Application of Bluetooth Technology for Mode-Specific Travel Time Estimation on Arterial Roads: Potentials and Challenges

Bahar Namaki Araghi[1]*, Lars Tørholm Christensen [2], Rajesh Krishnan[3], Harry Lahrmann[4]
1. PhD Student, Traffic Research Group, Department of Planning and Development, Aalborg University, Fibigerstræde 11, 9220 Aalborg Øst, Denmark,
E-mail: araghi@plan.aau.dk , Tel: +45-9940-8447,
2. Blip Systems, Aalborg, Denmark
3. Centre for Transport Studies, Imperial College London, United Kingdom
4. Associate Prof., Traffic Research Group, Department of Planning and Development, Aalborg University, Fibigerstræde 11, 9220 Aalborg Øst, Denmark,
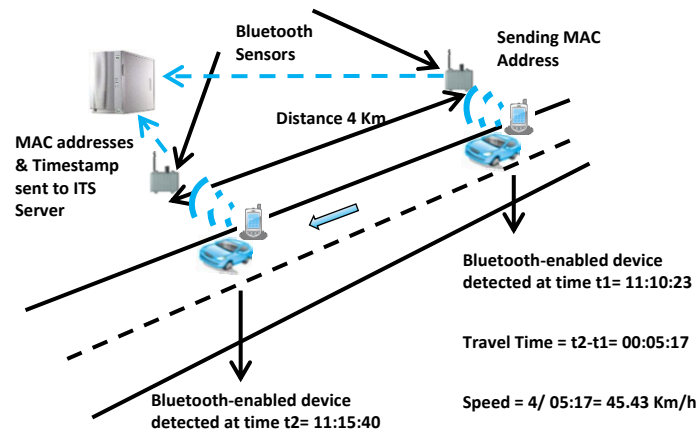
Abstract

The tendency to use Bluetooth Technology (BT) for travel time estimation is increasing due to growing number of Bluetooth-enabled devices among road users, anonymity of BT detections, flexibility of deployment and maintenance of Bluetooth sensors etc. Although Bluetooth has been demonstrated as a promising technology, there remain problems which affect the accuracy of the estimation such as difficulty of distinguishing between multiple travel modes (e.g. motor vehicles, bicycles and pedestrians). This study aims to examine the feasibility of estimating mode-specific travel time using BT data under uncongested traffic conditions. In this context, three clustering methods Hierarchical, K-means and Two-Step are used as the core techniques for classification. The results show that the methods can successfully distinguish between motor vehicles and bicycles from BT detection events, resulting in accurate travel time estimation for motor vehicles.

1. Introduction

Application of Bluetooth Technology (BT) as a new method of travel time estimation has been evolving in the recent past. The Bluetooth protocol uses an electronic identifier, or a tag, in each device called the Machine Access Control (MAC) address. The MAC address serves as a nickname for electronic devices to keep track of who is who during data communications. Vehicles carrying discoverable Bluetooth-enabled devices (i.e. navigators, mobile phones, headsets etc.) can be detected by Bluetooth sensors installed at multiple locations along the road network. The MAC address and its detection time are recorded by the sensors, which then can be used for travel time estimation (Haghani et al., 2010).

Figure 1- Application of Bluetooth technology for Travel Time estimation

Increasing number of Bluetooth-enabled devices among road users, anonymity of MAC addresses, flexibility of deployment and maintenance of Bluetooth sensors, along with the acceptable accuracy of travel time estimation using BT (Haghani et al., 2010; Puckett and Vickich, 2010; Barcelo et al., 2010) have resulted in an emerging interest for further evaluation of this technology as a cost-effective method for travel time estimation.

Due to the anonymity of MAC addresses (no current data sources link MAC addresses with personal information), external systems (i.e. Bluetooth sensors) tracking these electronic devices (i.e. Bluetooth-enabled devices) is more acceptable to the public compared to the Global Positioning System (GPS) or Automated Number Plate Recognition (ANPR). Furthermore, increasing numbers of Bluetooth-enabled devices not only do not impose any additional cost to the ITS infrastructure but also provide better opportunities for estimating travel time based on MAC address tracking. Larger numbers of Bluetooth-enabled devices increase the sampling size of the BT dataset, improving the accuracy and the reliability of BT based travel time estimations. Moreover, BT is less sensitive to the weather conditions (i.e. fog, darkness etc.) compared to technologies such as ANPR.

Although BT has been demonstrated as a promising technology, there remains problems such as over-counting of a single vehicle carrying more than one Bluetooth-enabled device (e.g. buses), low sampling rate and the difficulty of distinguishing between multiple travel modes (e.g. motor vehicles, bicycles and pedestrians).

Research to date has primarily focused on the application of BT for travel time estimation on motorways, where there is typically no need to distinguish between multiple travel modes. However, arterial roads and other city streets experience a mix of transport modes (i.e. motor vehicles, bicycles and pedestrians). Therefore, applying BT to estimate mode-specific travel time under heterogeneous traffic conditions remains a challenge. Heterogeneous traffic can affect both accuracy and reliability of travel time estimated using BT. This problem is particularly relevant in cities with a high volume of bicycle traffic.

This paper aims to examine the feasibility of estimating mode-specific travel time using MAC address tracking data under uncongested traffic conditions and highlight the challenges facing the application of BT for mode-specific travel time estimation.

2. Background

The implementation of Bluetooth MAC address for traffic monitoring began with the study by Ahmed et al. in 2008. In this study the Bluetooth MAC address matching algorithms for travel time data collection and measurements were developed and tested in a joint collaboration between Indiana Department of Transportation and Purdue University in research conducted by Wasson et al. (2008). The results of this research indicated that travel time trends were easy to discern by visual inspection and demonstrate the feasibility of using MAC address matching for travel time estimation. However, data gathered from arterial highways showed a significantly larger variance compared to data from motorways. This is due to the

impact of traffic signals and the noise in data caused by motorists diverting from the arterial road into side roads.

Tarnoff et al. (2009) presented the Bluetooth-based method as one of the most cost-effective approaches for travel time data collection. A series of studies was conducted by the Department of Civil and Environmental Engineering at the University of Maryland under supervision of Prof. Haghani to evaluate the application of BT as well as the accuracy of the travel time obtained. Haghani et al. (2010) introduced Bluetooth sensors as a new and effective means of freeway ground truth travel time data collection. Floating Car Data (FCD) was used as an estimate of the "ground-truth" to evaluate the quality of the highway travel time information generated by traffic surveillance systems. Results of the drive test data were employed to measure the quality of travel time data collected by the Bluetooth sensors. They showed that the ground-truth provided by the new sensors was not significantly different from the actual travel times.

Puckett and Vickich (2009) investigated the feasibility of utilising BT on arterial streets to collect travel time data. The demonstration of this technique as a viable method for estimating travel times in a corridor was proven on the basis of a single study conducted in Houston.

Malinovskiy et al. (2010) compared Bluetooth MAC address-based travel-time sensors with Automated License Plate Recognition (ALPR) sensors. The test results indicated that although the Bluetooth sensors tended to be biased towards slower vehicles, the measured travel time were representative of the ground truth travel time data obtained by the ALPRs.

The Pennsylvania Department of Transportation (PennDOT) (2010) conducted an independent study to evaluate the functionality of BT for travel time measurement. Travel time results, match rate, and cost of implementing BT were compared with the EZPass tag readers. It was determined that the travel times measured by BT were comparable to those obtained by tag readers.

Lahrmann et al. (2010) investigated the application of Bluetooth sensors for measuring travel time on motorways. A comparative study was carried out on the data obtained by BT and FCD. Results were compared regarding the travel time, coverage, and penetration rates. It proved that BT can be considered as an alternative technology for travel time measurements.

Saunier and Morency (2011) aimed to develop travel time reliability indicators through comparing the outputs of different data collection methods (i.e. FCD, GPS, ANPR and BT) used for monitoring the highway. Results of the comparative study for GPS and BT indicated that the both technologies agree during non-congested conditions but a similar assessment was not carried out for the peak period.

Based on what has been discussed, BT has shown acceptable accuracy for estimating the travel time of the motor vehicles under homogeneous traffic conditions. However in order to expand the application of BT for arterial roads with mixed traffic, feasibility of the application of BT for mode specific travel time needs further investigation.

3. Methodology

In order to obtain mode-specific travel time based on BT, it is necessary to classify and allocate the travel time observations depending on the types of the road users and their characteristics. The idea is to classify BT detection events into two groups: one for motor vehicles and the other for bicycles based on some underlying logic for uncongested traffic conditions. The proposed methodology is based on the assumption that in uncongested urban arterials the travel time of motor vehicles is statistically significantly shorter than the bicycles (in other words motor vehicles are usually faster than the bicycles).

Due to differences in functionality of Bluetooth-enabled devices (i.e. navigator, phones, computers etc.), it is expected that some groups of Bluetooth-enabled devices to be used only by specific modes. For example, navigators for route finding are used only by motor vehicles and not by pedestrians or bicycles. The information of type of Bluetooth-enabled device can be extracted from the MAC address. The 11-bits of the 48-bit MAC address contain the information of the type of device and the vendor company; refer to as Class of Device (CoD). The COD is mainly representing the type of device. For example using the COD, it is possible to distinguish whether the MAC address belongs to a mobile phone or a laptop. Hence, the information about the CoD can also be used to identify the type of vehicle carrying the BT device. These

two factors are jointly used to separate BT observations from cars and bicycles in this paper. The above factors are elaborated further in the next sub-section.

3.1 Hypothesis
In order to develop a clustering technique to distinguish between BT travel time observations from bicycles and cars, two hypothesis based on the insights explained earlier developed, as given below:

3.1.1 Hypothesis 1: Difference in travel time
First, it is assumed that the travel time of motor vehicles are lower than that of bicycles during uncongested traffic condition, and the difference is statistically significant. Therefore, it is expected that BT recorded travel time observations from an urban arterial with mixed traffic during uncongested traffic could be divided into two groups with different mean values. In order to test the first hypothesis, the null hypothesis H_0 and the alternative hypothesis H_1 are defined as follows:

$$H_0: \quad \mu_{Cluster(i)} = \mu_{Cluster(j)}$$
$$H_1: \quad \mu_{Cluster(i)} \neq \mu_{Cluster(j)}$$

Where:

$\mu_{Cluster(i)}$: Mean travel time in cluster i^{th}

H_0: indicates that mean travel times in both clusters are similar.

H_1 : indicates that mean value in both clusters are dissimilar.

The two groups can be created from all BT derived travel time observations using a suitable clustering method when the above hypothesis is valid.

3.1.2 Hypothesis 2: Difference in distribution of Bluetooth-enabled devices
Bluetooth MAC address contains the information about the CoD. The CoD field has a variable format which includes information of type of device. The COD represents the major and minor classes which are intended to define a general family of devices with which any particular implementation requests to be associated. Due to differences in application and functionality of Bluetooth-enabled devices (i.e. navigator, phones, computers etc.), it is expected that some major groups of CoDs to be used only by specific modes. In other words, some types of Bluetooth-enabled devices such as navigators used for route finding, radios and headsets for wireless communications will only be used in motor vehicles. Accordingly, it is anticipated that the devices belong to the major CoD group of audio/video devices including navigators, headsets and speakers can only be used within motor vehicles. Hence, after the clustering of Bluetooth based travel time, observations from BT devices with CoD of audio/video devices should belong to the cluster with shorter travel time. This step is mainly aimed to test the accuracy of the first hypothesis. In order to test this hypothesis, the distribution of CoD will be examined for each cluster.

3.2 Classification Method
In order to distinguish between bicycles and motor vehicles by BT data, clustering methods can be considered as one potential solution. However, note that any solution for this problem should be based on both the attributes of the recorded MAC address as input data and the expected output (i.e. mode-specified travel time estimation). For any detected MAC address, the information of sensor ID, timestamp (i.e. detection time), Radio Signal Strength Indicator (RSSI), Class of Device (CoD) are recorded in the feature vector. In this context, clustering methods are used as the core technique for classification. Although, there are limited numbers of attributes on feature vector, but clustering techniques are flexible in working with datasets with limited number of attributes of the variables. Moreover, these techniques have no assumption regarding the probability distribution of the underlying data (e.g.. assumption of normality) (Mooi and Sarstedt, 2011; Zhang, et al., 1996; Chiu et al., 2001). Three most applicable clustering mechanisms including Hierarchical clustering, K-means clustering and Two-step clustering techniques are employed. Implementation of the three methods also enables to have a comparison on the quality and capability of different methods. The three clustering methods are explained in the next section.

3.2.1 Clustering

Clustering techniques have been widely used in data mining and classification applications in transport studies. The objective of clustering is to classify the data into a number of clusters so that the data within a cluster is similar or related to one another and different or unrelated to the data belonging to other clusters. Different frames are considered to classify clustering mechanisms. One of the widely accepted frames is to classify clustering techniques into hierarchical clustering, partial clustering and hybrid clustering (Tayall and Raghuwanshi, 2010). The distinction between mentioned techniques is whether the set of cluster is nested or un-nested. The nested set permits clusters to have overlap or sub-clusters that are organised as a tree, whereas un-nested set consists of non-overlapping clusters such that each data is in exactly one cluster. Unlike traditional methods, hybrid clustering assigns any member of dataset a probability of belonging to each cluster. Then clusters will be shaped by assigning the members to the cluster with the highest probability (Anderson and Weiner, 2004). In this study, Hierarchical clustering, K-means clustering and Two-step clustering techniques are applied.

3.2.1.1 Hierarchical Clustering Algorithms

A Hierarchical algorithm is often displayed graphically using a tree-like diagram called a dendrogram. Dendrogram displays both clusters and sub-clusters relationships as well as the order in which the clusters were merged or split (Mooi and Sarstedt, 2011; Ramos et al., 2006). Two categories can be considered for Hierarchical algorithm, agglomerative (bottom-up) and divisive (top-down). In agglomerative procedure, first, the two most similar clusters (i.e., those with the smallest distance between them) are merged to form a new cluster at the bottom of the hierarchy. In the next step, another pair of clusters is merged and linked to a higher level of the hierarchy, and so on. In divisive all objects are initially merged into a single cluster, which is then gradually split up (Mooi and Sarstedt, 2011).

Similarity or dissimilarity between data points to generate a cluster is measured by an indicator called distance measure. A wide range of distance measures are used for determining the different clusters depending on the application and the data characteristics. In this study, the Euclidean distance measure is used to cluster BT data using Hierarchical clustering algorithm. This measure is only applicable if all variables are continuous and have the same physical unit (Tayal and Raghuwanshi, 2010).

$$d_{ij} = \left[\sum_{k=1}^{N_k} |x_{ik} - x_{jk}|^2 \right]^{\frac{1}{2}}$$

Where:
x_{ik}, x_{jk}: k^{th} element of the feature vector of data points i and j
d_{ij}: Distance between points i, j
N_k: Number of elements in the feature vector of a data point

3.2.1.2 K-means Clustering

The K-means clustering belongs to partial clustering techniques which classify the data to a number of clusters (K). K-value is a predefined value which needs to be specified beforehand. The K-means algorithm follows an entirely different concept compared to the Hierarchical methods. This algorithm uses the within-cluster variation as a measure to form homogenous clusters. This algorithm clusters the data in such a way that the within-cluster variation is minimised (Mooi and Sarstedt, 2011). The K-means algorithm aims to minimise a squared error objective function through an iterative process. The objective function is basically the (squared) distance from each observation to the centre of the associated cluster (i.e. centroid) shown below:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Where

$\left\| x_i^{(j)} - c_j \right\|^2$: A chosen distance measure between a data point $x_i^{(j)}$ and the cluster center c_j

J: An indicator of the sum square of the distance of the n data points from their corresponding cluster centres.

The centroids are representative of the clusters and usually the mean of a group of data for a continuous n-dimensional space dataset. The clustering process starts by randomly assigning the data to the pre-specified number of clusters (K). In an iterative process, the data are continuously reassigned to other clusters to minimise the objective function (i.e. within-cluster variation) (Mooi and Sarstedt, 2011).

3.2.1.3 Two-step Clustering

The Two-step clustering extracts the natural clusters within a dataset. There are a number of features employed by the Two-step clustering algorithm which differentiate it from traditional clustering algorithms. The features include ability to handle very large datasets, ability to handle both continuous and categorical variables, and ability to automatically determine the optimum number of clusters (Zhang, et al., 1996; Chiu et al., 2001). The Two-step clustering is based upon a sequential approach of creating clusters and then sub-clusters (Theodoridis and Koutroumbas, 1999). The name Two-step clustering is already an indication that the algorithm is based on a two-stage approach: Firstly, the algorithm undertakes a procedure that is very similar to the K-means algorithm. Secondly, the algorithm conducts a modified Hierarchical agglomerative clustering procedure by combining the objects sequentially to form homogenous clusters. This is done by building a so-called cluster feature tree whose "leaves" represent distinct objects in the dataset (Mooi and Sarstedt, 2011).

4. Experiment Setup and Data Collection

In order to explore the feasibility of BT for mode specific travel time estimation as well as to test the aforementioned method, a field experiment is conducted. In the field experiment, two sets of data are collected; Real-world BT data and the corresponding travel time data obtained by camera. Travel time data obtained by camera are used as ground truth. The final clustering techniques output of BT data are compared with the real travel time for the bicycles and motor vehicles captured by camera. More details are presented in the next section.

4.1 Site selection:

In order to minimise the impact of the noise caused by pedestrians on the analysis and reduce the geometric design impact on the cyclists speed, a number of criteria are considered. The criteria for site selection are as follows:

1. A link with mixed traffic and appropriate number of both motor vehicles and bicycles, in order to obtain a proper sample size.
2. A link with limited number of pedestrians, to minimise the disturbance from pedestrians carrying Bluetooth-enabled devices.
3. A link with normal topography which is not so steep to avoid significant influence on the bicycles speed compared to motor vehicles during off-peak.
4. A link with an appropriate distance from the roundabouts and signalised junctions to minimise the impact of the deceleration, acceleration, queuing, number of stops and associated delays.
5. A link with no or minimum number of access roads to residential or industrial places, to minimise the diversion of the users.

Considering the aforementioned criteria, a section of Sauersvej located between the intersection of Sauersvej-Sohngårdsholmsvej and Sauersvej- Scoresbysundvej, in south end of Aalborg was selected. As shown in Figure 2, this link is mainly surrounded by green areas on both sides with no access road connection. It has an acceptable distance from residential areas with minimum number of pedestrians plus a flat topography. The link functions as a connector with appropriate numbers of bicycles and motor

vehicles. The Bluetooth sensors were located on the ground at the shoulder of West-East Direction. The distance between the two Bluetooth sensors was 550 metres. The Locations of the Bluetooth sensors are shown in Figures 3 and 4.
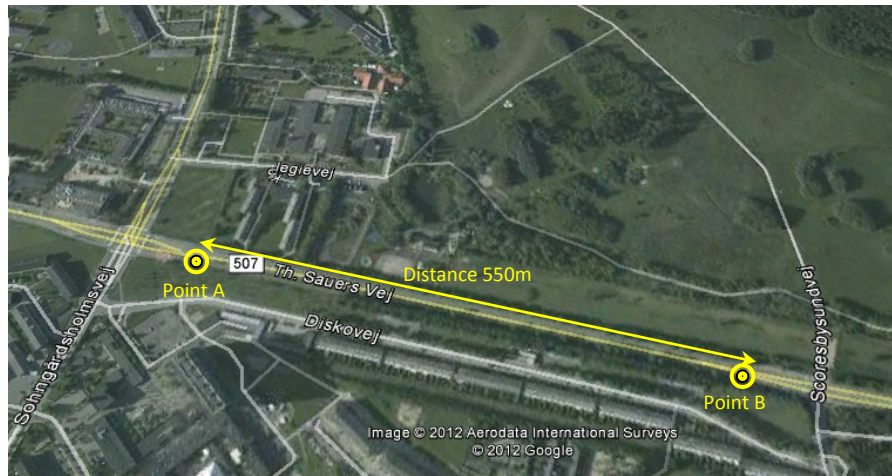

Figure 2- Selected link of Sauersvej, Aalborg, Denmark

Figure 3 - The Bluetooth sensor located on the ground at West-East Shoulder at Sauersvej-Sohngårdsholmsvej (Point A).	**Figure 4 - The Bluetooth sensor located on the ground at West-East Shoulder at Sauersvej-Scoresbysundvej (Ponit B).**

4.2 Bluetooth Sensor

The Bluetooth sensors used in this study are developed by company BLIP Systems, a Bluetooth sensor manufacturing company in Denmark. The sensors known as BlipTrack consist of Bluetooth detectors sending data via a 3G connection in real time and providing time stamped and direction of passing Bluetooth-enabled devices through a back-end server. The sensors have three antennae for optimal coverage of road segment. A directional antenna is pointed in each direction along the road, and a short range antenna is focused on area around the sensor (see Figure 5). The size of the detection zone varies between 70-200m on either side of the sensor along the road.

4.3 Data Collection:

In order to evaluate the accuracy of BT based mode-specific travel time, real traffic was recorded using video cameras simultaneously (i.e. the ground truth data). Data collection was conducted for the period of 6 hours (9:00-15:00) during off-peak (i.e. uncongested traffic).

Due to the wide detection zone of the BlipTrack, a single MAC address is likely to be detected more than once while it passes a sensor location. On the one hand, having a wide range of detection events increases the probability of detecting devices with high speed. On the other hand, this also increases the ambiguity and uncertainty about the location of detected devices at the time that they have been tracked by the sensor. Having a number of detection events for each device raises the question as to which values should be taken to obtain the most accurate estimate. In order to employ the information of multiple detections as well as defining accurate estimation methods, three different methods have been developed as alternatives to estimate travel time:

Enter-Enter: In this case, the first detection events recorded by upstream and downstream sensors are used for travel time estimation.
Leave-Leave: In this case, the last detection events recorded by upstream and downstream sensors are used for travel time estimation.
Peak-Peak: In this case, the detection event with the highest Radio Signal Strength Indication (RSSI) at upstream and downstream sensors is used for travel time estimation. The RSSI is an indicator representing the power level received by the antenna. The magnitude of the RSSI is highly dependent on the distance from the sensor location. In other words, if a detection event is recorded with a high level of RSSI, it is more likely that it is in a location closer to the sensor (i.e. minimum distance from the sensor location). Therefore, the hypothesis is that the detection event with maximum signal strength (maxRSSI) will result in higher accuracy for travel time estimation.

4.4 Data Filtering:

In order to remove the outliers from the BT based travel time data obtained by the *Enter-Enter*, *Leave-Leave*, and *Peak-Peak* methods, the data were filtered based upon an upper and lower speed level. The recorded speed higher than 2 times of the free flow speed (120 km/h) and lower than the speed of pedestrians (6 km/h) were disregarded. This is mainly based on the assumption that motorist vehicles are not allowed to drive faster than the free flow speed (60 km/h). In order to cover all the speed variations with a level of confidence, 2 times of the free flow speed (120 km/h) was considered as the upper level. It is also expected that a bicycle is not slower than a pedestrian with the average speed of 6 km/h.

3.3 Accuracy Metrics and Methods Evaluation

So as to evaluate the accuracy of mode-specific travel time estimations using BT methods, three accuracy measures including Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) are calculated. The MPE, MAPE and RMSE formula are given as follow:

$$MPE = 100 . \frac{1}{N} \sum_{i=1}^{N} \frac{(\widehat{T} - T_i)}{T_i}$$

$$MAPE = 100 . \frac{1}{N} \sum_{i=1}^{N} \frac{|\widehat{T} - T_i|}{T_i}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\widehat{T} - T_i)^2}$$

Where:

N: The total number of observed motor vehicles and cyclists

T_i: Real travel time for the ith record time period obtained using camera recording as the ground truth

$\widehat{T}$: Estimated mean travel time using BT for the ith time period

Note: for more information see Araghi (2010).

In order to evaluate the accuracy, the BT travel times were matched with the camera records when the difference between BT and the ground truth detection time is equal or less than +/- 1 second. At the speed of 60 km/h, +/- 1 second is equivalent to approx. +/- 17 metre location ambiguity along the road. This level of accuracy was deemed sufficient in this context using engineering judgement.

5. Data Analysis and Results

As can be seen in Figure 6, the BT data overlaps the records of motor vehicles and bicycles. Thus, this indicates the feasibility of BT to represent and estimate the mode-specific travel time; however, the accuracy of this method needs further investigation. The results of the hypothesis testing and accuracy analysis are presented in the following sections. Since the pattern observed for the both directions were similar, only the results for West-East direction are presented.
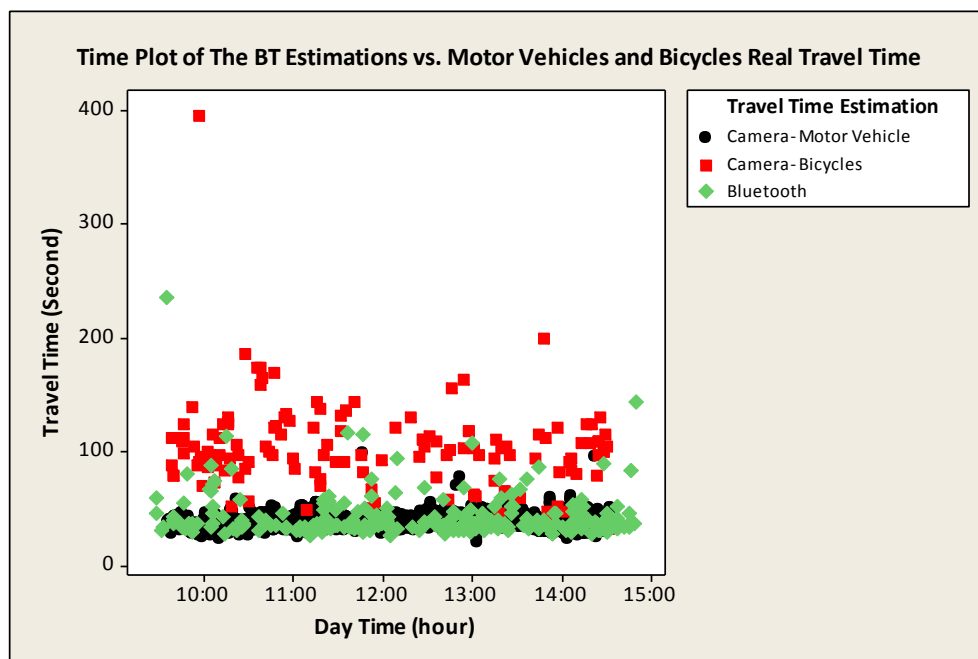

Figure 6- Time Plot of the Bluetooth Detection vs. Motor Vehicles and Bicycles Real Travel Time

5.1 Cluster Analysis - Difference in Travel Time (H_0)

Results of the cluster analysis for various travel time estimation methods for the West-East direction are summarised in Table 1. The results show that the three clustering techniques classify the data into two main classes; group 1 with short-travel time presenting motor-vehicles and group 2 with long-travel time representing bicycles. Thus, the first hypothesis is confirmed. Hence, it can be concluded that BT is capable to capture and represent the difference in travel time of motor-vehicles and bicycles during uncongested traffic situations.

The results of the cluster analysis using three methods of Hierarchical, k-means and Two-step were similar except for the *Enter-Enter* method with a small difference. The mean travel time estimated by *Enter-Enter*, *Peak-Peak* and *Leave-Leave* method for the cluster of motor vehicles are close to the real-travel time obtained by video cameras. However, the results for travel time estimated for bicycles differ significantly from the actual bicycle's travel time. There is an exception for *Peak-Peak* method and k-Mean cluster obtained by *Enter-Enter* method.

As shown in Table 1, the BT detection rate is significantly lower than the rate usually captured by BlipTrack (20% of the traffic flow) in an earlier study (Lahrmann et al., 2010). This is mainly due to the height and the BT sensor locations. Since the sensors were located on the ground, the size of the detection zone was reduced significantly resulting in a lower sampling rate. This again outlines the importance of the the impact of the location of the installed sensor, the height and the mount design. These factors affect Bluetooth device detection rate and hence travel time accuracy. These engineering aspects should be given due consideration in order to achieve reliable travel time estimates using BT sensors.

Table 1 – Results of the Cluster Analysis for West-East Direction

Data Collection	Method		Cluster	Count	Mean (sec)	SE-Mean	StDev	Min	Q1	Med	Q3	Max	Mode
Video Camera	Total		Motor Vehicle	3088	38.832	0.0891	4.954	21	36	38	41	99	40
			Bicycle	142	102.88	3.32	39.53	33	82.8	101	119.5	395	97
Bluetooth	Enter-Enter	Total without Clustering		241	44.26	1.35	20.96	27	34	37	46	236	36
		Hierarchical	Motor Vehicle	240	43.46	1.09	16.93	27	34	37	46	144	36
			Bicycle	1	236	N/A	N/A	236	N/A	236	N/A	236	N/A
		K-mean	Motor Vehicle	223	39.668	0.585	8.732	27	34	37	42	69	36
			Bicycle	18	101.17	9.13	38.75	72	76.8	87.5	114.25	236	76
		Two-step	Motor Vehicle	239	43.04	1.01	15.66	27	34	37	46	117	36
			Bicycle	2	190	46	65.1	144	N/A	190	N/A	236	N/A
	Peak-Peak	Total without Clustering		138	39.47	1.11	13.03	28	34	36	39	136	36
		Hierarchical	Motor Vehicle	134	37.597	0.522	6.048	28	34	36	39	61	36
			Bicycle	4	102.3	13.4	26.7	75	78.3	99	129.5	136	N/A
		K-mean	Motor Vehicle	134	37.597	0.522	6.048	28	34	36	39	61	36
			Bicycle	4	102.3	13.4	26.7	75	78.3	99	129.5	136	N/A
		Two-step	Motor Vehicle	134	37.597	0.522	6.048	28	34	36	39	61	36
			Bicycle	4	102.3	13.4	26.7	75	78.3	99	129.5	136	N/A
	Leave-Leave	Total without Clustering		245	41.6	1.76	27.59	24	34	37	40	301	37
		Hierarchical	Motor Vehicle	242	38.864	0.779	12.124	24	34	36.5	40	136	37
			Bicycle	3	262.3	19.8	34.2	236	236	250	301	301	N/A
		K-mean	Motor Vehicle	242	38.864	0.779	12.124	24	34	36.5	40	136	37
			Bicycle	3	262.3	19.8	34.2	236	236	250	301	301	N/A
		Two-step	Motor Vehicle	242	38.864	0.779	12.124	24	34	36.5	40	136	37
			Bicycle	3	262.3	19.8	34.2	236	236	250	301	301	N/A

1. SE-Mean: Standard Error Mean
2. StDev: Standard Deviation
3. Min: Minimum
4. Med: Median
5. Max: Maximum
6. Q1: First Quartile
7. Q3: Third Quartile

5.2 Cluster Analysis - Difference in CoD Distribution (H$_1$)

Results of the CoD distribution for cluster of bicycles are presented in Figure 7. According to the results, two main categories namely audio/video and phone are distinguished in the bicycle cluster. Based on the H$_1$ it was expected that the devices detected in bicycle cluster will not include the CoD category of audio/video. However, as shown in Figure 7, there are cases where the audio/video devices are clustered in the category of bicycles. After post-analysis, it was perceived that those BT devices were matched with the motor vehicles stopped for a longer period along the road side. This indicates that although the information of CoD can be used as a valuable source of information for classifying the users, but special cases similar to vehicles stopping should be carefully taken into account.
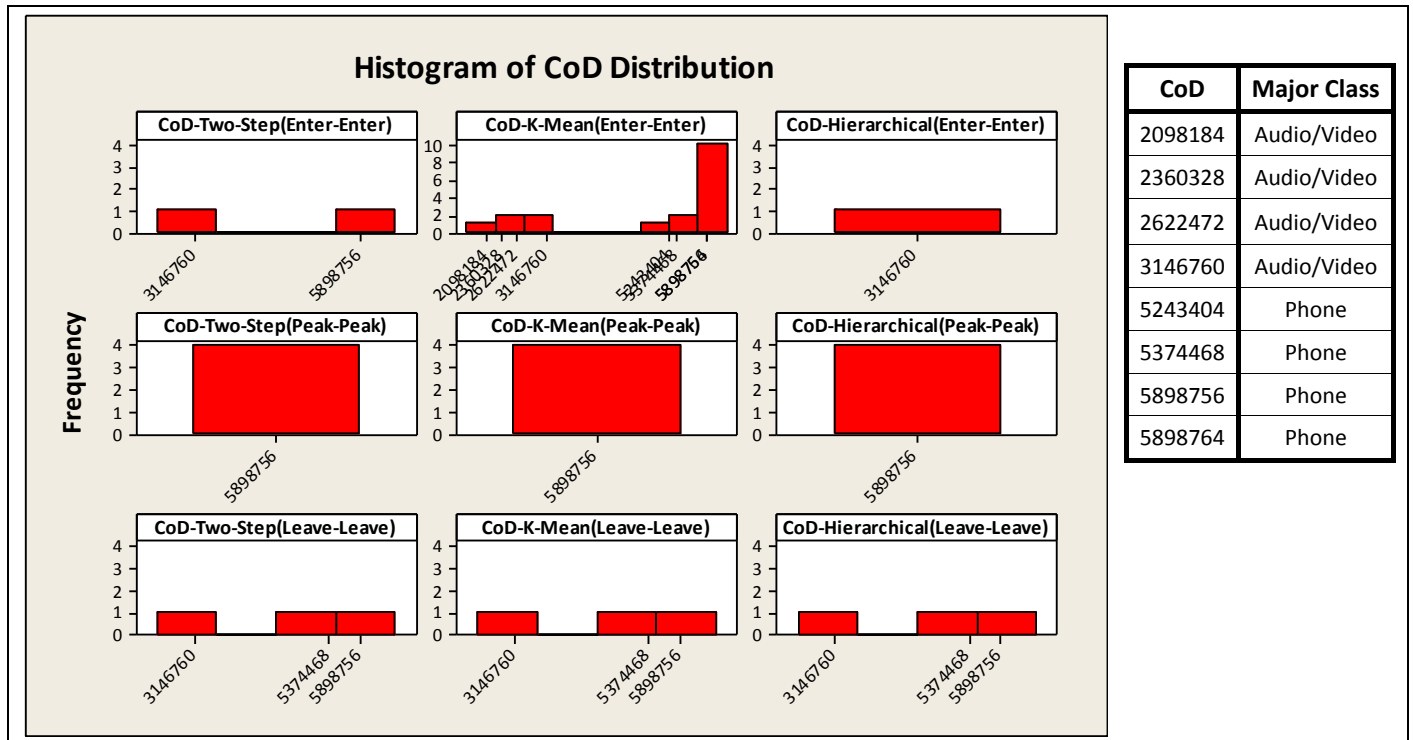

CoD	Major Class
2098184	Audio/Video
2360328	Audio/Video
2622472	Audio/Video
3146760	Audio/Video
5243404	Phone
5374468	Phone
5898756	Phone
5898764	Phone

Figure 7 - Histograms of CoD Distribution for Bike Clusters

5.3 Results of the Estimation Method

In order to evaluate the accuracy of the three estimation methods namely, *Enter-Enter*, *Peak-Peak* and *Leave-Leave*, distribution of the data around the ground truth values are presented in Figure 8. Compared to *Enter-Enter* and *Leave-Leave* methods, *Peak-Peak* produces closer distribution to the ground truth (Figure 8). As for the standard deviation shown both in Table 1 and Figure 8, *Peak-Peak* method results in smaller value. Thus, the accuracy of *Peak-Peak* method for travel time estimation is higher than the other two methods.

The accuracy metrics, MPE, MAPE, and RMSE, for the motor vehicle cluster are summarised in Table 2. The numeric values of the MPE, MAPE and RMSE conform with the pattern shown in Figure 8. Results of the RMSE for *Peak-Peak* and *Leave-Leave* methods show errors of 7 and 10.5 seconds for the travel time estimation, respectively. Similarly, for the MPE there are 3.8% and 1% error in travel time estimation obtained by *Peak-Peak* and *Leave-Leave* methods. MAPE accounts for 14% and 15.70% when it comes to *Peak-Peak* and *Leave-Leave* errors. Compared to the *Peak-Peak* and *Leave-Leave* methods, the *Enter-Enter* provides the least accurate travel time estimate. Comparing the results of the MPE, MAPE and RMSE show that in most cases the *Peak-Peak* method tends to be more accurate.

The accuracy metric analysis also emphasises the suitability of the proposed hypothesis of this study. Based on the results of this study, it can be concluded that BT is capable to provide an accurate estimate for motor vehicles travel time in mixed traffic under uncongested condition. However, it is necessary to use an appropriate clustering method to classify traffic modes.

Note that due to small sample size of bicycles in this study, performance of the accuracy analysis for this road user was not possible.
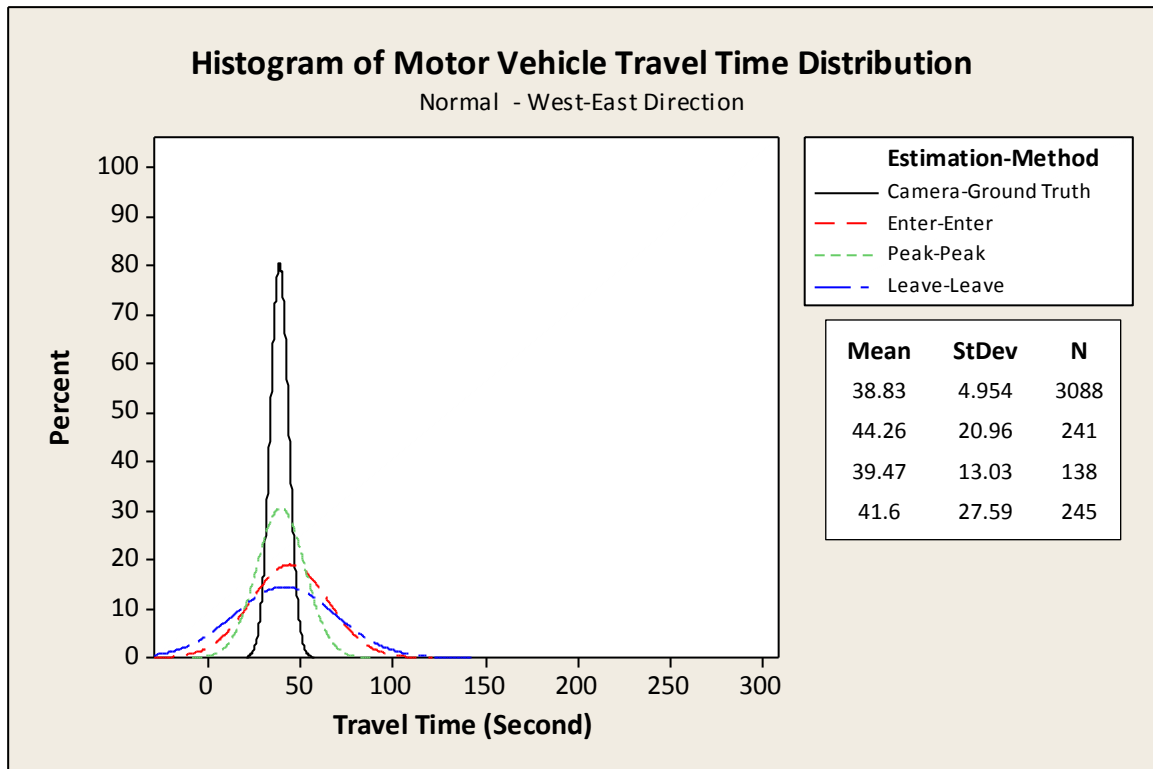

Figure 8 - Histograms of Motor Vehicle Travel Time Distribution Using Various Estimation Methods

Table 2 - Evaluating the Accuracy of Travel Time Estimated for Motor Vehicles Using Various Methods

Method	Accuracy Metric	Clustering Technique		
		Hierarchical	**K-means**	**Two-step**
Enter-Enter	MPE (%)	14.10	4.82	14.10
	MAPE (%)	28.94	21.08	28.94
	RMSE (Sec)	16.95	10.73	16.95
Peak-Peak	MPE (%)	-3.84	-3.84	-3.84
	MAPE (%)	14.13	14.13	14.13
	RMSE (Sec)	7.08	7.08	7.08
Leave-Leave	MPE (%)	-1.06	-1.06	-1.06
	MAPE (%)	15.70	15.70	15.70
	RMSE (Sec)	10.49	10.49	10.49

6. Conclusion

In order to evaluate the feasibility of BT for estimating mode-specific travel time in mixed traffic situation under uncongested conditions, a field experiment was conducted. The BT data and camera based data were collected to support the objectives of the study.

So as to classify the BT data, three statistical clustering techniques were employed namely, Hierarchical, K-means and Two-step clustering. The results of cluster analysis highlighted that the clustering techniques are able to classify BT data according to different travel modes satisfactorily. Since the output of the three techniques were very similar, there is not enough evidence to prioritise them. The output also emphasised

the fact that there is a statistically significant difference between the travel time of the motor vehicles and bicycles during uncongested conditions, resulting in the applicability of clustering techniques in this context. It is also perceived that there is a pattern in CoD distribution for various traffic modes; there is a tangible difference in the type of the BT devices used by road users. In this context, CoD can be used as an informative source of data for classifying the traffic modes in mixed traffic situations. However, the travel time estimation accuracy for bicycles could not be carried out due to low sample size of bicycles.

BT sensors capture multiple detection events for each passing vehicle. This paper explored the best way to make use of this data, and concluded that the *Peak-Peak* scheme outlined in this paper resulted in the most accurate travel time estimates among candidate methods.

This paper also identifies that CoD data captured by BT sensors provide vital information that can be used to distinguish travel modes. This is an area of future research where CoD information will be used to determine the travel time of motor vehicles during congested conditions.

References

1. Barcelo, J., Montero, L., Marques, L., (2010). 'Travel Time Forecasting and Dynamic OD Estimation in Freeways Based on Bluetooth Traffic Monitoring', *The 89th Annual Meeting Transportation Research Board,* Washington D.C.

2. Chiu, T., Fang, D., Chen, J. and Wang, Y., (2001). 'A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment', Proceeding of the International Conference On knowledge Discovery and Data Mining. San Fransico, USA.

3. Groenen P., J.F., Kaymak, U., Rosmalen J.V., (2006), 'Fuzzy clustering with Minkowski distance functions', Econometric Institute Report EI 2006-24.

4. Haghani, A., Hamedi, M., Sadabadi, K.F., Young, S.E., Tarnoff, P.J., (2010). 'Freeway Travel Time Ground Truth Data Collection Using Bluetooth Sensors', *The 89th Annual Meeting Transportation Research Board,* Washington D.C.

5. Haghani, A., Yang, S., Hamedi, M., (2007). 'Cellular Probe Data Evaluation, Case 13 Study: The Baltimore Multimodal Traveler Information System (MMTIS)', Technical 14 Report, Submitted to Maryland Department of Transportation.

6. Haghani A., Hamedi, A., Farokhi Sadabadi, K., (2009). 'I-95 Corridor Coalition 16 Vehicle Probe Project', University of Maryland, College Park.

7. Hamedi, M., Haghani, A., Sadabadi, F., (2010). 'Using Bluetooth Technology for Validating Vehicle Probe Data', *The 89th Annual Meeting Transportation Research Board,* Washington D.C.

8. Mooi, E., Sarstedt, M., (2011), A Concise Guide to Market Research, DOI 10.1007/978-3-642-12541-6_9, # Springer-Verlag Berlin Heidelberg.

9. Lahrmann, H., Skoven Pedersen, K. C., and Christensen, L.T., (2010). Bluetooth detektorer som ny cost-effektiv sensor i vejtrafikken. In Proceedings the Annual Transport Conference at Aalborg University, Aalborg.

10. Malinovskiy, Y., Wu, Y.J., Wang, Y., Lee, U.K., (2010), 'Field Experiments on Bluetooth-Based Travel Time Data Collection', *The 89th Annual Meeting Transportation Research Board,* Washington D.C.

11. Namaki Araghi, B. (2010). Using the k-NN Method for Incident Duration Prediction based on TPEG data, MSc Thesis, Centre for Transport Studies, Imperial College London, UK.

12. PennDot, 2010 'Bluetooth Travel Time Technology Evaluation Using the BlueTOADTM', Pennsylvania Department of Transportation and KMJ Consulting.

13. Puckett, D.D., Vickich, M.J., (2010), 'Bluetooth®-Based Travel Time/Speed Measuring Systems Development', Texas Transportation Institute, UTCM Project #09-00-17.

14. Ramos, S., Vale, Z., Santana, J., Rodrigues, F., (2006), 'Use of Data Mining Techniques to Characterize MV Consumers and to Support the Consumer-Supplier Relationship'. Proceedings of the 6th WSEAS International Conference on Power Systems, Lisbon, Portugal.

15. Saunier, N. & Morency, C., (2011), 'Comparing data from mobile and static traffic sensors for travel time assessment', *1st Congress of the Transportation and Development Institute of ASCE*.

16. Tarnoff, P.J., Bullock, D.M., Young, S.E., Wasson, J., Ganig, N., Sturdevant, J.R., (2009), 'Continuing Evolution of Travel Time Data Information Collection and Processing', *Conference The 88th Annual Meeting Transportation Research Board,* Washington D.C.

17. Tayal M.A., Raghuwanshi, M.M., (2010), 'Review on Various Clustering Methods for the Image Data', Journal of Emerging Trends in Computing and Information Sciences, Vol 2, pp.34-38.

18. Theodoridis, S. and K. Koutroumbas., (1999), 'Pattern recognition'. Academic Press, New York.

19. Zhang, T., R. Ramakrishnan and M. Livny., (1996), 'BIRCH: An efficient data clustering method for very large databases', Proceedings of the ACM SIGMOD Conference on Management of Data. Montreal, Canada.

20. Wasson, J.S., Sturdevant, J.R. & Bullock, D.M., (2008), 'Real-Time Travel Time Estimates Using Media Access Control Address Matching', *ITE Journal,* vol. 78(6) pp 20-23.

21. Xu, R. and Wunsch, D. II (2005). Survey of clustering algorithms. IEEE Transactionson Neural Networks, 16(3): 645–678.