Pedestrian Travel Analysis Using Static Bluetooth Sensors

Yegor Malinovskiy (Corresponding Author)
Graduate Research Assistant
Box 352700
Department of Civil and Environmental Engineering
University of Washington
Seattle, WA 98195
Tel: (206) 543-7827
Email: yegorm@uw.edu

Nicolas Saunier
Assistant Professor
Department of civil, geological and mining engineering
École Polytechnique de Montréal
C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7
Tel: +1 (514) 340-4711 ext. 4962
Email: nicolas.saunier@polymtl.ca

Yinhai Wang
Professor
Box 352700
Department of Civil and Environmental Engineering
University of Washington
Seattle, WA 98195-2700
Tel: (206) 616-2696
Fax: (206) 543-5965
Email: yinhai@uw.edu

Word count: 4898 + 250 (12 figures) = 7898 words

Submitted for Presentation and Publication in the 91st Annual Transportation Research Board Meeting in Washington D.C., January 2012.

Submitted on August 1st, 2011
ABSTRACT

Travel evaluation metrics have been historically biased towards motorized modes, which dominate land transportation choices and are partially responsible for numerous environmental and health issues facing our society today. Encouraging active travel solutions is seen as a means of improving sustainability, health and cohesiveness of a community. Unfortunately, information regarding volume, trip origin and destination, travel time and personal interactions is difficult to obtain due to a lack of sensor infrastructure and unrestricted movement of these modes. Therefore, information is often limited to annual surveys and model estimates which are insufficient to address the increasing needs of sustainable planning and large scale behavior studies. An automated, cost-effective approach to acquiring pedestrian data is desirable. The emergence of Bluetooth sensors as a means of gathering travel time data for traffic analysis presents an opportunity to use the same technology for pedestrian travel analysis. However, people generally carry more Bluetooth devices in their vehicles than on their person, making representative sample sizes a challenge. A study of pedestrian detection using Bluetooth is presented at two separate sites (Montreal and Seattle) to investigate the feasibility of using Bluetooth for pedestrian studies. The results indicate that given sufficient populations, high-level trend analysis can provide insights into pedestrian travel behavior.

Key words: pedestrian, Bluetooth, volume estimation, travel time, non-motorized
INTRODUCTION
Over sixty percent of the world’s population owns a mobile telephone, with many developed countries reaching near one hundred percent ownership rates (1). As mobile devices become more complex and their need to communicate with each other grows, there is a growing stream of information that is generated around each mobile device owner. The ubiquity of such devices creates interesting opportunities in human behavior and travel characteristics evaluations, a concept that is quickly gaining momentum in a number of scientific communities. For example, this flow of information is opening new means of analyzing public spaces such as shopping malls, zoos and airports (2), as well as entire towns (3). Of the several available communication protocols available, Bluetooth and Wi-Fi have become by far the most popular. Bluetooth is a short-range communication technology that has been widely used in daily life for mobile-to-mobile device communications. Wi-Fi has, in turn, ensured that many mobile devices have access to the Internet. Most personal electronic devices, such as personal digital assistants (PDAs), cell phones and smart phones, have embedded Bluetooth and Wi-Fi modules that can communicate with other peripheral electronic devices and the Internet.

Analogous to most communicating devices, each Bluetooth and WiFi device has a globally unique 48-bit Media Access Control (MAC) address. It is this unique identifier that provides potential for an easy means of obtaining travel characteristics of a given human-populated network (2, 4). Not only can the overall population of the system be estimated (1), but the travel times, routing choices and interactions can be analyzed to provide an overall understanding of spatial and temporal patterns. Moreover, this approach is as universal as the communication protocols that power it - locations that have no infrastructure or means for data collection can still benefit from this paradigm. This allows for a standard method of evaluation of urban core travel across the globe.

Relying on mobile devices for travel data acquisition also provides an opportunity to pursue information parity for sustainable modes. For nearly a century, transportation metrics have favored motorized means by focusing on measures like mobility (number of miles travelled) and vehicle volume counts, which do not describe the entirety of the transportation system (5). As a result, little attention has been paid to pedestrian and bicycle facilities until recent efforts have highlighted these modes as sustainable alternatives. The development of infrastructure-free information gathering techniques based on mobile devices allows communities to begin to collect data on these sustainable modes, on the basis of individuals and not their vehicles. However, the lack of a data collection framework and a systematic approach stymies these efforts.

This paper aims to create a precedent for monitoring pedestrian movements using static Bluetooth sensors. Basic parameters such as travel time and sample rates are discussed and presented for two study sites, located in Montreal, QC and Seattle, WA. First, an overview of existing efforts is given, followed by a discussion of the methodology followed in the study. Results from the two study sites are then presented and discussed, followed by concluding remarks.

STATE OF THE ART
In the transportation field, much of the focus for collecting data from ubiquitous devices has been on the MAC identifiers broadcast by the Bluetooth protocol. Many Bluetooth devices, such as headsets, are, by default, set in the discovery mode and can be
discovered by other Bluetooth devices inquiring for Bluetooth connections. In particular, the motorized transportation community has become increasingly interested in Bluetooth tracking for the collection of travel time data using dedicated, static sensors (2, 6, 7, 8, 9 and 10). Many research papers focused on utilizing Bluetooth for communications in Intelligent Transportation Systems (ITS) (11, 12 and 13). Ahmed et al. (2008) may be the first group that used Bluetooth MAC address for vehicle traffic monitoring. The Bluetooth MAC address associated with a probe vehicle was tracked by a Bluetooth and Wi-Fi-based mesh network. A Bluetooth MAC address matching method for travel time collection using static sensors was developed and tested by the Indiana Department of Transportation and Purdue University (14). Tarnoff et al. (2009) demonstrated and evaluated a Bluetooth MAC address detector developed by the University of Maryland at the 88th Annual Meeting of Transportation Research Board in 2009 (10). Evaluation of accuracy of Bluetooth-based travel time measurements using static sensors have also been conducted, with encouraging results: most travel times were well within 10% of the ground truth (15).

Research regarding pedestrian and bicyclist travel data collection via Bluetooth is far more limited. In one of the earliest studies, O’Neill et al. (2006) focused on correlating “gatecounts,” or trip-line counts of pedestrians and Bluetooth devices detected in the area of the count (16). In this study, it was found that about 7% of detected pedestrians were carrying Bluetooth devices. The number of devices detected grew linearly with the amount of pedestrians present. Network approaches to multi-modal data collection using Bluetooth have also received relatively little attention. A conference proceeding by Barberis et al. (2006) outlines the concept of Bluetown, a fully integrated data collection network based on Bluetooth beacons (17). The authors suggest creating an ad-hoc network of Bluetooth sensors that are tied into groups by central nodes, capable of relaying the acquired travel time info from each sensor into a main database. Although the possibility of collecting data from multiple modes is mentioned, the authors do not delve deeper. More recent work done by the Lausanne Data Collection Campaign in Switzerland continues to explore social spaces (18, 19). Working with Nokia, the group has been able to demonstrate social behavior metrics can be obtained from device activities (taking photos, sending text messages, sampling audio) and inter-device communication behavior (3). Chaintrreau et al. looked at the inter-contact time between devices as a measure of transfer opportunities, and found that the average number of Bluetooth contacts per day across three separate datasets was around 6 at the time of publication in 2006 (20).

Present data collection approaches include surveys, which are either administered on location or via a broad distribution, manual counts, which involve field data collection by personnel and automatic spot counts, achieved by either infra-red trip-line sensors, or, in the case of cyclists – inductance loops. Video-based data collection methods that are capable of counts as well as localized route choice are also under development (21, 22 and 23). Additional efforts using mobile GPS devices are also beginning to emerge, such as the CycleTracks system developed by the San Francisco County Transportation Authority (24). These efforts require the subjects to either carry a GPS-enabled logger or register their device (as is the case with CycleTracks) with a data collection service, which collects and aggregates individual trajectory information. Due to the costs of many of these approaches, communities often conduct studies on an annual basis, picking a
particular day of the year to act as a surrogate for overall performance (25). This approach may not only produce non-representative results due to weather variations and other unobserved factors, but also does not provide a clear trend line that can be analyzed for effective improvements in infrastructure or policy.

**METHODOLOGY**

In order to collect bypassing MAC addresses a MAC reader device was used. Regular Bluetooth devices are normally restricted to a range of about 10 m, but an increasing number of devices are becoming available that can read MAC addresses from up to 300 m away (4). Such a device, shown in Figure 1 a) has been developed by the Smart Transportation Applications Research (STAR) Lab at the University of Washington and was used in this study. The device was used in conjunction with a 7dBi omni-directional antenna to obtain a range of roughly 50 m. To ensure that the 50 m range was achievable, the device was mounted at a height of about 3 m. Although the range is 50m, many devices will not be detected until they have been in range for some time due to the random frequency hopping characteristics of the Bluetooth protocol. Furthermore, additional factors such as receiver antenna placement (and obstruction) as well as external noise will also affect the range. A typical mounting position is shown in Figure 1 b).

![FIGURE 1: a) Sensor used in tests](image)

Two locations were examined in this proof-of-concept study - a short, pedestrian-only corridor in downtown Montreal, QC (Montreal) and a section of the University of Washington (UW) campus in Seattle (Seattle), also restricted to pedestrian access. Both studies were conducted in the summer – testing in Montreal was performed August 18-19th 2010, and testing in Seattle was performed during the UW Graduation Ceremony weekend on June 11-12th, 2011. Data from both locations was truncated to 24 hours starting from 9:30am. The Montreal site consisted of a corridor that was about 100 m long, passing from point “A” in the middle of a block (where no vehicles were present) to point “B”, the end of the next block at an intersection with an arterial. The UW site consisted of a path about 350m long, leading from the Drumheller fountain (a popular picture spot) down to Husky Stadium, where the Graduation Ceremony was taking place. Figure 2 shows the maps of both locations, along with sensor placements (shown in blue) and shortest path outlines (shown in red).
The selected sites represent two common pedestrian areas, dense urban centers (downtowns) and campuses (malls, parks, etc.). There are a number of travel characteristics that are of interest at these types of locations. Count data (volumes) is the most basic data that is needed, and examining Bluetooth MAC address readings as a population estimate is of interest. Furthermore, the traditional paradigm of using Bluetooth for travel time data collection also has practical applications for non-motorized purposes. Finally, an additional metric, dwell time, can be collected to determine the amount of time an individual spends within a particular area. This value has implications for retail as well as public transportation purposes. However, prior to discussing the obtained results regarding these three metrics, it is important to establish the characteristics of each site with respect to device presence rates and types.

One of the reasons that research regarding pedestrian detection using Bluetooth has been limited is due to the relatively low sample sizes attainable. While vehicles may have a number of Bluetooth-equipped accessories, and are subject to hands-free laws which require drivers to wear headsets, pedestrians have a limited number of devices that they may be carrying in the current environment. Primarily, we can expect pedestrians to carry cell phones, headsets and computing devices such as tablets or laptops. Additional devices such as heart-rate monitors and pedometers are beginning to enter the market, but do not yet have a significant share of the total device market. The shares of each device manufacturer per site are shown in Figure 3. Overall, 2520 unique devices were seen in Montreal, while only 534 were seen in Seattle over a 24-hour period. Both sites had similar brand populations, comprising primarily of headset/handset manufacturers such as Nokia, Samsung and LG. These devices are especially prevalent due to the support of continuous discoverability (broadcast) in handsets. In contrast, Apple, Motorola and HTC handsets use a limited discoverability protocol, where the device remains visible for only up to 120 seconds. Other devices, however, such as tablets, can support continuous broadcast modes.
Short, one-hour manual counts were done within range of the installed sensors to determine the approximate sample rates at each location. In downtown Montreal, which had extremely high pedestrian flows of around 2687 pedestrians/h, the sensors were able to capture roughly 5% of the population. In the UW Seattle location 2.25% of the population was represented by a MAC address (26). These figures are significantly lower than the vehicle sample rates of roughly 10% that were obtained at the data collection points in other studies (e.g. 4 and 15).

**EXPERIMENTAL RESULTS**

*Dwell Time Analysis*

Dwell times at both locations were examined to see how these values may be of value in describing a particular location. Dwell times were calculated on the bases of continuous presence – a device had to check in every 60 seconds to be considered continuously present. This is a reasonable arbitrary threshold that attempts to filter out briskly passing individuals (over 5ft/s) so they would not be considered “dwelling” and pass the detection range in under the 60 second limit. The longest interval time with a continuous presence is considered to be the dwell time. In Montreal, the two locations chosen for sensor mounting were only 100 m apart but had different dwell time characteristics; with an average dwell time of 1.28 minutes (76.8 seconds) at location “A”, near the intersection and 1.70 minutes (102 seconds) at location “B”, in the middle of the alley. As expected, the sensor mounted at the intersection discovered significantly more unique Bluetooth MAC addresses - 832 versus the 573 unique MACs discovered at the mid-alley location where only pedestrians were present. Figure 4 shows the recorded dwell times at each location at the Montreal site. The data is presented by hour of day, displaying the variability encountered during each hour. The 9 o’clock hour is wrapped around two days, since testing began at 9:30 am. The box and whiskers plots show the middle two quartiles as a box, with the whiskers out at the first and third quartiles. The thicker horizontal lines represent the means. Hours that did not have any data are not shown. Looking at Figure 4, most pedestrians spend up to about 3 minutes passing through the detectable zone of
100 m. Median dwell times seem consistently larger at mid-alley than at the intersection, and have different temporal evolution, e.g., globally decreasing after 16:00 at the intersection while increasing from 15:00 to 18:00 before decreasing at mid-alley, which may reflect more restaurant activity near this location. It’s also apparent that the pace quickens at night when the numerous shops and stands are closed. The sample sizes collected during each hourly interval are also shown on the right axis. The two commuter peaks are clearly seen at both locations, but are more evident at the intersection, likely due to increased vehicle traffic. The morning peak appears in two stages as well, with a portion of devices appearing at 5am and the dominant majority at 9am.
Figure 4: Dwell times at the Montreal Chinatown location

To further examine the difference between the two ends of the study segment in Montreal, dwell time distributions (limited to 15 minutes) are shown in Figure 5 below. The two distributions are found to be significantly different, with a p-value of .001034 using the Kolmogorov-Smirnov test. This can be attributed to a significantly higher number of very low dwell times that are most likely detected in vehicles passing by. In Figure 5, it can
be seen that the 0-1 minute bin at the intersection (A) has roughly 60% more MACs than the same bin in the alley (B). This is 15% higher than the average increase of MACs experienced at the intersection, implying that this increase is not simply a result of higher device populations seen at the intersection.

**FIGURE 5: Dwell time distributions at the Montreal sensor locations**

The Seattle location exhibits significantly different trends due to the event-based nature of the pedestrian flows. Furthermore, the locations differ significantly in characteristics – location “A” is next to the Drumheller fountain, where many stop to take pictures as well as meet others, while location “B” is on an uneventful stairway leading to the UW Husky stadium. Thus, the dwell times at location “A” are notably higher – 2.14 vs. 0.81 minutes at location “B”. More devices (422) were also picked up at the fountain than at the stairwell (353). The dwell time readings are also obtained primarily during the beginning and end of the Graduation Ceremony, with very few readings at other times. Figure 6 shows the boxplots by hour of the dwell times seen at both sensor locations. Two peaks of detected device population samples can also be seen, as more people collect to make the Ceremony and later depart. The first population peak at the stairwell is sharper, taller and occurs about an hour later than the first peak at the fountain, suggesting that people spent about an hour on average strolling around campus prior to going down the stairwell to the stadium.
Figure 7 shows the dwell time distributions (limited to 15 minutes) at both ends of the study segment. The distributions vary significantly from one another (Kolmogorov-Smirnov test p-value less than $2.2 \times 10^{-16}$) – as expected, the dwell times at the fountain
are generally longer; as it itself is an attraction, unlike the stairwell. There are almost no dwell times over 2 minutes at the stairwell.

![Dwell Time Distribution at A (fountain)](image1)

![Dwell Time Distribution at B (stairwell)](image2)

FIGURE 7: Dwell time distributions at the Seattle sensor locations

**Travel Time Analysis**

Hourly summaries of travel times in each direction recorded in the Montreal corridor are shown in Figure 8. Travel times were calculated using the first-first paradigm (4), where the first sighting at each sensor is used to determine the travel time interval. Most travel times collected were under 10 minutes (median 4.27, average 11.48 minutes), with no apparent bias towards either direction (277 devices travelled from A->B and 279 travelled in the reverse direction). Since the section examined has multiple cafes, restaurants and shops, travel times higher than the necessary time to walk the whole distance are expected.
Figure 9 shows the distribution of the measured travel times, with a clear exponential trend. The most common travel times are under 5 minutes, which indicates that most
people do not stop for too long on their way through, although the speeds are still low, as indicated in the dwell time analysis.

FIGURE 9: Montreal travel time distributions in both directions

Travel times in Seattle are higher and display strong directional trends, as can be seen in Figure 10. At first, visitors began arriving at the Ceremony over an extended period of time and took time to take pictures or meet others and tour the campus. This is indicated by the higher travel time averages between the hours of 10:00 - 12:00. Once the Ceremony begins, travel times decrease significantly; as people simply rush to the stadium directly (they’re late!). Finally, a second peak occurs, now in the opposite direction, as people leave the stadium almost concurrently, this peak is narrower, lasting from 16:00 to 17:00. It can also be seen that no travel times were collected outside of the Ceremony due to low pedestrian volumes - data is in effect available only for the hours of 10:00 – 17:00.
Figure 11 shows the travel time histogram for the Seattle location – the most common travel time is around 10 minutes, which equates to a speed of roughly 0.58 m/s (1.91 ft/s),
indicating that most people took time to do other activities besides simply walking from the fountain to the stadium.

**FIGURE 11: Distribution of travel times to and from Husky Stadium in Seattle**

The graduation ceremony can be compared with the previous day 6/10/2011, which was the last official day of the quarter, during which no classes were held. No large events were held at Husky Stadium as well, a “quiet” day. In fact, only 17 devices were detected on the stairwell over 24hrs (starting at 9:30am), compared to 181 detected at the fountain. Still, even the fountain device number is less than half of what was seen during Graduation. Figure 12 shows the travel times found between the locations and the dwell times experienced at the fountain. The travel time data is very sparse, often with only one or no data points per hourly period. The dwell times at the fountain are well represented, but are shorter than those experienced during Graduation, averaging 1.91 minutes vs. 2.14 minutes.
FIGURE 12: Travel time and dwell time on a “quiet” day before Graduation on the UW Campus.
SUMMARY AND CONCLUSIONS

Summary
An attempt at characterizing pedestrian environments and walking behavior using Bluetooth is presented, examining two distinct locations, a downtown area in Montreal and a university campus in Seattle. Three types of data were examined – device types encountered, dwell times at individual sensor locations and the travel times between the mounted sensors. Both locations provided fairly low sample rates (between 2 and 5 percent); however, general trends could still be distinguished and interpreted, especially in event scenarios. Based on the acquired data, basic information about pedestrian activity, such as travel times and dwell times can be acquired. Low volume scenarios, however, often did not have sufficient samples to obtain usable data. Although the current sample sizes are low, current trends in personal device suggest higher levels inter-communicability, potentially increasing the sample sizes available. Furthermore, as more device types begin to communicate, additional information regarding the type will further enhance filtering and data resolution.

Concluding Remarks

Two additional concerns arise with the proposed data collection approach: privacy and bias. The increasing ease with which information can be recorded and shared has continued to impede on traditional values of privacy. From license plate readers to credit card transactions and store club cards, there is an overwhelming amount of information that is collected about individuals every day. The data collection methodology presented in this paper involves the use of MAC addresses, which, despite not being directly tied to individuals, have the potential to be misused. Precautions, including encryption and data expiration (deletion) must be taken. For example, for much of the collected data, matches that occur in intervals over 60 minutes are meaningless. Thus, addresses that have not been matched within this time period should be deleted. The addresses that have been matched can be replaced with a unique identifier that is derived from the order of arrival or another unrelated variable. This way, no permanent record of observed devices exists (27).

Bias is also of significant concern with the proposed methodology. It should be expected that more affluent, younger populations would carry more Bluetooth-enables devices, especially highly visible devices such as headsets. However, there are additional considerations, such as device type (some less expensive cellular phones are capable of continuous broadcast only) and technological savviness (forgot/unable to turn Bluetooth visibility off) that also play a role. Preliminary testing on bus stops in various Seattle neighborhoods suggest that there may actually be more visible devices present in less affluent neighborhoods, but additional work is necessary to determine the extent and direction of bias. Overall, Bluetooth appears to be capable of providing automatically and continuously basic pedestrian movement trend information which can be used for planning purposes. With more personal device data becoming available correctly interpreting and filtering the available data is becoming the most pressing issue in travel analysis. The locations chosen in this study relied on selective sensor location to filter pedestrians from other devices present in the area. Future work will include development of clustering and filtering techniques that can allow for mixed-mode data gathering.
ACKNOWLEDGEMENTS
The authors wish to acknowledge the funding support from Washington State Department of Transportation and Transportation Northwest (TransNow), USDOT University Center for Federal Region 10. We also appreciate the help of Ali El Husseini and Audrey Maillard, research assistants at Ecole Polytechnique de Montreal for their assistance with the data collection efforts.
REFERENCES


