

On the use of Bluetooth MAC Scanners for live reporting of the transport network

Ashish BHASKAR^a, Le Minh KIEU^b, Ming QU^c, Alfredo NANTES^d, Marc MISKA^e, Edward CHUNG^f

^{a-f} *Smart Transport Research Centre, Queensland University of Technology, Brisbane, QLD-4151, Australia*

^a *E-mail: ashish.bhaskar@qut.edu.au* [CORRESPONDING AUTHOR]

^b *E-mail: leminh.kieu@student.qut.edu.au*

^c *E-mail: ml.qu@qut.edu.au*

^d *E-mail: a.nantes@qut.edu.au*

^e *E-mail: marc.miska@qut.edu.au*

^f *E-mail: Edward.chung@qut.edu.au*

Abstract: Recently there has been significant interest of researchers and practitioners on the use of Bluetooth as a complementary transport data. However, literature is limited with the understanding of the Bluetooth MAC Scanner (BMS) based data acquisition process and the properties of the data being collected. This paper first provides an insight on the BMS data acquisition process. Thereafter, it discovers the interesting facts from analysis of the real BMS data from both motorway and arterial networks of Brisbane, Australia. The knowledge gained is helpful for researchers and practitioners to understand the BMS data being collected which is vital to the development of management and control algorithms using the data.

Keywords: Bluetooth, Traffic monitoring, Travel time, Travel profiles, Taxi Bluetooth data, Bus Bluetooth data

1 INTRODUCTION

Transport stakeholders are collecting traffic data from various sources, ranging from traditional loop detectors to state-of-the-art Bluetooth Media Access Control Scanner (BMS). In early 2000, researchers explored the use of Bluetooth (BT) technology for the automotive industry. Nusser and Plez (2000) presented the architecture of Bluetooth network as an integral part of in-car communication and information system. Researchers (Sawant et al., 2004, Murphy et al., 2002, Pasolini and Verdone, 2002) have tested the proof-of-concept for the use of BT for Intelligent Transport System services, and have verified that the BT equipped devices in moving vehicles could be discovered. It took almost a decade since its use was first explored, for large scale deployment of the Bluetooth technology for transport applications

The concept behind BMS is rather simple. BMS scans the Media Access Control Identifier (MAC-ID) of the *discoverable Bluetooth devices (BT)* within its communication range. We term this range as *zone*, which is generally around 100 m in radius and depends on factors such as BMS antennae characteristics. Most of the portable electronic devices such as mobile phones, car navigation systems, headphones, etc are equipped with BT and its usage is increasing. Installing time-synchronised BMSs on the road network has the potential to provide live reporting of the transportation of BT devices over the road network. Assuming the devices are transported by the vehicles, individual vehicle travel pattern can be easily obtained.

In literature, travel time from BMS data is compared with that from video camera's for motorways (Wang et al., 2011) and arterial (Mei et al., 2012) and promising results are

reported. For other studies (Haghani and Aliari, 2012) travel time obtained from traditional matching of BMS data is being considered as ground truth travel time. BT tracking is not only being explored for car travel times estimation, but also for other applications such as bicycle travel time (Mei et al., 2012), travel patterns of people movement in airports, shopping malls etc. (Bullock et al., 2010, Malinovskiy and Wang, 2012, O'Neill et al., 2006), work zone delays (Haseman et al., 2010), Origin-Destination estimation (Barceló et al., 2012, Blogg et al., 2010, Barceló et al., 2010), route choice analysis (Hainen et al., 2011, Carpenter et al., 2012), and freeway travel time variability (Martchouk et al., 2011).

The amount of data collected from a BMS station depends on numerous factors related to the penetration of BT in the traffic, software and hardware related with the BT protocol, etc. Researchers have experimented with the BMS antennae types, its position and number against the quantity of data collected. Difference in the quantity of data being collected by different antennae type has been reported. Porter et al., (2011) recommends vertically polarized antennas with gain between 9 to 12 dBi. Vo et al., (2012) and Click and Lloyd (2012) recommends to use of more than one BMS at the site can increase the quality of data collection.

As can be inferred from above, literature is mainly focusing on the applications of direct match of BT MAC-ID's at different BMSs, with limited understanding of the BMS data acquisition process and properties of the data being collected. This objective of this paper is to fill this gap. First, Bluetooth communication process is reviewed. Thereafter, data acquisition process and travel time from BMSs is modelled. Finally, the real data from Brisbane is analysed and paper is concluded.

2 BLUETOOTH COMMUNICATION AND DATA ACQUISITION PROCESS

2.1 Review of the Bluetooth communication process

Bluetooth is a short range communication protocol that operates in the unlicensed Industrial Scientific and Medical (ISM) band at 2.4 to 2.485 GHz. The Bluetooth devices are classified into three classes (see Table 1) that depends on the maximum radio frequency power output and corresponding communication zone. BMS is Class-1 type, and most of the portable devices, transported by the travelers (mobile phone, car navigation etc), are generally Class-2 type. High communication zone BMS is beneficial for transport application as it can provide sufficient time for BMS to discover the BT devices transported through the zone. BMS can only read the MAC-ID and there is no information about the spatial distance between BMS and the respective BT device. The device can be anywhere within the zone, resulting as the spatial error in the data being obtained by BMS.

Table 1: Bluetooth classification

Class	Zone radius (m)	Radio frequency power output (max)	Comment
Class-1	100	20 dBm	Primarily for industrial use
Class-2	10	4 dBm	Most commonly found in mobile phones, car navigation etc.
Class-3	1	0 dBm	For very short range devices such as keyboard, mouse, etc.,

A Bluetooth device (also termed as a Bluetooth unit) has two major states *standby* or *connection* state and seven substates. *Standby* implies no interaction with the other devices and *connection* implies that data is being transferred. The seven sub-states (modes) to

establish *connection* are: *inquiry*, *inquiry-scan*, *inquiry-response*, *page*, *page scan*, *slave-response* and *master-response*. Multiple devices can be connected, given one of them acts as a *Master* and the remaining as *Slaves*. The actual procedure for Bluetooth connection is complex but can be simply modelled as follows (refer Figure 1):

- i. The Master device has to be in *Inquiry* mode to enquire about the other devices (in *Inquiry-scan* mode) within the communication range by sending package containing its information (address and clock)
- ii. If the slave is in *Inquiry-scan* mode then it scans the inquiry sent by Master. Thereafter, Slave can switch to *Inquiry-response* mode to respond by sending its information (address and clock) for Master.
- iii. Master listens to the response from the slave(s) within its range and may switch to *Page* mode to page (hopping sequence and other information) the discovered slave device(s).
- iv. The Slave has to be in *Page-scan* mode to scan the page sent by Master, and may switch to *slave-response* mode to send its response (device access code).
- v. Finally, the Master has to be in *Master-response* mode to send further information to establish final connection between the two.

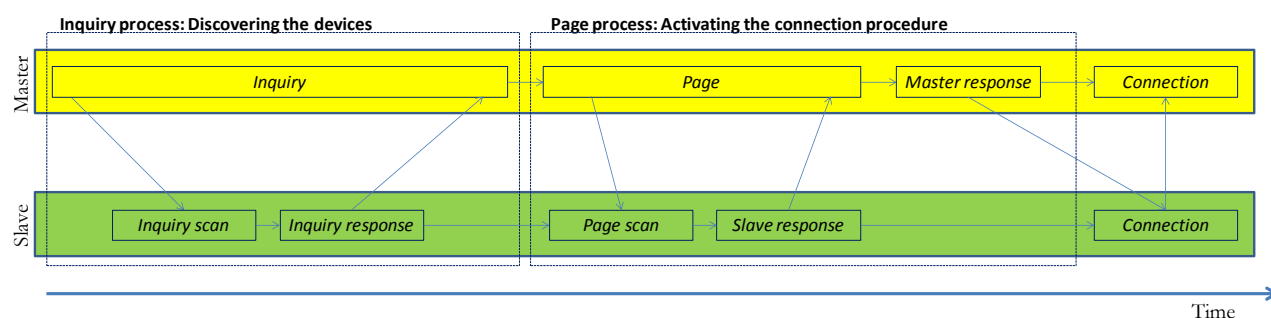


Figure 1: Simplified model for Bluetooth devices' connection procedure (Time axis not to scale; Inquiry process is more time and energy consuming than page process).

BMS is only interested in the inquiry process of discovering the devices where it (acting as a *Master*) should be able to acquire the MAC-IDs of the other devices (acting as a *Slave*) within its *zone*.

ISM band is shared by other wireless technologies (such as Wifi, Near Field Communication (NFC), cordless phone etc. Bluetooth communication performs *Frequency Hopping* to avoid interference between the wireless devices sharing ISM band. A device in a mode is transmitting and receiving information alternatively at a certain frequency defined for certain time slot, thereafter it hops to another frequency. Information exchange should be in the same frequency, i.e. if *Master* sends its inquiry at frequency k , only those *Slaves*, which at that particular time instance are scanning at the same frequency k , could scan this information. Moreover, in order to save power consumption, a unit in *inquiry-scan* mode only listens for a very short period of time (11.25 ms by default) and thereafter, enters standby mode for a longer period of time (1.28 seconds). This means, more than 99% of time the unit in *inquiry-scan* mode is not communicating. Hence, the discovery process (and connection process) is not instantaneous and requires time even in an ideal environment (where messages are not lost). Bluetooth protocol recommends a device to be in *inquiry* mode for 10.24 seconds (SIG, 2010).

2.2 Conceptual modelling of BMS data acquisition process

BMS is configured to be in continuous *inquiry* mode over a time period terms as *inquiry cycle* (C_I), where BMS are alternatively sending the enquiry messages and scanning the potential replies over the range of predetermined frequency channels. These *cycles* are repeated as a seamless train of *inquiry cycles* for uninterrupted discovery of the devices. Figure 2a conceptualizes the process by illustrating an inquiry train, where i^{th} inquire starts at time t_{i-1} and ends at time t_i and C_I is the time difference between the two. During an inquiry process, the device can be discovered at any time. Generally, the data acquisition software linked with BMS only provides the MAC-ID's scanned during an inquiry but not the exact time when it is discovered. All the discovered MAC ID's are linked to the time stamp corresponding to the end (or start) of the respective inquiry. This contributes to the temporal error in the data acquisition. For instance, Figure 2b represents an example where during an i^{th} inquiry, a device is discovered at time t ($t_{i-1} < t \leq t_i$) and is reported at time t_i . Temporal error in reporting the time when the device is discovered ($\beta = t_i - t$) is the difference of the time when the device is reported (t_i) and time when it is actually discovered (t).

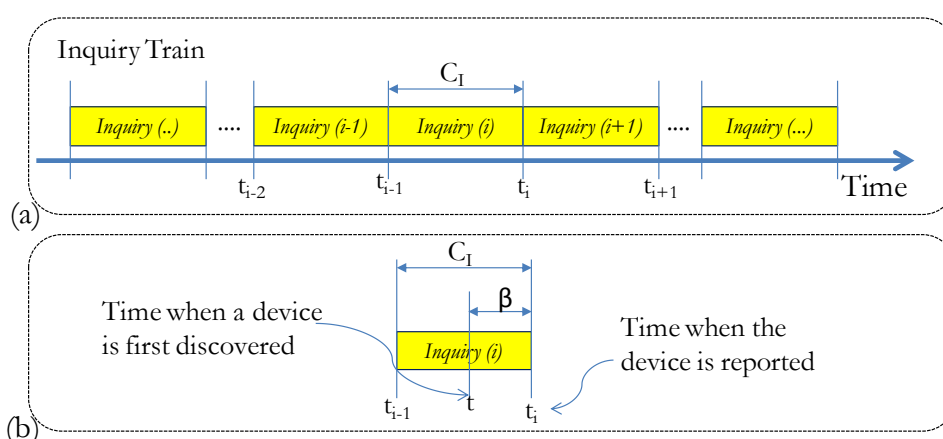


Figure 2: Illustration of an a) inquiry train and b) portion of the temporal error in data acquisition.

In order to reduce huge amount of data to be stored and transferred, data acquisition software generally reduces the data at the device level. For instance, in Brisbane, a record detected over consecutive *inquiry cycles* is stored only once with the timestamp corresponding to the time when it is first detected. Additionally, a field termed as *duration* is added to the data that corresponds to the time difference between the first and last detection (see Table 2). For instance, in Table 2, device ID 10 was detected for 6 consecutive *inquiry cycles*. Instead of storing the record for six times with different timestamps, the record is stored only once with duration of 120 seconds. Ideally, duration should be multiple of configured C_I but due to noise in C_I it may not be.

Table 2 Sample BMS data from BCC data, Brisbane Australia (Device ID is encrypted MAC ID of the BT device, the timestamp corresponds to the time when the device is first detection and duration is the time difference between the first and last identification of a BT device at a BMS location.

Number	Device ID	Intersection ID	Timestamp	Duration (seconds)
1	10	10087	2011/08/04 09:23:26	120
2	25	10087	2011/08/04 09:42:15	76
3	33	10087	2011/08/04 11:32:07	65

3 TRAVEL TIME MODELLING USING THE BMS DATA

Travel time estimation has long been topic of research. Especially on urban arterial it has been very challenging to estimate travel time. Models have been proposed to fuse data from mulitsources for estimating average travel time (Bhaskar et al., 2009, Bhaskar et al., 2010), movement specific travel time (Bhaskar et al., 2012) and travel time statistics (Bhaskar et al., 2011). BMS provides opportunities for individual vehicle travel time estimation that can be accumulated for estimation of average travel time and other statistics. Following are different ways we can model travel time using BMS data.

3.1 Travel time estimation models

BMS data is the data for an individual BT device discovered within the BMS *zone*. Considering the data from Brisbane as an example, following are the data fields:

- a) *Device-ID (m)*: MAC-ID or encrypted MAC-ID of the device discovered
- b) *BMS-Station ID (s)*: ID of the location where the BMS scanner is installed
- c) *Time stamp (t_{m,s})*: Time when the device *m* is first observed at *s*
- d) *Duration (d_{m,s})*: Time gap between the first and last observation of the device *m* at station *s*. This is also a proxy of the travel time of the device through the BMS *zone*.

Considering the above dataset, following three sections and travel time estimation models are defined:

3.1.1 Section En2En: Entrance to Entrance

Here the study section is defined from the entrance of the u/s BMS zone to entrance of the d/s BMS zone (see Figure 3). Travel time ($tt_{m,u/s,d/s}^{En2En}$), of the vehicle *m*, for this section is defined as:

$$tt_{m,u/s,d/s}^{En2En} = t_{m,d/s} - t_{m,u/s} \quad (1)$$

3.1.2 Section Ex2Ex: Exit to Exit

Here the study section is defined from the exit of the u/s BMS zone to exit of the d/s BMS zone (see Figure 3). Travel time ($tt_{m,u/s,d/s}^{Ex2Ex}$), of the vehicle *m*, for this section is defined as:

$$tt_{m,u/s,d/s}^{Ex2Ex} = (t_{m,d/s} + d_{m,d/s}) - (t_{m,u/s} + d_{m,u/s}) \quad (2)$$

3.1.3 Section P2P: Point to Point

Here the study section is defined from the point, *P_u*, within the u/s BMS zone to the point, *P_d*, within the d/s BMS zone (see Figure 3). Travel time ($tt_{m,u/s,d/s}^{P2P}$), of the vehicle *m*, for this section is defined as:

$$tt_{m,u/s,d/s}^{P2P} = (t_{m,d/s} + d_{m,d/s} - \Delta_{m,d/s}) - (t_{m,u/s} + d_{m,u/s} - \Delta_{m,u/s}) \quad (3)$$

Where $\Delta_{m,u/s}$ (or $\Delta_{m,d/s}$) is the time needed by the vehicle *m* to travel from point *P_u* (or *P_d*) to the exit of the u/s BMS zone (or d/s BMS zone). This can be expressed as a function of the duration of the vehicle within the respective zone. Refer to Bhaskar (2012) for details about the estimation of $\Delta_{m,u/s}$ (or $\Delta_{m,d/s}$) from duration.

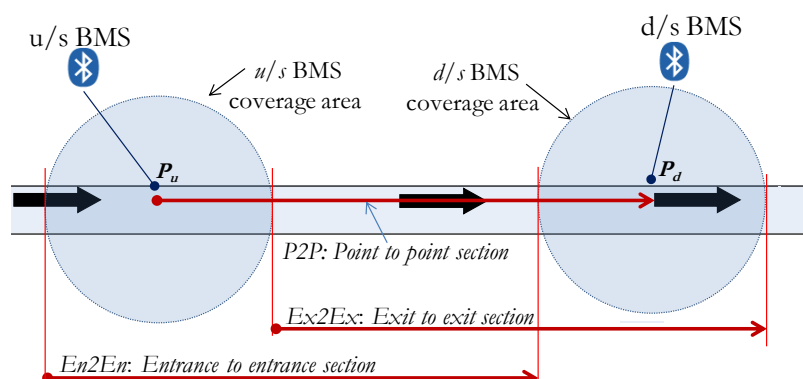


Figure 3: Systematic illustration of the three sections between upstream and downstream BMS zones

For arterial network, due to logistics reasons BMS is generally installed at the intersection, specifically in the signal controller box. For instance Figure 4 illustrates a photograph from one of the BMS stations on arterial network of Brisbane, Australia. The shark fin shape at the top is the antennae of the BMS which is installed in the controller. If BMS is installed at the intersection, then the above defined sections represents different travel time profiles on the networks as follows:

- a) *Section En2En*: The travel time of this section contains partial delay at the upstream and downstream intersections.
- b) *Section Ex2Ex*: The travel time of this section only contains delay at the downstream intersection, and no delay from upstream intersection, provided the further downstream intersection has not spilled over.
- c) *Section P2P*: Considering P_u and P_d at the stop-line of the upstream and downstream intersections, respectively, this section represents the travel time of the link defined between the two intersections.

For ITS applications, if one is interested to exploit the travel time from BMS as a feedback to the signal controller to optimise its parameter, then *Section Ex2Ex* or *Section P2P* should be considered.



Figure 4: Scanner antenna on top of the signal controlled box. The scanner is installed in the in the signal controller (Brisbane, Australia)

On the contrary, traffic on the motorway networks is not externally controlled by signals. Breakdown of traffic is mainly due to internal friction caused by vehicle to vehicle interactions. All the three sections should have similar travel time during freeflow and congested situation, except for situations (congestion buildup and dissipation periods) where the *zones* are partially congested. Nevertheless, while reporting the travel time from the BMS data it is important to specify how the sections are defined.

3.2 Noise in the individual vehicle travel time profile from BMS

The matched travel time data do contain noise due to reasons such as:

- a) *Unknown mode*: Obtained travel time is for the BT device transported by a traveller utilising any mode (car, bus, bicycle, pedestrian etc.) of transport. Different modes have different travel time depending on its operational and behavioural characteristics. If one is interested in car travel time, then presence of pedestrian or bicycle can result in unrealistic high travel time values and vice-versa. This issue is more dominate on the arterial networks.
- b) *No information outside the zone*: The estimate is only from the data available at zones. Hence the actual travel pattern of the vehicle between the zones is unknown. A vehicle can rest along the route or can take different route with significantly different travel time than that of the assumed route.
- c) *Multiple matches*: Especially on arterial network, a device can be observed at a zone and then it might take a detour, return to the same zone, and thereafter travel to the next zone. In such situation, device can be observed twice at first zone and only once at the second zone, resulting in two travel time values. Similarly, other combinations of multiple matches can occur resulting is noise.
- d) *Missed observation*: A BT device has a probability to be discovered at a zone and not all devices passing the zone are discovered. For instance, say a device travels twice between zones u and d . During its first trips the device was observed at u at time t_{u1} , however it was missed at d . During its second trip, it was observed at d at time t_{d2} but missed at u . Such observations will result in noisy travel time from u to d as $(t_{d2}-t_{u1})$. Similarly other combinations of observations can result in inaccurate travel time values.

In literature, filtering techniques such as, Moving Median Filter (Wang et al., 2011), Median Absolute Deviation (MAD) (Kieu et al., 2012), Box-and-Whisker filter (Tsubota et al., 2011) and other techniques utilising *Greenshield's model* and *least median of squares* (Van Boxel et al., 2011) and *multiple matched filter* (Kieu et al., 2012) have been utilised to reduce noise from the directly matched travel time values.

Moreover, recently an issue related to devices transmitting non-unique MAC-ID's has been observed. This is discussed further in section 5.5

4 ANALYSIS ON THE REAL DATA FROM MOTORWAY

4.1 Study site and data availability

The data used for this section is from 18 BMSs installed along 26 km long Gateway Motorway, Brisbane corridor between Nudgee and Mt. Gravatt (see Figure 5).

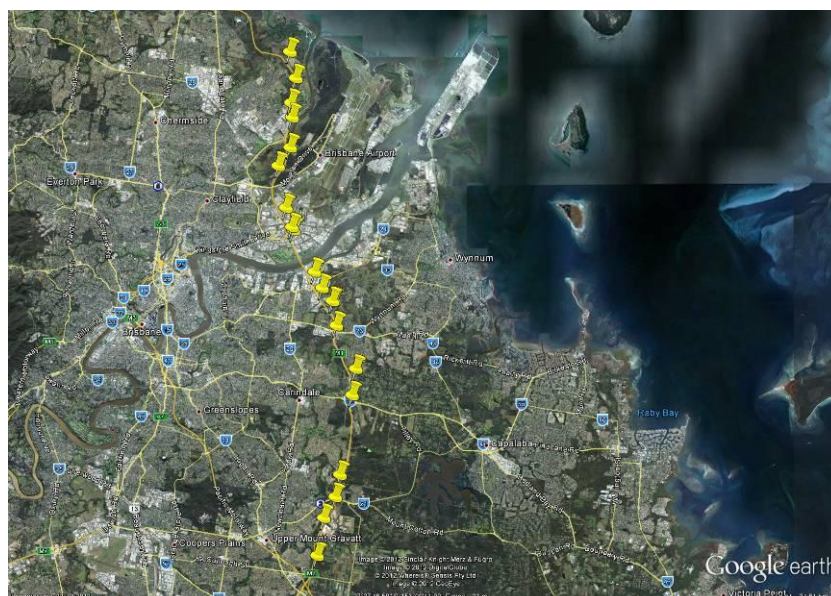


Figure 5 BMS scanners by TMR along the Gateway Motorway, Brisbane Australia

4.2 Types of Bluetooth devices

MAC-ID is a unique number stored in the hardware of the device. It is 48 bits long and is generally expressed as a sequence of twelve hexadecimal digits (six groups of two hexadecimal digits). For instance 00:22:CE:18:28:88 is a MAC-ID where six groups of two hexadecimal digits are separated by *colon*. The first six hexadecimal digits correspond to the manufacturer (or vendor or another organization's) unique identifier termed as Organizationally Unique Identifier (OUI). OUI is regulated by the standard organization. The last six hexadecimal digits correspond to the device's series number. For instance, 00:22:CE (first six digits of 00:22:CE:18:28:88) indicates the vendor of the device is *Cisco* (MAC-Vendor-Lookup).

Here, we map the first six digits of the MAC-ID with the available IEEE database (IEEE) of MAC-ID's and respective vendor/manufacturer to analyze the types of devices being extracted by the BMS. Figure 6 represents the pie-chart of the devices observed along the Gateway Motorway. Interestingly, TomTom and Garmin the car navigation systems only represent a small portion of 4.3% and 1.2%, respectively. Around 34% of the devices were registered with Nokia. Other mobile phones in decreasing order of observations are Samsung (9%), Sony (0.7%), LG (0.7%), Motorola (0.4%), and Apple (0.3%). The low capture rate of smart phone can be attributed to the fact that the devices have to be in the *discoverable* model to be detected by the BMS. Smart devices such as Apple I-phones, by default, are in *discoverable* mode only for 120 seconds, once discovery is initiated by the user. Hence, they have relatively low chances of being discovered by the BMS.

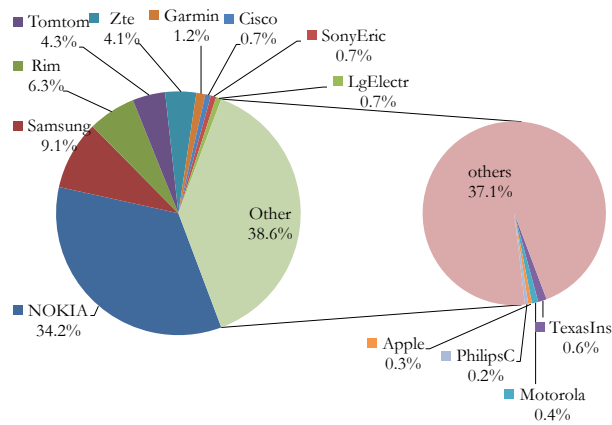


Figure 6: Pie chart of the different devices observed along the Gateway motorway

4.3 An insight on the matching between upstream and downstream

A Bluetooth device has a probability of observation at a BMS station. It may not be observed at each BMS locations it passes through. Figure 7 presents a sample of devices observed at 18 stations. Here rows and columns represent the MAC-ID and BMS locations respectively. The value in each cell is the time when the MAC-ID represented in the header (first) row is observed at the BMS location represented in the header (first) column. The green colour of the cell indicates the device is detected whereas, the black colour represents that the detection of the device is missed. White colour of the cell indicates no information can be inferred. For instance, MAC-3 (third row) is observed at BMS-1 to BMS-4; BMS-6 to BMS-9; BMS-12 and BMS-14. The traffic is on the motorway, so the device must have passed through BMS-5 and BMS-10; hence it is reported as missed. The device might have travelled further downstream but we cannot infer whether it was missed at BMS-15 to BMS-18. These observations clearly indicate that the capture of a Bluetooth device at BMS station is probabilistic in nature.

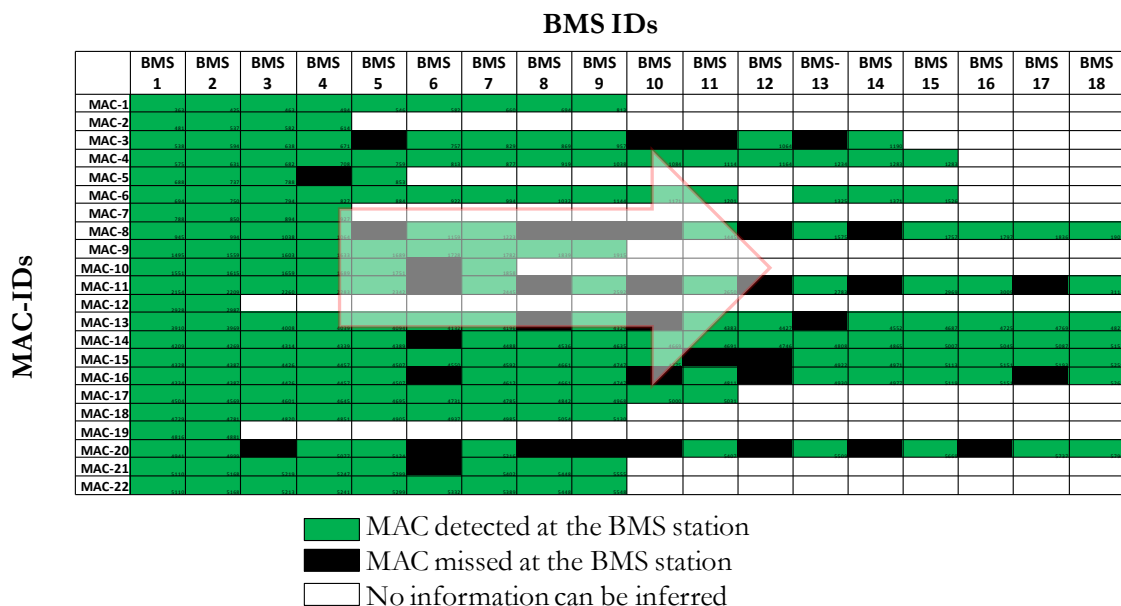


Figure 7: An overview of a sample of BT devices observed at respective BMS stations

4.4 Travel time profiles

Figure 8 represents a travel time profile for individual matches along a section of the Gateway

motorway. The peak and off peak periods can be easily differentiated. Interestingly, the individual vehicle travel time profile shows different layers of travel time values, the reasons for which are as follows:

- a) For the current section the scan cycle (C_1) of the BMS is 5 seconds and the BMS *zone* is 200 m in diameter. BMS provides the time when the vehicle is observed at the *zone* and not the exact time when the vehicle is detected at a point location. This leads to errors in reporting the time when the vehicle is observed at a respective location. Thus the reported travel time (time gap between the time when the vehicle is reported at upstream and downstream locations, respectively) will have errors, resulting in different layers of travel time values.
- b) Normally left lane (left hand driving) has slower traffic than right lane. Lane by lane difference in travel time will also contribute to the different layers of travel time values.

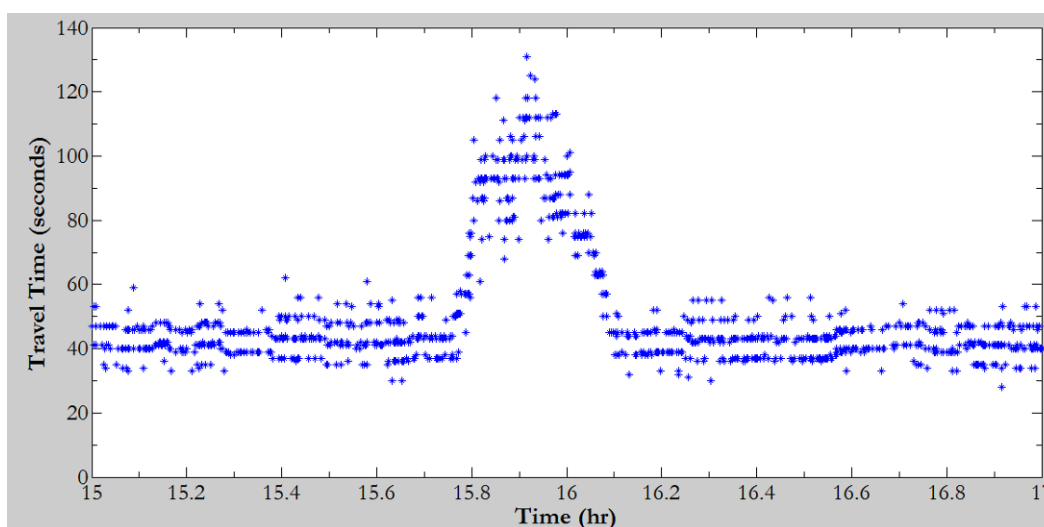


Figure 8: Travel time profile on a motorway

4.5 Percentage usable data

For travel time estimation, the BMS obtained from two stations are to be matched. As discussed in section 4.3, a BT equipped vehicle may not be discovered at both the BMS stations. Hence, we analyse the percentage of usable data as follows:

$$\% \text{ Usable Data} = \frac{(M12 + M21) * 2}{N1 + N2} * 100 \quad (4)$$

Where: $N1$ and $N2$ are the number of MAC observations at BMS-1 and BMS-2, respectively;
 $M12$: number of matches from BMS-1 to BMS-2; and
 $M21$: number of matches from BMS-2 to BMS-1

It is observed that along Gateway motorway % *Usable Data* ranges from 50% (BMS stations 5 km apart) to 80% (BMS stations 1.5 km apart). This depends on the proportion of traffic passing through both the stations.

5 ANALYSIS ON THE REAL DATA FROM ARTERIAL

5.1 Study site

The section analyses the data obtained from the BMS scanners installed on the Brisbane arterial network (see Figure 9). More than 150 scanners are installed in the signal controller boxes along the major intersections. Contrary to scanners on motorway, these BMS are tuned with C_I of 20 seconds.

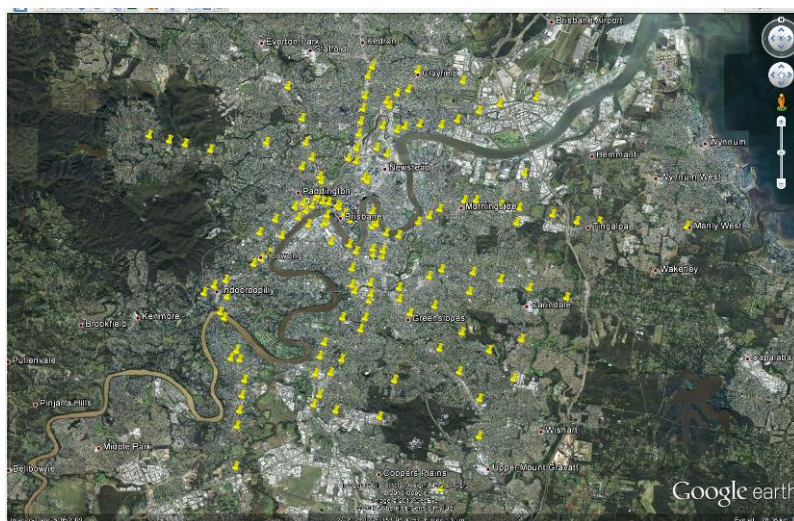


Figure 9 BMR stations on the arterial network of Brisbane, Australia

5.2 Travel time profiles

Here we present the travel time based on *Section Ex2Ex* (Refer to section 3.1.2) for a section along the Wynnum road, Brisbane. Figure 10 and Figure 11 presents the raw and cleansed individual vehicle travel time measured, respectively. Here, each column represents the day of the week (first column is Monday, last column is Sunday). For each sub-plot, X-axis is time (in hr) and Y-axis is travel time (in seconds). Green highlighted days are working days, whereas, red highlighted days are weekends or public holiday.

Figure 11 is obtained from the data of Figure 10 by applying a statistical filter, termed as Median Absolute Deviation (MAD) filter or Hampel identifier (Pearson, 2002). This filter removes outliers by comparing them with neighbor travel time observations within 10 minutes interval. For each minute, a window of 5 minutes before and 5 minutes after is considered, and this window is moved from the first to the last minute of the day. The outliers are identified if they are larger than the Upper Bound Value (UBV), or lower than a Lower Bound Value (LBV) of the current window as defined below.

$$UBV = median + \sigma f \quad (5)$$

$$LBV = median - \sigma f \quad (6)$$

Where σ is the standard deviation from the MAD, in which a normally distributed data can be approximated as $\sigma = 1.4826 * MAD$

$$MAD = median(|X_i - median(X_j)|) \quad (7)$$

The value of σf defines the scatter of data, where f is a scale factor which varies on a case by case basis. If f is small, the gap between UBV and LBV to the median value is small, and

vice versa. The value of f has been suggested by some authors to be from 1 to 5 (Davies and Gather, 1993, Pearson, 2002). For the current analysis, $f = 2$ is considered.

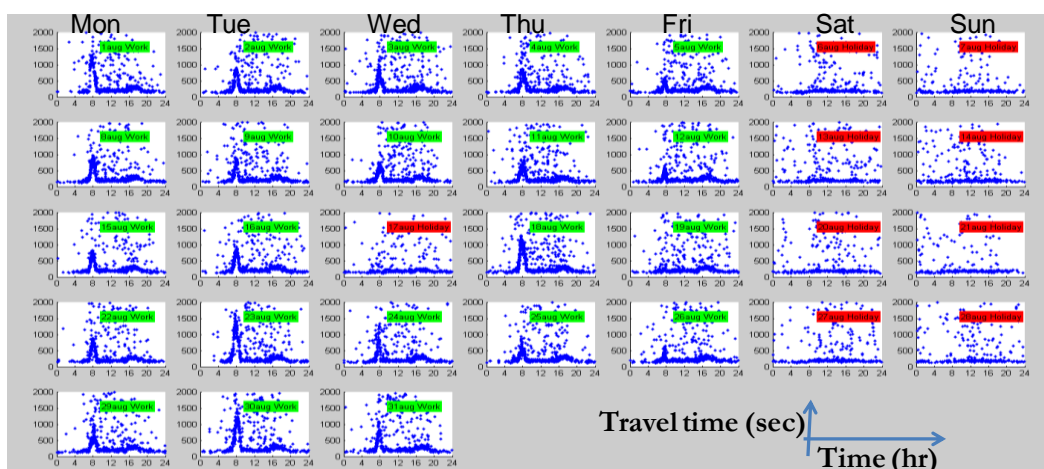


Figure 10: Raw individual vehicle travel time profile for a month along two BMS stations on Brisbane arterial network

Figure 11 provides a month snapshot of the travel time profiles along the study section. Weekday traffic is different from those of weekends and holidays. 16th August 2012 was a public holiday, the profile for that Wednesday is different from other Wednesdays of the month. Travel time profile for Friday is very different from that of Monday. These results clearly indicate that the BMS data has the potential to provide travel time profiles over the road network.

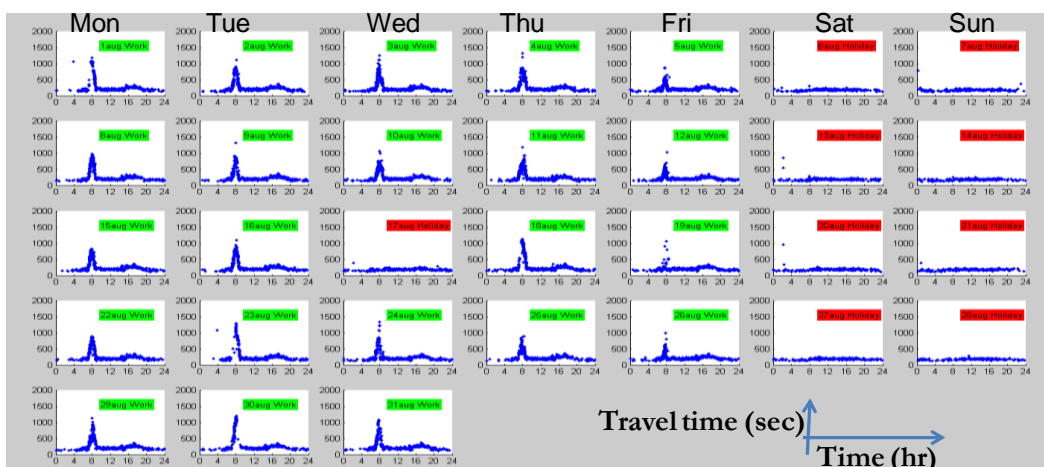


Figure 11 Cleansed individual vehicle travel time profile for a month along two BMS stations on Brisbane arterial network

5.3 Frequency of the travel time data points from BMS data

Figure 12 provides snapshot of the number of BT travel time points (after filtering) per minute during the morning peak periods (7 am to 9 am) of August 2012. It is observed that average number of BT travel time points vary from 1.23 to 3.6 BT points per minute during the working days. There are periods when no BT travel time point is available, for instance during 1st August around 7:20 to 7:40. Algorithms need to be developed to fill this gap.

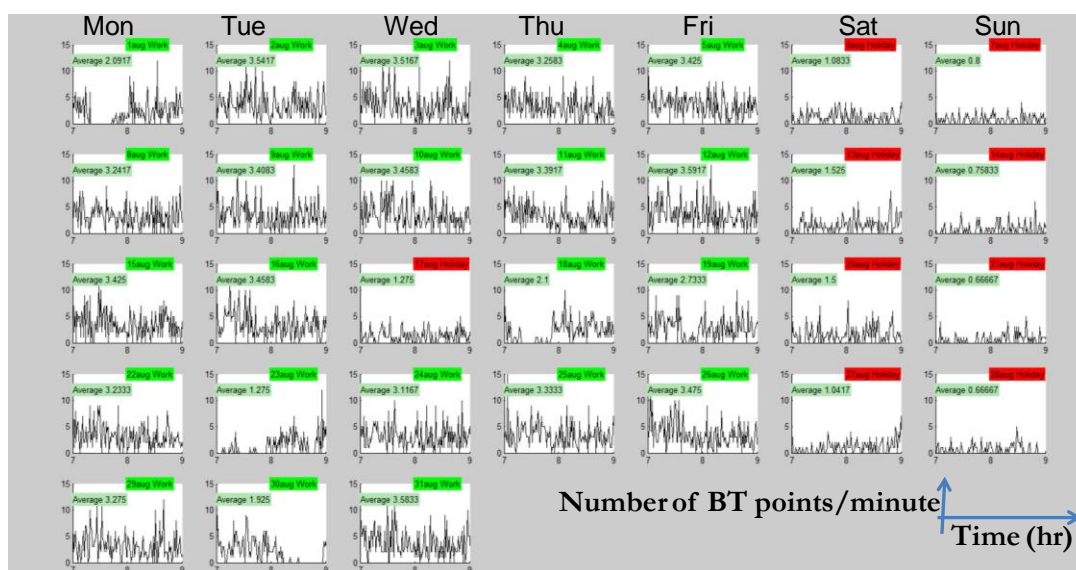


Figure 12: Number of BT point per minutes during the peak periods

5.4 Is bus overrepresented in the BMS database?

One of the concerns about the use of BMS data, especially from urban arterial, is the bias in the travel time estimates from multiple BT devices being transported by a vehicle. For instance, if a bus is transporting 20 passengers with BT equipped mobile phones, then the discovery of these mobile phones by BMS will be considered as 20 different vehicles, and average travel time along the corridor from the BMS data will be biased with the travel time from the bus. Here at Brisbane, we had opportunity to integrate Bus Vehicle Identification (VID) system with BMS network to explore such bias, if any.

VID is used to provide priority to the buses at the signalised intersections. It consists of a set of Radio Frequency Identification (RFID) sensors installed at upstream, at stopline and at downstream of the intersection where transit signal priority is to be provided. These sensors detect the presence of a bus by reading the RFID tag on the bus. Each bus is provided a unique tag and the system stores the time when the bus is detected at the VID sensor location. Matching the VID data at different VID sensor locations we obtain the individual bus travel time between intersections.

Overlaying the VID detector map over the BMS stations maps we identify pair of intersections where both VID and BMS data is available. For the current analysis we present the results from Wynnum Road, Brisbane. We estimate Bluetooth and bus travel time independently from BMS data and VID respectively. Appropriate filters are applied to filter the travel time profiles (Kieu et al., 2012). Thereafter, we integrate the bus travel time profile with the Bluetooth travel time profile. For instance Figure 13 presents a graph where travel time profiles from VID are overlaid over the travel time profiles from BMS, here blue dots represents travel time from BMS and black stars represents bus travel time from VID.

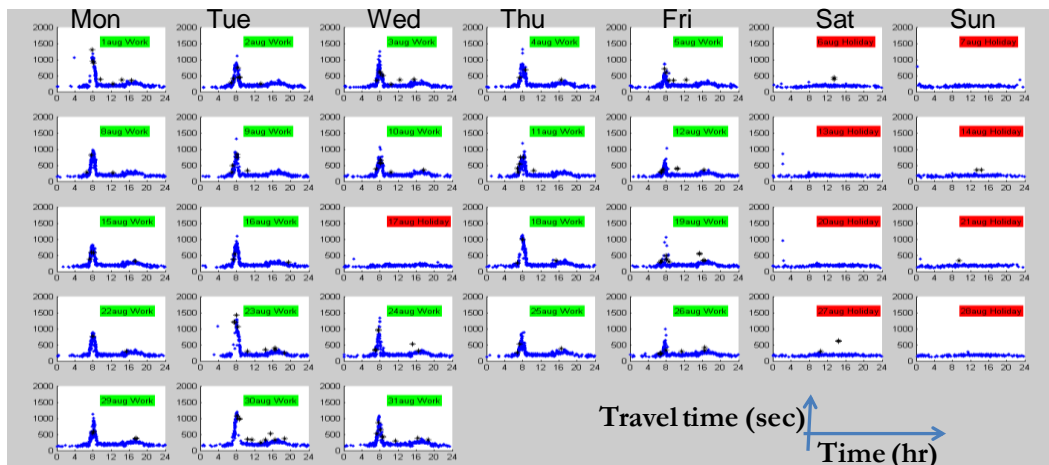


Figure 13 Results from the integration of BMS and VID dataset. Blue dots: Cleansed individual vehicle travel time profile for a month along two BMS stations on Brisbane arterial network; Black star: Bus travel time profile from VID data

If a bus is represented by multiple BT devices in BMS dataset, then matching these profiles we should observe multiple BT travel time points close to the bus travel time point. For each bus travel time point, we look at ± 1 minute window (see Figure 14), and count the number of BMS data points. These data points can be considered to be from the bus. The algorithm for this is as follows: Say for a given day and corridor, BMS dataset is represented as a list of time (t_{BMS}) and corresponding Bluetooth travel time (TT_{BMS}) values. Similarly, the VID dataset is represented as a list of time (t_{VID}) and corresponding bus travel time (TT_{VID}) values. Then for each data in VID dataset ($t_{VID}(j)$, $TT_{VID}(j)$) we look at the number of samples in BMS dataset satisfying the following conditions:

$$\forall i \in BMSdataset$$

$$t_{BMS}(i) \in [t_{VID}(j) - 1 \text{ min}, t_{VID}(j) + 1 \text{ min}] \quad (8)$$

and

$$TT_{BMS}(i) \in [TT_{VID}(j) - 1 \text{ min}, TT_{VID}(j) + 1 \text{ min}]$$

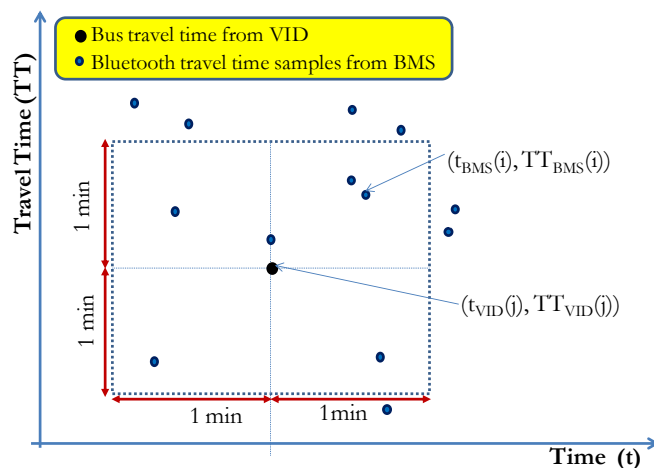


Figure 14 An example of a search window for the potential BMS travel time data points from a bus

Figure 15 presents the empirical cumulative probability of the number of BMS BT travel time points within the search window of each VID bus travel time point, obtained from the aforementioned analysis over six months of the data. It is observed that:

- a) The probability of a bus represented as more than one vehicle in the BMS database is between 25% to 50%
- b) The probability of a bus represented as more than three vehicles in the BMS database is less than 20%
- c) It is rare a bus is represented by more than six vehicles.

This analysis indicates that the bias from the bus in BMS travel time profiles is very low.

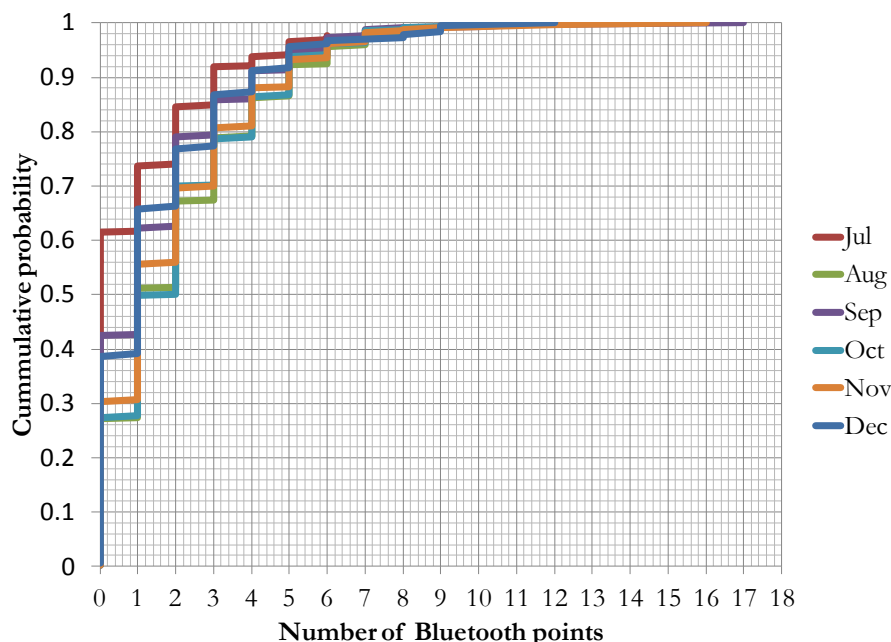


Figure 15 Empirical cumulative probability of number of BT samples from a bus (six months of data from Brisbane, Australia)

5.5 A discussion on the uniqueness of MAC ID

Ideally MAC address should be unique, but it can be *cloned* (Cherchali et al., 2008). This is not a normal practice, but it is observed that some Bluetooth devices carried by taxi fleet have devices which are *cloned* with the fleet operator requirements. The availability of MAC-ID from these *cloned* devices can result in unambiguous results from the Bluetooth MAC matching. To demonstrate this, we present the results from the analysis of one month (January 2012) of data from Brisbane. Here, similar MAC addresses observed at different BMS locations (more than 10km in Euclidean distance) within a given time window of 1 minute and 5 minute are filtered. On urban arterials it is not possible for a vehicle to travel more than 10km in 5 minutes (approximately 120 km/h in space mean speed) or 1 minute (approximately speed of 600 km/h). Hence these similar observations should be from *cloned* Bluetooth devices. Figure 16 illustrates an example of a BT device ID 10755 (encrypted) is observed at seven BMS stations within a short time window of 5 minutes. The geo-locations of these seven stations is presented on Google earth map (Figure 16a). Figure 16b illustrates the respective air-trajectory of the BMS device from one location to other, where the number in the figure represents the order in time in which the device is observed at the respective BMS station. One can clearly see that it is impossible for the device to travel through these BMS within 5 minutes.

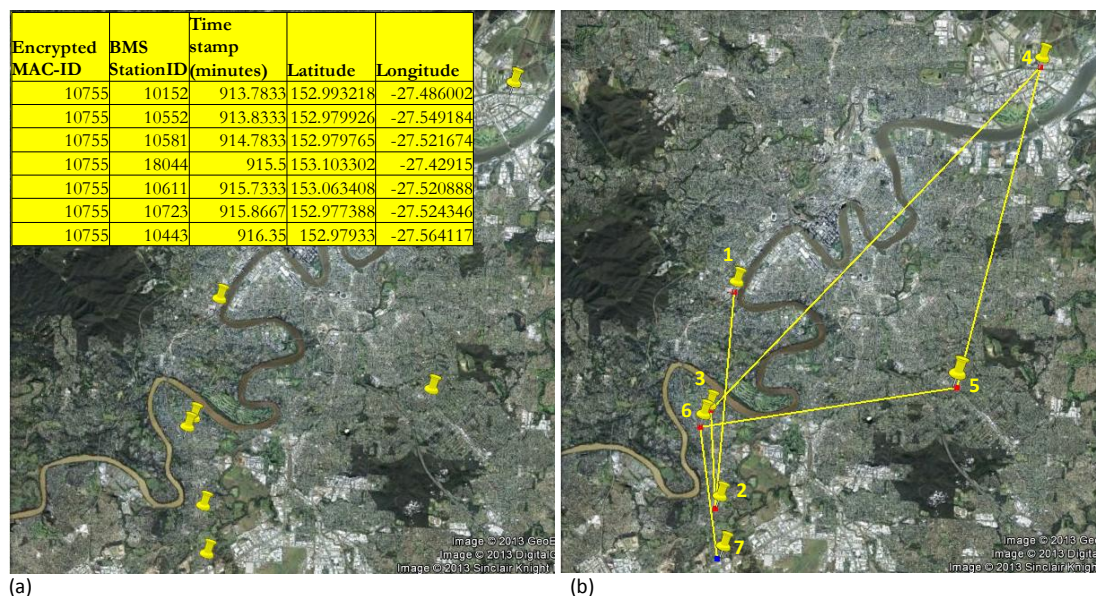


Figure 16 An example of a cloned BT device observed at different locations on the network within a very small time period

The Figure 17 presents the results for the number of “duplicated” MAC-ID and their proportion to the total number of BMS data points within the same day. Here, X-axis is the day of the month (January 2012); primary Y-axis is the number of times similar MAC-ID’s are observed within a given time window (5 minutes (blue solid line) and 1 minute (red solid line)) at two BMS locations far apart (more than 10 kms in Euclidean distance), we terms this as “duplications per day”; and secondary Y-axis is the proportion of such duplications to the total daily MAC-ID observations (5 minutes: Blue dotted line; and 1 minute: Red dotted line) from all the BMS stations. It can be seen that such observations are quite low with probability of occurrence less than 0.025%. 18th January 2012, had highest duplication for the month. Although the number of duplication is 39 (see Figure 17 table on the right), but they are represented by only 7 unique MAC-ID’s.

From the above analysis we can conclude that currently the *cloning* of MAC-IDs is not a big issue for the use of BMS data for traffic monitoring. The percentage of such observations is negligible compared to the massive data collected by BMSs. Unrealistic high or low travel time values should be identified as an outlier by any standard filtering algorithm. Moreover, one can think of analyzing the historical database for identification of MAC-IDs which are potentially *cloned* (using the aforementioned procedure) and “black list” them for future applications.

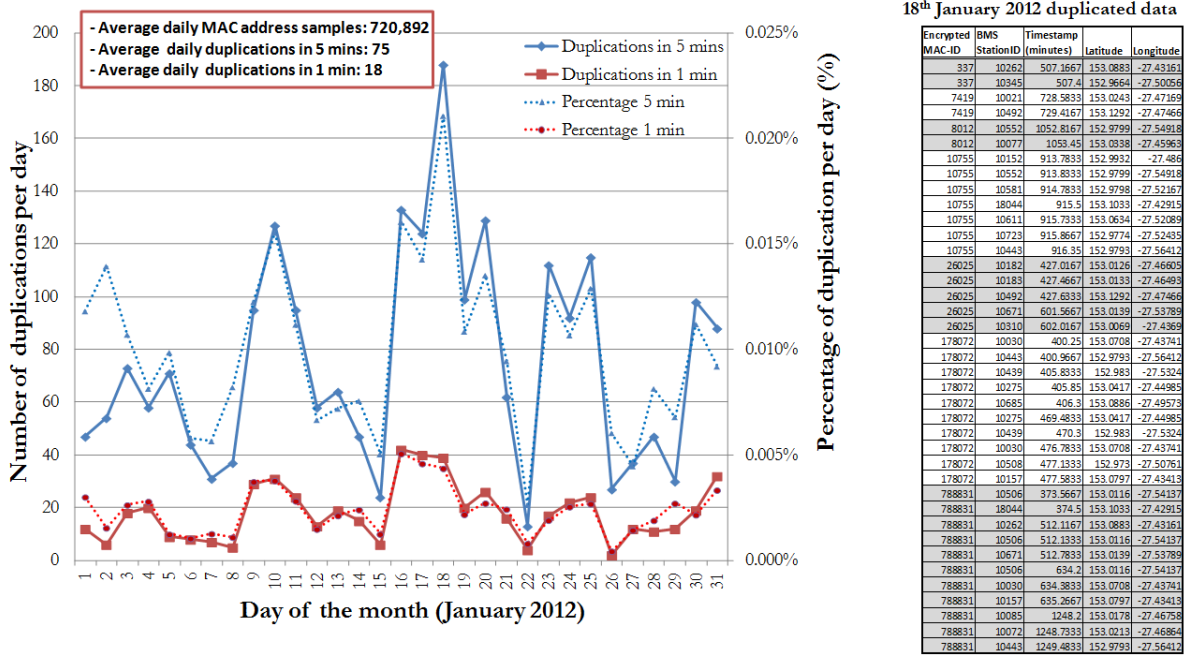


Figure 17 Empirical probability of number of BT samples from a bus (six months of data from Brisbane, Australia)

6 CONCLUSION

This paper contributes to the increase in understanding of the Bluetooth protocol and its application in transport engineering. Conceptual model for the Bluetooth communication process is reviewed and the data acquisition process of the Bluetooth MAC scanner (BMS) is modelled. Theoretical models for travel time estimation from BMS, especially on urban arterial, are proposed. The sources of errors from the BMS data are also identified. The conceptual discussions and modelling is backed with the analysing of the real Bluetooth data from both motorway and arterial networks from Brisbane, Australia which discovers some interesting facts about the use of BMS data. It is concluded that:

- While reporting the travel time from BMS, the definition of section (*Exit to Exit, Entrance to Entrance, Point to Point*) should be clearly mentioned, especially for arterial networks where different sections will have different travel time values.
- The capture of a BT device passing through the BMS zone is probabilistic in nature. Smart phones have relatively low chances of being detected by BMS.
- BMS has the potential to provide an excellent overview of the traffic profiles over the BMS networks. It can provide a snapshot of the daily traffic conditions where congested and uncongested conditions can be clearly identified. However, looking into details, the average number of travel time data points from BMS is not very high: during peak periods we only observe around 1 to 4 points per minute, and there are periods when there is no data point. Although this number is high compared to the other data sources such as probe vehicles, but might not high enough for real time applications (such as real time feedback to the signal controller) on arterial networks. On arterial network, there is significant variability of individual vehicle travel time that depends on the delay of the vehicle at the signalised intersection (a vehicle hitting green has lower delay than the one which has to stop over the red phase) which increases the requirement for larger samples for accurate representation of actual travel time (Bhaskar et al., 2011, Bhaskar et al., 2009).
- Generally practitioner and researcher do wonder how buses are represented in the

BMS data as vehicles carrying multiple Bluetooth devices should be overrepresented. Our analysis on the integration of the bus VID and Bluetooth BMS data shows that the empirical probability of bus represented (in BMS dataset) as:

- i. Less than or equal to 1 vehicle is between 30%-60%
- ii. Less than or equal to 2 vehicles is between 50%-75%
- iii. Less than or equal to 3 vehicles is between 70%-85%
- iv. Less than or equal to 4 vehicles is between 80%-90%
- v. Less than or equal to 5 vehicles is between 85%-95%
- vi. More than 5 vehicles is less than 5%

This indicates that buses do are overrepresented in the BMS dataset and it is rare to have overrepresentation by more than 5 vehicles.

- e) Few instances are reported where a BT device is observed at multiple BMS stations within a very short period. This is due to the *cloning* of the BT devices. The analysis shows that this is not quite frequent. The findings do answer some of the unambiguous observed travel patterns observed from the BMS data. However, it does raise a question, what if in future the use non-unique MAC-ID increases. If so, it can have significant impact on the use of BMS data.

ACKNOWLEDGEMENTS

This research contributes to the *Integrated Traveller Information* research domain at Smart Transport Research Centre, Queensland University of Technology, Brisbane, Australia. We acknowledge Brisbane City Council (BCC) and Department of Transport and Main Roads (TMR), Brisbane, Australia for providing the data used for this research.

REFERENCES

- BARCELÓ, J., MONTERO, L., BULLEJOS, M., SERCH, O. & CARMONA, C. A Kalman Filter Approach for the Estimation of Time Dependent OD Matrices Exploiting Bluetooth Traffic Data Collection. Transportation Research Board 91st Annual Meeting 2012.
- BARCELÓ, J., MONTERO, L., MARQUÉS, L. & CARMONA, C. 2010. Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. *Transportation Research Record: Journal of the Transportation Research Board*, 2175, 19-27.
- BHASKAR, A. 2012. Fundamental understanding on the use of Bluetooth MAC Scanner as a complementary transport data source. Brisbane, Australia: Smart Transport Research Centre, Deliverable 8, Real time and predictive traveller information project.
- BHASKAR, A., CHUNG, E. & DUMONT, A.-G. 2009. Estimation of Travel Time on Urban Networks with Midlink Sources and Sinks. *Transportation Research Record: Journal of the Transportation Research Board*, 2121, 41-54.
- BHASKAR, A., CHUNG, E. & DUMONT, A.-G. 2010. Analysis for the Use of Cumulative Plots for Travel Time Estimation on Signalized Network. *International Journal of Intelligent Transportation Systems Research*, 8, 151-163.
- BHASKAR, A., CHUNG, E. & DUMONT, A.-G. 2011. Fusing Loop Detector and Probe Vehicle Data to Estimate Travel Time Statistics on Signalized Urban Networks. *Computer-Aided Civil and Infrastructure Engineering*, 26, 433-450.
- BHASKAR, A., CHUNG, E. & DUMONT, A.-G. 2012. Average Travel Time Estimations for Urban Routes That Consider Exit Turning Movements. *Transportation Research Record: Journal of the Transportation Research Board*, 2308, 47-60.

- BLOGG, M., SEMLER, C., HINGORANI, M. & TROUTBECK, R. Travel time and origin-destination data collection using Bluetooth MAC address readers. 33rd Australasian Transport Research Forum, 2010 Canberra, Australia.
- BULLOCK, D. M., HASEMAN, R., WASSON, J. S. & SPITLER, R. Anonymous Bluetooth Probes for Airport Security Line Service Time Measurement: The Indianapolis Pilot Deployment. 89th Annual Meeting in Transportation Research Board, 2010 Washington, DC, USA.
- CARPENTER, C., FOWLER, M. & ADLER, T. J. 2012. Generating Route Specific Origin-Destination Tables Using Bluetooth Technology. *Transportation Research Board 91st Annual Meeting*
- CHERCHALI, A., GUDELIS, M. J., LESTER, W. G. & MCLAUGHLIN, R. J. 2008. *Technique for automated MAC address cloning*. US patent application 11/033,930.
- CLICK, S. M. & LLOYD, T. Applicability of Bluetooth Data Collection Methods for Collecting Traffic Operations Data on Rural Freeways. Transportation Research Board 91st Annual Meeting 2012.
- DAVIES, L. & GATHER, U. 1993. The Identification of Multiple Outliers. *Journal of the American Statistical Association*, 88, 782-792.
- HAGHANI, A. & ALIARI, Y. 2012. Using Bluetooth sensor data for ground-truth testing of reported travel times. *Transportation Research Board 91st Annual Meeting*. Washington D.C., USA.
- HAINEN, A., WASSON, J., HUBBARD, S., REMIAS, S., FARNSWORTH, G. & BULLOCK, D. 2011. Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 2256, 43-50.
- HASEMAN, R., WASSON, J. & BULLOCK, D. 2010. Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking. *Transportation Research Record: Journal of the Transportation Research Board*, 2169, 40-53.
- IEEE. <http://standards.ieee.org/develop/regauth/oui/public.html> [Online]. January 2013].
- KIEU, L. M., BHASKAR, A. & CHUNG, E. 2012. Bus and car travel time on urban networks: Integrating Bluetooth and Bus Vehicle Identification Data. *25th Australian Road Research Board Conference*. Perth, Australia.
- MAC-VENDOR-LOOKUP. <http://www.macvendorlookup.com/> [Online]. February 2013].
- MALINOVSKIY, Y. & WANG, Y. Pedestrian Travel Pattern Discovery Using Mobile Bluetooth Sensors. Transportation Research Board 91st Annual Meeting 2012.
- MARTCHOUK, M., MANNERING, F. & BULLOCK, D. 2011. Analysis of Freeway Travel Time Variability Using Bluetooth Detection. *Journal of Transportation Engineering*, 137, 697-704.
- MEI, Z., WANG, D. & CHEN, J. 2012. Investigation with Bluetooth Sensors of Bicycle Travel Time Estimation on a Short Corridor. *International Journal of Distributed Sensor Networks*, 2012.
- MURPHY, P., WELSH, E. & FRANTZ, J. P. Using Bluetooth for Short-Term Ad-Hoc Connections Between Moving Vehicles: A Feasibility Study. IEEE Vehicular Technology Conference 2002 2002 Birmingham, AL. 414-418.
- NUSSER, R. & PELZ, R. M. Bluetooth-based wireless connectivity in an automotive environment. Vehicular Technology Conference, 2000. IEEE VTS-Fall VTC 2000. 52nd, 2000 2000. 1935-1942 vol.4.
- O'NEILL, E., KOSTAKOS, V., KINDBERG, T., SCHIEK, A., PENN, A., FRASER, D. & JONES, T. 2006. Instrumenting the City: Developing Methods for Observing and

- Understanding the Digital Cityscape *In*: DOURISH, P. & FRIDAY, A. (eds.). Springer Berlin / Heidelberg.
- PASOLINI, G. & VERDONE, R. Bluetooth for ITS? Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on, 27-30 Oct. 2002 2002. 315-319 vol.1.
- PEARSON, R. K. 2002. Outliers in process modeling and identification. *Control Systems Technology, IEEE Transactions on*, 10, 55-63.
- PORTER, D. J., KIM, D. S., MAGAÑA, M. E., POOCHAROEN, P. & ARRIAGA, C. A. G. 2011. Antenna characterization for bluetooth-based travel time data collection. *Transportation Research Board 90th Annual Meeting*
- SAWANT, H., JINDONG, T., QINGYAN, Y. & QIZHI, W. Using Bluetooth and sensor networks for intelligent transportation systems. Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on, 3-6 Oct. 2004 2004. 767-772.
- SIG. 2010. *Specification of Bluetooth system* [Online]. Available: <http://www.bluetooth.com/pages/smart-logos-manufacturers.aspx> [Accessed December 2012].
- TSUBOTA, T., BHASKAR, A., CHUNG, E. & BILLOT, R. 2011. Arterial traffic congestion analysis using Bluetooth duration data. *Australasian Transport Research Forum 2011*.
- VAN BOXEL, D., SCHNEIDER, W. & BAKULA, C. 2011. Innovative Real-Time Methodology for Detecting Travel Time Outliers on Interstate Highways and Urban Arterials. *Transportation Research Record: Journal of the Transportation Research Board*, 2256, 60-67.
- VO, T., SUH, W., GUENSLER, R., GUIN, A., HUNTER, M. P. & RODGERS, M. O. 2012. Assessment of multi-antenna array performance for detecting bluetooth enabled 2 devices in a traffic stream. *Transportation Research Board 91st Annual Meeting*
- WANG, Y., MALINOVSKIY, Y., WU, Y.-J. & LEE, U. K. 2011. Error Modeling and Analysis for Travel Time Data Obtained from Bluetooth MAC Address Matching. Department of Civil and Environmental Engineering University of Washington.