

XII Conference on Transport Engineering, CIT 2016, 7-9 June 2016, Valencia, Spain

The potential of e-ticketing for public transport planning: the Piedmont region case study

Maurizio Arnone^{a*}, Tiziana Delmastro^a, Giulia Giacosa^a, Mauro Paoletti^b, Paolo Villata^b

^a*SiTI - Higher Institute on Territorial Systems for Innovations, Via Pier Carlo Boggio 61, Torino 10138, Italy*

^b*A.T.I. Trasporti Interurbani S.p.A., Via Circonvallazione 19, Saluzzo (CN) 12037, Italy*

Abstract

In the Piedmont region (Italy) the electronic ticketing system called BIP, is currently active across much of its territory, and the data collected in the Province of Cuneo since the full activation of the system (2014) provide today a sound source of information. Two different travel documents are available, travel passes and pay-per-use, with different validation rules: check-in only for travel passes and check-in and check-out for pay-per-use. Data produced by this electronic ticketing system employing smart cards allow to perform a detailed analysis of each user's behaviour, and calculate time and space distributions of each passenger trip. In detail, data originating from smart card transactions allow to trace back the trip chains, establish journey origins and destinations, and produce a "travel diary" for each passenger. Based on this data, performance indicators (i.e. load factor) as well as user mobility patterns and origin-destination matrices can be calculated in an automated and reliable way. This article presents a methodology for assessing the quality of the data collected when information about boarding and alighting stops is available from the (on board) validation system. It also presents an algorithm to assign a destination for each trip where only the boarding information is available. In the case study of the Province of Cuneo, it was found that 91% of the pay-per-use journey data are reliable and can be used for further analysis, whereas with the use of the proposed algorithm it was possible to estimate the destinations for 82% of the travel pass trips.

|

1. Background and objectives

The rural public transport service in the Province of Cuneo (Italy) is made up by more than 100 bus lines and a total of 2,500 daily services on a standard winter working day (data referring to 2015). The service is carried out by the Grandabus consortium, which includes 14 local transport companies. Grandabus offers an electronic ticketing system that allows the use of different buses of any of its 14 operators with a single ticket. All tickets have to be loaded on the BIP smart card or alternatively single tickets can be purchased on board and are also recorded by the system. The BIP smart card can hold travel passes (e.g. weekly, monthly, annual) or pay-per-use credit. These two travel documents have different validation rules: travel pass holders only have to check-in whereas holders of pay-per-use credit have to both check-in and check-out in order to enable the system to identify the trip carried out and deduct the corresponding fare from the smart card. In 2014 more than 200,000 travel passes were sold and more than 100,000 smart cards with pay-per-use credit were issued. 12,000,000 ticket validations were carried out (90% travel passes and 10% pay-per-use credit) along the public transport lines operated by the Grandabus consortium.

The aim of the present research was to explore the potential of the BIP system for the planning, programming and management of public transport in the Piedmont region (Italy) building on the rural transport data available for the province of Cuneo. In particular, the research had two main objectives:

1. Check the quality of the data regarding the pay-per-use credit (where both check-in and check-out are required) and estimate the percentage that is free of errors, shortcomings and/or inconsistencies and therefore can be directly used.
2. Develop, test and evaluate the reliability of a methodology for estimating the check-out for travel passes (for which only the check-in is required) in order to identify the alighting bus stop and thus recreate the entire trip carried out by public transport.

In the case of pay-per-use credit it was possible to reconstruct 91% of the trip stages (sequence of check-in and check-out). In particular, 71% of the data did not require any elaboration whereas 20% showed errors that were corrected with algorithms developed ad hoc. Only 9% of the trip stages could not be reconstructed.

In the following sections the paper will focus on the second objective and the case of travel passes, detailing the methodology developed and the results obtained, as it presents a more complex and challenging case.

2. Literature review

The problem of estimating the alighting bus stop of transit passengers using smart cards, who are only required to carry out check-in, has already been addressed in the existing literature and similar methods have been applied at international level in different contexts (urban, rural), for different network types (surface transport, underground, railway) and for different quantities of recorded data (from metropolitan areas to small-medium towns).

Trépanier et al. (2007) proposed a model to estimate the destination location for each individual boarding a bus with a smart card; experiments were carried out on the mid-size public transport system of Gatineau, Quebec, and showed a success of 66% for destination estimation, reaching around 80% at peak hours. Cui (2006) presented the development of an algorithm to estimate a bus passenger Trip Origin-Destination (OD) Matrix based on automatic data collection system archived data including Automated Fare Collection data, Automatic Passenger Count data, and Automatic Vehicle Location data. The data used in the case study pertained to the morning peak time period (6-9 am) of 6 routes managed by the Chicago Transit Authority for five weekdays. Munizaga et al. (2011) estimated an OD public transport matrix from smart card and GPS data for Santiago, Chile, where the contactless smart card bip! is the only available payment system on buses and by far the most important in the Metro (99%). The method was applied to 36 million observations corresponding to one week of operation of the large scale multimodal public transport system of the city and it allowed to obtain 79.9% success in the estimation of trip stage alighting points. Li et al. (2011) presented an algorithm to estimate, within a satisfying computational time, passenger trajectories with various transfer distances, using pre-processed smart card data in Jinan city, China. In Jinan the number of daily bus

passengers is up to 2.5 million and the amount of card swiping is above 1.5 million; the OD estimation showed a 70% success in matching with the actual OD pairs. Wang et al. (2011) explored the application of the trip-chaining method to infer bus passenger OD from smart card transactions and automatic vehicle location data in London. Five bus routes, including two connecting suburban areas, were selected to test the OD inference procedure and about 60% of bus passenger trips using Oyster cards had both correct origin and destination estimated. The validation was carried out by comparing the results with manual passenger survey data (conducted by Transport of London every five to seven years). Lee et al. (2012) analysed a sample of transit users with different smart card types in the combined network of the Minneapolis/St. Paul metropolitan area (including bus, light rail and trains) in order to infer their origin and destination from transaction data and to better understand their travel behaviour at an individual level. Results showed a correct identification of boarding and alighting location in about 83% of cases. Moreover the authors found that activity and travel patterns differ significantly across the different fare cards types. A more recent study (Lee, Hickman, 2014) aimed at exploring and implementing a potential method of deriving transit passengers' trip purpose or activity from the automated fare collection data.

Three key assumptions are at the base of most of the algorithms used for the estimation of virtual check-outs:

- 1) users do not use private vehicles between two stages of a trip carried out with public transport;
- 2) users do not cover long distances on foot to carry out a transfer;
- 3) users carry out a tour trip, that is at the end of the day they return at the bus stop where they started their trip in the morning.

3. Case study and methodology

Four rural lines in the Province of Cuneo were examined with an average of 130 services per day and a total of 303 bus stops and 415 route segments (segments between two bus stops). In the lines examined there are 3,600 validations on a typical working day.

The first step of the analysis included the creation of the “travel diary”, that is the list of all the trips carried out by a user on a certain day, in a chronological order, and including all the relevant information (recorded or calculated), such as: date, service, line, vehicle, operator, boarding bus stop (identified by the GPS system at the check-in), boarding time (check-in time), alighting bus stop (estimated in the case of travel passes), alighting time (calculated for travel passes according to the estimated alighting bus stop and the travel time between boarding and alighting), kilometres travelled (length of the trip between the boarding and alighting bus stop) and time on board (difference between the alighting and boarding time). Three different databases were used for the creation of the “travel diary”: (1) validation (validations carried out and associated data), (2) programmed public transport service and (3) network (data on routes, route segments and bus stops).

The travel diary was analysed in order to distinguish between trips composed by one stage and trips with multiple stages (Fig. 1). Based on the structure of the service offered, which foresees an average interval of 30 mins between two consecutive buses that serve the same bus stop, the following assumptions were made:

- a trip is direct if the next BIP transaction takes place after 30 mins or more (the interval between check-out and check-in is longer than 30 mins);
- a trip is composed by different stages if the check-out/check-in interval is less than 30 mins.

In the first case the user interrupts his trip chain to carry out activities (e.g. work, studying, shopping) whereas in the second the user carries out a transfer between two vehicles (usually between 2 different lines).

An ad hoc algorithm (virtual check-outs algorithm) based on the trip chaining method was developed in order to estimate the alighting bus stop of travel pass users, who are only required to carry out check-in when boarding the bus. The developed algorithm is based on the calculation of the distance (evaluated, in our case study, using the values of latitude and longitude) between the bus stops of each route considered in the analysis. In this context a

route is the geographical object of a line. Routes have one direction only and can include different services with the same sequence of bus stops.

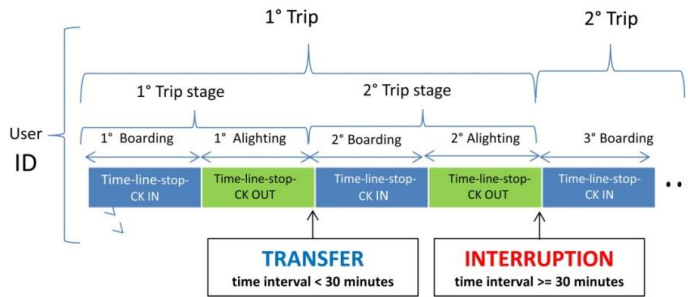


Fig. 1. Travel diary and its components

Let us define s_{igr}^j the r -th stop of the j -th route (corresponding to the j -th trip stage) of user i for day g , and $d(a,b)$ the distance between the locations a and b . The alighting bus stop, s_{igA}^j of the j -th route of user i for day g (Fig. 2) is estimated as the bus stop closest, when not exactly coinciding, to the boarding stop, s_{igB}^{j+1} , of the next route, $j+1$:

$$s_{igA}^j = \min_r d(s_{igB}^{j+1}, s_{igr}^j), \quad d(s_{igB}^{j+1}, s_{igr}^j) < D \quad (1)$$

A tolerance value of distance D is also defined (in our case study, 1 km). For the last route of the day ($j=N_g$, where N_g is the number of routes travelled by user i within the day g), following the third key-assumption listed in the literature review, the boarding stop of the next route is assumed to be the first boarding stop of the day:

$$s_{igA}^{N_g} = \min_r d(s_{igB}^1, s_{igr}^{N_g}), \quad d(s_{igB}^1, s_{igr}^{N_g}) < D \quad (2)$$

All trips whose destinations cannot be estimated with equations (1) or (2) (single trips with only one boarding stop within the day, or trips where the distance between the potential alighting stops and the boarding stop of the next route is higher than the tolerance value of distance) are defined “other trips”. In such cases, it is necessary to introduce other assumptions (e.g. the user has a recurring behaviour on different working days) or tap into other types of data (e.g. data related to the route segment that the travel pass is associated with, if available).

Given the significant amount of data, an algorithm was developed to calculate the distances between bus stops and store useful information about the potential alighting stops in a data structure defined “Look up Table” (Wang et al., 2011), which can be used by the virtual check-outs algorithm when needed.

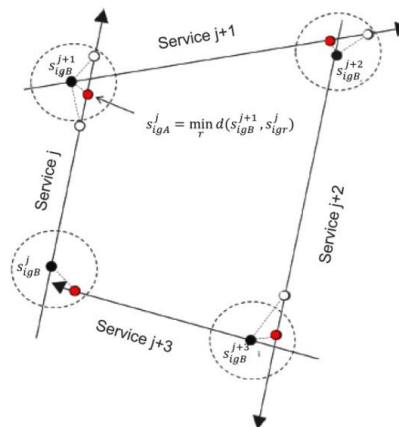


Fig. 2. Presentation of the model for the estimation of the virtual check-outs based on distance

The "Look up Table" (Table 1) is a reference table in which at each stop of a certain route (rows) are associated all the stops of the other routes considered in the analysis (columns) located at minimum distance and within a radius D (in our case study D corresponds to 1 km) from it. This algorithm runs only once (only when the network is subject to changes), distances between stops are evaluated through GIS (geographic information system) and the Look up Table is built in the programming language Python using a DICTIONARY structure that associates each couple (stop, route) with the list of its potential alighting stops and the corresponding distance.

Summarising, the virtual check-outs algorithm follows several steps:

- 1) for each BIP smart card ID, a list of the validations and boarding stops sorted by time of check-in is created. Each check-in is associated with the corresponding service so that it is possible to identify the sequence of services used and boarding stops for each trip chain ("travel diary");
- 2) services are matched with their route;
- 3) the BIP smart card ID corresponding to "single trips", whose destination cannot be estimated by equations (1) or (2), are identified and eliminated from the starting database;
- 4) for each BIP smart card ID corresponding to a normal trip and for each consecutive couple of boarding stop and previous route used, the Look up Table is consulted: the entry of the row corresponding to the boarding stop of the next route (s_{igB}^{j+1}) and the column corresponding to the previous route (j -th route) is the estimated alighting stop of user i transferring from the j -th to the $(j+1)$ -th route in day g ;
- 5) the alighting stops estimated through the Look up Table represent the virtual check-outs and are inserted in the travel diary associated to each BIP smart card ID.

Table 1. Example of a Look up Table for a route

Stop ID	Route 30 (line 115)	Route 81 (line 91)	...
695	695	/	
20045	20045	20045	
21434	21434	21516	
...			

4. Results

The analysis has revealed that, for various reasons, the data collected automatically do not allow to trace 100% of the trip stages. In the case study it was possible to estimate, with the use of the algorithm developed, the virtual check-outs for 82% of the trip stages. For 13% (e.g. single trips and trips where the distance between consecutive stops is higher than 1 km) it was necessary to use another estimation method (OD matrix expansion using the matrix obtained through the check-outs algorithm as base matrix) while 5% presented problems that made impossible the estimation of the check-out bus stop. The success rate achieved in our case study (82% of destination estimation) proved to be in line and even higher than the rates showed by the other case studies examined, probably due also to the fact that the rural public transport system analysed in Piedmont is simpler than urban or multimodal systems.

The second result of the analysis is related to the applicability and reliability of the virtual check-outs estimation method. In particular, the reliability of the model, developed for travel pass users, was evaluated by applying the same methodology to a data sample of the pay-per-use credit (where both check-in and check-out is required) and comparing the estimated alighting bus stops with those recorded automatically. In the case study more than 70% of the estimated alighting bus stops matched the one recorded or was at a maximum distance of 500 metres from it (Fig.3), which indicates a high reliability of the model for planning purposes (where a near zero error is not necessary). Nonetheless, in the case of travel passes better results should be expected as travel pass users tend to have a more regular travel behaviour (and therefore easier to simulate) than pay-per-use credit users that use public transport more randomly.

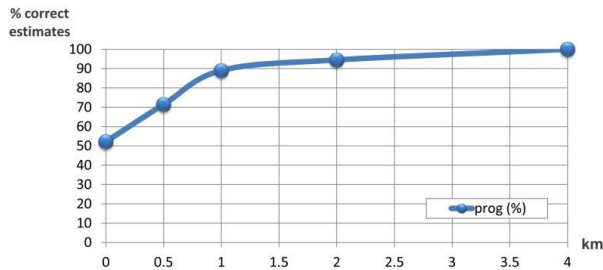


Fig. 3. Percentage of correct estimates with the variation of permissible error in terms of distance between the estimated and recorded bus stop

Thanks to the BIP system, it is possible to collect in an automatic and inexpensive way all the information regarding the use of public transport services that previously could only be supplied through costly sample surveys, the reliability of which was difficult to assess.

In addition, detailed data offer the possibility to reconstruct the daily trip chain for each user and associate it with all the relevant information regarding the travel document purchased and the user itself (e.g. age, employment, place of residence), duly maintaining user anonymity. In this way we move from the vague passenger concept, for whom we knew nothing, to the (anonymous) user concept, to whom various information can be related. For example, previously it was not possible to identify the same user on two different services and therefore he/she was considered as two different passengers. Today it is possible to know if the same user boarded two vehicles one after another, the length of the transfer and the distance between the bus stops where the transfer takes place.

Fig.4 shows the bus stop-bus stop OD matrix of a service at a single user level. Every person boarding the service is represented by a horizontal line whose ends indicate the boarding and alighting bus stops (numbered from 1 to 75). It is possible to distinguish between travel pass and pay-per-use credit users. The sum of lines between bus stops indicate the number of on board passengers during that trip stage.

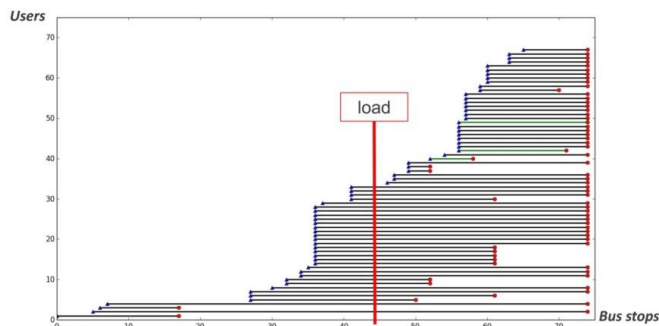


Fig. 4. Example of an OD matrix disaggregated per user and service

Identifying the users in every step of their daily travel diary provides information, for example, on how much they travel per day, week and month, and how many similar trips they carry out. In addition, it provides information on the use they make of their travel documents (e.g. km travelled, operators used) and whether they combine it with other ones. For example, it is possible to quantify the number of times the users buy a single ticket in order to travel on routes not included in their travel pass. It is also possible to define the catchment area of each bus stop (e.g. the distance between the bus stop and where the users are located, the distance over which a bus stop is no longer appealing).

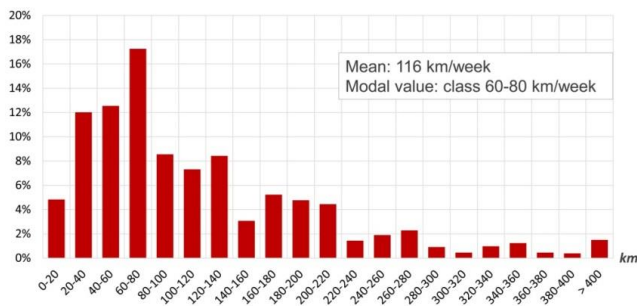


Fig. 5. Frequency of the distances covered by travel pass users within a reference week

Fig.5 shows the frequency of the distances covered by travel pass users within a chosen week (6-12 October 2014) using the public transport lines of the case study. Thanks to the data associated to the travel diary, this information can be disaggregated per type of travel document or type of user (e.g. students, elderly people) up to the individual user, allowing to calculate the price paid for each km travelled and providing, for example, useful data in the case of fare restructuring.

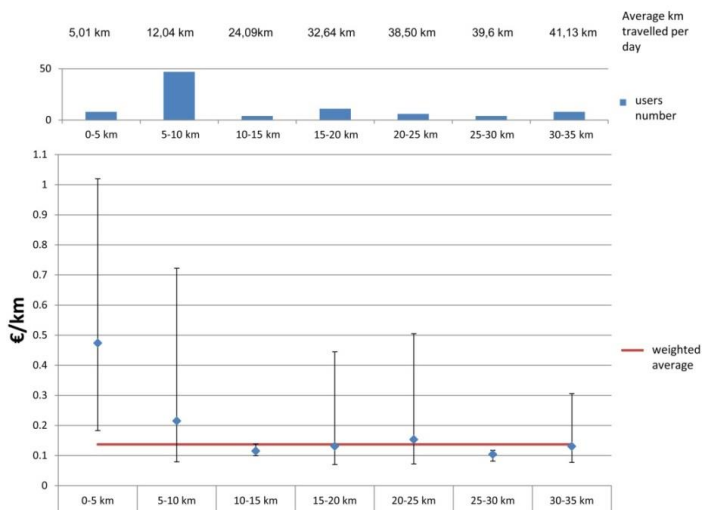


Fig. 6. Cost of the trip per km and distribution of weekly student travel pass users per distance band

Fig.6 shows the number of users owning a weekly student travel pass split into the seven distance bands on which the fare system of our case study is based. The inference of alighting stops allows to estimate the average distance per day travelled by students on the services included in the analysis (normally distances travelled are higher than the upper limit of the corresponding band since most of the users carry out round trips) and to evaluate the maximum, minimum and average cost per kilometre band. Results demonstrate that the lower distance bands (0-5/5-10 km) pay more per travelled kilometres than the other bands which show similar costs per trip.

5. Conclusions

Based on the BIP card validations database, collected during a week in October 2014 in 4 lines of the rural public transport service in the Province of Cuneo, cross referenced with the information on the programmed public transport service and network (routes, route segments and bus stops), different analyses were carried out in order to assess the quality of the data collected and explore its potential use in public transport planning and programming. Furthermore, in order to estimate the alighting bus stop in the case of travel passes (where check-out is not required), a methodology was developed, tested and its reliability evaluated.

Results revealed that the percentage of valid data regarding pay-per-use credit is very high (in the case study it was possible to reconstruct 91% of the trip stages). Furthermore it was also possible to estimate the check-out for 82% of the check-ins in the case of travel passes with the use of a distance based estimation model. The estimation model developed proved to be of high reliability. In particular, when it was applied to pay-per-use credit (in order to allow a comparison between the estimated and recorded check-outs) more than 70% of the estimated alighting bus stops coincided with the recorded ones or were within a distance of 500m from them.

The availability of reliable, detailed and continuous data regarding the use of services and bus stops facilitates the planning and programming of the public transport service, both for operators and public administrations. First, thanks to the data collected with the BIP system it is possible to carry out analysis that previously were costly, both in terms of time and money, and of limited reliability (e.g. average passenger load of a service/route/line, passenger-km, mean number of km travelled on board). Second, the data available enable analyses that previously were not possible. For example, through ad hoc algorithms, it is possible to extract OD matrices for a service, line, operator, region, etc. They could be aggregated (without distinguishing between different users) or for a single user. Furthermore, once the kilometres travelled by each user are estimated (on a service, in a day, in a month, etc.) it is possible, by associating the type of travel document used, to calculate the price paid per km travelled per user or user category. In addition, the information about the real travelled distances, conveniently combined with user categories, travel behaviours and travel documents used, could be the starting point for the feasibility study of a potential more sustainable and fair fare system where passenger pay exactly for the services they use.

The examples mentioned are just few of the possible uses of the data made available by BIP but allow to understand the significant potentials of this technology for continuous and detailed monitoring of the services, an important element, even if often underestimated, for public transport planning, programming and management.

References

- Cui, A. (2006). Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems, M.S. Thesis, Massachusetts Institute of Technology, MA.
- Lee S. G., Hickman, M. (2014). Trip purpose inference using automated fare collection data. *Public Transport* 6 (1-2), pp. 1-20.
- Lee SG, Hickman MD, Tong D (2012) Stop aggregation model (SAM): development and application. In: *Transportation research record* 2276, TRB. National Research Council, Washington, pp 38–47.
- Li, D., Lin, Y., Zhao, X., Song, H. and Zou, N. (2011). Estimating a Transit Passenger Trip Origin-Destination Matrix Using Automatic Fare Collection System. In: Xu, J., Yu, G., Zhou, S. and Unland, R. (eds.) *Database Systems for Advanced Applications. 10th International Conference, DASFAA 2005, Beijing, China, April 17-20, 2005, Proceedings*. Springer-Verlag, Berlin Heidelberg, pp. 502-513.
- Munizaga, M., Palma, C. and Fischer, D. (2011). Estimation of a Disaggregate Public Transport OD Matrix from Passive SmartCard Data from Santiago, Chile. *XV Congreso Chileno de Ingeniería de Transporte, 4-6/10/2011*. Sociedad Chilena de Ingeniería de Transporte, Chile.
- Trépanier, M., Tranchant, N. and 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), pp. 1-14.
- Wang, W., et al. (2011). Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems. *Transportation* 14 (4), pp. 131-150.