Determining the public transport demand by validation data of the electronic tickets

A Fadeev and S Alhusseini
Siberian Federal University, 79, Svobodny Avenue, Krasnoyarsk, 660041, Russia

E-mail: fai@ak1967.ru

Annotation. Currently, the definition of transport demand is relevant and not fully solved problem. Existing methods of transport demand calculating provide for on-site examination of passenger flows, which requires a lot of time and resources. Indirect methods of studying transport demand, based on the collection, integration and analysis of large and heterogeneous data generated by various sources in the spaces of human activity: mobile phones, vehicles, etc. (Urban computing [10], Big data [11, 12], internet of things, IoT) represent a special perspective. The article considers a method to determine the transport demand according to the results of the electronic tickets validation (electronic travel tickets): formulated the problem of determining the traffic demand according to system of accounting for electronic tickets and the system of satellite automobile transport monitoring (Global navigation Satellite System), the principle of its decision, the issues of obtained results adequacy.

1. Introduction
Transport demand is a set of passenger correspondence and traffic flows data in the network. Transport demand is influenced by a large number of factors; modelling and forecasting of transport demand is a non-trivial problem, the correct solution of which depends on the effectiveness of transport planning [1, 2, 3, 4].

Determination of transport demand is carried out by the four-step transport model (four-step transportation forecasting), consisting of the following procedures [4, 5, 6, 7, 8]: determining the number of trips (trip generation), the calculation of the correspondence matrix (trip distribution), the division of correspondence by type of movement (mode split, mode-share), distribution of correspondence over the network (traffic assignment, route assignment, route choice).

On-site examination of passenger flows is carried out to study the parameters of transport demand: distribution of passenger flows in directions, changes in time, evaluation of transport demand calculated methods [5]. The main drawback of on-site examination of passenger flows is the high cost of time, resources and high complexity [9]. In part, this problem is solved by equipping vehicles with automated passenger accounting systems. However, such equipment requires additional costs and is not provided in all cases.

Currently, indirect methods of studying transport demand are of interest. They based on the collection, integration and analysis of large and heterogeneous data generated by various sources in the spaces of human activity: mobile phones, vehicles, etc. (Urban computing [10], Big data [11, 12], Internet of things, IoT). This approach is used to address the main problems faced by cities: air pollution, energy consumption, traffic, etc.
Let’s consider a method of studying the demand for public transport by means of transactions intellectual analysis, travel documents (tickets) by validating a smart card, transport card, magnetic card, mobile phone or other electronic gadget, with fixing the details of the card or device in an automated transportation management system. We will call such magnetic cards and devices the electronic travel tickets.

2. Problem description
Depending on the current tariff system, the validation of the electronic ticket is performed once or twice. Double ticket validation (at the beginning and end of the trip) is carried out in cases when the cost of the trip depends on the distance or number of stops (zones), for example, on suburban and intercity bus lines. On urban passenger transport, in most cases, the fare for the trip is applied, which does not depend on the distance, the validation of the electronic ticket is made once at the beginning or end of the trip. As a result, information about one point of the trip (start or end) is recorded. The opposite point of the trip (final or initial) is unknown.

The transport management system records the ticket, the vehicle and the time of the validation operation. In modern validation devices, other data are also generated, for example, coordinates (latitude and longitude), however, in order to ensure the solution of the problem in the presence of devices of the previous generation, we will assume that the additional data of the validation system does not provide.

The problem of determining transport demand using smart cards is considered in [9 - 24]. In [13, 14, 17, 20, 22] the transport demand of metro, [15, 18, 19] - railway, [14, 22] - ground passenger transport (bus, tram) were investigated. As a result of the analysis of these works the following unresolved issues were identified:

a) the proposed algorithms for calculating passenger correspondence do not take into account the option of paying for travel at the end of the trip (before exiting the vehicle);
b) determination of trips is carried out on the basis of the one-day data analysis. In this case, the last trip is interpreted on the assumption that the passenger returns to the starting point of the first trip. The possibility of determining trips based on the analysis of validation operations carried out over a long period of time has not been considered;
C) algorithms for determining passenger correspondence do not provide recognition of a certain number of validation operations (the initial or final point of the trip remains unknown). It is required to determine the order of accounting for uncertain transactions in the calculation of the passenger correspondence matrix;
d) it is necessary to formulate a methodology for assessing the representativeness of the e-ticket validation operations sample of the trip’s general population (often travel on e-tickets is less than 50% of all traffic).

3. Problem solution
Let’s formulate the problem of determining the transport demand according to the system of accounting of electronic tickets.

In the process of performing the validation operation, the following information is recorded: the ID of the travel ticket (magnet card), the time of the operation, the vehicle ID. Thus, the set of validation operations can be represented as a relation:

\[ R(D,S,A,T) \]  \hspace{1cm} (1)

having the following attributes:
\( D \) – the operation identifier;
\( S \) - electronic travel ticket identifier;
\( A \) - vehicle identifier;
\( T \) - the operation time, which, depending on the type of fare (at the beginning or end of the trip) determines the point of loading or unloading of the passenger.
It is required to calculate the route passenger correspondence: route number, start and end points of the passenger’s trip.

In some cases, the passenger transfers during the process of moving by urban passenger transport, i.e. moves over sections of the network. It is necessary to establish the initial and final points of such a movement, i.e. network passenger correspondence. Network correspondence consists of one or more routing.

In the dispatch control system for passenger transport, navigation data packages are generated containing the vehicle identifier, the number of the travel route, the time the package was formed, and coordinates (latitude and longitude). Having processed these packages, we will form the trajectory of the vehicle along the route network. The procedure for calculating the vehicle trajectory along the route network is described in [25]. We will use this trajectory to determine the stopping point of a vehicle’s location at a given point in time: validating an electronic ticket, loading or unloading a passenger at a known stopping point.

The trajectory of the vehicle on the route network is described by relation:

\[ W(A, M, K, I, T^a, T^d) \]  

having the following attributes:
- \( M \) - route number;
- \( K \) - direction of travel along the route (forward or reverse);
- \( I \) - stopping point identifier;
- \( T^a, T^d \) - arrival time at the stopping point and departure time, respectively.

Thus, by the time of the validation operation, the vehicle identifier must determine the tuple \( W \) corresponding to the validation operation of the electronic ticket. Figure 1 illustrates the principle of determining the stopping point for loading or unloading a passenger, depending on the time of validation of the electronic ticket. In the figure, the i-th stop point includes all operations performed after arriving at this point before arriving at the next i + 1-st stop point, i.e. the i-th stop includes the j-th and j + 1-th operations of the fare payment. j + 2-nd, j + 3-rd and j + 4-th operations relate to the i + 1-st stop point.

We transform many electronic ticket validation operations into the following relation:

\[ R^i(D, S, T, W^i) \]  

where: \( W^i \) - is the identifier of the point of the vehicle trajectory.

---

**Figure 1.** Illustration of the operation stop point definition:
- \( t_j^l \), \( t_{j+1}^l \) - the time taken to validate the electronic ticket;
- \( t_j^a, t_{j+1}^a \) - time of the vehicle arrival at \( j, j+1, \ldots \) stopping point, respectively;
As mentioned above, one of the trip points was recorded in the electronic ticket validation operation. In the process of paying for travel at the trip beginning, this is the starting point, at the end of the trip is the final point. The opposite travel destination (final or initial) is determined based on the adjacent validation operation. Adjacent is the following operation when paying at the beginning of the trip or the previous operation, when paying at the end of the trip, meeting the conditions of connectedness. Adjacent operations are related (sequential) if the validation of the electronic ticket is performed at the beginning of the trip or at the end of the trip in both operations. Trips during which the electronic ticket validation is carried out in a mirror manner (at the beginning of the trip in the first and at the end of the trip in the second or vice versa) are incoherent. They do not allow to determine the route correspondence of the passenger: the initial or final point of the trip, one or both validation operations remain unrecognized (interpreted).

In addition, connected operations must meet the following conditions:

a) paying at the trip beginning, a subset of the stopping points of the current operation route located within walking distance from the validation point of the next trip;

b) paying at the end of the trip, there should be a stopping point within walking distance from the previous validation point in the subset of stopping points located in front of the route validation point of the current operation.

The algorithm for calculating passenger correspondence is considered in more detail in [9]. The result is a relational set of route correspondence:

\[ P(D, I^b, I^e, T^b, T^e, L) \]

where: \( D \) - is the identifier of the electronic validation operation;
\( I^b, I^e \) - identifier of the initial and final points of correspondence, respectively;
\( T^b, T^e \) - start and end time of the trip; \( L \) - is the length of the trip.

Network correspondence is one or more interconnected route correspondence, i.e.:

\[ h_j = \{ p_x, p_{x+1}, \ldots \} \]

The following expressions are proposed for calculating the passenger correspondence matrix:

\[ \lambda_{ij} = \sum_{k=1}^{m} p_{kij} \alpha_k^{'} \varphi_i \varphi_j \]

where: \( \lambda_{ij} \) - the estimated number of trips between points \( i, j \);
\( \alpha_k^{'} \) - coefficient of payment by electronic tickets for travel on the \( k \)-th route;
\( p_{kij} \) - the number of recognized trips between points \( i, j \) along \( k \)-th route \( (k = 1, m) \);
\( \varphi_i, \varphi_j \) - balancing coefficients of the \( i \)-th point for departure and the \( j \)-th point on arrival, respectively \( (\varphi_i >= 1; \varphi_j >= 1) \).

E-ticket fare factor:

\[ \alpha_k^{'} = \frac{R_k}{Q_k} \]

where: \( R_k, Q_k \) - the number of passengers transported by electronic tickets, and the total number of passengers on the \( k \)-th route.

The coefficients of balancing:
4. Conclusion

The method for determining transport demand considered in this article is implemented in a computer program by the Delphi programming language using the relational DBMS MS SQL Server. The effectiveness of the proposed algorithm was verified by calculating the passenger correspondence of public transport in the city of Krasnoyarsk (Russia). Krasnoyarskgortrans municipal government institution (MKU) provided 6.2 million electronic ticket validation operations for the month of October 2016 and 6.5 million operations for April 2019, as well as 144.6 million navigation data packages.

The process of calculating passenger correspondence through the program mentioned above is carried out in two stages:

- determination of the actual trajectory of vehicles by navigation of passenger correspondence;
- interpretation of passenger correspondence.

Table 1 shows, as an example, a list of itinerary passenger correspondence received as a result of processing validations of an electronic travel ticket (id = 100010328) during the month of April 2019. During the April, 11 trips were made on the electronic ticket. Each validation operation has a unique identifier (id), which is listed in the first column of the table. Trips are sorted by time, which illustrates the passenger movement chain: the identifier of the departure point (id departure stop) corresponds to the final destination (id arrival stop) of the trip in the previous row of the table. In the 8-th column of the table under consideration, the length of the trip (route correspondence) is given, in the 9-th – the distance between the stopping points of arrival and departure (which should not exceed the distance of pedestrian accessibility).

The last two rows of the table do not contain information about the passenger’s arrival points: these are unrecognized correspondence.

Table 1. Route passenger correspondence of an electronic travel ticket received based on the results of processing its validations during the April 2019 (passenger transport of the city of Krasnoyarsk).

| Id validation | # of the routes, direction (A,B) | Id boarding stop | Id arrival stop | Boarding time | Arrival time | L of the trip, km | L between the stops, km |
|---------------|----------------------------------|------------------|----------------|---------------|--------------|-------------------|------------------------|
| 13119620      | A                                | 930              | 478            | 10.04 11:40   | 10.04 12:38   | 17.38             | 0.36                   |
| 13119629      | B                                | 454              | 929            | 10.04 13:33   | 10.04 14:15   | 17.18             | 0.13                   |
| 13119625      | A                                | 930              | 815            | 12.04 13:12   | 12.04 13:52   | 13.37             | 0.08                   |
| 13119621      | B                                | 814              | 929            | 12.04 15:24   | 12.04 16:09   | 13.45             | 0.13                   |
| 13119626      | A                                | 930              | 1118           | 14.04 15:05   | 14.04 15:08   | 1.43              | 0.05                   |
| 13119630      | B                                | 938              | 926            | 14.04 16:15   | 14.04 16:18   | 0.67              | 0.00                   |
Figure 2 shows the dynamics of electronic travel tickets validations by the days of April 2019. It can be seen from the figure that transportation by electronic tickets is stable, there are clearly expressed drops in the number of validations on weekends, which correspond to changes in the total volume of transportation (passengers with electronic tickets and other methods of payment for travel).

The developed algorithm allows interpreting up to 80% of electronic travel tickets validations (figure 3). By the end of the period for which the calculation of passenger correspondence is carried out, there is a decrease in the proportion of decrypted trips. This effect is explained by the fact that some passengers travel with a break of several days. Such trips in the considered example were carried out next month, the validation operations of electronic tickets of which were not considered. In practice, the data for the last 5 billing periods is not recommended due to the absence in the result set of a part of passenger correspondence; the billing period should be accepted with overlap of at least 5 days.

Thus, the results of processing validations of electronic tickets are a sample from the general population of transport demand correspondence. It is necessary to assess the representativeness of this passenger flows general population sample.

To assess the representativeness of the sample in question, let’s compare it with the results of selective automated passenger metering provided by the Krasnoyarskgortrans Municipal Budgetary Institution. Inspection of passenger flows was carried out using special equipment mounted in vehicles. Passenger counting is provided by infrared detectors differentially at the route stop points. The volume of automated accounting amounted to more than 281 thousand passengers; 6938 flights examined for the month of April 2019.

The results of the statistical analysis established that the passenger correspondence received from validation of electronic tickets, sufficiently representative of the passenger flows of public transport, i.e. can be applied to estimate parameters of the general population passenger correspondence within the tolerance.

It is established that the coefficients of the use of electronic tickets $\alpha'_r$ (7) required for the calculating of the matrix of passenger correspondence for different routes have different mathematical expectation: these coefficients should be normalized separately for each route.

| 13119623 | 61 | B | 926  | 929  | 16.04 15:46 | 16.04 15:47 | 0.69  | 0.13  |
|-----------|----|---|------|------|------------|------------|-------|-------|
| 13119624 | 63 | A | 930  | 815  | 18.04 10:48 | 18.04 11:28 | 13.37 | 0.08  |
| 13119628 | 63 | B | 814  | 1217 | 18.04 13:51 | 18.04 14:19 | 11.44 | 0.00  |
| 13119627 | 63 | B | 1217 |     | 23.04 14:51 |           |       |       |
| 13119622 | 60 | B | 833  |     | 25.04 11:32 |           |       |       |
References

[1] Martin W A and McGuskin N A 1998 Travel Estimation Techniques for Urban Planning National Cooperative Research Program Report 365 (National Research Council: Washington DC)

[2] Shtotskaya A A and Mikhailov A Y 2017 Assessment of transport mobility based on disaggregated models Bulletin of Irkutsk State Technical University 21(5) 199-207

[3] The Online TDM Encyclopedia Retrieved from: www.vtpi.org/

[4] Yakimov M R 2013 Transport planning: creating transport models of cities (Moscow: Logos)

[5] Semenov V V and Ermakov A V 2015 Historical analysis of transport processes and transport infrastructure modeling (KIAM)

[6] Ortuzar J D and Willumsen L G 2011 Modelling transport (John Willey & Sons)

[7] Lee D 1973 Requiem for large-scale models Journal of the American Institute of Planners 39(3)

[8] Atkins S 1986 Transportation planning models: What the papers say Traffic Engineering and Control 27(9)

[9] Fadeev A, Alhusseini S and Belova E N 2018 Monitoring Public Transport Demand Using Data From Automated Fare Collection System Advances in Engineering Research, volume 158 Proceedings of the International Conference "Aviamechanical engineering and transport" (AVENT 2018) pp 5-12

[10] Zheng Y, Capra L, Wolfson O and H Yang 2014 Urban computing: Concepts, methodologies, and applications ACM Trans. Intell. Syst. Technol. 5(3) 1–55

[11] Chen M, Mao S, et al. 2014 Big Data. Related Technologies, Challenges, and Future Prospects (Spinger)

[12] Blankstein I M, Fadeev A I, Fedorov A V, Shadrin N V and Mahova E G 2015 Justification of the feasibility of studying the transport mobility of the population based on monitoring of mobile subscribers Journal of the Siberian Federal University: Technics and Technologies 8(2) 254-63

[13] Barry J J, Newhouser R, Rahbee A and Sayeda S 2002 Origin and destination estimation in New York City with automated fare system data Transportation Research 1817 183–7

[14] Zhao J, Rahbee A and Wilson N 2007 Estimating a rail passenger trip origin–destination matrix using automatic data collection systems Computer Aided Civil and Infrastructure Engineering 22 376–87

[15] 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 11 1–14

[16] Devillaine F, Munizaga M A and Trépanier M 2012 Detection activities of public transport users by analyzing smart card data Transportation Research Record 2276 48–55

[17] Munizaga M and Palma C 2012 Estimation of a disaggregate multimodal public transport OD matrix from passive smartcard data from Santiago, Chile Trans-portionation Research 24 9-18

[18] Bagchi M., White P. What role for smart-card data from bus systems? // Munic. Eng., vol. 157, 2004, pp.39–46

[19] Nassir N, Khani A, Lee S G, Noh H and Hickman M 2011 Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system Transportation Research Record 2263 140–50

[20] Cui A 2006 Bus passenger origin–destination matrix estimation using automated data collection systems (Massachusetts Institute of Technology)

[21] Wang W, John P and Nigel H M 2011 Bus passenger origin–destination estimation and travel behavior using automated data collection systems in London Journal of Public Transportation 14(4)

[22] Alsger A, Mesbah M, Ferreira L and Safi H 2015 Use of smart card fare data to estimate public transport origin–destination matrix Transportation Research Record 2535 88–96

[23] Barry J J, Freimer R and Slavin H L 2009 Use of entry-only automatic fare collection data to
estimate linked transit trips in New York City Transp. Res. Rec. J. Transp. Res. Board 2112 53–61
[24] Mahrsi K E, Côme E, Oukhellou L and Verleysen M 2017 Clustering smart card data for urban mobility analysis IEEE Transactions on intelligent transportation systems 18(3)
[25] Fadeev A I 2019 The task of determining the actual routes in the traffic control system using satellite navigation 2019 IOP Conf. Ser.: Mater. Sci. Eng. 537 022043