UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

UTTRI

POTENTIAL OF PASSIVE DATA SOURCES

2016-10-23

Transportation Tomorrow Survey 2.0

Jeremiah Chan, Siva Srikukenthiran
Khandker Nurul Habib & Eric J. Miller

EXECUTIVE SUMMARY

The Transportation Tomorrow Survey (TTS) is the largest and most comprehensive regional household travel survey conducted in Ontario, performed every five years surveying 5% of households in the Greater Golden Horseshoe region. However, like many other transportation surveys, there have been issues regarding the validity of the resulting data. The main complication identified was the decline in the use and ownership of landlines, especially in the younger population. This results in an under-coverage issue when using a landline phone directory as the sample frame for household travel surveys. Other issues/limitations include data being collected retrospectively and by proxy (one person responding on behalf of all household members).

Prior reports highlighted the use of passive data sources as a possible method to supplement survey data and mitigate these issues. Passive data are information gathered from respondents without their active knowledge. The nature of the data allows for the analysis of "revealed preference," or actual behavior, as opposed to the "stated preference" data which are obtained from typical survey methods. Furthermore, the potential access to continuous data sources raises the potential ability to update current survey data sets and O-D matrices. The goal of this report was to outline passive data sources, their current usage around the globe and the potential for their use as part of the TTS 2.0 project. These data sources include smart cards, Bluetooth collection systems, and cellular data.

In smart card automated fare collection systems (SCAFC), payment is accomplished through the contact of cards on a reader, rather than cash and tickets. Smart cards have shown a potential to gather large amounts of passive data, using infrastructure already in place for fare collection, while allowing transit agencies to track travel patterns over long periods of time. Due to this latter reason, data can be compared between multiple time frames and over longer periods of time than traditional survey methods. Of the technology investigated, smart card data holds the most promise, largely because of the implementation of the PRESTO card in the GTHA transit systems. Although the region is made up of eleven different transit agencies, each with their own fare system, Metrolinx aims to have the PRESTO Card as a uniform payment method for all of them before the 2021 TTS. However, although smart cards have a great amount of potential in the GTHA, its biggest drawback is that it is limited to only public transit services and is difficult to extend to other modes. As a result, it does not stand on its own in providing travel information, and needs to be integrated within a larger data collection framework for future TTS.

Bluetooth technology is arguably the most accurate of the available technologies, having smaller detection ranges and higher detection rates. However, in order to maximize its effectiveness, the stations must be placed closer together in order to limit the areas where there is no coverage. This leads to an increase in cost, especially if this technology is used in a large area, such as the GTHA. There is the potential to use it in smaller dense areas such as downtown Toronto; however, a large number of data collection stations would still be required. Therefore, it is not recommended that this type of passive technology be used as part of the TTS. However, should data fusion methods be further researched, it may be possible to use this technology in more focussed satellite collection efforts.

In contrast to Bluetooth, cellular technology is less accurate but works better over larger areas. Even with the issues in accuracy, it is more appropriate for acquiring data over a large region like the GTHA. Furthermore, due to its extremely high penetration rate, this technology offers a large data set that can be collected and studied over large periods of time. However, the drawback of cellular data is that it lacks accuracy,

providing only approximate locations of users within tower ranges (Wang, et al., 2013). While some attempts have been made to mitigate this hindrance by collecting data from multiple towers, coupled with cellular signal traces of routes, the accuracy and the acquisition cost of such data needs further investigation.

Due to the continuous nature of passive data, there is the potential to use this data to update existing O-D matrices. This was shown in data gained using smart card technologies in New York and Bluetooth technologies in Indiana. With this potential revealed, it may be possible to use passive data to update O-D matrices, should they not be used as the main source of information collection. However, additional research is required to develop methods of fusing this data with other sources.

Moving forward, the greatest potential of incorporating passive data in future TTS lies with the use of smart card data, because of the existence of the PRESTO card. Furthermore, although the cost of acquisition is unknown, this data would encompass transit users in a very large area of the GTHA. With regards to the core-satellite design, the PRESTO card data can provide some key travel information, such as trip patterns, origins, and destinations. Nevertheless, several steps are needed to further investigate and incorporate this data. These include acquiring PRESTO card data from Metrolinx, determining its feasibility in tracking transit trips, researching algorithms for creating O-D matrices and fusing them with those produced by traditional survey methods, exploring geographic coverage, and identifying and research ways to mitigate issues in the collected data. The initial analysis of the PRESTO data is expected to occur in summer of 2017, with research into data fusion methods beginning this summer and continuing into next year.

POTENTIAL OF PASSIVE DATA SOURCES

TRANSPORTATION TOMORROW SURVEY 2.0

Table of Contents

1 INTRODUCTION

Transportation surveys have long been used to collect information regarding traveller origins, destinations, and patterns. They allow transit agencies to collect the critical information required for transportation planning, developing forecasting models which can help predict and respond to changes in daily travel patterns. (Griffiths, et al., 2000)

Traditionally, these data have been collected through Computer Aided Telephone Interviews (CATI), where surveyors directly call a respondent's landline and input answers into a computer. The Transportation Tomorrow Survey (TTS), being a regional household survey of the Greater Toronto and Hamilton Area (GTHA), is one survey that has used this method. Dating back to 1986, the TTS has been administered every five years, collecting travel data on individuals in approximately 5% of households (Data Management Group, 2014).

However, like many other transportation surveys, there have been issues regarding the validity of the resulting data, as outlined in a major study done by the Transport Association of Canada (TAC) (Transportation Association of Canada, 2012). The main complication identified was the decline in the use and ownership of landlines, especially in the younger population. This results in an under-coverage issue when using a landline phone directory as the sample frame for household travel surveys. Other issues/limitations include data being collected retrospectively and by proxy (one person responding on behalf of all household members). When calling over the telephone, the respondent is asked to report trips from the previous day; this often results in missing trips and details. The same problems also occur when collecting data via proxy, as work and school trips are often under-reported if the traveller is not the respondent. This TAC report aimed to address these issues by investigating a broad range of alternative designs, and proposing several possible new data collection methods. One recommended area of research was the use of passive data sources to supplement survey data.

Passive data are information gathered from respondents without their active knowledge. This differs from active data collection, which involves the informed consent of the respondent who answers questions voluntarily (Esomar World Reserach Codes & Guidelines, 2009). The nature of the data allows for the analysis of "revealed preference," or actual behavior, as opposed to the "stated preference" data which are obtained from typical survey methods. Furthermore, the potential access to continuous data sources raises the potential ability to update current survey data sets and O-D matrices, as is explored in later sections (Miller & Habib, 2015).

Furthermore, another issue with traditional survey methods is the rising cost and difficulty of collecting data, as well as contacting, recruiting and retaining households. For example, methods such as household or roadside surveys are expensive and time-consuming to conduct, a burden on the respondent, and only provide a "snapshot" of traffic patterns and behaviors (White & Wells, 2002). Passive data, in many cases, can be a solution to this, making use of sensor infrastructure that already exists. Although this method of data collection reduces many issues, such as bias and proxy, it comes at the expense of interaction with the respondents, resulting in a loss of information about the traveller, their trip purposes and other details (Xu, et al., 2014) (Huntsinger & Donnelly, 2014). Although there have been attempts at mitigating these concerns, such as inferring trip purpose through combining data from concluded destinations and land use, they have not been reliable (Widhalm, et al., 2015).

With this in mind, the TTS 2.0 Proposal Plan outlined several areas of research that could provide insight into the possibility of passive data streams being a viable source of travel survey data: (Miller & Habib, 2015)

- The potential usage of PRESTO Card data in providing an opportunity to better observe and understand transit usage on a continuous basis.
- The usage of open source and public sector data sets to gather travel related behavior data.
- The cost, availability and quality of such data, and how they could impact the TTS 2.0 core-satellite approach.

In addition to these main areas of study, the TTS 2.0 also proposes a "core-satellite" approach, in which passive data may be used as a "satellite" data collection to aid the current survey process. To do this, several ideas must be further investigated, including:

- The application of passive data sources to form Origin-Destination (O-D) matrices
- The possibility of using these data to continuously update current O-D matrices in real time

The derivation of O-D matrices is the ideal result of passive data analysis, as they give insight into travel patterns, as well as allow transit agencies to plan for the future. They generally provide an estimate of the number of vehicles travelling between points in a network over a period of time; however, they can be extended to model movements of transit users given the right type of data. In addition, passive data may also be fused with other data sets, such as traditional survey methods, to form these matrices. Furthermore, accurate origin-destination matrices are essential for the development of the assignment models used for traffic planning (White & Wells, 2002). These data can then by organized and displayed on a "mobility map" for a visual representation of traffic (Caceres, et al., 2007).

This report outlines several passive data streams, their current usage around the globe and the potential for their use in the GTHA as part of the TTS 2.0 project. **Section 2** describes smart card data usage in current transportation systems, and how the data have been used for the derivation of origin-destination (O-D) matrices. **Section 3** outlines Bluetooth data collection systems and how they have been applied around the world. **Section 4** analyzes the interpretation of cellular data and how they may be used to assist transportation surveys. **Section 5** describes several other passive data sources and how they may be used to collect transportation data. Finally, **Section 6** outlines the core-satellite approach and the role that passive data may have with it.

2 SMART CARD DATA

In smart card automated fare collection systems (SCAFC), payment is accomplished through the contact of cards on a reader, rather than cash and tickets. This technology was introduced and incorporated into some public transit systems, such as Tokyo and Washington in the late 1990s. Since then, its popularity has spread to other areas like Chicago and London and, in many cities, it is becoming the most popular payment method. (Munizaga, et al., 2010)

There are some practical reasons that countless public transport systems in the world use smart card systems for their buses and subways. The use of smart cards is environmentally conscious because of the reduction in paper waste compared to traditional fare. Furthermore, they allow transit companies to track and log more information on traveler trips as large amounts of detailed data are collected. (Zhang, et al., 2016; Pelletier, et al., 2011)

Generally, smart cards work using a “tap in” system. Upon, or prior to, entering a transit vehicle or station, the user taps the card on a card reader, and the cost of the trip is deducted from the total balance on the card. Data are collected and recorded from these taps, as well as being matched to a unique serial number on the card to more easily track individual users. This data can include variables such as time, location, bus route, bus station or subway station. This, in turn, is collected and stored by transit agencies, to be processed and organized into useful data. (Trépanier, et al., 2007)

There can be variations in these systems, the most notable one being a “tap off” option for transit vehicles. This alternative is geared towards systems that charge by distance, rather than a flat rate. While this variation allows for the collection of additional data, it is often not included for other reasons such as requiring extra infrastructure or slowing down bus operations. Furthermore, in most cases, this data are simply not needed for the purpose of the transit agency. (Trépanier, et al., 2007)

Another important consideration is the privacy concern involved in the collection of smart card data. Although some systems require personal information upon registration, many users are uncomfortable with this data being used by planners. To account for this, the available data for analysis often are only tied to serial numbers, with no ties to the transit user themselves. Personal information must be derived or inferred, depending on trip patterns. This can be done by comparing general demographic information associated with the determined origins and destinations. (Chapleau, et al., 2008)

Consequently, even with the increased usage of passive smart card collection data, there have been two issues that repeatedly surface, calling for solutions in different forms:

1. **Transfers**. Due to the discontinuous and indirect nature of transit travel, many users must transfer vehicles in order to get from origin to destination. To identify and catalog these transfers, there have been several methods proposed, many of which use an estimated time frame. With an assumed transfer time, if there are no possible routes within this time frame, the trip is presumed to have been completed and, therefore, the following transaction would represent the start of the next trip.
2. **Alighting Location**. For systems lacking the “tap off” option, the alighting location must be determined in order to form O-D matrices. The most common method to find this location involves linking trips together, using the origin of one trip as the destination of the next. This is an assumption made by most studies and has proven highly accurate, as will be displayed in section 2.1. (Barry, et al., 2001)

Unavoidable, however, is the main limitation for data collection through smart cards, namely that the information collected from this method only applies to public transit systems, with little to no data on vehicular transit. While some information can be inferred through the analysis of bus movement, this is not enough to develop useful models. This means that whatever data or O-D matrix is derived from smart cards, however, useful it may be, can only be used to assist in transit planning, and other methods are required to understand travel across all modes. (Trépanier, et al., 2007)

The following sections outline several smart card applications and how this type of data may be used towards the TTS 2.0 project.

2.1 Current Applications of Smart Card Data Around the World

Smart card fare collection systems have already been used in various places around the world, mostly with high degrees of success. The systems that have been analyzed in this report focused on field tests done in Santiago, New York City, Brisbane, Gatineau, Chicago and London. As shown in Table 1, each system varies slightly in methodology and results, but many follow the same general strategy and assumptions:

1. The destination of a trip is the origin of the next trip. (Munizaga & Palma, 2012)
2. Passengers return to their original origin at the end of the day. (Alsger, et al., 2014)

The following sections describe the various applications of passive smart card data for the development of O-D matrices in the different cities indicated in Table 1.

TABLE 1: COMPARISON OF BASIC COMPONENTS OF SMART CARD FARE COLLECTION SYSTEMS IN DIFFERENT LOCATIONS

Location	Company/Transit Authority	Alighting Transactions?	Estimated Walk Time/Distance	Additional Comments
Santiago, Chile* **(Munizaga & Palma, 2012)**	Transantiago	No	400 meters or 5 minutes	-the system uses both metro and bus
New York City, New York* **(Barry, et al., 2001)**	Metropolitan Transit Authority (MTA)	No	N/A (Subway only)	-compared exit tallies to exit turnstile counts
Brisbane, Australia* **(Alsger, et al., 2014)**	TRANSlink	Yes	15 to 90 minutes (range)	-estimated transfers rather than alightings
Gatineau, Quebec **(Trépanier, et al., 2007)**	Societé de transport de l'Outaouis	No	N/A	-no attempt to deal with data outliers; they were simply removed and ignored
Chicago, Illinois **(Zhao, et al., 2007)**	Chicago Transit Authority (CTA)	No	400 meters or 5 minutes	-integrated AFC and AFL data
London, United Kingdom **(Wang, 2010)**	Transport for London (TfL)	Subway Only	5 minutes	-alighting estimation used only for buses

*Described in greater detail in following sections

2.1.1 Smart Card Application in Santiago, Chile

The Transantiago transit system in Santiago, Chile utilizes a smart card system for its buses and subways. The SCAFC has an extremely high penetration in this city, with a 97% usage rate, creating a great opportunity to test the potential of smart card data to generate O-D matrices. Therefore, in March 2009 and June 2010, Munizaga & Palma (2012) conducted two tests on data collected from approximately 8.5 million trips (DICTUC, 2003).

The Transantiago smart card system uses a “tap on” method. Travelers are only required to tap their bip! cards upon boarding. When a transaction occurs by tapping, information is collected on the card ID, time, date, amount of money paid and type of transit. Locations were only recorded on metro trips as this was the only time the exact location was known. For bus trips, a GPS tracking system was installed in each vehicle which would provide an estimate of when the bus was at each stop. Then data processing, therefore, was just a matter of matching the time stamp of each transaction to the time that the bus was logged at a stop. (Munizaga & Palma, 2012)

During processing, the data were separated into three main categories, metro, bus and bus station, which was later analyzed separately. The metro data were the easiest to sort, as the boarding locations were directly known. The bus data, on the other hand, were more complicated to organize as the GPS data were not continuous, but only recorded every 30 minutes. The remaining data were estimated using the approximate times that the bus was at each stop, matched against transactions at these times. Processing the bus station transactions were further complicated by the fact that a person could take multiple buses while at a station. These data were assessed with a more iterative method, by examining the next transaction (following the assumptions made in Section 2.1), comparing all possible routes through that bus station, and assuming that the user took the route with the lowest travel time. (Munizaga & Palma, 2012)

Furthermore, when assessing the distance that transit users were willing to travel for transfers, Munizaga et al (2010) assumed that this distance was 400 meters or 5 minutes. It was noted, however, that these values may have varied between people, cities and weather conditions. This was used when determining all the possible routes in the analyses and algorithms.

With this methodology, Munizaga & Palma (2012) were able to provide estimates for over 80% of the destinations, as well as develop O-D matrices for each of the two-time frames, by aggregating the data based on areas (Table 2 and Table 3). However, for the remaining trips where the alighting destination could not be determined, the data were simply removed from the matrix. Additionally, due to the fact that data was available in both space and time coordinates, O-D matrices had the potential to be created at any disaggregate level, allowing for more in-depth analysis if required.

TABLE 2. AGGREGATE O-D MATRIX FROM MARCH 2009 IN SANTIAGO (MUNIZAGA & PALMA, 2012)

	North	West	East	Center	South	South-East	D_j
North	157,950	36,389	51,489	78,906	22,988	18,087	**365,810**
West	34,164	294,670	116,217	162,561	37,041	30,525	**675,177**
East	49,382	112,436	317,606	173,157	70,812	150,056	**873,450**
Center	74,593	167,516	160,132	171,932	103,127	96,399	**773,700**
South	22,222	34,877	73,977	104,116	189,216	54,216	**478,624**
South-East	18,379	30,450	158,839	97,234	55,614	250,057	**610,572**
O_i	**356,690**	**676,338**	**878,261**	**787,906**	**478,798**	**599,339**	**3777,333**

TABLE 3. AGGREGATED O-D MATRIX FROM JUNE 2010 IN SANTIAGO (MUNIZAGA & PALMA, 2012)

	North	West	East	Center	South	South-East	D_j
North	157,950	36,389	51,489	78,906	22,988	18,087	**365,810**
West	34,164	294,670	116,217	162,561	37,041	30,525	**675,177**
East	49,382	112,436	317,606	173,157	70,812	150,056	**873,450**
Center	74,593	167,516	160,132	171,932	103,127	96,399	**773,700**
South	22,222	34,877	73,977	104,116	189,216	54,216	**478,624**
South-East	18,379	30,450	158,839	97,234	55,614	250,057	**610,572**
O_i	**356,690**	**676,338**	**878,261**	**787,906**	**478,798**	**599,339**	**3777,333**

2.1.2 Smart Card Application in New York City, New York

Similar to Santiago, New York City has a SCAFC system for their public transit system called the "MetroCard"; however, it is only used for subway travel. In New York, the smart card market share is not quite as large (78%) so this study was only a 100-person sample, used for verification before actual pilot tests were run. The test data were compared to diary information collected from the New York Metropolitan Transportation Council. This data comparison confirmed assumptions made in Section 2.1 on trip destination (Barry, et al., 2001). With these assumptions, O-D matrices were formed for the transit network in New York, similar to how they were created in Santiago. Additionally, Barry et al (2001) used multiple methods to compare O-D trips to ridership counts, the most important being comparing the data with tallies generated from exit turnstiles.

As one of the earlier examinations of the application of passive smart card data applications, the pilot test performed in New York illustrated the feasibility of developing station to station O-D matrices for public transportation systems (Ma, et al., 2012). Barry et al (2001) were able to develop matrices and trip tables for both weekdays and weekends, as well as during construction periods (Table 4), to assist with planning.

TABLE 4. SAMPLE TRIP TABLE GENERATED DURING THE CONSTRUCTION OF A NEW SUBWAY LINE IN NEW YORK (BARRY, ET AL., 2001)

		Destination Line and Station							
		West End Line							
Origin Line	Origin Station	55 St	71 St	79 St	18 Ave	20 Ave	Bay Pky	25 Ave	Bay 50 St
Brighton	Ave H	0	1	7	3	1	1	1	12
	Ave J	4	3	6	3	3	4	1	19
	Ave M	1	4	9	13	11	9	11	4
	Kings Hwy	15	8	11	15	7	11	14	26
	Ave U	4	8	8	10	11	7	3	6
	Neck Rd	0	4	8	6	4	4	0	3
	Sheepshead Bay	10	10	19	15	18	40	3	23
	Brighton Beach	41	15	32	16	28	37	10	27
	Ocean Pky	11	3	6	2	9	9	8	9
Culver	Kings Hwy	4	1	4	3	4	1	0	3
	Ave U	4	0	1	3	1	9	3	3
	Ave X	4	1	3	0	3	10	10	4
	Neptune Ave	4	0	2	2	6	4	2	4

2.1.3 Smart Card Application in Brisbane, Australia

The transit system in Brisbane, Australia varied somewhat from the previously mentioned cities, as their buses subways require an additional “tap” upon alighting from the vehicle. Users would repeat these steps for every vehicle they used, essentially giving transit agencies twice the amount of data for conducting tests. In addition, the smart card, or “go card,” system is fully automatic and offers discounted fare prices from the paper tickets. (TRANSlink, 2016)

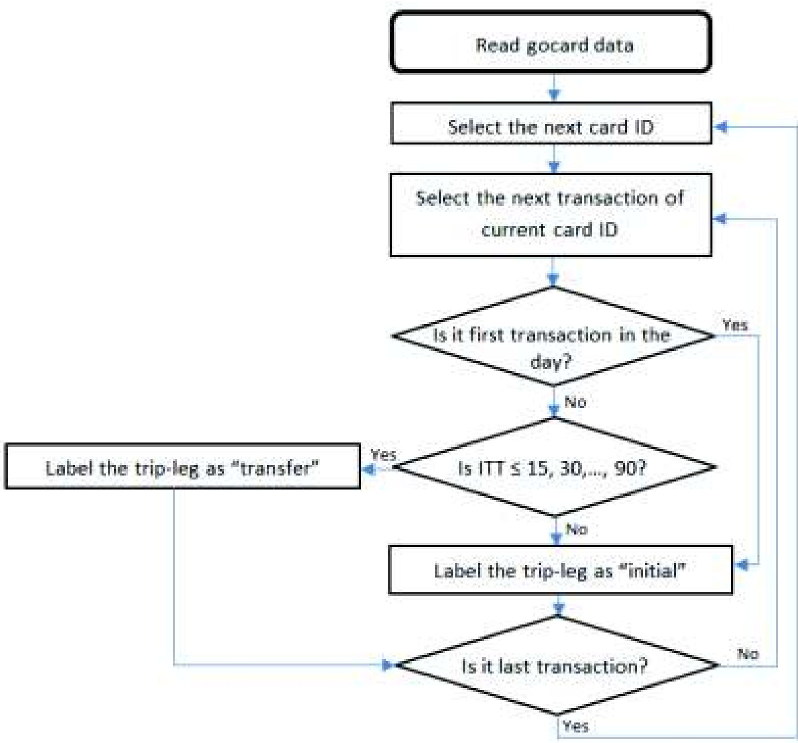

FIGURE 1. O-D ESTIMATION ALGORITHM IN BRISBANE (ALSGER, ET AL., 2014)

With the acquisition of destination data, there is no longer a need to estimate the location of alighting based on subsequent trips in order to develop O-D matrices. Instead, Alzger et al (2014) conducted multiple tests altering the remaining variable: transfer time. Collecting and processing data in the same way as the aforementioned cities, they analyzed the time that passengers took between trip legs, developing an algorithm for deriving trip segments (Figure 1). The data produced by this algorithm were then used to estimate O-D matrices.

This method distinguished transfer points from trip end points by examining inter-transaction time (ITT). This factor was varied in 15-minute intervals, starting at 15 minutes and ending at 90 minutes. For example, when ITT was tested for 15 minutes, in cases where transfer time was measured to be less than this, the algorithm labelled the trip leg as a transfer. In cases where transfer time was greater than 15 minutes, a destination was established at that location, and the second trip was assumed to be the initial leg of the subsequent trip. This process was repeated to generate multiple matrices, allowing for the comparison of the number of trips measured for each ITT value. (Figure 2)

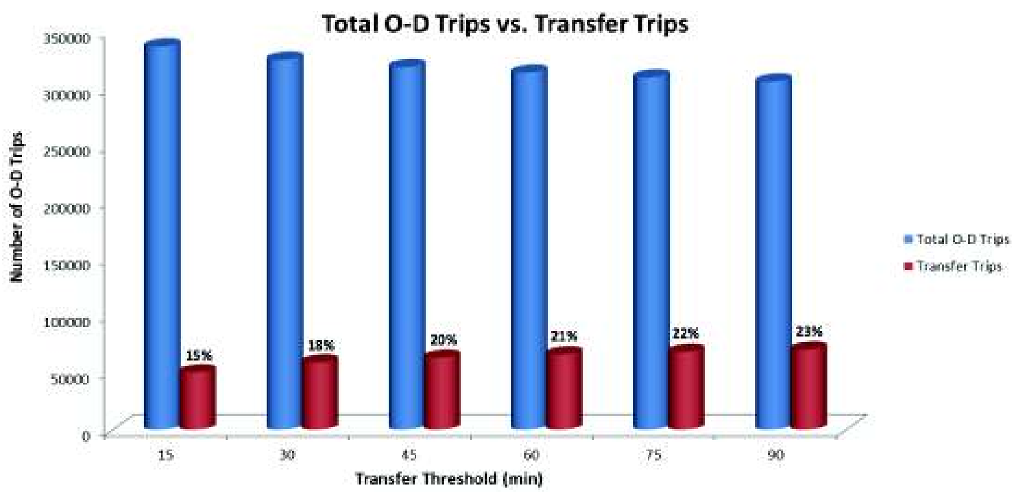

FIGURE 2. COMPARISON OF NUMBER OF TRIPS TAKEN WITH VARYING TRANSFER TIMES IN BRISBANE (ALSGER, ET AL., 2014)

From this experiment, it was deduced that most passengers in Brisbane did not transfer vehicles in their daily trips. Through changing the ITT, there was minimal impact on the number of trips taken. However, this may not have been true in all locations, so assumptions ranging from 400 to 800 meters of walking distance were safe to make (Alsger, et al., 2014). Unfortunately, these assumptions may have failed to track shorter stops, an issue that will be discussed in Section 2.2.

2.1.4 Smart Card Application in Other Locations

The three cities studied in the prior sections provided an illustration of the major features of smart card data applications and how they may be used or processed in different areas. Further studies have been carried out in cities such as Gatineau, Chicago, and London, each following a similar method but varying slightly in the quantity of data collected and the organization of these data.

In Gatineau, Quebec, Trépanier et al (2007) inferred the destination of passengers using both smart card data and GPS readers on buses. By connecting time stamps with the position of the bus at those times, the researchers were able to estimate the boarding stop of all the passengers. This was accomplished by noting that GPS locations and times were recorded when the bus was leaving the stop; therefore, all smart card transactions before this time were assumed to be at that stop, with some exception for time stamps recorded just after the doors close. However, due to Gatineau having a bus only system (Société de transport de l'Ouaouais, 2012), Trépanier et al (2007) argued that these data were difficult to use for transportation planning models as the data were aggregated and synthesized in a way that diluted information quality.

Among the studies that were done in Chicago, Illinois were similar to Gatineau, with changes because the system had subway lines and automatic passenger counters. Zhao et al (2007) focused mostly on the processing of data, noting that unlike O-D matrices developed from other methods, smart cards allowed for the collection of individual, rather than aggregate, data. This meant that the data allowed transit agencies to track individual users, as well as patterns from the general population. Following the same methods as in Santiago and New York City, the study performed by Zhao et al (2007) further proved that inferring O-D matrices from origin-only fare collection systems could still provide the fundamental information needed for analysis of transit operations. Furthermore, it was found that there may be potential to obtain ridership information based on user details tied to each card; however, this would require the consent of the cardholder.

Finally, a smaller scale case study was done in London, United Kingdom, where only five bus routes were selected, and both smart card, or "oyster card," and GPS data were collected in a similar method to Gatineau. This study took a small sample of the population and used ticket counts from electronic ticketing machines to expand them to the entire population. Due to the fact that the ticket counts encompassed a large amount of the population, this was a simple way to gather general information on the entire system. Although this may not be as representative as previous tests, Wang introduced the use of geographical information systems (GIS) to improve bus locations. In most cases, this information is already collected by transit agencies so the data only needs to be account for in the algorithms (Wang, 2010).

2.2 Advantages and Limitations of Smart Card Data and the Extent to Which They Can Be Used with Current Survey Methods

2.2.1 Advantages of Smart Card Data

One of the major advantages of using smart card data over traditional methods is the ability to track data over long periods of time (Trépanier, et al., 2007). With CATI surveys, data are collected in a cross sectional fashion, but with smart cards, data can be collected in a more continuous manner (ex. weeks or months). This expands the temporal coverage available for analysis (Miller, et al., 2012). Therefore, rather than having one O-D matrix that describes behavior on one day, the data can be compared over the entire period of collection (Bagchi & White, 2005).

Next, smart card data allows for access to larger sets of data, compared to traditional methods. With the lack of respondent burden, as with all passive methods of data collection, a larger amount of data are collected. However, smart card data offers the additional ability to collect individual data, as each card has an identifier which is tied to an individual user. This allows trips to be matched to the same user, assuming that these identifiers can be acquired and matched. (Trépanier, et al., 2007)

Furthermore, the smart card infrastructure already exists in cities, meaning these data are readily collectable for analysis. In some locations, penetration rate is extremely high. For example, in Santiago, over 97% of transit users use smart carts as a primary form of payment, providing information on a very large representation of the population (Munizaga & Palma, 2012). In cities where this infrastructure does not exist, there is typically a high initial investment in time and cost for system installation, debugging and full operation. Nevertheless, there is a relatively low incremental cost for additional data collection. (Zhao, et al., 2007)

2.2.2 Disadvantages and Limitations of Smart Card Data

The most obvious disadvantage of passive smart data is that they, by nature, can only be used for transit analysis. Therefore, their applications are limited to public transportation models, requiring pairing with other forms of data collection to provide information on overall city transportation patterns. Although there is some impact on street traffic from vehicles such as buses and streetcars, the primary usage of smart card data is for transit analysis (Trépanier, et al., 2007). In response to this, Bagchi & White (2005) recommended that smart card data be used in conjunction with travel diaries to allow for greater coverage on vehicular traffic. However, data fusion methods are required to correctly combine these data sets. (Miller, et al., 2012)

Another disadvantage of most SCAFC systems is the absence of collection of destination data. Most transit systems have a flat fare rate; as a result, there is no need to collect this type of data as money is collected on the entry tap. Furthermore, this would also require the installation of additional infrastructure at the exits of buses and subway systems, increasing the initial implementation cost of passive smart card systems (Trépanier, et al., 2007). Although it is possible to estimate destination locations from the method described in Section 2, systems that collect alighting information, such as in Brisbane or London, are significantly more accurate in O-D estimation and matrix derivation (Alsger, et al., 2014).

Next, it is important to note that smart card systems have not been implemented or designed for data acquisition, but instead to automate fare collection. Although they provide useful information on traveler patterns, their original purpose was limited to services like revenue accounting. Therefore, many systems may not have all the infrastructure to collect the data necessary for transit planning (Zhao, et al., 2007).

Finally, due to the process of assuming a maximum transfer time, trips with shorter times spent at a destination may be overlooked. As most cards are tapped on only the first leg of a trip, short times spent at a destination may lead to them being ignored and instead counted as a transfer. (Munizaga & Palma, 2012)

2.2.3 Derivation of O-D Matrices

As seen in Section 2.1, origin-destination matrices can be derived from passive smart card data by comparing different trips, and inferring destinations from previous trips in cases where they are not recorded. Although the average transfer time can change from location to location based on the design of the transit system, varying this time in calculations has had minimal impact on the derived matrix (Alsger, et al., 2014). Through the pilot tests completed in Santiago, New York, Brisbane and many other cities, the assumptions listed in Section 2.1 regarding alighting location and returning to the original location have been found generally accurate and passive smart card data have proven to be a useful source of traveler information (Bagchi & White, 2005).

Passive smart card data also allows for the potential of both continuous data collection and updating O-D matrices. As mentioned in Section 2.2.1, not only do SCAFC systems provide spatial coverage but temporal information as well. Because data are constantly being collected as public transportation is being used, they can be analyzed and compared at any point in time (Munizaga & Palma, 2012). This idea applies also to other passive methods and is further discussed in Section 6.2.

2.3 Potential for Smart Card Applications in the GTHA

This section describes the potential for using smart card data as a form of passive data collection in the GTHA. Specifically, the possibility of using PRESTO card data supplied by Metrolinx is explored.

In terms of sources and availability, passive smart card data are accessible, and can only be collected, in cities and areas where the smart card infrastructure already exists. Given this prerequisite, there is a low incremental cost to collect and process data once a proper algorithm is written for analysis. The cost would only vary depending on the frequency of data collection. Additionally, if the agency collecting information does not provide open data to the public, there may be an initial purchase cost for companies wishing to obtain these data (Zhao, et al., 2007).

For cities in which smart card systems do not already exist, an implementation may range depending on city size and transit establishment. For example, in 2013, TransLink in Vancouver started the implementation of their SCAFC system, covering all of 47 rail stations and buses in Greater Vancouver. The total estimated cost of installation over 2700 km^2 area was $194 million in Canadian dollars (TransLink, 2016; TransLink, 2014). However, although there was a high implementation cost for SCAFC, the additional cost for data acquisition and analysis are minimal compared to other methods (Zhao, et al., 2007).

In comparison to Vancouver and other previously mentioned cities, the PRESTO Card system being implemented in the GTHA consists of both areas where alighting is recorded and where it is not. For some GO Transit owned systems, such as the GO Train, users tap their card at both entrance and exit, in order for the company to correctly charge based on distance. In other areas, such as the TTC subways or streetcars, PRESTO Cards are only used as payment upon boarding. This leads to a varying degree of the type of available data, where destination must be estimated for a subset of the trips when deriving O-D matrices. Both of these situations are not new, as previous examples of processing these data have been provided in Section 2.1. However, further research should be done to determine methods of fusing the two sources of data, and the sample size and frame required to provide adequate information for transportation planning.

Usage of PRESTO Card data allows for several specific benefits compared to smart cards in other cities. In terms of the available data, Metrolinx offers them in both raw and processed forms. This allows for more options when combining, processing or analyzing data. For example, should smart card data be used as an addition to other survey methods, the acquisition of raw data would be more useful than already processed data because of their flexibility. Furthermore, the PRESTO Card offers the ability to register a card to a user, allowing agencies to gain some information on traveler demographic. However, due to privacy concerns with sharing this information, certain limitations will mostly like be in place in order to share these data.

Based on this analysis, PRESTO Card data are a highly recommended source of passive data for the next generation TTS as they cover both a large region and a large population. Currently, the GTHA has eleven different transit agencies, each with their own respective fare system. The PRESTO Card aims to connect these transit systems and become their preferred means of payment, a goal Metrolinx hopes to achieve within the next few years (Metrolinx, 2016). Should this be accomplished, PRESTO Card data will be extremely valuable by providing data on the transit users of a large region. Finally, as the smart card in the system already exists in the GTHA, there is no additional cost involved to implement the technology for the purpose of using it for the TTS.

3 BLUETOOTH DATA

As of 2010, over 60% of the world's population owned a cell phone, including almost 100% of those in more developed countries. This high ownership rate indicates a potential for these devices to be tracked for transportation planning uses. Although cellular data are a clear form of collectable information (which will be discussed in Section 4), many of these devices also have Bluetooth capabilities. This opens up another data collection avenue for transit analysis. (Malinovskiy, et al., 2012)

Bluetooth technology is a short-range wireless communication specification which conveys information via Machine Access Control (MAC) addresses. These MAC addresses are 48-bit electronic identifiers that are unique to each digital device. Depending on the power rating of each device, Bluetooth enabled devices can communicate up to a range of 100 meters. Mobile phones can be detected by other transceivers if their system is set to "discovery mode", meaning that they are visible to all other Bluetooth enabled devices. Furthermore, even when connected to a different Bluetooth device, inquiries are still made by the device, which can be both collected and logged by transceivers. (Haghani, et al., 2010; Haghani & Hamedi, 2013)

The general approach in using passive Bluetooth technology to collect transit information is to periodically install transceivers along a roadway or system which detect Bluetooth devices at successive stations. The MAC address of these devices are time stamped and recorded at each location, creating a log of the path taken by the vehicle. Distance travelled could also be measured by taking the gap between the two transceivers. Analysis of the time and distance data would then help determine travel times, space mean speeds and potentially O-D matrices. (Bachmann, et al., 2013; Haghani & Hamedi, 2013)

One key limitation is that Bluetooth must be switched on in order for sensors to detect the device. Unfortunately, this is not always the case as the Bluetooth function is not always needed. A study completed by Haghani et al (2010) found that the proportion of vehicles which contained an active Bluetooth device was less than 4% of the full traffic stream (Haghani, et al., 2010). These numbers, in order to account for the full population, can be adjusted using the *Travel Time Data Collection Handbook* (Day, et al., 2012). Once these conditions are verified, each passenger vehicle essentially becomes a Bluetooth probe vehicle, passively collecting data both spatially and temporally (Bachmann, et al., 2013).

The following sections describe the current use of Bluetooth detectors for traffic analysis around the world, as well as their advantages and limitations. Furthermore, although there have been few prior applications, the potential for using this type of data for the derivation of O-D matrices is also discussed.

3.1 Current Applications of Bluetooth Data Around the World

Many studies done on Bluetooth data have come out of Indiana, from Purdue University, and the National Cooperative Highway Research Program (NCHRP). These studies have investigated the viability and potential for the use of passive Bluetooth data for traffic analysis. Although the goal for most of these systems is not to infer O-D matrices, the methods still provide useful information for transit planners.

3.1.1 Bluetooth Application in Indiana

Of the few origin-destination studies performed, one of them was conducted in West Lafayette, Indiana, in October of 2011. The data collection occurred on a night when the Purdue University football team was playing, monitoring the local fans travelling towards the stadium. Often times it is challenging to collect data for special events or non-recurring activities, but the information is often needed as they may have a substantial effect on the transportation of local or regional areas. (Day, et al., 2012)

To complete this study, Day et al (2012) set up 12 monitoring stations along major corridors to detect passing Bluetooth devices on their way to the game (Figure 3). Data on inbound route choice, outbound route choice, location and magnitude of congestion were measured, processed and documented. By detecting the unique MAC address of each passing device, Day et al (2012) were able to identify the position of each vehicle as they passed by the monitoring stations. O-D matrices were then inferred, with origins and destinations being the regions where the initial and final stations were located.

FIGURE 3. LOCATIONS OF BLUETOOTH MONITORING STATIONS SURROUNDING THE FOOTBALL STADIUM (DAY, ET AL., 2012)

Assuming that the rate of increase, when applying this data to the full population, was roughly equal for all these routes when applying this data to the full population, volume was expected to be proportional to the traffic volumes in the derived O-D matrix (Table 5). Furthermore, it is important to note that the method proposed by Day et al (2012) only applies to the major and most common routes. Additional monitoring stations are required to deduce volumes on the smaller routes. Nevertheless, the proposed method illustrates the effectiveness of using passive Bluetooth technology as a means of collecting data.

TABLE 5. O-D MATRIX GENERATED FROM THE WEST LAFAYETTE FOOTBALL GAME (DAY, ET AL., 2012)

		Destination Station											
	STA	01	02	03	04	05	06	07	08	09	10	11	12
Origin Station	**01**		225	135	63	47	190	5	14	5	14	13	30
	02	89		305	92	42	76	11	11	1	5	13	27
	03	49	199		237	109	23	16	17	6	17	14	17
	04	75	79	263		360	16	18	46	14	69	26	1
	05	90	50	152	455		55	17	35	18	135	47	1
	06	362	123	72	28	47		11	12	8	41	23	11
	07	6	21	25	29	23	11		3	2	4	1	5
	08	14	4	29	72	72	14	1		18	5	41	1
	09	6	1	11	20	27	9	0	26		26	0	0
	10	23	15	33	82	186	44	4	6	9		22	0
	11	56	27	40	78	62	63	1	42	0	23		1
	12	46	34	30	5	0	10	5	0	0	0	1	

Elsewhere in Indiana, Bluetooth technology was also tested for its potential use in measuring travel times and speeds. Before the study completed by Day et al, Haseman & Bullock (2009) had placed Bluetooth sensors along the I-65 highway in Northern Indiana to test their viability in measuring changes in congestion levels. The transceiver units included the Bluetooth sensor, a wireless link antenna and a solar power panel, all connected to the message signs (Figure 3). During a 12-week long period, 1.4 million travel records were detected along the highway. Travel time was estimated by matching the time stamped Bluetooth MAC addresses at each station; the delay data was then displayed in real time to drivers via message signs. This study displayed the potential for continuous data updating with Bluetooth technology. (Haseman & Bullock, 2009)

3.1.2 Bluetooth Application in Montreal and Seattle

The studies completed in Seattle and Montreal differed from Indiana by being based on pedestrian, rather than vehicular, travel and by focusing on dwell time instead of travel time. Two experiments were conducted by Malinovskiy, Saunier and Wang (2012) comparing the two locations in Seattle and Montreal, a school campus, and Chinatown, respectively, to establish proof that Bluetooth technology can effectively track traveller movement.

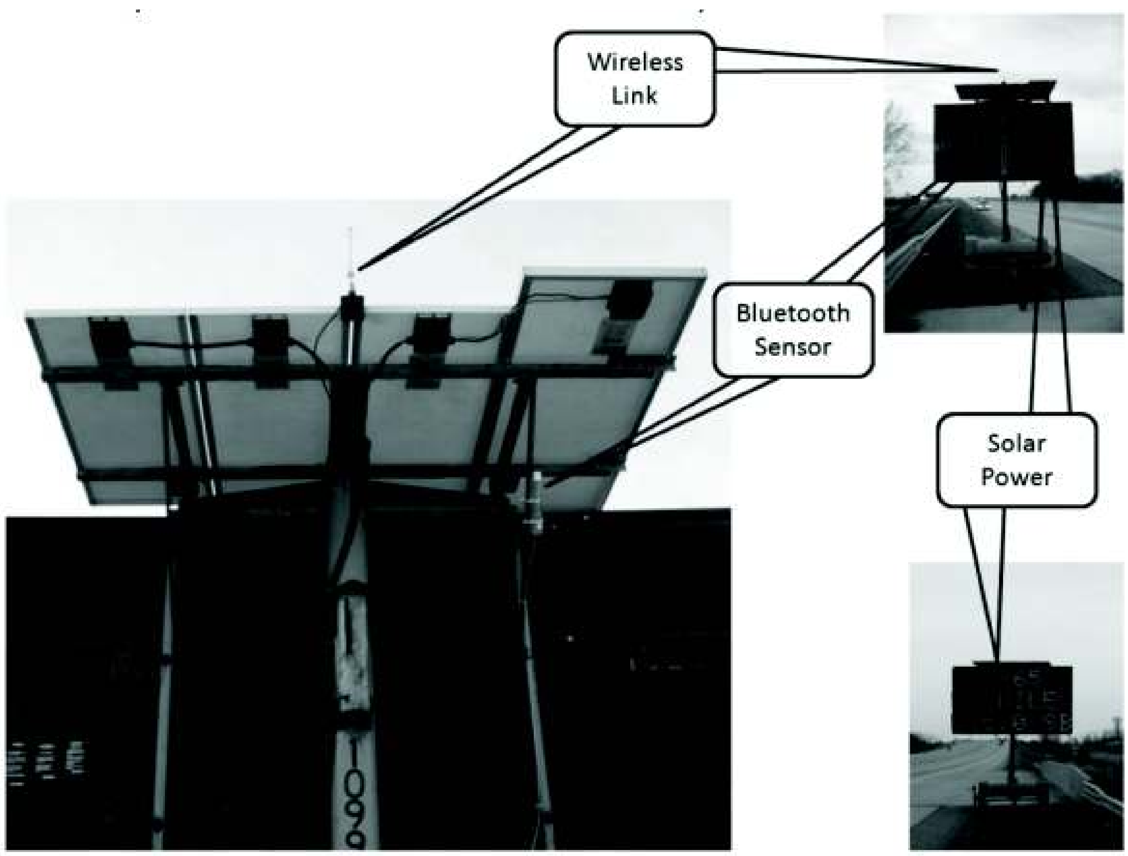

Figure 3. Sample monitoring station set up with a solar power supply (Haseman & Bullock, 2009)

Using omnidirectional antennae, two transceivers were installed at each location at a height necessary to ensure maximum range. Due to the random frequency hopping of the Bluetooth protocol, some time is required in order for devices to be detected by the sensors. Furthermore, obstacles, placement, and other signals can disrupt the frequency and range of each transceiver. Therefore, it was important for the sensors to be mounted at a height of at least 3 meters as well as be clear from obstruction. Additionally, when expanding Bluetooth technology past pedestrians to vehicular transit, expanding the range (by increasing the power level) can increase the probability of detection by increasing the time spent in the range of the transceiver. (Malinovskiy, et al., 2012)

Both studies produced a low average detection of the population when compared to actual pedestrian flows, at roughly 5% and 2.5% in Montreal and Seattle, respectively. This value was lower than the vehicular detection (10%) but it may have been due to the addition of more obstructions blocking the signal in urban areas. This is a key consideration to take into account when applying the concept to cities (Malinovskiy, et al., 2010). In addition, improvements can be made in the process by incorporating analysis of dwell times during generation of O-D matrices; this may provide a better estimate of the time a user spends at a location (Malinovskiy, et al., 2012).

3.1.3 Bluetooth Application in Other Locations

Another study on the accuracy of passive Bluetooth data collection was performed on the I-65 highway. Monitoring stations were installed along 1500 miles of freeway and 1000 miles of arterials in the states of New Jersey, Pennsylvania, Delaware, Maryland, Virginia and North Carolina. The data were collected, processed and stored in a similar manner to Haseman & Bullock (2009) on the I-65 highway, by time stamping Bluetooth detections and using the distance between transceivers to calculate speeds. These data were then compared to floating car[1] data, to decipher the accuracy of speed calculations, resulting in very similar velocities. When compared to actual traffic volumes, it was found that this method only sampled between 2% and 3.4% of all vehicles. From this study, Haghani et al (2010) inferred that passive Bluetooth technology could provide a high quality of ground truth data. (Haghani, et al., 2010)

In Portland, a similar study was completed on a 2.5 mile signalized arterial road over a period of 27 days. By comparing GPS and Bluetooth in this study, Quayle et al (2010) suggested that the larger data set from Bluetooth data more effectively captures performance characteristics. Furthermore, transceivers should be placed in strategic locations to capitalize on the power supply, and intersections should be avoided, as the traffic interactions may affect the resulting data. Several issues, however, were brought up with the popular method of data collection, mostly related to synchronization and detection. These issues included clock synchronization, stationary devices, detection cycles as well as privacy protection; these are further discussed in Section 3.2.2. (Quayle, et al., 2010)

A small study was also done in Toronto, assessing how Bluetooth technology may be used simultaneously with induction loop detectors. Bachmann et al (2013) suggested these technologies as they complement each other in both space and time. Bluetooth technology while accurate, however, has a poor sampling, covering more spatially but less temporally. Comparatively, loop detectors provide large amounts of temporal data as they detect every car that passes, yet spatial coverage is limited by the distance between detectors. (Bachmann, et al., 2013)

Finally, an example of creating O-D matrices from Bluetooth data was presented by Ahmed, El-Darieby, Morgan and Abdulhai (2008) in which multiple transceivers were interlinked via Wi-Fi connection to create a "wireless mesh network." This system would work by having the Bluetooth routers collect time-stamped MAC

[1] Vehicles are equipped with GPS locators which communicate to a server which exacts location, movement and speeds via data processing algorithms (Cohn & Bischoff, 2012)

addresses, and communicate them to other routers via a Wi-Fi connection. Through interlinking the routers, travel paths and, consequently, O-D matrices could be inferred. (Ahmed, et al., 2008)

3.2 Advantages and Limitations of Bluetooth Data and the Extent to Which They Can Be Used with Current Survey Methods

3.2.1 Advantages of Bluetooth Data

One of the major advantages of using passive Bluetooth data is the accuracy of the collection. Due to there being very few parameters considered, such as time and MAC address, little data are lost through processing. This accuracy has been verified when the data has been compared against probe vehicle data. As a result, Bluetooth has the potential to be used as a validation tool for other data collection methods. (Haghani, et al., 2010; Haghani & Hamedi, 2013)

The second major advantage is the relatively low cost associated with passive Bluetooth data collection. The only financial expenses are those suffered during the implementation of collection stations and data processing, as information is collected from devices that users already own (Ahmed, et al., 2008). Additionally, because readers collect data automatically, larger amounts of data can be collected at a low incremental cost. With already existing infrastructure, any additive costs would only be generated from processing and equipment maintenance (Quayle, et al., 2010).

3.2.2 Disadvantages and Limitations of Bluetooth Data

A disadvantage of passive Bluetooth data is that the data collection stations require power to operate for periods longer than 2 to 4 hours. This means that long-term data collection is not possible without a sufficient power supply (Quayle, et al., 2010). This brings up another issue in areas with interrupted traffic flow such as cities or urban spaces. In these areas, the most logical location for the transceivers is at intersections, to take advantage of existing power supply. However, this results in data that are heavily influenced by traffic at intersections and signal timings. Therefore, ideally, Bluetooth data collection stations should preferably be placed in mid-block locations to acquire higher quality data. Because of this issue, it is difficult to find adequate locations for transceiver placement (Day, et al., 2012).

Another limitation of passive Bluetooth data is the inability to collect vehicle trajectories and point speeds at each station. This is because Bluetooth technology cannot accurately detect the distance of a device from the transceiver, only whether it is within range and the signal strength. In O-D derivations this has little impact; however, it is a limiting factor in other transport applications. Furthermore, Bluetooth technology also requires slightly more complex statistical analysis methods to filter outliers as multiple points may be collected at the same position. This makes it difficult to determine whether there is congestion at this point or the device was logged close to both entering and exiting the transceiver's detection range. (Quayle, et al., 2010)

Compared to other passive methods, Bluetooth also has had a relatively lower penetration, ranging from 2% to 8% representation of all vehicles in 2009 (Haseman & Bullock, 2009). Although many users carried cell phones, very few had Bluetooth capabilities enabled, simply because they did not have any devices of which to connect; however, more recent and Toronto specific data on penetration is needed.

3.2.3 Derivation of O-D Matrices

There have been some examples of O-D derivation from the studies analyzed in Section 3.2 which may be useful to the TTS 2.0. Day et al (2012) created an O-D matrix using Bluetooth transceivers located at major routes surrounding a football game in West Lafayette. Using the flows to and from the game, MAC addresses were recorded and logged, later to be processed. Using this process, an O-D matrix was formed describing the travel patterns during a special event, a technique that can be expanded for use in other situations.

Although this study was done in a small area, the scope can be broadened and developed to cover a larger area. This, however, would increase the cost of implementation as many more transceiver stations would need to be installed. Furthermore, the study only covered the major routes such as highways and arterials. For a more comprehensive data set, monitoring stations should be sparsely set up along smaller roads as well.

The end result would be similar to a method used by Ahmed et al (2008), who proposed a "wireless mesh network" to collect O-D data from Bluetooth devices. This network would be connected though a Wi-Fi network, having the ability to gather information in real time for transportation services. This "mesh network" would have transceivers spread out and interconnected to create a "net," which would gather the required information for origin-destination data for processing.

3.3 Potential for Bluetooth Applications in the TTS

This section analyzes the potential for the use of Bluetooth technology in collecting passive data for transportation planning. Although strengths were found, such as high accuracy of detection, it also has drawbacks in both cost and power requirements. Due to this and a lack of prior applications in the GTHA, it is may not be a viable option for the TTS.

First, the main expenditures would be from data collection and power generation. As noted in Section 3.2.1, due to the fact that Bluetooth readers collect data automatically, a large amount of information can be collected at a low incremental cost. Furthermore, as previously mentioned, power generation is a large issue with Bluetooth technology, being depending on electricity to run the interconnected system of both Wi-Fi and Bluetooth (Quayle, et al., 2010). However, locations where electricity is not readily available, power generation costs can be reduced through the use of solar panels (Figure 3) (Day, et al., 2012; Haseman & Bullock, 2009).

Currently, there are solar powered sign boards used for displaying traffic messages at congested intersections or highway exits. Bluetooth stations could be powered in a similar fashion, using mostly solar power to run. Traffic & Parking Automation North America Inc., a Bluetooth traffic monitoring company, estimates that the spacing requirement between stations is approximately 250 meters. In order to cover a smaller area such as only the city of Toronto (an estimated area of 630 km^2), about 10,000 stations would be required. However, although it is known that this technology tends to be more expensive, pricing is not available at this time so further investigation and experimentation are required to determine its viability. (TPA North America Inc., 2016; Bruce, 2016) (Bruce, 2016)

Overall, due to its high accuracy in detection, Bluetooth is best implemented in more densely populated locations such as downtown or largely inhabited areas. This allows for better detection of devices and more accurate estimations of origins and destinations. It also limits the number of stations required, therefore reducing operating costs. However, the application of passive Bluetooth technology as the main source of data collection is unlikely as, in general, installation costs would most likely be too great for it to be a viable option.

4 CELLULAR DATA

When in search of methods to collect passive data for transportation analysis, cellular data provides a reliable source for data collection on a large sample of the population. As with Bluetooth technology, cellular telecommunication has a high potential for information collection due to the immense penetration of cell phones in today's population. (Wang, et al., 2013)

Passive cellular data works by collecting time-stamped traces of mobile signals via cell towers. These towers have an average radial detection range of approximately 320 meters, creating a cell. However, there is some overlap between each individual cell where multiple towers may be in range. When a phone is used within a cell, it sends a signal to the nearest tower, which gives an estimate of the approximate location of the phone within that area (Wang, et al., 2013). These estimates may range in accuracy from 500 meters in urban areas to 2.5 kilometers in rural areas (Horn, et al., 2014).

Unfortunately, data generation through this method is limited, as it can only be collected during specific times:

- **Calling**. When the user is using their mobile device to call, their location can be tracked with the towers for the duration of the call. In some cases, only the start and end times of the call are recorded (White & Wells, 2002). In these cases, location is also recorded when a phone crosses a cell boundary and the signal is passed from one tower to another, called a **handover** (Bekhor & Shem-Tov, 2015).
- **Texting**. Whenever the user sends or receives a text, location data are sent to the cell towers. (Calabrese, et al., 2011)
- **Data Usage**. By using mobile data to connect to the internet from cell phones, the devices send and receive signals from the towers, which log time and location. (Calabrese, et al., 2011)
- **Paging**. If no data has been recorded for a while (ex. 2 hours), the company sends out a signal to record the position of the device (Bekhor & Shem-Tov, 2015).

Although continuous data collection is possible and is the ideal result, it is usually not implemented as the information is not necessary for the purpose of the cellular company. As a result, continuous data are not readily available for transportation analysis (White & Wells, 2002).

When using cellular communication for passive data analysis, time logged location data are collected from the towers and processed for trip generation and O-D matrix derivation. Using the time and approximate location information of each cell phone, vehicle numbers and users can be tracked in their daily travel patterns. Furthermore, due to the continuous nature of data collection, a key objective is to automatically, and in real-time, create O-D matrices to provide immediate response to traffic problems. (Caceres, et al., 2007)

Finally, data collection is inexpensive, as the cellular infrastructure already exists and cell phones are immensely popular. The only costs are for the acquisition and processing of data (Caceres, et al., 2007). However, due to privacy concerns, additional data may be introduced to obscure information, including encrypted identifier numbers (White & Wells, 2002) or random network switching[2] (Horn, et al., 2014).

4.1 Current Applications of Cellular Data Around the World

Cellular networks around the world rarely differ greatly from the system described in the prior section. The main differences are the strength of cell signals and the cellular range of the cell towers (Andrews, et al., 2011). However, there have been some differences in the methods used for analysis of the collected data, or

[2] Temporary mobile subscriber identity (TMSI) randomly switches the network so that a user cannot be harmfully tracked (Horn, Klampfl, Cik, & Reiter, 2014)

for separating areas covered by each cell tower. Studies on cellular network systems for transportation planning completed in Israel, North Carolina, and Massachusetts have provided insight on the potential uses of passive cellular data. Other studies from China, Kansas, Vienna and Austria also provide valuable information. These studies are discussed in the following sections.

4.1.1 Cellular Applications in Israel

During a 16-week period from March 7th to July 2nd in 2007, Bekhor & Shem-Tov (2015) completed a study on using passive cell phone data to estimate travel patterns. Over 4 million records from 10,000 devices were collected weekly, with the sample set changed at the end of each week. In this study, locations were tied only to towers, rather than specific locations within the tower area. Therefore, the O-D data were validated at an aggregate level, with cell users being associated with the tower that was currently serving each respective device. (Bekhor & Shem-Tov, 2015)

In terms of data collection, the raw information that was collected included a sensor identifier (separated from the personal phone number), location of the cellular antenna serving the phone at that moment, and the time. In addition, if the device had not sent or received a signal in 2 hours, a locater signal was sent out by the company to determine its whereabouts. (Bekhor & Shem-Tov, 2015)

Due to locations being tied directly to towers, several issues were brought up regarding accuracy and a "zig-zag" phenomenon. Due to this low spatial accuracy, locations were taken and recorded as the antenna location, rather than the point of the cell phone itself. Although data was available on the distance that a device would be from said antenna, this was inaccurate and therefore not used. Data were aggregated into locations based on towers themselves, rather than specific locations within the tower areas. However, although this provided more confident data sets, the drawback of doing this was that precise locations could not be given as origins and destinations. Next, the handovers that occur when a phone moves from one area to another also caused problems. This was especially true when there was a large volume of cell phones tied to one tower, at which point handovers would occur to spread the load out amongst adjacent towers. When collecting data recorded to each antenna, this resulted in a "zig-zag" movement pattern, even if the user was not moving. To adjust for this, handovers with a circular movement, meaning those which return to the same adjacent tower, were recommended to be ignored when analyzing data. (Bekhor & Shem-Tov, 2015)

Finally, Bekhor & Shem-Tov (2015) proposed methods, related to time spent in each zone, for differentiating between different modes of travel. However, due to the low accuracy of cellular data within each cell tower area, it was very difficult to estimate slower modes of transportation such as walking or biking. As these modes spend longer times connected to a single antenna, it was challenging to determine if the user is staying put or simply travelling at a slower pace. However, when determining O-D matrices, this information is not as important, but is still needed for trip generation. (Bekhor & Shem-Tov, 2015)

4.1.2 Cellular Applications in North Carolina

A similar study was completed in Raleigh, North Carolina in which the Capital Area Metropolitan Planning Organization (CAMPO) used cellular data for speed and travel time validations. CAMPO, similar to many other American organizations that study cellular applications, established a contract with AirSage to collect data. AirSage, with agreements through specific carriers (specifically Sprint Mobile), provided cellular data which were then extracted, aggregated and mapped, while removing personal information. (Huntsinger & Donnelly, 2014)

In their study, Huntsinger & Donnelly (2014) compared passive cellular data from AirSage with the data acquired from the Triangle Regional Model (TRM), an advance trip-based model estimated and calibrated from travel survey data collected in 2006. The data were collected in a similar manner to the method in Israel, varying only by the use of traffic analysis zones (TAZs) rather than tower cells. TAZs are simple areas defined by population, demographics or geography (TMIP, 2010). By using these areas instead of cells defined by towers, in combination with other data sources, Huntsinger & Donnelly (2010) were able to infer information regarding traveler details and demographics. The end result showed that the AirSage data were similar and comparable to the TRM; however, there was a slightly smaller sample volume for the former.

4.1.3 Cellular Application in Other Locations

Aside from Israel and North Carolina, passive mobile data collection has also been used in cities such as Kansas, Boston, and Vienna. The methods used in these locations differed slightly, but were similar in that they illustrated different ways that cellular data could potentially assist in data collection for transportation planning.

In 2013, Wang et al completed a study tracking data from the Kansas Metro corridor, a series of highways that connect Topeka, Lawrence and Kansas City. This experiment addressed the fact that location derivation from cell phone data have low spatial accuracy. To correct this issue, the researchers suggested that this type of data be used for longer distances, such as between cities. Furthermore, they recommend that a longer observation period and a large sample size are required to reduce bias and obtain more stable data. With this larger time frame, a "dynamic" OD matrix could be formed, changing the length of time depending on what the study demands. (Wang, et al., 2013)

Finally, a similar study was completed in Boston, where movement of cell phones was passively analyzed in eight counties in Massachusetts, totaling over 5.5 million people (Calabrese, et al., 2011). As with Wang et al (2013), the study used a large sample size and long observation period, collecting data over the entire day. Furthermore, by grouping trips with the same destination matrices at different temporal windows (ex. hourly, daily, weekly, etc.), different matrices could be generated for specific cases (Calabrese, et al., 2011).

4.2 Advantages and Limitations of Cellular Data and the Extent to Which They Can Be Used with Current Survey Methods

4.2.1 Advantages

As previously stated, cell phone data has the highest penetration rate of passive devices. This is due to the fact that, in developed countries, an extremely large percentage of the population have a cell phone in their possession. This, in turn, leads to a large sample population from which travel data can be collected. Furthermore, since cell phones are tied to individuals, these data can be used for both vehicular and traveler tracking. In contrast, a method such as smart cards can only collect data from public transit. Therefore, the ability to track all people independent of their mode of travel is a valuable attribute that can be harnessed by transportation analysts. (Wang, et al., 2013; Bekhor & Shem-Tov, 2015)

In addition to high penetration, cellular data also provides the potential to continuously update O-D matrices. This potential comes from the somewhat continuous acquisition of data, as cell phones collect location information whenever they are used. Furthermore, as data are collected at all times of the day, depending on when the devices are used, this expands the possible time frames for which data can be collected and analyzed. (Wang, et al., 2013) (Caceres, et al., 2007)

4.2.2 Disadvantages

However, despite cellular data having strengths in both sample size and temporal range, they lack in accuracy, providing little spatial resolution. While cell phone towers, with their locations fixed, allow for precise knowledge of their locations, the estimated location of the cellular devices within the tower area are not accurate. However, this issue can be somewhat mitigated with the use of multiple towers. If a cell phone is within range of more than one tower, the location can be better estimated based on an average of each reading. However, this process requires greater amounts of data processing and, therefore, can be more expensive. (Calabrese, et al., 2011) (Ward, 2011)

In addition, although cell phones are more ubiquitous than other forms of passive sensors such as Bluetooth stations or GPS devices, their use is spread across service providers. It is possible that this may introduce some demographic bias, possibly due to differences in advertising approach or age of the company, but further research is required to investigate the extent of this bias. Therefore, in order to capture a representative population, data must be acquired from additional providers, increasing the cost. (Ward, 2011)

4.2.3 Derivation of O-D Matrices

A study was performed in Boston, where passive cellular data was used for the derivation of O-D matrices (Calabrese, et al., 2011). By using location data determined from cell towers, origins, destinations and trips were estimated. To account for the low spatial accuracy of this type of data, a one-kilometre radius was used for each tower area when determining the location of devices. This meant that if a user stayed within an area of one-kilometre from when the towers first estimated their location, they were recorded at the location of the tower. This resulted in trip origins and destinations being generalized to tower areas, which had a drawback of missing the smaller trips. Data were then collected from the cell towers and processed in similar methods to other passive sources to create O-D matrices (Calabrese, et al., 2011). Although cellular data has been used to track the movement of users elsewhere, to develop O-D matrices, most other studies, such as the one completed by Bekhor & Shem-Tov (2015) in Israel, follow the same method as used in Boston.

4.3 Potential for Cellular Applications in the GTHA

This section analyzes the potential for the use of cellular technology in collecting passive data as a part of the TTS.

As Toronto, and Canada in general, is a highly developed area, cell phones are a very common possession; therefore, cellular data are more prevalent in the GTHA. In addition, there are many different carriers available, allowing for more options to source this data. Because of these two factors, cellular data potentially available in the GTHA are less likely to exhibit large demographic bias.

As mentioned in Section 4.1.3, cellular data excels in longer distance travel, due to their low spatial accuracy. By taking data at locations that are further away from each other, this issue can be ignored, at the cost of less fine data sets. (Wang, et al., 2013) Due to the fact that much of the GTHA is made up of people who commute from smaller urban regions, such as Mississauga and Scarborough residents working in downtown Toronto, these conditions are ideal for the use of cellular data. By setting up multiple Bluetooth detection stations in each urban region and several more downtown, data can be collected on daily travel patterns from a lot of commuters.

Next, an issue that should be addressed when considering cellular data as part of the TTS is any bias towards certain cell phone carriers based on age or demographic. For example, larger companies such as Bell and Rogers provide high-quality data service and networks but may cost more per month to use. In contrast, other companies such as Wind and Koodo offer cheaper plans but with less serviceability and network stability. Factors that may be affected by this range in carrier traits could be age or income. However, this is a topic that must be further researched.

POTENTIAL OF PASSIVE DATA SOURCES

Cellint is one provider of traffic flow and origin-destination data who could be considered as a source of cellular data for the TTS, with a contract with Rogers to get access to their location data. Although pricing was not available, as they differ by the situation, Cellint may prove to be a useful source of passive travel data. In contrast to some of the other methods described in Section 4.1, this company attempts to increase accuracy by providing street level precision by detecting devices using multiple towers. Furthermore, they offer processed data integrated into service modules such as Google Maps and Google Earth. Additionally, Cellint also offers a platform called NetEyes, which helps manage and distribute data, displaying it through reports and excel spreadsheets. (Cellint Traffic Solutions, 2016)

Great potential currently exists for the use of cellular passive data in the GTHA due to the high penetration of cell phones and the abundance of available data. Should this source be used, most likely as a supplement to other survey methods, further research needs to be performed to determine the sample size required, which cell providers should be used and ways to fuse these data with other sources.

5 OTHER SOURCES OF PASSIVE DATA

Aside from smart cards, Bluetooth, and cellular data, there are many other forms of passive data collection that may prove useful for the TTS 2.0 to use in future surveys. However, there is limited information from past studies regarding these methods. One of these sources that is discussed in this section is GPS data. However, due to the limited examples available, no recommendation can be made the time on its possible application in the TTS.

5.1 GPS Data

With the current growing advancement in technology, it is clear that GPS, or global positioning system, is one traffic data collection method that should be examined. There are many types of GPS surveys out there, including on-person or wearable technologies, which are often combined with travel diaries. This section, however, focusses only on passive forms of data collection.

Unlike past technologies, GPS is more of a continuous source of data, with information often being collected at very short time intervals. It is also highly accurate, using satellites to triangulate positions. Furthermore, when compared to traditional travel survey methods where self-reported trips are underreported, GPS captures 1.29 times more trips. This demonstrates that these data can be used for ground truth analysis and provide information on the revealed preferences of travelers. (Dumont, et al., 2012)

In the GTHA, and North America in general, several companies such as INRIX and Tomtom offer GPS products which can provide real-time traffic updates for travellers or transportation agencies. These devices provide the user with travel routes, while also collecting information on the location and speed of these travellers. This information can then be used to assist transportation planning. (INRIX Inc, 2016) (TomTom International BV, 2016)

GPS, as with other technologies has its own benefits and drawbacks. One benefit is that it is highly accurate and provides many waypoints. This allows for precise trip tracking with the additional waypoints helping to better determine stops and destinations. However, as it currently stands, GPS data are heavily biased towards freight vehicles. Additionally, there is limited tracking over long periods of time due to the nature of the data, namely being privatized and scrambled for privacy reasons. (Hard, et al., 2016)

Studies on passive GPS data are limited but they show that the technology has great promise in the tracking of taxi travel patterns. Due to the fact that many taxi companies track their fleets through GPS, this data can be useful in trip generation and determine congestion patterns. In a study done in Berlin, Nuremberg and Vienna, GPS data from taxi fleets were treated as floating cars and analyzed to develop O-D matrices (Schäfer, et al., 2002). These examples display that, although it has not been as commonly used, GPS data have potential and further research should be done to investigate the extent and viability of their use.

6 CORE-SATELLITE APPROACH

6.1 The Role of Passive Data in the Core-Satellite Approach

The core-satellite approach is a data collection paradigm where data are collected and combined from different sources. It aims to be flexible, accepting a multitude of data types, and applying data fusion methods to create a cohesive platform of data. More important data sources are considered the "core", while less important sources are considered as "satellites". For example, if telephone surveys were used to collect primary information such as trips and paths taken while mail-out surveys were used for gathering personal information, the combination of these two survey methods and data sets would describe the core-satellite paradigm. (Miller, et al., 2012)

However, the role that passive data plays in this approach is not as defined and can change depending on the extent of data collection. For example, depending on the penetration or impact of the passive technology, it may be possible that enough data can be collected for it to be considered a "core" approach. However, this may not always be the case, and researchers may wish to use passive data to instead fill gaps in other surveying methods. For example, if their primary method has a deficiency in collecting data from public transit, the addition of smart card data as a satellite source would be of interest to them.

For passive data sources be used in the TTS 2.0 core-satellite approach, additional research needs to be done on methods of data fusion, and the extent to which each particular passive data source can be used. Some sources such as smart card data and cellular data have great potential as a satellite source, if not a core data source, whereas Bluetooth data may only be used as a satellite source in specific circumstances.

6.2 Using Passive Data to Update Origin-Destination Matrices

Due to the continuous nature of most passive data sources, there is potential to use this type of data to update already existing O-D matrices. These matrices need not be originally formed from passive data, and the ability to update matrices with this type of data would allow for more up-to-date traffic analysis. Using current survey methods provides no temporal data as it provides a cross-sectional representation of data at the time it is collected. Additionally, it takes time and resources to collect information and process it, delaying the time between initial data collection and final analysis. On the other hand, with passive data collection methods, the time stamps of collection events are recorded, allowing for temporal analysis.

There have been several instances where passive data was used to update existing origin-destination matrices. In New York, data were collected from smart cards and were used to create and update trip tables. The original matrices were created as usual, by estimating origins and destinations based on previous assumptions. To update these matrices, Barry et al simply repeated the test at different time frames, taking advantage of the data that was continuously available. By doing this, the researchers were able to determine different trends in travel patterns (Barry, et al., 2001). GPS data was also used to continuously update O-D matrices in a study done with taxi fleets in Portugal. In this case, an algorithm was developed to sort and process data into trip tables and additional data was incrementally layered on top to update it (Moreira-Matias, et al., 2016).

Although these methods have been used to update existing matrices, the original matrices were generated using the same type of data (i.e. smartcard, GPS, cell data). Should passive data be used to update other survey methods, additional research must be done to discover methods for data fusion.

7 CONCLUSIONS

This report aimed to investigate the potential for the use of passive data sources to replace or assist in the acquisition of traveler data for incorporation in future TTS. The use of passive data collection technology has intrinsic benefits, having little respondent burden and the potential to continuously update O-D matrices. With 'passive' meaning that the data are collected without the awareness of the user (complying with privacy concerns), by nature, there is very little burden on the respondent when providing data. It also allows for the acquisition of "revealed preference" as opposed to the "stated preference" that is supplied by current survey methods (Miller & Habib, 2015). Furthermore, the issue of collecting data by proxy is eliminated as data are collected separately for each individual.

The main passive data collection methods that were researched were smart cards, Bluetooth and cellular, with some study done on GPS technologies. Each of the major methods studied was found to have its own benefits, as well as drawbacks, as outlined in Table 6.

TABLE 6. BENEFITS AND DRAWBACKS OF MAJOR PASSIVE TECHNOLOGIES

Passive Technology	Benefits	Drawbacks
Smart Card	-Readily available in the GTHA -Data are available in both processed and unprocessed forms -tracks data over long periods of time -very large data sets	-can only analyze transit systems -requires some assumptions in determining destinations and transfer times -merging different types of data from PRESTO cards
Bluetooth	-high accuracy of location -low incremental cost to operate	-data collection stations require power to operate -high initial costs as it is not currently existing -difficult to use in larger regions
Cellular	-extremely high penetration -strength over large regions -continuous data collection	-dependent on cell phone use -low spatial accuracy -the source of data is not uniform (i.e. different service providers)

Due to the continuous nature of passive data, there is the potential to use this data to update existing O-D matrices. This was shown in data gained using smart card technologies in New York (Barry, et al., 2001) and Bluetooth technologies in Indiana (Haseman & Bullock, 2009). With this potential revealed, it may be possible to use passive data to update O-D matrices, should they not be used as the main source of information collection. However, additional research is required to develop methods of fusing this data with other sources.

One of the more promising passive technologies for use in the next TTS is smart card technology. This is largely due to the fact that the GTHA transit systems will have a smart card system, the PRESTO card, in place. Although the region is made up of eleven different transit agencies, each with their own fare system, Metrolinx aims to have the PRESTO Card as a uniform payment method for all of them before the 2021 TTS

(Metrolinx, 2016). Should they achieve this goal, data collected from PRESTO will provide a large and valuable data set of the travel of transit users in the GTHA.

After looking at examples from other cities such as New York, Santiago, and Chicago, smart cards have shown a potential to gather large amounts of passive data. These systems revealed the strengths of using smart card data, namely that the required infrastructure is often in place, and that they allow transit agencies to track travel patterns over long periods of time. Due to this latter reason, data can be compared between multiple time frames and over longer periods of time than traditional survey methods (Trépanier, et al., 2007). However, although smart cards have a great amount of potential in the GTHA, its biggest drawback is that it is limited to only public transit services and is difficult to extend to other modes. Due to this, many researchers recommend that it be combined with other survey methods, such as travel diaries (Bagchi & White, 2005). Overall, should passive data be used in conjunction with the TTS 2.0, the PRESTO card shows the most potential out of all the researched technologies. With regards to the other two researched technologies, each has their own strengths and weaknesses.

Bluetooth technology is arguably the most accurate of the three, having smaller detection ranges and higher detection rates. However, in order to maximize its effectiveness, the stations must be placed closer together in order to limit the areas where there is no coverage (Day, et al., 2012). This leads to an increase in cost, especially if this technology is used in a large area, such as the GTHA. There is the potential to use it in smaller dense areas such as downtown Toronto; however, a large number of data collection stations would still be required. Therefore, it is not recommended that this type of passive technology be used as part of the TTS. However, should data fusion methods be further researched, it may be possible to use this technology in more focussed satellite collection efforts.

In contrast to Bluetooth, cellular technology is less accurate but works better over larger areas (Calabrese, et al., 2011). Even with the issues in accuracy, it is more appropriate for acquiring data over a large region like the GTHA. Furthermore, due to its extremely high penetration rate, this technology offers a large data set that can be collected and studied over large periods of time. However, the drawback of cellular data is that it lacks accuracy, providing only approximate locations of users within tower ranges (Wang, et al., 2013). While some attempts have been made to mitigate this hindrance by collecting data from multiple towers, coupled with cellular signal traces of routes, the accuracy and the acquisition cost of such data needs further investigation.

Moving forward, the greatest potential of incorporating passive data in future TTS lies with the use of smart card data. This is due to the fact that the PRESTO card is already in use. Furthermore, although the cost of acquisition is unknown, this data would encompass transit users in a very large area of the GTHA. With regards to the core-satellite design, the PRESTO card data can provide some key travel information, such as trip patterns, origins, and destinations. Nevertheless, several steps are needed to further investigate and incorporate this data. Similar steps would be required for analyzing the use of cell phone data:

- Acquire PRESTO card data from Metrolinx and determine its usability for tracking travel paths
- Research and determine algorithms for sorting data and creating matrices that are reusable for future data sets
- Explore the geographic extent to which the data covers and what additional data are required
- Research possible forms of bias that may exist with PRESTO card data
- Investigate data fusion methods to fill gaps in other methods and vice versa
- After experimentation, identify any issues with the collection and processing of smart card data, and research ways to mitigate these problems

After comparing the three different technologies, it can be concluded that although there is potential for passive data to be used in the TTS 2.0 via the derivation of O-D matrices, the accuracy of such methods is an issue. For example, due to the number of assumptions required when using smart card data, such as information only available regarding boarding, alighting stations and locations need to be inferred. Furthermore, with cellular technology, large amounts of location data can be gathered, although it may not be as precise as other collection methods. On the other end of the spectrum, Bluetooth data are highly accurate but are more expensive. Although Bluetooth may be more accurate, other technologies such as smart cards and cell phones are cheaper and more feasible in the context of the TTS 2.0. Additionally, there may be other technologies such as GPS that can provide sources of passive data but few examples exist at this time.

8 BIBLIOGRAPHY

Ahmed, H., EL-Darieby, M., Morgan, Y. & Abdulhai, B., 2008. *A Wireless Mesh Network-based Platform for ITS,* s.l.: IEEE.

Alsger, A. A., Mesbah, M., Ferreira, L. & Safi, H., 2014. *Public Transport Origin-Destination Estimation Using Smart Card Fare Data,* s.l.: s.n.

Andrews, J. G., Baccelli, F. & Ganti, R. K., 2011. A Tractable Approach to Coverage and Rate in Cellular Networks. *IEEE Transactions on Communications,* 59(11), pp. 3122-3134.

Bachmann, C., Roorda, M. J., Abdulhai, B. & Moshiri, B., 2013. Fusing a Bluetooth Traffic Monitoring System With Loop Detector Data for Improved Freeway Traffic Speed Estimation. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* May, 17(2), pp. 152-164.

Bagchi, M. & White, P., 2005. The potential of public transport smart card data. *Road User Charging: Theory and Practices,* 12(5), p. 464–474.

Barry, J. J., Newhouser, R., Rahbee, A. & Sayeda, S., 2001. Origin and Destination Estimation in New York City with Automated Fare System Data. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 1817.

Bekhor, S. & Shem-Tov, I. B., 2015. Investigation of travel patterns using passivecellular phone data. *Full Terms & Conditions of access and use can be found athttp://www.tandfonline.com/action/journalInformation?journalCode=tlbs20Download by: [University of Toronto Libraries]Date: 29 June 2016, At: 11:04Journal of Location Based Services,* 9(2), pp. 93-112.

Bruce, R., 2016. *Potentials of Bluetooth Technology*. Toronto: s.n.

Caceres, N., Wideberg, J. & Benitez, F., 2007. Deriving origin–destination data from a mobile phone network. *IET Intelligent Transport Systems* , March, 1(1), pp. 15-26.

Calabrese, F., Lorenzo, G. D., Liu, L. & Ratti, C., 2011. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing,* 10(4), pp. 36-44.

Cellint Traffic Solutions, 2016. *NetEyes*. [Online]
Available at: http://www.cellint.com/neteyes/
[Accessed 3 September 2016].

Chapleau, R., Trépanier, M. & Chu, K. K., 2008. The Ultimate Survey for Transit Planning: Complete Information with Smart Card Data and GIS. *Data for Public Transit Planning, Marketing and Model Development.*

Data Management Group, 2014. *Transportation Tomorrow - Design and Conduct of the Survey*. [Online]
Available at: http://dmg.utoronto.ca/pdf/tts/2011/conduct2011.pdf
[Accessed 30 May 2016].

Day, C. et al., 2012. *Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data,* West Lafayette: s.n.

DICTUC, 2003. *Actualización de encuestas Origen Destino de viajes, V Etapa*. s.l.:Informe Final a Sectra.

Dumont, J., Shalaby, A. & Roorda, M. J., 2012. A GPS-aided survey for assessing trip reportingaccuracy and travel of students without telephoneland lines. *Transportation Planning and Technology,* 35(2), pp. 161-173.

Esomar World Reserach Codes & Guidelines, 2009. *Passive Data Collection, Observation and Recording.* [Online]
Available at: https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Codes-and-Guidelines_Passive_Data_Collection-Observation-and-Recording.pdf
[Accessed 31 May 2016].

Griffiths, R., Richardson, A. & Lee-Gosselin, M. E., 2000. Travel Surveys. *Transportation in the New Millenium.*

Haghani, A. & Hamedi, M., 2013. Application of Bluetooth Technology in Traffic Detection, Surveillance, and Traffic Management. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* May, 17(2), pp. 107-109.

Haghani, A. et al., 2010. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 2160, pp. 60-68.

Hard, E., Chigoy, B. & Songshitruksa, P., 2016. *TMIP Webinar.* s.l.:s.n.

Haseman, R. J. & Bullock, D. M., 2009. *Real Time Measurement of Work Zone Travel Time Delay and Evaluation Metrics Using Bluetooth Probe Tracking,* Indianapolis: Purdue University.

Horn, C., Klampfl, S., Cik, M. & Reiter, T., 2014. Detecting Outliers in Cell Phone Data: Correcting Trajectories to Improve Traffic Modeling. *Transportation Research Record: Journal of the Transportation Research Board,* pp. 49-56.

Huntsinger, L. F. & Donnelly, R., 2014. *Reconciliation of Regional Travel Model and Passive Device Tracking Data,* Washington: Transportation Research Board.

INRIX Inc, 2016. *INRIX Analytics.* [Online]
Available at: http://inrix.com/products/
[Accessed 10 October 2016].

Malinovskiy, Y., Saunier, N. & Wang, Y., 2012. Analysis of Pedestrian Travel with Static Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 2299, pp. 137-149.

Malinovskiy, Y., Wu, Y.-J., Wang, Y. & Lee, U. K., 2010. *Field Experiments on Bluetooth-Based Travel Time Data Collection* , Washington: Transportation Research Board.

Ma, X.-l., Wang, Y.-h., Chen, F. & Liu, J.-f., 2012. *Transit smart card data mining for passenger origin information extraction,* s.l.: University of Washington.

Metrolinx, 2016. *Presto.* [Online]
Available at: http://www.metrolinx.com/en/projectsandprograms/presto/presto.aspx
[Accessed 9 September 2016].

Miller, E. J. & Habib, K. M. N., 2015. *TTS 2.0: An R&D Program to Develop the Next Generation Passenger Travel Survey Methods for the Greater Golden Horseshoe,* Toronto: Data Management Group.

Miller, E. J. et al., 2012. *Changing Practices in Data Collection on the Movement of People,* s.l.: Lee-Gosselin Associates Limited.

Moreira-Matias, L. et al., 2016. Time-evolving O-D matrix estimation using high-speed GPS data streams. *Expert Systems with Applications,* February, Volume 44, pp. 275-288.

Munizaga, M. A. & Palma, C., 2012. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies,* October, Volume 24, pp. 9-18.

Munizaga, M., Palma, C. & Mora, P., 2010. *Public Transport OD Matrix Estimation from Smart Card Payment System Data.* Lisbon: s.n.

Pelletier, M.-P., Trépanier, M. & Morency, C., 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies,* August.14(4).

Quayle, S., Koonce, P., DePencier, D. & Bullock, D., 2010. Arterial Performance Measures with Media Access Control Readers: Portland, Oregon, Pilot Study. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 2192, pp. 185-193.

Schäfer, R.-P., Thiessenhusen, K.-U. & Wagner, P., 2002. *A Traffic Information System by Means of Real-Time Floating-Car Data,* Berlin: Research Gate.

Société de transport de l'Ouaouais, 2012. *Area served.* [Online]
Available at: http://www.sto.ca/index.php?id=services&L=en
[Accessed 17 June 2016].

TMIP, 2010. *Defining Traffic Analysis Zones.* [Online]
Available at:
http://www.fhwa.dot.gov/planning/tmip/publications/other_reports/technical_synthesis_report/page01.cfm
[Accessed 4 July 2016].

TomTom International BV, 2016. *Navigation - Get there faster.* [Online]
Available at: https://www.tomtom.com/en_ca/drive/car/
[Accessed 10 October 2016].

TPA North America Inc., 2016. *BlueFax Traffic Monitoring System.* [Online]
Available at: http://tpa-na.com/traffic_monitoring.html
[Accessed 23 August 2016].

TransLink, 2014. *Compass Project Update: TransLink to deliver Compass Card to post-secondary students.* [Online]
Available at: http://www.translink.ca/en/About-Us/Media/2014/October/Compass-Project-Update.aspx
[Accessed 21 June 2016].

TRANSlink, 2016. *About go card.* [Online]
Available at: https://translink.com.au/tickets-and-fares/go-card/about-go-card
[Accessed 16 June 2016].

TransLink, 2016. *SkyTrain Station & Accessible Entrance Maps.* [Online]
Available at: http://www.translink.ca/en/Schedules-and-Maps/SkyTrain/SkyTrain-Station-and-Elevator-Maps.aspx
[Accessed 21 June 2016].

Transportation Association of Canada, 2012. *Changing Practices in Data Collection on the Movement of People,* Toronto: Lee-Gosselin Associates Limited.

Trépanier, M., Tranchant, N. & Chapleau, R., 2007. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* 11(1).

Wang, M.-H., Schrock, S. D., Boek, N. V. & Mulinazzi, T., 2013. Estimating Dynamic Origin-Destination Data and Travel Demand Using Cell Phone Network Data. *International Journal of Intelligent Transportation Systems Research,* May, 11(2), pp. 76-86.

Wang, M.-H., Schrock, S. D., Broek, N. V. & Mulinazzi, T., 2013. Estimating Dynamic Origin-Destination Data and Travel Demand Using Cell Phone Network Data. *International Journal of Intelligent Transportation Systems Research,* April, 11(2), pp. 76-86.

Wang, W., 2010. *Bus Passenger Origin-Destination Estimation and Travel Behavior Using Automated Data Collection Systems in London, UK,* Massachusetts: Massachusetts Institute of Technology.

Ward, D. K., 2011. *Using Cell Phone Technology to Collect Travel Data*. North Carolina: CAMPO.

White, J. & Wells, I., 2002. *Extracting Origin Destination Information From Mobile Phone Data* , s.l.: TRL Ltd & Highway Agency.

Widhalm, P. et al., 2015. Discovering Urban Activity Patterns in Cell Phone Data. *Transportation: Planning, Policy, Reserach, Practice,* July, 42(2), pp. 597-623.

Xu, Y. et al., 2014. *Understanding Aggregate Human Mobility Patterns Using Passive Mobile Phone Location Data: a Home-Based Approach,* New York: s.n.

Zhang, F., Zhao, J., Tian, C. & Xu, C., 2016. Spatiotemporal Segmentation of Metro Trips Using Smart Card Data. *IEEE Transactions on Vehiular Technology,* 65(3).

Zhao, J., Rahbee, A. & Wilson, N. H. M., 2007. Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. *Computer-Aided Civil and Infrastructure Engineering,* July, 22(5), p. 376–387.