_A^{A2}$. In the same way, the last Wi-Fi observation can be detected either before the bus departure (case B1) or after the departure (case B2). Another time window based on bus departure time can be described as well.

Given an observed passenger i with arrival time obs_A^i and departure time obs_D^i , a Wi-Fi record j will be

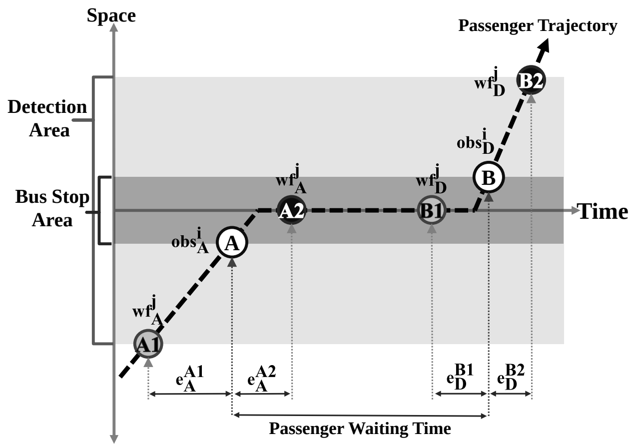

Fig.4. Time window constraints for identifying potential candidates

TABLE III
THE CASE OF TIME WINDOW CONSTRAINTS

Case	Target Time Window	MAC Detection Time	Factors for Determining Time Windows
A1	Passenger arrival	Before passenger arrival	Walking speed
A2	Passenger arrival	After passenger arrival	Detection frequency
B1	Bus departure	Before bus departure	Detection frequency
B2	Bus departure	After bus departure	Bus speed

considered as a potential candidate of the observed passenger when the entry_time and exit_time are within the following constraints:

$$obs^i_A - \frac{R-r}{v_{ped}} \le wf^j_A \le obs^i_A + f^j \tag{2}$$

$$obs^i_D - f^j \le wf^j_D \le obs^i_D + \frac{R-r}{v_{bus}} \tag{3}$$

The notations are described as follows:

obs^i_A : The observed arrival time of a passenger i at the bus stop.

obs^i_D : The observed departure time of the boarded bus.

wf^j_A : The entry_time attribute of a Wi-Fi record j.

wf^j_D : The exit_time attribute of a Wi-Fi record j.

R : The radius of the monitor's detection area (m).

r : The distance from the monitor to the end of the passenger waiting area (m).

v_{ped} : Average walking speed (m/s).

v_{bus} : Average bus speed from bus departure to the end of detection area (m/s).

f^j : The estimated maximum detection period between Wi-Fi observations (s) which is determined by

$$f^j = \begin{cases} wf^j_{\max} & ;CV > 0.5 \\ wf^j_{mean} + 2wf^j_{sd} & ;otherwise \end{cases} \tag{4}$$

where wf^j_{mean}, wf^j_{sd}, $wf^j_{\max}$ denote the attributes observation_period_mean, sd, and max of a Wi-Fi record j. The value of f^j is dependent on the coefficient of variation CV which is the ratio of the standard deviation to the mean of the period between observations. Identifying f^j is challenging due to the variability in the detection period. On the one hand, the value should be sufficient to include a potential Wi-Fi record in the time windows. On the other hand, the time windows should not be too large to prevent the inclusion of excessive noise data in the candidate set.

In general, the value of f^j can be determined using the maximum period between Wi-Fi observations. In some cases, the maximum period may not be well-represented. For example, a passenger may use a mobile device during the wait. Since the device is active during that time, the detection period could be very short resulting in the insufficient maximum period. In the case A2 in Fig.4, the first Wi-Fi observation may be derived once the device is active after the passenger arrival. Also, the device could be idle a significant of time before the bus departure for case B1. In such cases, the maximum period derived from the active device could be too short for including the Wi-Fi records in the time windows in (2) and (3) since the value of f^j is incomparable to e^{A2}_A, and e^{B1}_D. A way to solve this issue is extending the maximum period.

As only the mean and standard deviation are available, the distribution of detection periods could be assumed for estimating f^j. Firstly, it is assumed that the detection periods are not significantly varied when the mean is twice the standard deviation ($CV \le 0.5$) and the detection periods are normally distributed. Consequently, f^j can be determined by adding twice the standard deviation to the mean and assuming that the maximum period is in the normal distribution.

4) A Modified Bipartite Matching Method

A bipartite graph $G = (U,V,E)$ is demonstrated in Fig.5. The vertices are divided into two disjoint sets for observed data U, and Wi-Fi records V. Each edge $e = (i, j), i = 1,...N; j = 1,..., M$ corresponds to a potential match between an observed passenger and a Wi-Fi record. The weight associated with each edge is defined:

$$\varepsilon(i, j) = \frac{|wt_i - at_j|}{wt_i} \tag{5}$$

where a function $\varepsilon(i, j)$ evaluates the percentage error between the observed waiting time of a passenger i, wt_i and presence duration of a Wi-Fi record j, at_j. The presence duration is derived from the entry_time and exit_time attributes.

Two modifications can be applied to the bipartite graph. First, an edge $e = (i, j)$ is removed from the graph if a Wi-Fi record is not a potential candidate of a waiting passenger. Second, a maximum weight threshold γ_{wg} can be applied as a filter to reduce the probability of misclassification. The threshold is used to ensure that the difference between actual waiting time and the presence time duration of a Wi-Fi record is acceptable. In this study, the threshold is 0.4 based on empirical analysis of repeated MAC-IDs. All the edges with undesirable weight are removed from the graph, as well as the nodes without a linking edge. This implies that some

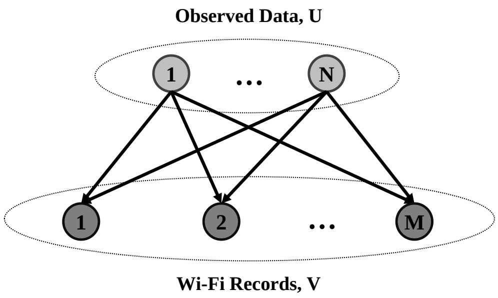

Fig.5. Bipartite graph representation

observed passenger may not be assigned a Wi-Fi record since Wi-Fi data may not be captured from every passenger.

Finally, in order to find an optimum solution for the graph where each observed passenger node matches with the most potential Wi-Fi record, a minimum-weight bipartite matching can be formulated as follows:

$$\min_{\psi} \sum_{i=1}^{N} \sum_{j=1}^{M} \varepsilon(i,j)\delta(\mu(i)=j) \tag{6}$$

s.t. $$\delta(\mu(i)=j) \in \{0,1\}$$

$$\forall i \in \{1,2,...,N\},\ j \in \{1,2,...,M\} \tag{7}$$

$$\sum_{j=1}^{M} \delta(\mu(i)=j) = 1, \forall i \in \{1,2,...,N\} \tag{8}$$

$$\sum_{i=1}^{N} \delta(\mu(i)=j) \le 1, \forall j \in \{1,2,...,M\} \tag{9}$$

The matching solution is adopted from the vehicle matching problem proposed by Wang *et al.* [16] as the nature of the matching problem is equivalent when the modified bipartite graph is applied. The objective function (6) aims to find a solution with the minimum of overall weight. Herein the minimum weight implies the minimum of overall percentage errors between observed waiting times and device presence times. The constraint (7) ensures that the $\delta(\bullet)$ provides binary integers. A retained passenger node is required to match with a Wi-Fi record by the constraint (8). Finally, duplicate matching is not allowed for a Wi-Fi record. Only one device is assumed to be detected from a passenger. A Wi-Fi record is allowed to have at most one match by the constraint (9).

The classification features of a Wi-Fi record p which is matched to an observed passenger is recognized as the classification profile X_{train}^{p} of the waiting passenger class $(y_{train}^{p} = passenger)$, while the features of a Wi-Fi record q which is never considered as a candidate of any observed passenger are recognized as the profile X_{train}^{q} of the noise data class $(y_{train}^{q} = noise)$. Moreover, another parameter which can be learnt during the training process is the minimum of RSSI_mean for the waiting passenger class, γ_{RM}^{WP}. This feature could facilitate the two-stage classification.

C. Two-stage Classification

Classification processes need to handle a large number of Wi-Fi records due to the large proportion of noise in Wi-Fi data. Here, a two-stage classification is therefore proposed to reduce computational time. First, the trained parameter γ_{RM}^{WP} initially classifies some records in which the RSSI_mean is noticeably low. It is assumed that the classification profile of such Wi-Fi records tend to represent noise data rather than waiting passengers.

Next, the K-Nearest Neighbor (KNN) classifier is adopted to classify a Wi-Fi record based on the classification feature by selecting k minimum distance values $d(B,C)$ for a majority vote. The distance function $d(\bullet)$ measures the similarity between two classification features: the classification features of a testing dataset B and a classification profile C. Euclidean distance can be applied as the function. The value of k is determined using an elbow method. A validation accuracy curve can be plotted against various values of k to find an elbow point where the smallest k gives the nearly highest accuracy.

The KNN method is selected for classification due to the challenges in cross-validation processes. There is no evidence to ensure that the classification results are correct without the direct request of MAC-IDs from waiting passengers. Multi-day observations are necessary for assuming the repeated MAC-IDs to be waiting passengers. Since the number of observations is limited, AWT estimation is problematic in the early state of the system. However, training KNN classifier can be performed with limited datasets. Also, the training process requires short computational time which allows the classifier to be trained regularly. The training can be performed daily at the end of the day when more repeated MAC-IDs are available.

D. Average Passenger Waiting Time

AWT at a bus stop can be estimated from the potential Wi-Fi records of waiting passengers. The estimation can be performed to derive AWT during a time interval τ which is denoted by:

$$AWT^{\tau} = \frac{1}{S}\sum_{j=1}^{S} wf_D^{\,j} - wf_A^{\,j};\ wf_D^{\,j} \in \tau \tag{10}$$

where $wf_A^{\,j}$ and $wf_D^{\,j}$ are the entry_time and exit_time of a Wi-Fi record j, and S is the total number of records during the time interval.

VI. Empirical Studies

Wi-Fi data were collected from a bus stop on Chatham North Road, in front of The Hong Kong Polytechnic University Phase 8. Two-hour observations were conducted for four weekdays during the evening peak time period. The bus stop is served by 9 bus lines, including common bus services that use the same bus stops. Both high-frequency and low-frequency services were operated at the bus stop. To elaborate the bus service frequency, the headway at the bus stop can be calculated from the difference between a bus departure time and the departure time of the previous bus from the same bus line. Twenty-eight percent of bus headways were less than 5 minutes, 41% between 5 and 10 minutes, and 31% more than 10 minutes.

During the observations, passenger loading on the buses was usually high due to travel demands during the evening peak periods. The road traffic was mainly congested. Also, considerable pedestrian flows were observed nearby the bus stop area. The average number of waiting passengers at the bus stop during a two-hour observation was 58. Manual observations of individual passengers were recorded for evaluation purposes including passenger arrival times at the bus stop area, the boarded bus numbers, and the bus departure times. Finally, the four-day dataset was separated for training the classifier and testing the system. Using 3-day datasets for training and another dataset for testing, multi-day evaluation was performed.

A. Data Filtering

Fig.6 is plotted from a two-hour observation. The massive noise in Wi-Fi data were significantly reduced after performing four core processes. First, the data from stationary devices, about 25% of the raw data, were removed through primary filtering. Next, feature extraction compressed the output from the primary filtering process into 13%. Also, ten percent of the Wi-Fi records were filtered out through data cleansing as they were considered unreliable for AWT estimation. Finally, the classifier identified 35 Wi-Fi records which accounted for 60% of the observed passengers.

The final result shows that the system cannot identify all waiting passengers. A basic reason for this can be that the devices are undetectable; either the Wi-Fi function was disabled or the passengers did not carry any Wi-Fi devices. Furthermore, some valid data could be lost due to the filtering mechanisms. For example, the Wi-Fi data from the lucky passengers who were able to board the bus without having to wait tended to be filtered out during the data cleansing process since the duration of their presence at the bus stop was too short. In contrast, some invalid observations might be classified as waiting passengers. According to manual observations, it was found that some people can behave like a waiting passenger e.g. people who waited for a bus but finally left the bus stop without boarding a bus. The Wi-Fi data from those people are then invalid. To validate the filtering results, a survey of passengers' MAC-IDs is necessary. Since the survey has not been conducted at this stage, AWT estimation accuracy is considered based on the main objective of this study.

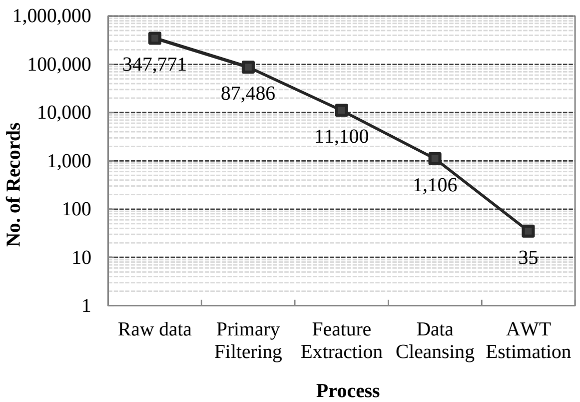

Fig.6. Data filtering performance

B. Estimation Results

The system performance is evaluated in terms of two measures: the mean absolute error (MAE) and the mean absolute percentage error (MAPE).

$$MAE = \frac{\sum \left| AWT_{obs}^{\tau} - AWT_{est}^{\tau} \right|}{Z} \tag{11}$$

$$MAPE = \frac{1}{Z} \frac{\sum \left| AWT_{obs}^{\tau} - AWT_{est}^{\tau} \right|}{AWT_{obs}^{\tau}} \tag{12}$$

where Z is the number of data points, AWT_{obs}^{τ} and AWT_{est}^{τ} are the observed and estimated AWT at the bus stop during a time interval τ.

To analyze the system performance, three datasets are used for AWT estimation: (i) bus headway data, (ii) probe request (PR) packets in the raw data, and (iii) all Wi-Fi data. The first dataset is used for a half-headway method as the baseline performance. For the second dataset, only PR packets in the raw data are used for evaluation. Since the PRs are broadcasted to all nearby devices, not to a specific one (e.g. a Wi-Fi access point), the effects of available Wi-Fi services nearby the bus stop can be investigated. The last dataset is raw Wi-Fi data which include all captured packets.

To analyze data resolution effects on AWT accuracy, estimations were performed to provide AWT information for different estimation time intervals, i.e. every 15, 30, 60, and 120 minutes. Fig.7 shows the evaluation results in terms of (a) MAPE and (b) MAE based on different estimation intervals. The overall performance indicates that the proposed system improves AWT estimation from the baseline accuracy. The system can perform AWT estimation with 5-25% of MAPE and 18-86 seconds of MAE.

There are two interesting points that should be highlighted here. First, the availability of Wi-Fi services nearby the bus stop improves the estimation accuracy. The proposed method performs better using all Wi-Fi data. The accuracy is decreased by up to 10% when only PR packets are considered. It should be noted that, in the PR dataset, not only are the Wi-Fi data from stationary devices removed but all the data involving nearby access points (i.e. data frames) are also discarded. Therefore, the detection period of passengers' devices in the PR dataset can be increased resulting in lower accuracy. Second, the accuracy is improved when the estimation interval is extended. The effects of the available Wi-Fi services and the estimation

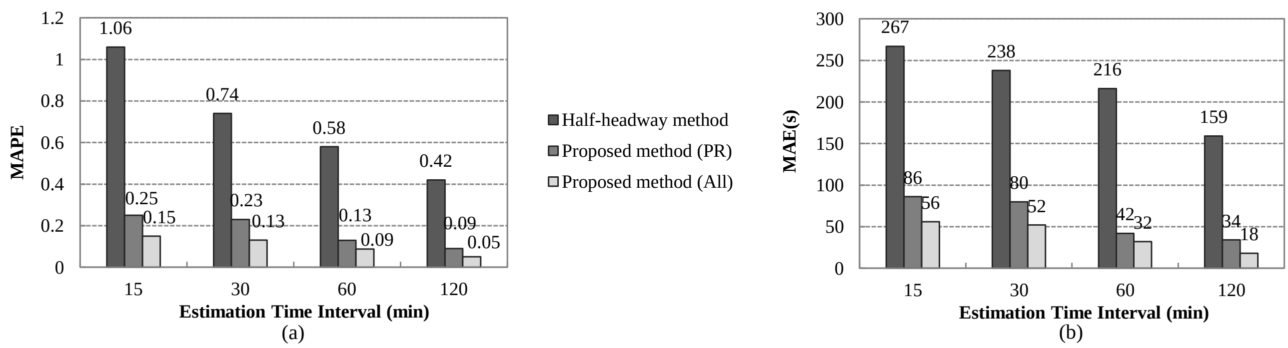

Fig.7. AWT estimation accuracy in terms of (a) MAPE, (b) MAE

interval are further discussed in the following sub-sections.

C. Estimation Errors in Passenger Waiting Time

By assuming passenger waiting time from the presence time of a Wi-Fi record, a number of factors can affect the accuracy of individual waiting times. According to the time window constraints for identifying potential candidates (2) – (3), a Wi-Fi record can be detected at any point in the time windows. The range of estimation error for a Wi-Fi record j can be assumed based on the time windows.

$$RoE(j) = \left[-(2f^{j}), \frac{R-r}{v_{ped}} + \frac{R-r}{v_{bus}} \right] \quad (13)$$

The estimated maximum period between Wi-Fi observations f^{j} provides negative errors which result in underestimation of waiting time, whereas a long monitor detection range causes overestimation from positive errors. Since the variation of other parameters may not be critical, a crucial parameter here can be f^{j}.

The Wi-Fi records which are classified as waiting passengers were further analyzed to understand the typical detection period of a MAC-ID. Each Wi-Fi record consists of the average and standard deviation of the period between observations. Here, initial statistics of the typical detection period can be estimated. Firstly, the typical period could be assumed as the mean of the average values from passengers' Wi-Fi records. Also, the variation of the typical detection period is assumed from the mean of standard deviations.

For the dataset with only PR packets, the typical detection period is 59 seconds with 43 seconds of standard deviation. In other words, an observation can be detected from a MAC-ID every 59 seconds on average. The typical detection period is shorter for the dataset with all Wi-Fi data. An observation can be detected every 35 seconds with 32 seconds of standard deviation. This can explain the lower estimation accuracy when only PR data are considered. Filtering other packets results in a longer period between Wi-Fi observations and the range of error in (13) is extended. Moreover, the longer period could result in the higher probability of missed detection since the system requires more time to detect a waiting passenger.

Evaluation of the time for detecting a waiting passenger is valuable since it can describe the missed detection of short waiting times. Without the availability of passengers' MAC-IDs, one way to estimate the passenger detection time is to make assumptions based on Wi-Fi data characteristics. In the proposed system, at least three observations are needed so as not to filter out a valid Wi-Fi record during the data cleansing process. As a result, the average time duration for detecting a waiting passenger can be estimated using double the typical period between observations. To be more precise, the walking time to the bus stop area should be deducted since the monitor detection range is larger than the bus stop area. The relationship can be denoted by:

$$TD = 2TP - \left(\frac{R-r}{v_{ped}} \right) \quad (14)$$

where TD is the average time for detecting a waiting passenger, and TP is the typical period between observations of a Wi-Fi record.

D. Estimation Errors in Average Passenger Waiting Time

Since AWT is calculated from individual Wi-Fi records during an estimation interval, the accuracy is reliant on the distribution of estimated waiting times. Table IV presents a comparison of four passenger waiting time distributions during a two-hour observation. The distribution from manual observation shows that the majority of the waiting time is in the 1-10 minute range. According to the manual observation of the bus lines boarded, the passengers who waited for more than 10 minutes usually boarded bus on uncommon routes with long headways.

Performing the proposed method using all Wi-Fi data provides waiting time distribution that is comparable to the observed one. The probability is slightly higher than the observation for the waiting times of up to 1 minute. The distribution could be shifted down due to the effects of the Wi-Fi detection period which results in the underestimation of waiting times. The distribution is even more different when the PR dataset is used. The probability of waiting times during 1-5 minutes is decreased while the probability of waiting times during 5-10 minutes is increased. It can be assumed that the lower accuracy is also affected by higher chances of missed detection, apart from the detection period. Since missed detection is generally encountered for short waiting times, the higher probability of the waiting times with a duration of 5-10 minutes could be reasonable.

For the half-headway method, the probability of waiting times with a duration of 1-5 minutes is lessened compared to the observed one whereas the probabilities are higher for the waiting times of more than 5 minutes. Based on the manual observation of passenger behavior, the accuracy of the half-headway method could be affected by the common bus routes. Since several bus lines are serving common stops, passengers can have multiple choices and they may decide to catch an alternative service depending on which bus arrives first. Also, the availability of bus arrival time information could be a reason. Several studies have found that the assumption regarding uniform passenger arrival can be violated, since passengers may time their arrival even for the frequent bus services [17], [18], [19].

The factors affecting the accuracy of AWT can be determined based on the waiting time distributions. Firstly, overestimation can be caused by missed detection. As discussed in the previous sub-section, the waiting time could be too short, especially for lucky passengers who are able to catch their bus quickly.

Next, the accuracy can be decreased when the distribution of estimated waiting times is incomparable to the distribution of actual waiting times. Both overestimation and underestimation can be encountered. As can be seen from Table IV, using all Wi-Fi data provides more accurate waiting time distribution than is possible from the PR dataset. The result corresponds to the AWT estimation

TABLE IV
PROBABILITY DISTRIBUTION OF PASSENGER WAITING TIMES

Dataset	Up to 1 Minute	1-5 Minutes	5-10 Minutes	More Than 10 Minutes
Manual observation	0.11	0.49	0.25	0.15
Proposed method (all)	0.14	0.46	0.26	0.14
Proposed method (PR)	0.10	0.45	0.30	0.15
Half-headway method	0.07	0.22	0.41	0.30

accuracy in Fig.7 which shows that using more Wi-Fi data results in better accuracy. In addition, as discussed in the data filtering results misclassification can also occur. Invalid waiting times might be included in the distribution.

Finally, to estimate AWT during a time interval in (10), a Wi-Fi record might be included in an incorrect estimation interval due to the range of errors described in (13). In such a case, more errors are introduced to the distribution of estimated waiting times during the interval, and the lessened accuracy. This implies that a shorter estimation interval can cause higher errors since a Wi-Fi record has a greater chance of being assigned to an incorrect interval. The results in Fig.7 show that extending the estimation interval can reduce the errors in AWT estimation.

VII. Conclusion and Future Works

An alternative method of AWT estimation is proposed in this paper. The system exploits the ubiquitous Wi-Fi enabled devices carried by bus passengers for use in AWT estimation. Methodologies are developed in order to handle the major challenges in the passive Wi-Fi data captured from bus stop environments. The system is evaluated using Wi-Fi datasets collected from a bus stop in Hong Kong. The results show that passive Wi-Fi data can be considered as an alternative data source for AWT estimation. The factors affecting estimation accuracy are discussed. It can be concluded that the greater detection period of Wi-Fi devices can significantly improve the estimation accuracy.

Future research will be focused on improving passenger classification algorithms when passengers' MAC data are available. Furthermore, AWT estimation for individual bus routes at a bus stop can be developed so as to enhance the contribution to transportation development. The proposed system can provide a fundamental framework for estimation when Wi-Fi data from strategic bus stops are available.

References

[1] V. Guihaire, and J. K. Hao, "Transit network design and scheduling: a global review," *Transp. Res. Part A*, vol. 42, no. 10, pp. 1251-1273, Dec. 2008.
[2] Y. Liu, J. Bunker, and L. Ferreira, "Transit users' route choice modelling in transit assignment: a review," *Transport Rev.*, vol. 30, no. 6, pp. 753-769, Nov. 2010.
[3] P. I. Welding, "The instability of a close-interval service," *J. Oper. Res. Soc.*, vol. 8, no. 3, pp. 133–142, 1957.
[4] L. A. Bowman, and M. A. Turnquist, "Service frequency, schedule reliability and passenger wait times at transit stops," *Transp. Res. Part A*, vol. 15, no. 6, pp. 465-471, Nov. 1981.
[5] M. R. Amin-Naseri, V. Baradaran, "Accurate estimation of average waiting time in public transportation systems," *Transport. Sci.*, vol. 49, no. 2, pp. 213-222, Aug. 2014.
[6] R. Fernández, "Modelling public transport stops by microscopic simulation," *Transp. Res. Part C*, vol. 18, no. 6, pp. 856-868, 2010.
[7] V. Kostakos, E. O'Neill, A. Penn, G. Roussos, and D. Papadongonas, "Brief encounters: Sensing, modeling and visualizing urban mobility and copresence networks," *ACM Trans. Comput.-Hum. Interact.*, vol. 17, pp. 1–38, Mar. 2010.
[8] N. Abedi, A. Bhaskar, and E. Chung, "Tracking spatio-temporal movement of human in terms of space utilization using Media-Access-Control address data," *Appl. Geogr.*, vol. 51, pp. 72-81, 2014.
[9] A. Danalet, B. Farooq, M. Bierlaire, "A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures," *Transp. Res. Part C*, vol. 44, pp. 146-170, July. 2014.
[10] Y. Wang, J. Yang, Y. Chen, H. Liu, M. Gruteser and R. P. Martin, "Tracking human queues using single-point signal monitoring," in *Proc. 12th Int. Conf. Mob. Syst. Appl. Serv.*, USA, 2014, pp. 42-54.
[11] J. Manweiler, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "Predicting length of stay at wifi hotspots," in *Proc. IEEE INFOCOM*, Italy, 2013, pp. 3102-3110.
[12] H. Shu, C. Song, T. Pei, L. Xu, Y. Ou, L. Zhang, and T. Li, "Queuing time prediction using wifi positioning data in an indoor scenario," *Sensors*, vol. 16, no. 11, Art. no. 1958, Nov. 2016.
[13] C. Yan, P. Wang, H. Pang, L. Sun, and S. Yang, "CELoF: WiFi dwell time estimation in free environment," in *Proc. Int. Conf. Multimedia Modell.*, Iceland, 2017, pp. 503-514.
[14] T. V. Le, B. Song, and L. Wynter, "Real-time prediction of length of stay using passive Wi-Fi sensing," in *Proc. IEEE ICC*, France, 2017, pp. 1-6.
[15] A. B. M. Musa, and J. Eriksson, "Tracking unmodified smartphones using wi-fi monitors," in *Proc. 10th ACM Conf. Embedded Netw. Sens. Syst.*, Canada, 2012, pp. 281-294.
[16] J. Wang, N. Indra-Payoong, A. Sumalee, and S. Panwai, "Vehicle reidentification with self-adaptive time windows for real-time travel time estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 540-552, Apr. 2014.
[17] M. Luethi, U. A. Weidmann, and A. Nash. "Passenger Arrival Rates at Public Transport Stations," in *Proc. 86th Transp. Res. Board Annu. Meet.*, USA, 2007.
[18] M. Frumin, and J. Zhao. "Analyzing Passenger Incidence Behavior in Heterogeneous Transit Services Using Smartcard Data and Schedule-Based Assignment," *Transp. Res. Rec.-J. Transp. Res. Board*, vol. 2274, pp. 52–60, 2012.
[19] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: a smart card data analysis," *Transp. Res. Part C*, vol. 90, pp. 292–306, May. 2018.

Piyanit Wepulanon received the B.Eng. degree in computer engineering from King Mongkut's University of Technology Thonburi, Bangkok, Thailand, in 2010. She is currently working towards the Ph.D. degree in transportation engineering in the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

Her research interests include intelligent transportation systems, mobile crowd sensing, and public transport.

Agachai Sumalee received the B.Eng. degree in civil engineering from KMITL-Ladkrabang in 1999 and the M.Sc.(Eng.) and Ph.D. degrees in transportation planning and engineering from the University of Leeds, Leeds, U.K., in 2000 and 2004, respectively. He is a Professor with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

His research areas include transit planning, intelligent transport systems, network modeling and optimization, and transport economics. Prof. Sumalee is currently the Editor-in-Chief of *Transportmetica B: Transport Dynamics*, the Editor of the Journal of *Transport Policy*, an Associate Editor of *Networks and Spatial Economics*, and an Editorial Board Member of several prestigious journals.

William H. K. Lam received the B.S. and M.S. degrees from the University of Calgary, Calgary, AB, Canada, in 1978 and 1980, respectively, and the Ph.D. degree in transportation engineering from the University of Newcastle upon Tyne, Newcastle, U.K., in 1992. He is currently a Chair Professor in civil and transportation engineering with the Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong. He is also with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China.

His current research interests include transport network modeling, traffic simulation, public transport, and pedestrian studies. Prof. Lam is a member of the International Advisory Committee of the *International Symposium on Transportation and Traffic Theory (ISTTT)* and a member of the International Scientific Committee of the *International Symposium on Transportation Network Reliability (INSTR)*. He is also the Founding Editor-in-Chief of *Transportmetrica A* and the Coeditor-in-Chief of the *Journal of Advanced Transportation*.