EVALUATION OF USING WIFI SIGNALS TO ESTIMATE INTERSECTION TRAVEL TIME

Sajad Shiravi
PhD Candidate, Univeristy of Waterloo, Canada

Kamal Hossain
PhD, University of Waterloo, Canada

Liping Fu
PhD, Professor, University of Waterloo, Canada

Amir Ghods
PhD, SMATS Traffic Solutions Inc., Canada

ABSTRACT

Travel time is an important measure in traffic engineering and planning with many applications including identification of network bottlenecks, plan to improve traffic mobility, providing commuters with travel time information, and traffic signal control evaluation and control. Currently a number of technologies can provide travel time information such as GPS enabled probes and identifying vehicles with Bluetooth or Wi-Fi devices. The later method detects and matches unique Media Access Control (MAC) address of the Bluetooth or Wi-Fi activated devices to calculate travel time information. This method is a non-intrusive and cost effective and has gained a lot of attention in the past few years. Extensive research has been done on evaluating the accuracy, application, and market penetration rate of using Bluetooth technology for travel time estimation for both urban arterials and highways. However, the application of Wi-Fi MAC address detection and matching for travel time estimation at urban arterials has not been adequately studied. The limited available studies are contradicting with significant differences in terms of travel time accuracy, and penetration rates. This study intended to evaluate the accuracy and reliability of using the combination of Bluetooth and Wi-Fi based travel time estimates through a case study. A sample size analysis is conducted and the expected statistical sampling errors are compared with that of obtained from comparing the Bluetooth and Wi-Fi data with ground truth information which is collected through video footage. The results of this study show that Wi-Fi signals can also provide reliable travel time information. When combined with Bluetooth travel times data, which significantly increases the market penetration rate comparing to using Bluetooth alone. The combination of using Bluetooth and Wi-Fi signal provided penetration rates up to 8% with errors less than 10% compared to ground truth data.

Keywords: Wi-Fi data, Bluetooth data, Travel time, Accuracy, Sample size

1. INTRODUCTION

Travel time information is an important measure of traffic performance which is widely used by transportation practitioners to evaluate and quantify the performance of the roads and plan for improvements. For this purpose transportation agencies traditionally collect travel time information through dedicated probe vehicles equipped with Global Navigation Systems (GPS). Travel time information from GPS probes has been deemed as having a very high cost per data point collected with limited spatial and temporal coverage. In the recent years non-intrusive traffic data collection systems such as those based on Media Access Control (MAC) address matching, like Bluetooth technology, has gained a lot of attention. Accuracy and cost efficiency are the main characteristics that have made this data source attractive to the transportation industry.

Bluetooth traffic data collection technology provides useful traffic information such as travel time, average space mean speed for measuring the performance of roadways as well as origin-destination data for transportation planning. This information is obtained by matching the unique MAC addresses of passing vehicles with activated Bluetooth devices in two or more predetermined locations. MAC addresses are unique identifiers defined by IEEE as a communication protocol for wireless Bluetooth and Wi-Fi communications. This unique code, in most wireless enabled digital devices, is a string of 6 groups of 2 hexadecimal digits in the format of XX:XX:XX:XX:XX:XX which could be a combination of numbers or letters. In the process of Bluetooth MAC address detection the advertiser device sends an advertise packet to announce its presence; while, the scanner device searches for the presence of an advertiser. The scanner receives the advertise packet that contains the advertiser device's MAC address. In the context of Bluetooth based traffic data collection, the advertiser is a Bluetooth equipped device such as a cell phone located inside the passing vehicle with its Bluetooth module in advertise mode, and the scanner is the Bluetooth data collection device installed close to the road.

In addition to Bluetooth technology, nowadays, a large number of portable and in-vehicle devices also contain Wi-Fi communication components. For detection purposes unlike Bluetooth, which requires the device to be in "discovery (advertise) mode", technically, Wi-Fi signals includes the MAC address of the device in their communication packet; hence, only needs to be left turned on to be detected. This results in substantially increased chance of detections comparing to Bluetooth. Higher detection rates can potentially increase the chances to obtain matches among the detected MAC addresses thus, increasing the sample size (i.e. market penetration rate) required for travel time estimation and improving the reliability of collected traffic data. However, Bluetooth technology has a probe request rate in range of a few seconds while devices with Wi-Fi only emits probe requests once every 30 to 45 seconds in sleep mode (Cunche, 2013) depending on the device. The communication can be more frequent if the screen is on or connected to an in-vehicle access point.

The available studies regarding the use of wireless communication technology for vehicular traffic data collection have been mostly limited to Bluetooth data. Bluetooth and Wi-Fi are different in their design, detection range, communication protocols and popularity and very few studies have investigated the application of Wi-Fi data for traffic data collection in terms of feasibility, filtering, accuracy, market penetration rates, and detection zones.

The application of wireless communication technology for traffic data collection has been previously evaluated for pedestrian activity and movement tracking. In particular, the application of Bluetooth data has been extensively evaluated for both pedestrian and vehicular traffic data collection. A review of the available literature shows that the matching rate of Bluetooth data in urban arterials (or market penetration rate) compared to the actual traffic volumes are between 2-5% with more than 90% individual reading accuracy and between 80% to 90% accuracy in average travel time estimations (which is mainly an issue of sample size) depending on the time period considered (Vo, 2013), (Liu, 2013), (Araghi, 2015), (Wang, 2014), (Haghani, 2010), (Yucel, 2013), (Malinovskiy, 2011).

Abedi (2013) is among the very few studies which have evaluated the application of Wi-Fi data for crowdsourcing purposes and compared its functionality with Bluetooth data in terms of discovery time, architectural differences, signal strength, range, antenna properties and popularity. It is also observed that 90% of the unique MAC addresses found in a field test were from Wi-Fi, without indicating what portion belongs to phones or other devices. The study proposes two multirange and overlapping approaches to narrow down the detection location. The study did not evaluate Wi-Fi data specifically for vehicular traffic data collection.

Musa et., al (2012) proposed a trajectory estimation method in a grid network deployment using Wi-Fi MAC address matching based on Viterbi's method in which the vehicle's trajectory was estimated depending on the sensors in the network which had detected the device. A comparison of the proposed method with ground-truth trajectory information from GPS data for individual vehicles showed an average of 14% error in location detection. However, the study did not contain any information on the market penetration rates of MAC matches. Moreover, the proposed outlier filtering method was limited to "blacklisting" based on MAC address OUI (Organizationally Unique Identifier) characteristics. Due to lack of a comprehensive ground truth data the accuracy remained unknown for travel time estimation over the evaluation time period. In another study by Wang (2014) the accuracy of several different types of detectors including Bluetooth and Wi-Fi detectors were evaluated in an urban environment. The collected travel time information was then compared with ALPR (Automatic License Plate Reader) data which serves as the ground truth travel time information for their study. Since ALPR is not a volume measuring system, due to lack of real-world traffic volume data, an accurate estimate of market penetration rate was not available in the study. The study reported an average of 40% MAPE (Mean Absolute Percentage Error) in estimating hourly travel time using the hybrid Bluetooth and Wi-Fi detectors. Root mean standard errors up to 140 seconds were observed in the study reporting higher travel time estimation errors during peak hours.

A review of the existing studies in evaluating the usage of MAC address matching for travel time estimation reveals that the studies on using Wi-Fi data for traffic travel time estimation are very limited. Moreover, Wi-Fi data accuracy was usually evaluated on an hourly basis without providing any information on the accuracy of the average travel time estimates in shorter time periods that are required for real-time applications such as traffic signal control purposes. In addition, no accurate estimation of Wi-Fi typical market penetration rates are reported in the current literature. Overall, the travel time estimation using Wi-Fi data is a less explored field in transportation and many gaps still exist in this area. The motivation of this research is to fill in some of the gaps mentioned above in this field by evaluating the accuracy and penetration rate of Wi-Fi based travel time estimation using ground truth information from video data and conducting comparisons with Bluetooth data.

2. TEST SETUP AND STUDY SITE

The data collection site was an intersection, Columbia-Westmount, located in the northern part of Waterloo, Ontario, in a residential area close to the University of Waterloo Campus with a very simple setting (e.g. no nearby stores, drive-through restaurants and minimal pedestrian activity). Each approach of the signalized intersection contains two through movement lanes and a separate left turn movement. The intersection operates on a time of day pretimed signal plan. A map of the study site illustrating the intersection under study and sensor locations is shown in Figure 1. As can be seen, at each approach of the intersection one low and one medium range Wi-Fi detector is installed to capture all movements at the intersection. The low range detector is measured to have about 50 meter range and the medium range is measured to have 100-150 meter range. The video cameras were only installed on the West and East approaches to capture the ground truth travel time information for comparison purposes. The field test was conducted on a weekday from 3:40pm to around 6:00pm in order to capture both the peak and off-peak traffic conditions.

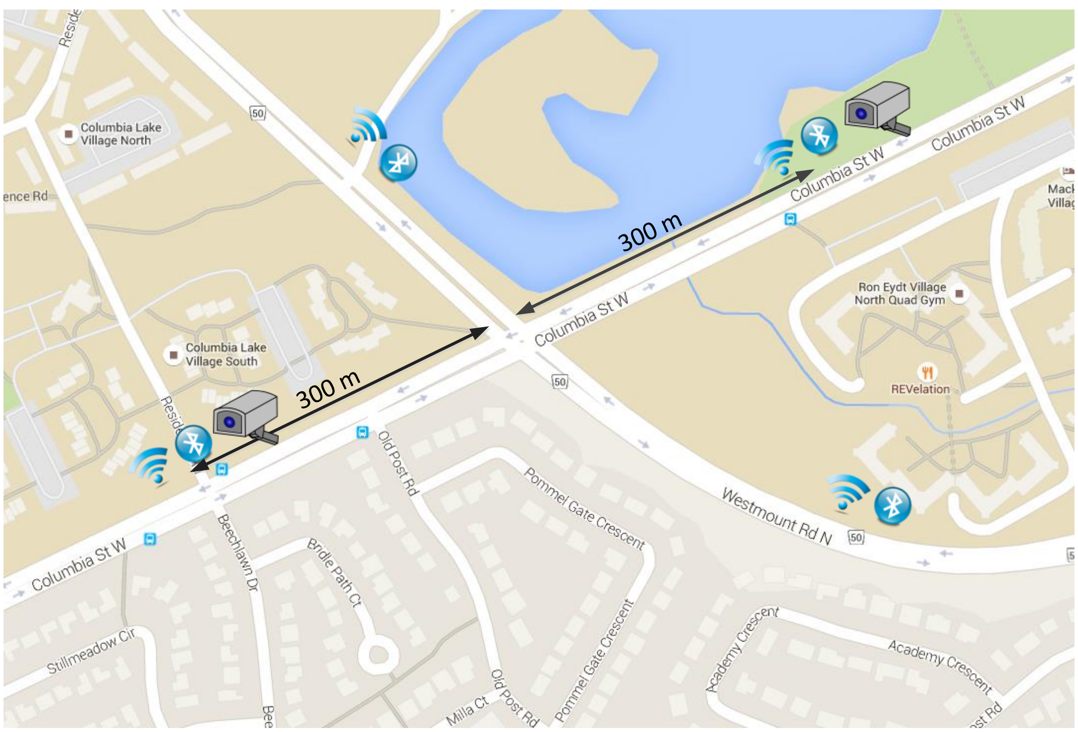

Figure 1: Map of the Selected Field Test Location with Sensor Locations

The Bluetooth/Wi-Fi scanner used in this study was a tablet device with an Android base application, called TrafficTab^{TM} manufactured by Smats Traffic Solutions Inc. (SMATS, 2016), shown in Figure 2. The device uses the internal Bluetooth capability of the Android tablet itself with an addition of an external Wi-Fi dongle connected to the tablet (Figure 2). The internal Bluetooth contained a class 2 radio with an average 50m and the short-range Wi-Fi module with around the same range. A second Wi-Fi antenna was also used with a range of around 150m during the field tests. The scanner device has a Bluetooth scanning duration of around 12 seconds after which another scan is run right away. In each Bluetooth scan, a MAC address can be detected and listed only one time. However, the same device can be detected again in the next 12-s round of Bluetooth scan if it is still in the vicinity of the scanning range. The scan duration of the Wi-Fi dongle, however, is as low as 1-2 seconds such that a Wi-Fi device can be detected many times within a short period of time. The Android application output file provides detailed information including the detected MAC address, the first and last detection time stamps of a single MAC address, how many times the MAC address were detected while the device was in the range as well as type of the detection: Bluetooth vs. Wi-Fi. Table 1 provides a sample of the detection outputs.

Table 1: Sample Outputs of Bluetooth and Wi-Fi Detections

Detected MAC address	First detection timestamp	Last detection timestamp	Detection Frequency	Detection type
E4F8EF1E124C	16:54:32	16:55:02	2	Bluetooth
FCA19E11D362	16:57:20	16:57:20	1	Bluetooth
4C7C5F8A395C	16:58:47	16:59:55	6	Wi-Fi
489D242D96D8	16:59:31	16:59:45	3	Wi-Fi

Figure 2: Portable Bluetooth and Wi-Fi scanner device use in the study

Video camera data were also collected for post-processing the traffic and determining the ground truth travel time and the traffic volume within the testing period. For this purpose, Miovision Scout video detection devices manufactured by Miovision Technologies were mounted on the East and West approaches, on the same locations as the Bluetooth and Wi-Fi scanners were placed.

3. DATA ANALYSIS AND RESULTS

In this section the data filtering process is explained and different methods are compared. Once the valid data points have been obtained a statistical sample size analysis is conducted to compare the minimum sample sizes required to the available samples for different time windows and for each of the mid-range and low range antennas. By using the valid data points and the suitable time window, obtained in the previous steps, an error analysis is conducted to compare the BT/Wi-Fi travel times with the expected sampling error and the ground truth information in each time window. The analysis also compares the use of first, median and last detection travel times.

Figure 3 provides a visual representation of the East-West matched MAC addresses (combined Bluetooth and Wi-Fi) for travel times less than 300 seconds for both low and medium range detectors overlaid with ground truth travel time data obtained from video data. Overall as can be seen there is a reasonable visual harmony between the two data sets. However, the data requires filtering to eliminate outliers that may have been generated due to pedestrian or cyclist activity or other sources of errors such as nearby Wi-Fi access points.

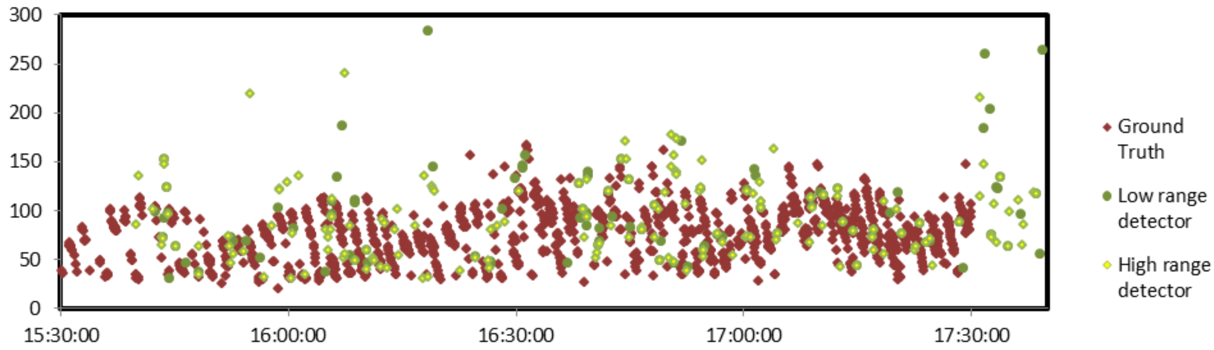

Figure 3: Bluetooth and Wi-Fi travel times overlapped with ground truth data

Analyzing the collected Wi-Fi and Bluetooth data primarily required the effective removal of outliers to prevent inaccurate travel time reports. Usually in roadways that experience low variability in travel times a simple upper and lower bound is sufficient to eliminate the outliers. However, in this study in order to accurately eliminate outliers, a three level outlier elimination algorithm was applied. The first filtering level involved eliminating records corresponding to irrelevant MAC addresses. This included MAC addresses with irregular formats such as: “FF:FF:FF:FF:FF:FF” or “10050” or “00:00,..”. The second level involved eliminating MAC addresses that were detected for unreasonable number of times. Considering that vehicles are moving objects that are exposed to detection for a limited amount of time, any data record with number of detections (i.e. hits) of more than 100 was eliminated. A more detailed look into the data revealed that a large portion of MAC addresses related to operating systems with randomization capability have an OUI (Organizationally Unique Identifier) of “05:00:00”. For the final level of filtering, a reputable outlier detection method proposed by Dion (2013), Transtar and Transguide algorithms were applied to the data. For outlier elimination of travel time data at arterial intersections several filtering algorithms have been developed and introduced. Transtar and Transguide filters determine the valid window range of the next dynamic or fixed window based on the mean and standard deviation of the travel times of the previous time interval. The Transguide algorithm requires two user defined parameter including the window size and a validity threshold. There is no fixed best value for these parameters and depending on the situation, market penetration rates and other factors their optimal value can vary. Figure 4 presents the effects of an 80% percent validity threshold compared to 60% threshold. As can be seen the 60 percent threshold fails to account for the sudden increase in travel times around 16:30:00 which the 80% thresholds manages to capture.

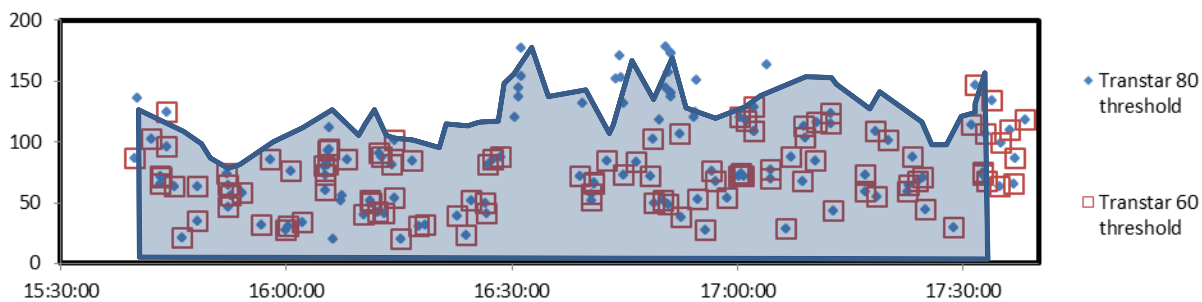

Figure 4: Visual representation of the effect of filtering threshold on the validity windows

A visual comparison is conducted to identify the most suitable threshold compared to ground-truth information. Visually, the Transguide algorithm creates the most suitable validity windows with thresholds from 60% to 80% as . Table… provides a quantitative analysis that compares the accuracy of the filtering methods with different thresholds compared to ground truth data using the APE (Absolute Percentage Error) measure defined in Equation 1.

[1] $$APE\% = \left| \frac{TT_{MAC} - TT_{GT}}{TT_{GT}} \right|$$

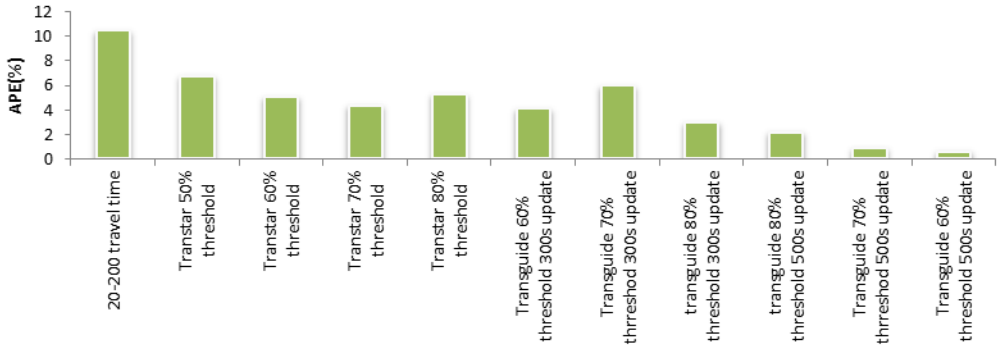

Figure 5: A comparison of the accuracy of different filtering methods compared to ground truth data

As can be seen from the results in Figure 4 either of the filtering methods with thresholds from 50% to 80% provide absolute percentage errors less than 10% for the two hour period. Moreover, for the Transguide algorithm, using 500s update interval provides more accurate results compared to the 300s update interval. Depending on the signal control parameters such as the cycle lengths and allocated green times, travel time variability may differ which significantly affects the suitable threshold that captures all the valid travel time readings in each time window.

3.1. Minimum Required Sample Size Analysis

Statistically it is possible to calculate the minimum number of samples required based on a pre-defined confidence interval and error. In this study by using the 15 minute standard deviation values of the ground truth travel times and Equation 1 the minimum required sample size are calculated. The coefficient of variation (CV) is obtained using a 15 minute moving window time frame of the ground truth data to represent the population standard deviation and average.

[2] $$n = \left[\frac{Z\alpha/_2 \; CV}{E}\right]^2$$

The available numbers of sample sizes are then calculated based on 5, 15 and 30 minute time intervals considering 10% error and alpha equal to 0.05. As can be seen in Figure 6a, using Bluetooth data alone does not provide sufficient sample size to estimate average travel times in 30 minute intervals. Wi-Fi data however, provides a larger sample size compared to Bluetooth data which reaches the minimum required sample size in almost all time periods under study. It is clear that the combination of Bluetooth and Wi-Fi travel times (from the medium range detector) provide the highest sample size. As shown in Figures 6b and 6c the combined travel times of Bluetooth and Wi-Fi data (from the medium range detector) does not provide the minimum sample size required in the considered 5 and 15 minute time intervals. For the short range detector, as shown in Figure 6d unlike the medium range detector, in the 30 minute time window the number of observed samples do not meet the minimum sample size requirement in a considerable portion of the study period. It is essential to mention that this analysis is only based on day of of travel time readings. To increase the sample size it is possible to aggregate the travel time information over several days in the same hours of the day which more or less resemble similar traffic conditions.

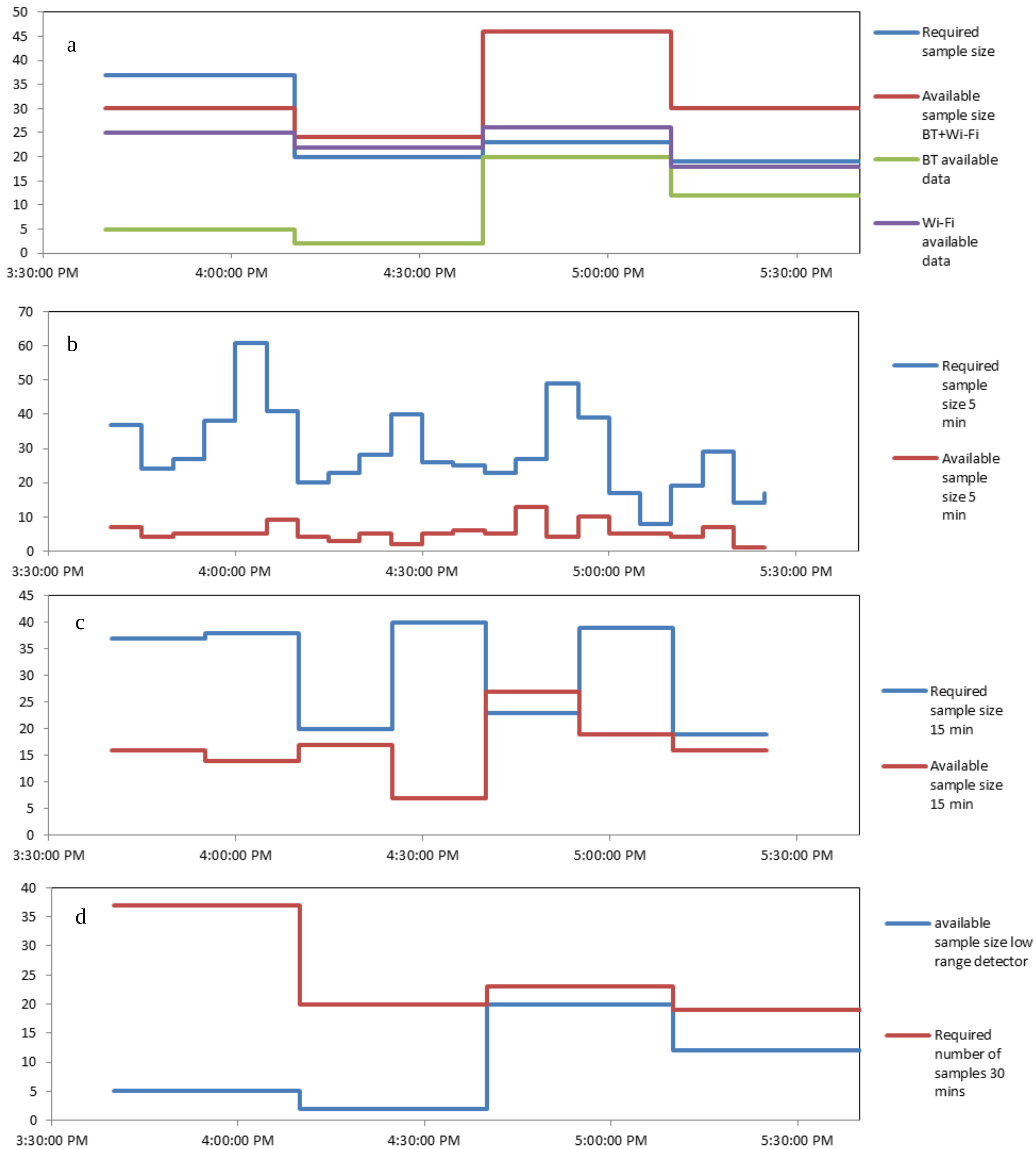

Figure 2: Minimum required versus the available number of travel time samples. a: 30 minute time interval sample size comparison for BT, Wi-Fi and hybrid BT/Wi-Fi travel times. b: 5 minute time interval sample size comparison for hybrid BT/Wi-Fi travel times. c: 15 minute time interval sample size comparison for hybrid BT/Wi-Fi travel times. d:30 minute sample size comparison for low range vs medium range detectors.

3.2 Travel Time Error Analysis

Statistically, using Equation 1 it is possible to calculate the expected error value given the number of samples available in each time interval. As shown in Figure 7 based on 30 minute intervals the expected sampling error is calculated and presented. Further, the errors for each 30 minute time window is then compared with the ground truth travel times for all first, last and median travel time values. As previously explained, depending on the circumstances a MAC address may be detected several times during the detection. The travel time values can then be calculated using either the first detections, last detections or the median detections. In Figure 7 the travel times obtained from these methods are compared with each other for the medium range detectors. As can be seen, the travel times obtained from the last detection provides the most accurate results followed by the first detection. For median detection, errors up to 30% were observed and therefore is not recommended.

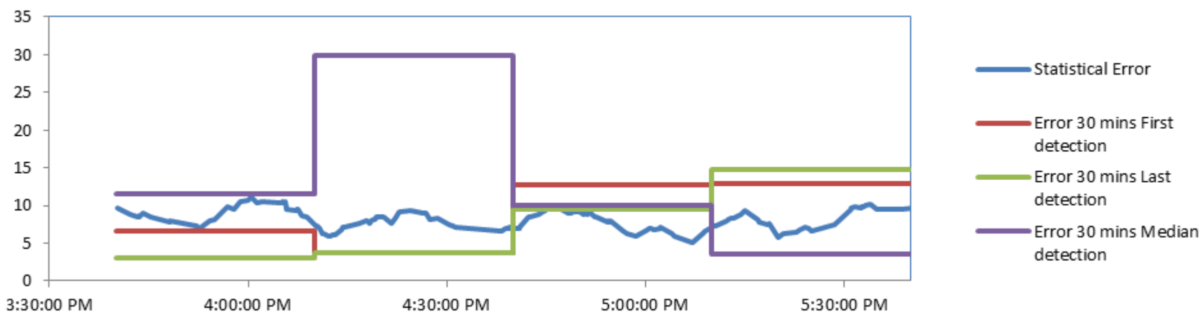

Figure 3: Error analysis in each 30 minute interval for first, last and median travel times

3.3 Market Penetration Evaluation

Market penetration rate provides an estimate of the portion of the actual traffic flow that was detected by the sensors. This requires knowing the actual flow of vehicles within the testing time period. In this study the traffic volume was obtained through video footage at the intersection and the market penetration rates were calculated by dividing the detected samples from the sensors by the counts obtained from video detection during each 3 cycles (120 second cycles) as shown in Figure 8. Overall, for each time interval, the hybrid Wi-Fi/Bluetooth method using the medium range detectors provide an average of 8% market penetration rate while Bluetooth alone was only able to capture less than 3% of the total traffic.

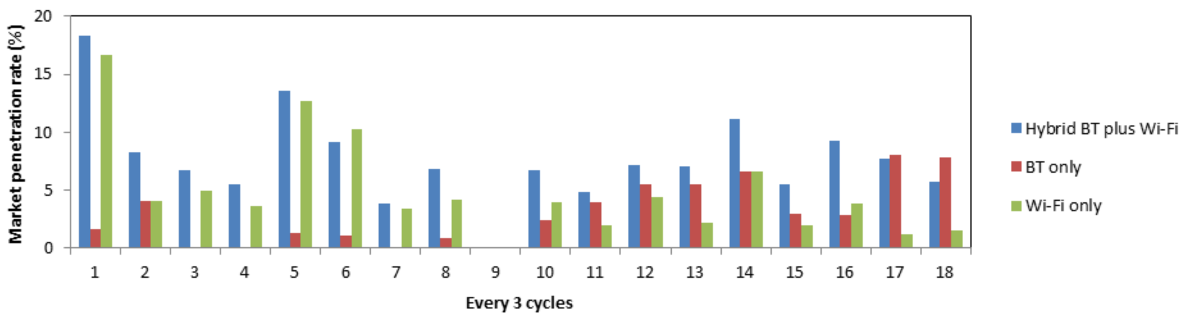

Figure 4: Comparison of market penetration rates for each 3 cycle intervals for Bluetooth, Wi-Fi and the hybrid method

4. CONCLUSION

This study aimed to evaluate the accuracy and market penetration rates of intersection travel times obtained from Wi-Fi and Bluetooth data. The results obtained in this study clearly showed that the combination of travel times obtained from Wi-Fi and Bluetooth data provide larger sample sizes compared to solely using Bluetooth data. Therefore, as expected from the sampling error analysis, the absolute percentage error compared to ground truth data is less than 10% for the combined dataset. This analysis showed that the combination of Bluetooth and Wi-Fi travel times resulted in around 8% market penetration rate as compared to Bluetooth data alone which showed less than 3% penetration rate. Statistically, regardless of the accuracy of the individual readings, lower sample size can cause higher sampling error in which the combination of Bluetooth and Wi-Fi data prevent from happening. For future research it is necessary to evaluate the sensor technologies in more densely located intersections.

REFERENCES

Abedi, Naeim, Ashish Bhaskar, and Edward Chung. "Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement." (2013).

Araghi, Bahar Namaki, et al. "Reliability of Bluetooth Technology for Travel Time Estimation." Journal of Intelligent Transportation Systems 19.3 (2015): 240-255.

Cunche, Mathieu. "I know your MAC Address: Targeted tracking of individual using Wi-Fi." Journal of Computer Virology and Hacking Techniques 10.4 (2014): 219-227.

Dion, F., and H. Rakha. Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. Transportation Research Part B, Vol. 40, August 2005, pp. 745-766.

Haghani and S. Young, "Bluetooth Traffic Monitoring - Technical Attributes and Applications," in North American Travel Monitoring Exposition and Conference, 2010.

Malinovskiy, Yegor, et al. "Field experiments on bluetooth-based travel time data collection." Transportation Research Board 89th Annual Meeting. No. 10-3134. 2010.

Musa, A. B. M., and Jakob Eriksson. "Tracking unmodified smartphones using wi-fi monitors." Proceedings of the 10th ACM conference on embedded network sensor systems. ACM, 2012

SMATS Traffic Solutions. TrafficTab Portable Wi-Fi and Bluetooth Scanner Manual. SMATS Traffic Solutions, Ottawa, Manual 2015.

Vo, T. An Investigation of Bluetooth Technology for Measuring Travel Times on Arterial Roads a Case Study on Spring Street. Georgia Institute of Technology, Atlanta, Master's Thesis 2011.

Wang, Yinhai, et al. "Error modeling and analysis for travel time data obtained from Bluetooth MAC address matching." Department of Civil and Environmental Engineering, University of Washington (2011).

Wieck, Michael. "Use of bluetooth based travel time information for traffic operations." 18th ITS World Congress. 2011.