Pers Ubiquit Comput
DOI 10.1007/s00779-013-0652-4

ORIGINAL ARTICLE

Towards proximity-based passenger sensing on public transport buses

Vassilis Kostakos · Tiago Camacho · Claudio Mantero

Received: 10 January 2012 / Accepted: 1 September 2012

Abstract While substantial research on intelligent transportation systems has focused on the development of novel wireless communication technologies and protocols, relatively little work has sought to fully exploit proximity-based wireless technologies that passengers actually carry with them today. This paper presents the real-world deployment of a system that exploits public transit bus passengers' Bluetooth-capable devices to capture and reconstruct micro- and macro-passenger behavior. We present supporting evidence that approximately 12 % of passengers already carry Bluetooth-enabled devices and that the data collected on these passengers captures with almost 80 % accuracy the daily fluctuation of actual passengers flows. The paper makes three contributions in terms of understanding passenger behavior: We verify that the length of passenger trips is exponentially bounded, the frequency of passenger trips follows a power law distribution, and the microstructure of the network of passenger movements is polycentric.

Keywords Public transport · Passenger sensing · Bluetooth · Origin/destination matrix · Mobile and ubiquitous computing

V. Kostakos (✉)
University of Oulu, 90014 Oulu, Finland
e-mail: vassilis@ee.oulu.fi

T. Camacho
Queensland University of Technology, Kelvin Grove QLD, Brisbane, QLD 4059, Australia
e-mail: tiago.diascamacho@qut.edu.au

C. Mantero
Horários do Funchal, Transportes Públicos S.A, 9020-242 Funchal, Portugal
e-mail: claudiomantero@horariosdofunchal.pt

1 Introduction

Recent advances in intelligent transportation systems mostly overlook the opportunities presented by the increasingly capable mobile devices that passengers carry daily. While Internet connectivity of such devices can help passengers plan and manage their travel, there still exists great potential in using these devices in a truly contextual manner: by taking full advantage of the fact that such devices in addition to connecting to the Internet can also communicate with nearby onboard systems in real time.

In this paper, we present one way in which the increasing penetration of smartphones and wireless devices can be taken advantage of transport operators. This paper presents a low-cost system that exploits proximity-based Bluetooth technology to capture and reconstruct bus passengers' origin and destination with higher granularity than electronic ticketing systems.

To achieve this objective, a novel methodology is presented for detecting both origin and destination of passengers onboard buses, along with a validation of the system's accuracy and performance in a real-world setting. While a number of demonstrator systems have in the past exploited direct communication with passenger devices, little work has systematically measured the performance and accuracy of such systems for the purposes of detecting passenger trips. Here, we present a system that relies on Bluetooth discovery mechanisms to infer the presence of individual passengers onboard busses. Using localization techniques, these detection events can be translated into "board" and "disembark" events that in turn can be used to reconstruct passenger trips. The results show that our technique can infer origin and destination points at the granularity level of individual bus stop simply by installing standalone equipment on buses. Furthermore, the results

Published online: 07 April 2013

show that the collected data can be used to analyse network usage and identify trends in passenger behavior.

It is important to note that even though the conducted studies focused on Bluetooth, the principles and engineering of this work are applicable to any proximity-based technology that passengers may carry, including Wi-Fi devices. While Bluetooth is currently the most popular proximity-based technology on passenger's mobile devices, this may change in the future. Even so, the results and principles presented here remain valid.

2 Related work

A number of projects have attempted to accurately reconstruct mobility patterns by exploiting people's mobile devices. In the past, mobile phone tracking has been used as an approach to measure the flows of passengers between parts of a city [7, 10]. The results, however, have low spatial resolution and are most appropriate for long-distance segments such as highways. In addition, Bluetooth or Wi-Fi traces have been used to analyze people's mobility [3, 8, 9, 14, 16, 18], albeit not in the context of public transportation. These studies suggest that Bluetooth is a useful technology for capturing individual mobility traces, mainly due to its popularity and widespread usage. In Table 1, we highlight how such technology compares to traditional trip counting approaches.

Indicatively, a recent market study with 1,000 consumers in Portugal indicated that 88 % of the population owns a mobile phone and that 22 % makes active use of Bluetooth/Infrared technologies [15]. The study suggests that men are more likely to use these technologies (26 %) than women (18 %) and that its use increased by 2 % from 2006 to 2007. While the study showed no significant effect of class status (measured by annual income), it did show that young people are more likely to use these technologies. The percentage of people using these technologies per age group was 10–14 (40 %), 15–24 (50 %), 25–34 (31 %), 35–44 (16 %), 45–54 (7 %), 55–64 (2 %), and 65+ (1 %). Hence, teenagers and young adults are more likely to use these technologies, even though it is expected that this may change in the future.

It is important to note that Bluetooth devices may operate in non-discoverable mode, and hence not be detectable. This means that only a subset of existing Bluetooth devices is technically observable. Estimates show the ratio of observable to non-observable Bluetooth devices reported previously in literature range between 2 % for Bremen, Germany to 7 % for Bath, and UK [17, 20]. These results show that while potentially a great subset of the population has Bluetooth-capable devices, *ceteris paribus* only a small portion keeps their Bluetooth devices in discoverable mode. Note that while this 7 % ratio is expected to increase over time, as more Bluetooth-capable devices appear in the market. Also, while this ratio is not necessarily a big portion of the population, nevertheless, it larger than the approximate 3 % of the population that traditional transport surveys cover in any particular region.

Bluetooth has also been used for traffic monitoring in the context of highways and major transport arteries, where the deployment of Bluetooth scanners at strategic locations allows for the approximation of macro-travel behavior [25]. Similarly, Barceló et al. [4] made use of statistical methods (e.g., Kalman Filtering) in order to be able to estimate traveling time and O/D matrices in highways. This method allows to integrate both Bluetooth information with more traditional data sensors such as inductive loops. The usefulness of Bluetooth sensors to estimate time traveling has been further corroborated through the BlueTODA™ system, where results obtained are said to be comparable to the E-ZPass tag readers, an electronic toll system that provides automated vehicle counts at specific corridors [13]. Such applications of Bluetooth in highly volatile environments demonstrate its usefulness beyond its original purpose as a simple proximity-based communication technology.

The system presented here reconstructs O/D data at the granularity of individual bus stops and can also record O/D data for individual passengers in real time. Similar work has been done using origin-only electronic ticketing data since most bus transit systems do not capture both origin

Table 1 Comparison of passenger trip detection using Bluetooth versus electronic ticketing and surveys

	Method of OD estimation		
	Bluetooth detection	Electronic ticketing	Survey
Sample size	~10 %	>50 %	~3 %
Spatial accuracy of destination data	High	Relies of inferencing (which introduces bias)	High (Explicitly stated by respondend)
Representativeness and Sample bias	Demographic bias of technology adoption	Bias if not all passengers swipe ticket	Bias due to sampling technique, human memory, and self-selection of respondents
Passenger effort	Enable Bluetooth	Swipe ticket	Answer questionnaire

and destination information (even though subways do capture such data). Hence, previous work has focused on deriving accurate estimations from incomplete data [1] and estimates of the flow of the passengers for unobserved parts of the network [12]. A notable analysis using origin-only data was conducted for the Chicago Transit Authority rail system which collects origin-only data [27]. This analysis is based on the assumptions that (1) there is no private transportation mode trip segment (car, motorcycle, bicycle, etc.) between consecutive transit trip segments in a daily sequence; (2) passengers will not walk a long distance to board at a different rail station from the one where they previously alighted; and (3) passengers end their last trip of the day where they began their first trip of the day. These assumptions were introduced due to the lack of further data about passengers' movements. Using the system we developed, it is actually possible to detect such passenger transitions, even across multiple modes of travel.

More recently, transit authorities have begun deploying sophisticated onboard passenger-counting systems that rely on a variety of techniques such as pressure sensing, light beams [11], and image interpretation [23]. While these entail high operational, installation, and maintenance costs, they fail to provide individual passenger detection, but rather operate on crude passengers counts. A recent study assessed the feasibility of using RFID readers to remotely sense passengers while boarding and exiting buses [19]. Although the results show good identification values, the cost of hardware is currently the order of the thousands of dollars, making it a high investment if to be deployed at a fleet level. On the other hand, Bluetooth hardware can be purchased off-the-shelf at much lower cost.

3 Materials and methods

The system presented here was developed and tested in collaboration with the regional public transportation operator, Horarios do Funchal. The operator is based in Madeira, Portugal, and has over 160 buses serving about 30 million passengers per year, across 1,400 bus stops. The buses are equipped with an onboard localization system relying on GPS, digital odometers, and door sensors. In addition, all buses are equipped with an RFID ticketing system.

The purpose of the evaluated system is to capture end-to-end passenger trips on buses. A study was conducted to evaluate the system in realistic conditions by installing it onboard buses. The study assessed the accuracy of the system in capturing onboard passenger flows by comparing the collected data against the ground truth represented by electronic ticketing data.

In this study, the mechanism for sensing passengers was developed by taking advantage of the Bluetooth protocol. Continuous scans for Bluetooth devices were performed for a duration of 5.12 s, and no friendly names were requested from the devices. Although the Bluetooth standard defines that a device must spend 10.24 s in a single inquiry to discover all devices within range [5], studies suggest that this value may be excessive and that the use of 5.12 s can discover 98 % of devices [21]. Also, it has been reported that when the number of devices increases within the vicinity, there is little gain in using longer inquiry periods, as there is the chance of devices blocking each other out [22].

3.1 Onboard passenger sensing

This study aimed to assess how accurately an onboard Bluetooth transceiver can capture passenger flows. Prior to deployment a series of preliminary tests were conducted to determine the optimum location to install the Bluetooth transceiver onboard the bus (see Fig. 1), including verification by a human that the system worked as expected. In order to minimize possible interferences, this transceiver worked solely as a scanning unit with its Bluetooth discoverable mode being disabled.

Figure 2 shows an overview of how end-to-end individual passengers trips were detected. The onboard Bluetooth transceiver continuously scanned for Bluetooth devices within the bus, and this data were first transformed into device trips. A device trip is defined by the time that a device stayed in range of the scanner while in the bus. To compensate for potential interference or for devices becoming temporarily unreachable, an empirically derived threshold of 120 s was used to determine the end of a device trip. Furthermore, devices that appeared solely once were automatically discarded due to the high possibility that they belonged to users outside the bus.

Subsequently, each trip's start and end times were geolocated using the existing bus localization database which stores the exact trail of every bus. Specifically, an event is recorded when the bus reaches the bus stop and the driver opens the doors. Therefore, it is possible to identify the exact bus stop when a device first appeared (hence the passenger boarded the bus) and when the device was last seen (hence the passenger alighted the bus).

The system collected data for 2 weeks, and during the last week, the actual electronic ticketing data were obtained to serve as ground truth. An analysis of the two datasets was conducted to assess the extent to which the data collected by the Bluetooth transceiver can be used to describe actual fluctuations of passengers as recorded by the electronic ticketing system. In addition, an O/D matrix was reconstructed from the collected data and evaluated by the operator's domain experts.

Fig. 1 Installation of the onboard Bluetooth transceiver. **a** A Gumstix computer along with a 24–5 volt converter used to power the computer with the bus electric circuit. **b** A bus being rewired. **c** The final installation consisted of a protective plastic case attached to the roof of the main cabin. **d** The system (indicated by an *arrow*) is installed near the center of the main cabin. **e** The control center where real-time data are gathered from the buses localization system and the whole operation is overseen

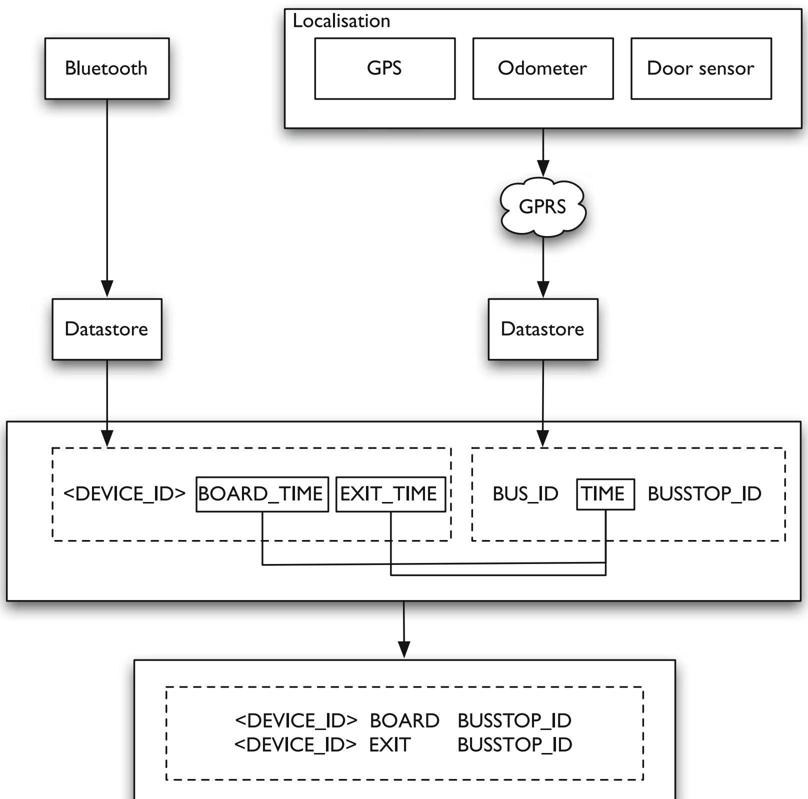

Fig. 2 Correlating the Bluetooth data gathered by the onboard Bluetooth transceiver with the bus localization data. By combining these two datasets, the bus stops where a device boarded and exited are determined

4 Results

The collected data include device trips, location of the bus at any given time using the Automated Vehicle Location (AVL) system, and tickets validated on the bus using the onboard Automated Fare Collection (AFC) system. As all passengers detain a fare which uses RFID which has to be presented at entrance of the bus, the ticket data served as the ground truth against which the performance of the system was assessed.

During this study, 4,701 tickets were validated on the bus, while 601 Bluetooth devices were detected, suggesting that about 12.8 % of the passenger population carries

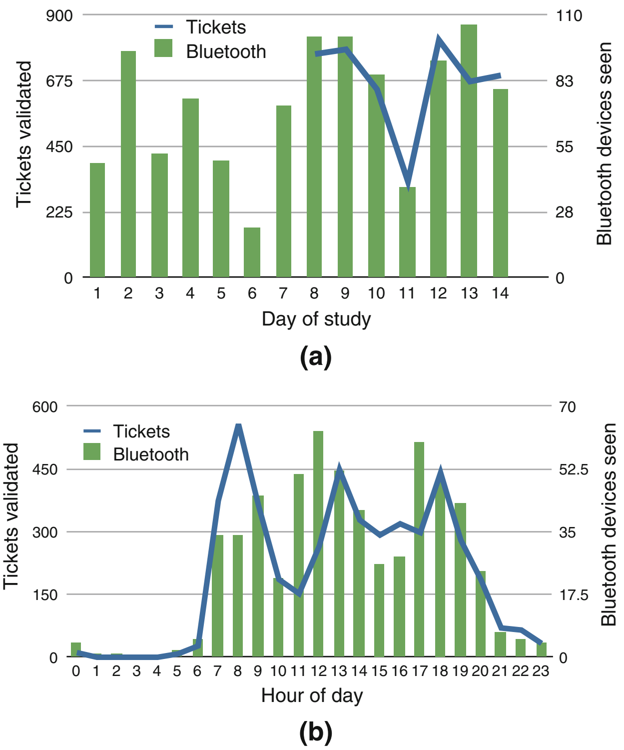

Fig. 3 Total number of tickets validated and Bluetooth devices captured according to **a** day of study and **b** hour of day. Ticketing data were only available for the last 7 days of the study

Bluetooth-enabled devices. The number of tickets validated onboard the bus, either by hour or by day, is shown in Fig. 3 as a solid line. In the same graph, the bars represent the number of trips recorded via Bluetooth for the respective periods.

Further analysis assessed the accuracy of the system by determining the level of correlation between the seen Bluetooth devices and tickets validated using three distinct sampling strategies (Fig. 4). The sampling strategies were chosen as representative of varying levels of windows size that the analysis can consider (hourly, 12-h, and daily).

A novelty of the system is that unlike traditional onboard systems, it is able to record the duration of a passenger's trip. The duration of all trips recorded during the study is shown in Fig. 5a and follows an exponential decay with exponent -0.002, ($R^2 = 0.96$). We note that this figure contains noise (seen as a small cluster in the top-left of the distribution), which as we describe later we filter out. Here, we show the raw data to indicate that noise can be easily identified, and when removed the correlation becomes $R^2 = 0.99$. In addition, it is possible to calculate the number of times each individual device boarded the bus since Bluetooth IDs are unique. The frequency of trips per individual device shown in Fig. 5b and follows a power law distribution with exponent $a = -2.2$, ($R^2 = 0.96$).

Finally, the trip data were used to reconstruct an O/D matrix as shown in Fig. 6, both as segments on a map and

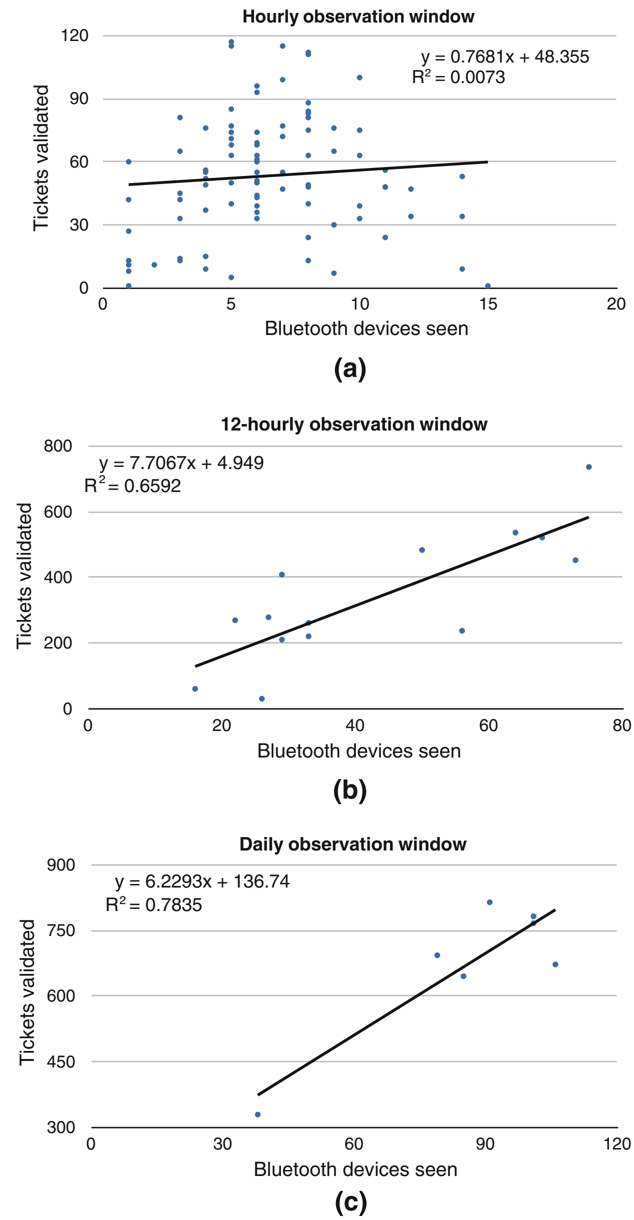

Fig. 4 Correlation between trips captured by the onboard Bluetooth transceiver (x-axis) and tickets validated (y-axis). Data points resemble **a** 60-min intervals, **b** 12-h intervals, and **c** 24-h intervals. The standard error per collected data point is ±26, ±12, and ±3.5 % for each chart, respectively

as a graph. The segments on the map show the start and end point of individual trips connected by a straight line. The graph depicts bus stops as nodes and links together bus stops which were involved as start and end in any individual trip. The thickness of edges indicates the volume of trips recorded between the specific bus stops.

4.1 Questionnaire

In addition to the quantitative studies, a questionnaire was used to collect passengers' feedback about their usage of

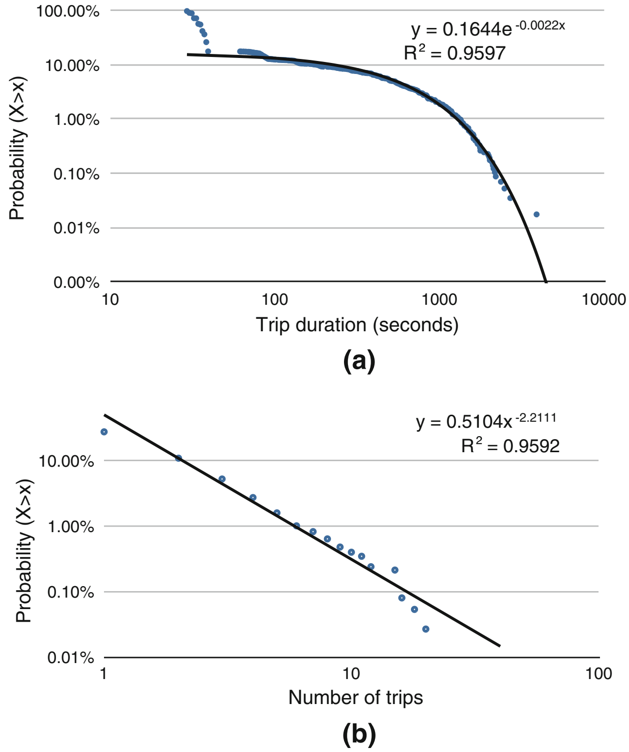

Fig. 5 **a** Cumulative distribution probability of trip duration. **b** Cumulative distribution probability of individual device detection frequency

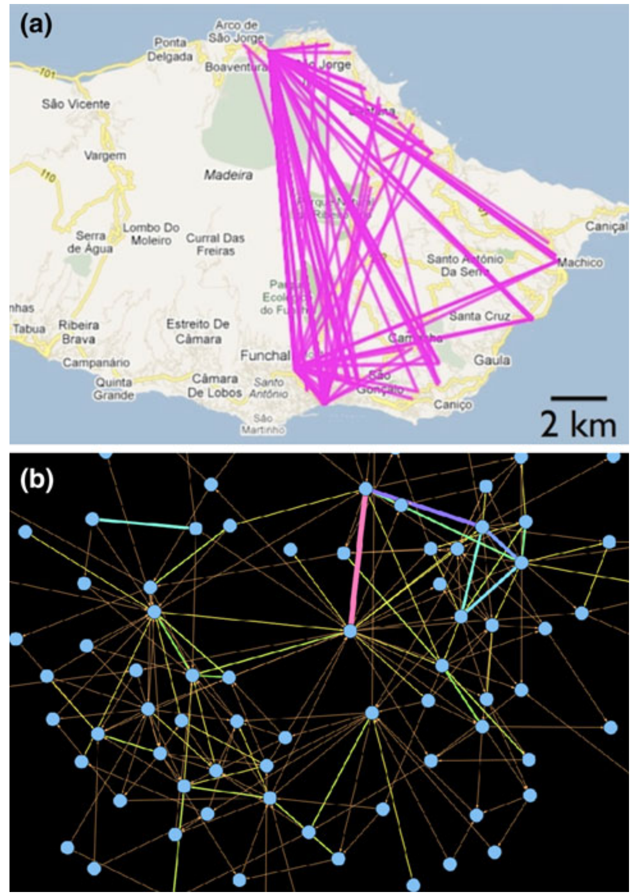

Fig. 6 O/D for the whole population as captured by the system, and shown **a** as segments on a map, and **b** as a graph where nodes designate bus stops and connections designate trips (connection *color* and *thickness* indicates the popularity of a segment). The bus served on four separate routes during the study

wireless technologies, as well as their perceptions about the use of proximity-based technologies. A total of 105 respondents answered a questionnaire at four different locations (51 females and 54 males) prior to system deployment. The results show that most respondents used public transportation more than twice a day (59 %), and 73 % reported they wait between 5 and 15 min for the bus to arrive.

The majority of respondents claimed to use their portable communication equipment while waiting for the bus (80 %). Overall, 60 % claimed to use their devices while waiting for messaging purposes, 41 % for making phone calls, 35 % for entertainment (e.g., games, music), and 5 % for online access. While most responders had Bluetooth-capable devices (73 %), only 11.5 % had their Bluetooth set to discoverable, which is similar to the ratio found in our study. Of those who had explicitly disabled their Bluetooth 50 % claimed security concerns and 37 % power consumption concerns, while 13 % gave no reason.

Regarding their preferred way of receiving digital content on their mobile devices, 31 % claimed they wanted to retrieve the content (i.e., pulling content), 32 % claimed they wanted to be pushed information if they had explicitly registered beforehand, 4 % wanted to be pushed information without prior registration, and 33 % claimed that they did not care as long as the information was relevant to them.

A chi-squared analysis revealed a significant association between demographics and device usage practices. Specifically, those more likely to use their mobile phones for texting while waiting at a bus stop are women ($\chi^2(1, N = 105) = 6.507, p < 0.05$) and those aged between 10 and 20 ($\chi^2(4, N = 105) = 21.69, p < 0.01$). In general, participants aged 10–30 were much more likely to use their mobile device while waiting for the bus ($\chi^2(4, N = 105) = 11.04, p < 0.05$). Further analysis revealed a significant association between waiting time and use of mobile devices, with those waiting between 5 and 10 min less likely to use their devices ($\chi^2(4, N = 105) = 12.83, p < 0.05$).

5 Discussion

This paper has presented quantitative results to assess the performance and accuracy of a proximity-based wireless systems for sensing passengers. In addition, a qualitative study was conducted to assess passengers' perceptions of this system and evaluate its potential in relation to passengers. This section provides an interpretation of the results and offers an assessment of the system's performance.

5.1 Sensing passengers

During our study, four different service routes were covered at different times of the day. Independently of the route, Fig. 3 presents the number of devices that were detected and the number of validated tickets, indicating a 12.8 % ratio of passengers carrying Bluetooth-enabled phones.

The existence of electronic ticket data provides the ground truth against which the performance of the system's accuracy in detecting passengers can be assessed. However, the underlying processes are dynamic, and therefore, any correlation between these quantities needs to adopt a sampling or binning strategy. Figure 4 shows the correlation between tickets and detected Bluetooth devices following three sampling strategies: hourly, 12-h, daily. Since ticketing data were available only for the last 7 days of the study, the plot for the hourly strategy consists of 7 (days) $\times$ 24 (h) data points, the 12-h consist of 7 $\times$ 2 data points, and the daily consists of 7 data points.

We verify that the linear regression model is not adequate for the hourly sampling strategy, as the coefficient of correlation approximates zero. In relation to the other two strategies, we verify a correlation coefficient of approximately 66 % for the 12-hourly sampling strategy and of about 78 % for the daily sampling strategy. Further analysis of these two models has indicated the adequacy of the linear model, as the residuals follow a normal distribution and 95 % are contained within the bounds defined by the standard error. These results indicate an increase in correlation strength as the sampling windows becomes larger. Note that it has been previously shown that relying on *averages* per time period leads to spurious auto-correlation [24]. However, in this case, the sampling window is adjusted to calculate the correlation between tickets and Bluetooth devices, hence the data are not averaged but aggregated. Therefore, all 3 strategies are valid and increasing the window size does *not* introduce bias since the data are aggregated, not averaged [24].

The apparent relationship between observation window size, data correlation, and sampling error suggests that short Bluetooth observation periods, or indeed observation periods with little activity, are likely to produce unreliable results. Hence, it is advisable that observation periods with adequate activity should be used in such analyses. In this study, a 24-h window of Bluetooth observations accounts for up to 78 % of the variation observed in electronic ticket data.

However, as seen in Fig. 3, the system was *at times* particularly bad at capturing an accurate representation of the true flow of passengers, especially during the morning rush hours. One explanation for these consistent deviations is that due to the high passenger density at these times of day, the scanner's ability to discover surrounding devices is diminished due to increased interference cause by passengers' bodies. Another factor contributing to the discrepancies between detected Bluetooth devices and ticket counts is that morning passengers may be of older age, therefore being less prone to carry Bluetooth-enabled devices. This latter hypothesis is mainly corroborated by the results presented in [15] and also by the results of the questionnaire issued to passengers.

5.2 Detecting trip length

Our study provided further evidence of the system's ability to record information about trips and flows on the network. The results show that a single onboard Bluetooth transceiver cannot *fully* cover the transportation network, which is evident by the fact that most passengers in the study were only detected once (Fig. 5b). Given a larger number of onboard Bluetooth transceivers, we can expect a more complete picture of the network. However, it is still possible to make inferences about travel behavior by considering trip duration, since the system did capture entire trips.

With the calculation of device trips for unique Bluetooth IDs, the cumulative distribution for the duration of the trip could be inferred, as seen in Fig. 5a. In these results, there exists a cluster on the top left, or head, of the distribution, indicating a high number of devices with an apparently short trip duration. Subsequent filtering on the data, however, discarded these short trips as noise. This noise was due to captured devices from outside the bus. To detect and discard such data, all devices that were spotted when the bus was en-route and not nearby a bus stop were discarded. Additionally, all devices that appeared and disappeared at the same bus stop were also removed from the analysis.

The bulk of the trip duration distribution in Fig. 5a exhibits an exponential decay curve with $R^2 = 0.96$ (when noise is removed then $R^2 = 0.99$), which is a distribution often used in transportation modeling to approximate passenger behavior [6]. This shows that the obtained data are consistent with previous findings even though in this case it has been collected using a novel approach. This trip duration distribution indicates that it is increasingly unlikely for individuals to make long trips. This finding is also consistent with the relatively bounded size and steep nature of the terrain in the city of Funchal, suggesting that indeed most passengers only take short trips.

5.3 Reconstructing the O/D matrix

The combination of device trips and vehicle location enabled the reconstruction of the flow of passengers in the form of an O/D matrix, as shown in Fig. 6a. The figure depicts every single trip recorded during our study as a straight line segment on a map. Due to the large number of overlapping segments, the O/D matrix is (partly) shown as a graph in Fig. 6b. Because the transport operator did not have an O/D matrix of comparable granularity, it was not possible to directly compare our derived O/D matrix to another existing one. In fact, given that traditional survey techniques produce O/D matrices at the level of "zones," such a comparison would not be possible since our approach generates O/D data at the granularity of bus stops. However, the quality of the results was assessed by experts of the transport operator. They verified that bus stops that they empirically know to be popular, such as certain popular bus stops attracting tourists, were reflected as such in our dataset. In addition, well-known routes known to be busy, such as those linking public housing areas to work and commercial regions, were similarly reflected in our dataset. However, a direct numerical comparison was not possible at this stage, but this is a clear topic for further work.

In Fig. 6b, popular segments of the network can be identified by the intensity of their color and thickness, while popular bus stops can be identified by the number of incoming and outgoing links on the graph. As expected, the most popular bus stops are near the capital of the island. Further analysis of the graph shows the structure of the network to be polycentric, with centers at the south, north, and south-east of the island. Each center serves as an attractor to nearby regions, resulting in a large number of relatively short trips between each of the centers and the nearby regions. This was further validated by considering the actual routes that the bus served during the study.

The O/D data presented here have higher granularity than what conventional techniques can generate, which in the case of this city consists of about 20 individual zones. In addition, the data are near-live, and hence can be used to improve real-time passenger services. Furthermore, such granular data can be used to identify locations that attract many passengers and combined with the relative distance between those locations, identify optimal ways of linking such locations. Furthermore, another use of such frequently updated data would be to optimize network simulators, typically used by public transit companies [26].

5.4 Novel services

The system presented here can act both as a passenger-counting apparatus as well as a direct communication channel to passengers. This novel characteristic suggests that as long as services of value are provided to passengers, they are more likely to keep their Bluetooth transceivers activated. This is supported by the results of our qualitative analysis, which shows that passengers tend to use their mobile devices when waiting for the bus, and would be interested in services that they can access with their phone. Indeed, our results show that users are increasingly likely to engage with their mobile device if they wait for longer periods for the bus stop (more than 10 min). In addition, we found that younger passengers are more likely to use their mobile phone at bus stops and therefore may form a specific target group for transport operators.

Several kinds of services are enabled by the system, some of which do not even require content dissemination. An example of such service is a point-rewarding scheme that rewards passengers according to potential inefficiencies of the public transit company. By recording passengers' presence at the bus stops, the system could then determine how long the passenger had to wait before the bus arrived. This information could then be used to award points to passengers in a scheme that would exchange points for tickets or other types of rewards. In addition, personalization services could be developed that take into account individual passengers' behavior and provide appropriate information or promotional offers.

Moreover, the analyses presented here allow for a high-frequency update of the network's O/D matrix. Previous studies indicate that up to 70 % of passenger trips in the USA are related to work and school commute and are therefore highly predictable [2]. Nevertheless, many events may influence normal function of the network, and the use of conventional methods fail to reflect these in a timely manner. An excellent example are sporting events, demonstrations, or even floods such as the ones in Funchal, which caused major road blockage and overall infrastructure destruction. These had a great impact on the bus network, and consequently, redesign of several routes was inevitable. Deployment of the system presented here would allow for a more rapid understanding of the magnitude of the changes and help with the calibration of resources.

5.5 Limitations

A key limitation of this system is that it relies on a technology (i.e., Bluetooth) that does not capture the whole population, especially older passengers. Even though more than 80 % of the population has Bluetooth-capable phones, only about 12 % of the population activates this functionality. This is most likely because they currently have no use for it. It is expected that over time the age-based differences of Bluetooth usage are to be alleviated, and furthermore, the introduction of services that offer real value will likely motivate passengers to activate their Bluetooth transceivers. Additionally, the analysis techniques described here should apply to any proximity-based technology that is popularized in the future, such as Wi-Fi or ZigBee.

Another aspect that needs to be taken into account is that passengers may enable or disable their phones while onboard. If this actually happens, then it will lead to skewed data as the system would collect inaccurate information about people's entry and exit from the bus. The use of a Bluetooth transceiver in bus stops could be an answer to this, as it would increased the probability of passengers enabling their Bluetooth while still at the bus stop. In theory, this would also mean that the number of captured devices would increase. Furthermore, our system cannot cope with passengers carrying more than one Bluetooth-enabled device. While we expect that few such passengers exist, the presence of such a passenger would result in multiple "passengers" being detected by our system.

Finally, the use of Bluetooth has privacy implications which are increasingly becoming apparent to consumers. The system presented here tracks for individual passengers behavior over time and by consequence records precise information about people's location. Such information can be used for harmful intents and purposes. For instance, a culprit may use knowledge of the fact that Alice is currently on the bus to infer that she is not at her flat and rob it. Hence, bus companies need to make sure that trip data are secured. However, the same threats are posed by magnetic and RFID tickets, since they too can be used to infer the approximate location of passengers. A distinguishing characteristic of Bluetooth, however, is that passengers can choose to enable it and disable as they please.

The issue of privacy in public transport is an important one. We believe that as the public is increasingly made aware of privacy issues of Bluetooth (or other similar technologies such as RFID), they are indeed more likely to stop using that technology. However, we argue that as new services become available to users of these technologies, people will be motivated to adopt these technologies and services on a daily basis. The issue comes down to the tradeoffs between potential privacy issues versus the obtained benefits. Hence, it is important to highlight that the system presented here offers a great opportunity for service personalization and one-to-one interactions with passengers, features which are likely to attract more people to adopt this technology.

6 Conclusion

This paper presented the real-world deployment of a system that captures public bus passenger behavior. The results obtained from this deployment demonstrate that passenger behavior can be captured on a highly granular level, enabling for the reconstruction of an Origin/Destination matrix at the detail of a bus stop and individual passengers. The system performs well, capturing almost 80 % of the fluctuation of passenger volume on a daily basis. Analysis of the results shows that the length of passenger trips are exponentially bounded, the frequency of passenger trips follows a power law distribution, and the microstructure of the network of passenger movements is polycentric. In addition, the deployed system can easily be reconfigured to provide additional services to passengers by proactively contacting their mobile phones devices.

References

1. Abrahamsson T (1998) Estimation of origin-destination matrices using traffic counts—A literature survey. IIASA Interim Report IR-98-021/May
2. American Transportation Association. http://www.apta.com. Last Accessed 23 Feb 2010
3. Balazinska M, Castro P (2003) Characterizing mobility and network usage in a corporate wireless local-area network. Proc 1st Int Conf Mob Syst Appl Serv MobiSys 03:303–316
4. Barcelo M, Marqués LLL, Carmona C (2010) Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. Transportation Research Board: Journal of the Transportation Research Board 2175:19–27
5. Bluetooth SIG. Specification of the bluetooth system, core version 2.1, http://bluetooth.com. Last Accessed 14 Feb 2010
6. Bovy PHL, Bliemer MCJ, van Nes R (2006) Transportation modeling: lecture notes. CT4801. Delft University of Technology. http://ocw.tudelft.nl/fileadmin/ocw/courses/TransportationandSpatialModelling/res00018/transportation4801.pdf
7. Caceres N, Wideberg J, Benitez F (2007) Deriving origin-destination data from a mobile phone network. Intell Trans Syst IET 1(1):15–26
8. Chaintreau A, Hui P, Crowcroft J, Diot C, Gass R, Scott J (2007) Impact of human mobility on opportunistic forwarding algorithms. IEEE Trans Mob Comput 6(6):606–620
9. Eagle N, Pentland A (2006) Reality mining: sensing complex social systems. Pers Ubiquit Comput 10(4):255–268
10. González M, Hidalgo C, Barabási A (2008) Understanding individual human mobility patterns. Nature 453:779–782
11. Hansen M, Qureshi M, Rydzewski D (1994) Improving transit performance with advanced public transportation systems technologies
12. Hazelton M (2003) Some comments on origin-destination matrix estimation. Transp Res Part A Policy Pract 37(10):811–822
13. Consulting KMJ (2010) Bluetooth travel time technology evaluation using the BlueTOAD™. Technical report, Pennsylvania Department of Transportation
14. Kostakos V, Nicolai T, Yoneki E, ONeill E, Kenn H, Crowcroft J (2008) Understanding and measuring the urban pervasive infrastructure. Pers Ubiquit Comput 13(5):355–364
15. Marktest. http://www.marktest.com. Last Accessed 23 Feb 2010
16. McNett M, Voelker GM (2005) Access and mobility of wireless PDA users. ACM SIGMOBILE Mob Comput Commun Rev 9(2):40
17. Nicolai T, Kenn H (2007) About the relationship between people and discoverable Bluetooth devices in urban environments.In: Proceedings of the 4th international conference on mobile

technology, applications, and systems and the 1st international symposium on computer human interaction in mobile technology—Mobility, p 72
18. Nicolai T, Yoneki E, Behrens N, Kenn H (2006) Exploring social context with the wireless rope. On the move to meaningful internet systems, pp 874–883
19. Oberli C, Torres-Torriti M, Landau D (2010) Performance evaluation of UHF RFID technologies for real-time passenger recognition in intelligent public transportation systems. IEEE Trans Intell Transp Syst 11(3):748–753. doi: 10.1109/TITS.2010.2048429
20. ONeill E, Kostakos V, Kindberg T, Schiek A (2006) Instrumenting the city: developing methods for observing and understanding the digital cityscape. UbiComp, pp 315–332
21. Peterson B, Baldwin R, Kharoufeh J (2006) Bluetooth inquiry time characterization and selection. IEEE Trans Mob Comput 5(9):1173–1187
22. Siegemund F, Rohs M (2003) Rendezvous layer protocols for Bluetooth-enabled smart devices. Pers Ubiquit Comput 7(2):91–101
23. Wilson N (2006) Public transportation service and operations planning: lecture notes. 1.258J/11.541J/ESD.226J. MIT Open Courseware. http://bazzim.mit.edu/oeit/OcwWeb/Civil-and-Environmental-Engineering/1-258JSpring-2006/CourseHome/index.htm
24. Working H (1960) Notes on the correlation of first differences of averages in a random chain. Econometrica 28(4):916–918
25. Young S (2009) The continuing evolution of travel time data information—Collection and processing: utilization of bluetooth traffic monitoring. In: Transportation engineering and safety conference
26. Zhao J, Rahbee A, Wilson N (2007) Estimating a rail passenger trip origin-destination using automatic data collection systems. Comput Aided Civil Infrastruct Eng 22(5):376–387
27. Zhao L (2006) Heuristic method for analyzing driver scheduling problem. IEEE Trans Syst Man Cybern 36(3):521–531