Monitoring Public Transport Demand Using Data From Automated Fare Collection System

A. Fadeev
Siberian Federal University (SibFU)
Krasnoyarsk, Russian Federation
fai@ak1967.ru

S. Alhusseini
Siberian Federal University (SibFU)
Krasnoyarsk, Russian Federation
Eng.sami20143@gmail.com

E.N. Belova
Siberian Federal University (SibFU)
Krasnoyarsk, Russian Federation

Abstract—The paper considers the methodology of determining the public transport demand using the data from the non-cash fare collection system. The passenger's trip from the origin stop to the destination stop is determined by analyzing the sequence of payment transactions, which were recorded and stored by an automated fare collection system. In earlier works the task of determining the trips' boarding and alighting stops was not discussed for transportation vehicles with an exit-only AFC system - the passenger taps off or uses his/her smart card only when he/she gets off the vehicle - which is used in Russia. In addition, the determination of trips' boarding and alighting stops was based on the analysis of AFC data of one day only. In this case, the last trip's alighting stop is determined, based on the assumption that the passenger returns to the first trip's boarding stop of the day. In this paper, the task of determining passenger trips is based on the analysis of smart card transactions data not only for one day but also during the subsequent periods (up to 5 days).

The analysis of the efficiency of the proposed algorithm was carried out with the use of 6.2 million non-cash transactions (for October, 2016), provided by the Municipal Public Institution of Krasnoyarsk city (MSE) "Krasnoyarskcitytrans".

Keywords—*transport demand, passenger flows, passenger trip, passengers' origin–destination trips.*

I. INTRODUCTION

The basic concept of the public transport theory, which characterizes the population transport needs, is passenger correspondence or trips' O-D. Correspondence refers to the movement of people associated with working, social and cultural purposes. Movements can be completed in various ways: by walking or using different types of transport. The movements can be considered as correspondence if they were done by using public or individual transport.

Studying the patterns of population's mobility is a golden key to solve a wide range of transport challenges and town planning problems such as: the development of the streets network, the development and optimization of transport routes

and others. Currently in megapolises and large cities there is an urgent need in the mobility of the population, therefore, studies of this kind seem topical and relevant.

To determine the mobility of the population, survey methods of questioning the population or examining passenger flows were used. Therefore, defining the transport demand by using the data from the AFC system is very effective because it significantly reduces the costs of resources to study the mobility of the population.

The automated fare collection system is widely used on public transport. Nowadays, AFC systems are also used in large cities of Russia. In the transaction which is done by using non-cash payment, the parameters of the passenger's trip are possible to determine based on the information acquired.

Payment by smart card (non-cash payment) can consist of one or two operations, depending on the type of fare. At fixed fare (which does not depend on the distance of the trip), the payment is made in one operation, during which the place and time of the operation are stored, as well as the route and transport vehicle in which the trip is carried out. And if the fare depends on the trip's distance, the payment consists of two operations at boarding and at alighting, i.e. in this case, all the necessary information about the trip is recorded (the time and stop of boarding and alighting).

Thus, using the AFC data for the study of transport demand, compared with other transport demand evaluation methods, shows significantly lower cost, and also provides a continuous (non-stop) study of transport demand, which improves the quality of it.

II. ASSIGNMENT FORMULATION

To determine transport demand according to smart card data, it is necessary to form a set of passenger correspondence N. Passenger correspondence is formed from trips: the passenger's correspondence can consist of one or several trips.

Thus, the determination of transport demand through smart cards transactions is divided into two tasks:

1.Formation of passenger trips (boarding and alighting stops).

2.Determination of passenger correspondence (passengers' O-D trips).

Forming passenger trips: this task is relevant for cases when payment by smart card consists of one operation. In this case, the following information is recorded by the AFC system:

- ID of the ticket (smart card);

- time of the transaction;

- stopping point;

- route and direction;

- transport vehicle registration number.

The payment operation can be made both at the beginning and at the end of the trip.

To determine the trip, it need to estimate the stopping point and the time of beginning or the end of the trip. And to estimate passengers' O-D trips their trip-legs should be chained together for each smart card holder, Without forgetting that some trips consist of only one leg (without transfer).

III. PRIOR RESEARCH

The results of studying transport demand by using data from AFC systems were discussed in many studies [1...12]. In the works, based on the data from AFC systems, the public transport demand was inferred for subway [1, 2, 5, 8, 10], for rail [3, 6, 7], and for land transport (bus, tram) [2...10].

In previous studies, the passenger's trip from the origin stop to the destination stop is determined by analyzing the sequence of payment transactions. In the AFC entry-only system, where the passenger required to tap off or use his/her smart card only when he/she boards the transport vehicle, i.e. the parameters of boarding stop are known and recorded in this system. The alighting stop for this trip can be inferred based on assumption, that it is located within walking distance from the following boarding stop.

Thus, the alighting stop for this trip is among the stops of the route, which was specified by current payment transaction. This stop is selected from the current rout's stops, which were located after the boarding stop and within walking distance (buffer zone) from the following boarding stop. If there are several stops in the buffer zone, the choice of the alighting stop is carried out in two ways:

a) The stop providing the most favorable condition for passenger transfer, if the stops of alighting and following boarding are located on different routes [10];

b) the closest stop to the following boarding stop.

In some cases, for example, when there are circular routes, the alighting stop which is selected as the closest does not correspond to the actual situation. For example, in Fig.1 Stop n has the smallest distance from Stop i, and depending on the direction of movement along the circular routes, the passenger can finish the trip at Stop m and transfer to Stop i. For such cases in [10], it was suggested to take into account the time of the alighting process that should be shorter than the time when the following boarding occurred.
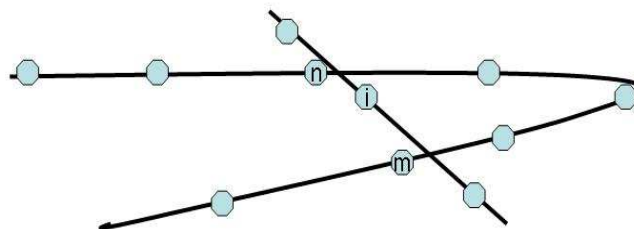

Fig. 1. An example of a circular route

In most studies, payment transactions are tracked and analyzed only for a single day. That's why the last trip alighting stop is determined based on the assumption that the passenger returns to the first boarding stop on the same day or to the nearest stop located within walking distance from the first boarding stop.

In [1], these assumptions were tested by using travel diary information collected by the New York Metropolitan Transportation Council (NYMTC), and it was found that the accuracy of these assumptions for subway passengers is about 90%.

Payment records that are not usable include the following cases:

- One trip per day was done by using smart card and it makes three percent of all subway trips which were completed by smart card.

- The first and last transactions of the day are made at the same station: the passenger return stop in this case remains unknown, the last trip is uncertain.

- Two or more consecutive transactions occur at the same stop, which is possible when two or more people travel together using one smart card, in this case only the last of these transactions is usable.

Study [2] focused on the estimation of an origin⬜destination for rail passenger trips in the Chicago city. The authors assumed that the sequential payment transactions do not interrupt with the private transportation mode trip-segment like car, motorcycle, bicycle, etc. In addition, the buffer zone was limited to 400 m in the study.

In the reviewed studies [1..12] in the phase of calculation of the passenger O-D matrix, almost identical algorithms to infer passenger trips were used based on the trip-chaining method, with minor misclosure due to the differences and variations in the structure of the AFC system in each city.

The main difference between algorithms in different studies is in the following parameters:

- Walking distance (buffer zone).

- Allowable transfer time (the interval of time between the alighting stop and the next boarding stop on which it depends whether to consider the trip as a transfer or not).

The definition of passenger correspondence was carried out by identifying transfer movement by analyzing the sequence of passenger movements. To consider a movement as transfer, there are the following conditions that should be met between this movement and the following movement:

Alighting stop of the first movement is located in a buffer walking distance from the next movement boarding stop;

the time between the alighting in the first movement and the boarding in the second movement does not exceed the allowable transfer time.

Thus, none of the above-mentioned reviewed studies discussed the procedure to determine the trip's alighting and boarding stops in the situation where the AFC system type is exit-only. Nowadays in Russia the exit-only AFC system type is used, especially in buses with small capacity. Also, the possibility of inferring last trips alighting stop based on the analysis of the payment transactions which were made on the following day has not been considered. And the analyzed transactions were limited by one day.

IV. METHODOLOGY

In this paper, we will consider the previous remarks while building the algorithm for inferring the alighting or boarding stop as well as the algorithm for estimating passengers' O-D trips, and the results of application of these algorithms for determining the public transport demand in the city of Krasnoyarsk will be presented.

The following information is recorded by the AFC system when payment transaction occur:

- ID of the ticket (smart card);

- time of the transaction;

- stopping point;

- route and direction;

- transport vehicle registration number.

As mentioned previously, in public transport in Krasnoyarsk city, there are two options for non-cash fares payment:

a) when boarding the transport vehicle;

b) while alighting from the transport vehicle.

The second option of payment is operated in buses with small capacity.

Thus, a set of transactions is a relation $R(D,K,G,A,V,T)$, that has the following variables:

D - transaction ID (unique transaction number);

K - the ID of the ticket;

G- route number;

A - vehicle model;

V - ID of the stop;

T - time of transaction.

A transaction is a relationship tuple (an instance of an object), which is denoted by $r(d,k,g,a,v,t)$.

The route consists of an ordered set of stopping points and the coordinates of each stop location (x, y) are known, so the distances between any stops can be determined, which means that determining the distance between origin and destination stops is possible.

In general, a route is considered as movement line from the origin to the destination route stop, so the main and reverse directions of the route will be considered as separate routes. This to some extent simplifies the task, because the problems associated with an unequal number of stopping points and different distances between stops along the two directions of the route are eliminated.

The trips set is $P(D,K,G,A,V^b,T^b,V^e,T^e,L)$ where:

V^b , V^e - the ID of the trip's boarding and alighting stops;

T^b , T^e - time at boarding and alighting;

L is the trip distances; variables D, K, G and A are recorded by the AFC system.

Passenger correspondence can consist of one or several trips i.e. $h_i = \{p_i^1,...\}$. The algorithm for formation of the correspondence based on the trip-chaining method was explained in many studies, for example, in [10], so it will not be addressed in this paper.

The presented algorithm will estimate trips from the set of passenger payment transactions. The algorithm is based on the analysis of the set of passenger payment transactions, which had been time-sorted. For transactions' set, the algorithm is repeated for each of them.

As mentioned earlier, the transaction recorded one of the trip's stop - the boarding stop, if the AFC system in the transport vehicle is an enter-only system, or the alighting stop, if it is an exit-only system. The missing trip stop (boarding or alighting) is inferred based on the adjacent transaction. The adjacent transaction is the next transaction when the payment is made at a boarding stop or the previous transaction, when the payment is made at an alighting stop. To infer the missing stop of the trip, it should be determined whether the adjacent transactions are connected (sequent) or not, i.e. the movement of the passenger was carried out using the public transport. If the sequence of trips is interrupted, for example, for one of the trips the passenger used a taxi, a not usable trip is recorded, for which the missing stop remains unknown.

Thus, consecutive transactions are to meet one of the following conditions:

a) From the set of the route's stops, which were located beyond the payment point when the payment transaction occurred at the boarding stop, there is a stop located within a walking distance from the next payment point, provided that the next payment transaction also occurred at the boarding stop.

b) From the set of the route's stops which were located before the payment point when the payment transaction occurred at the alighting stop, there is a stop located within a walking distance from the previous payment point, provided that the previous payment transaction also occurred at the alighting stop.

Thus, the condition of the consecutive transactions can be written as follows:

a) $l(v_j^+, v_{j+1}) < L_p$

b) $l(v_j^-, v_{j-1}) < L_p$

where: L_p - walking distance (buffer zone);

v_{j+1} , v_{j-1} - the payment point of the next or previous transaction, respectively;

v_j^+ - the route's stops, which were located after the payment point of the current transaction;

v_j^- - the route's stops, which were located before the payment point of the current transaction.

Since the stops of the current and next/previous transaction can belong to different routes, their global coordinates will be used to calculate the distance between them.

To determine the type of consecutive transactions, the possible options of them should be considered, which are shown in Figure 2. Here:

a) payment for the transaction (i) at the boarding stop, and for the transaction (i + 1) - at the alighting stop or v.v., for (i) at the alighting stop, and for (i + 1) at the boarding stop. In such cases, the recorded stops by transactions i and (i + 1) are within a walking distance. Obviously, this will not allow inferring the missing stop for one of these transactions;

b) b) the current and next transaction with payment at the boarding stops. For such cases, in the route of the current (i) transaction there is a stop beyond the payment point and it is within walking distance from the payment point for the transaction (i + 1);

c) current (i) and previous (i-1) transactions with payment at the alighting stops. For such cases, in the route of the current (i) transaction there is a stop, before the payment point and located within walking distance from the payment point for the previous transaction previous (i-1).
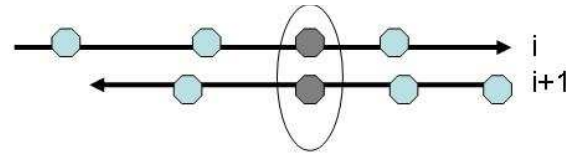
a) consecutive transactions with different types
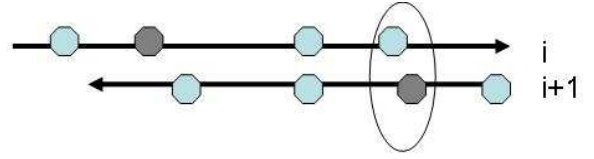
b) current and next transaction with payment at the boarding stop
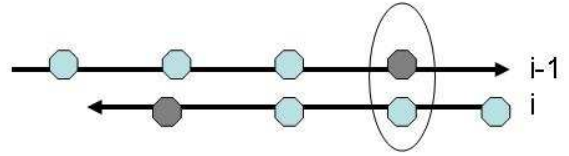
c) current and previous transaction with payment at the alighting stop

Fig. 2. Possible consecutive transaction options

Based on what is said above, an algorithm for formatting the passenger trips has been formulated.

The set of non-cash payment transactions of one passenger has a chronological order for a certain period of time.

V. DESCRIPTION OF THE ALGORITHM

Step 1. Let us select the current (j) transaction. At the beginning of the calculation, the current transaction is the first transaction in the studying period. In later stages, the current transaction is the one following the (j) transaction from the previous stage.

Based on the data of the current transaction, the passenger (j) trip is formed $p_j(d_j, k_j, g_j, a_j, v_j^b, t_j^b, v_j^e, t_j^e, l_j)$.

If no current transaction is selected (all transactions are processed), complete the calculation. If the (j) transaction exists and is selected, proceed to Step 2.1.

Step 2. Formation of the passenger's trips.

2.1. Let us consider that the payment for the current transaction was made at the boarding stop. Let us choose the next transaction (j + 1). If the next transaction does not exist, proceed to step 2.2.

Let us verify the previous assumption that the payment occurs at the boarding stop, which is true if the following conditions are met:

a) The payment points in the current and next transactions are not located in one buffering zone (the distance between them is bigger than a walking distance) $l(v_j, v_{j+1}) > L_p$;

b) in the route of current transaction there is a stop following the payment point, which is within walking distance from the payment point of the next transaction $l(v_j^+, v_{j+1}) < L_p$.

The boarding stop and time of the(j) trip (v_j^b, t_j^b), if these conditions are met, are the time and stop of payment for the current transaction (v_j, t_j) and the alighting stop (v_j^e) for (j) trip is selected from the set of the (j) transaction route's stops that has the shortest distance to the next transaction payment point (j + 1). The alighting time (t_j^e) of the (j) trip can be obtained from the database of the monitoring the movement of the public transport along the routes. In this paper, in order to

test the algorithm, the alighting time was determined by using the average speed of the transport vehicle on the route.

The trip distance (l_j) is calculated as the sum of the distances between the stops along the route (g_j) from the boarding stop (v_j^b) to the alighting stop (v_j^e).

Proceed to step 1.

2.2. Let us consider that the payment for current transaction was made at the alighting stop. Let us choose the previous transaction (j - 1). If the previous transaction does not exist, proceed to "the (j) trip remains uncertain" as one cannot infer the boarding stop, proceed to step 1. If the previous transaction (j - 1) exists, verify the condition of consecutive
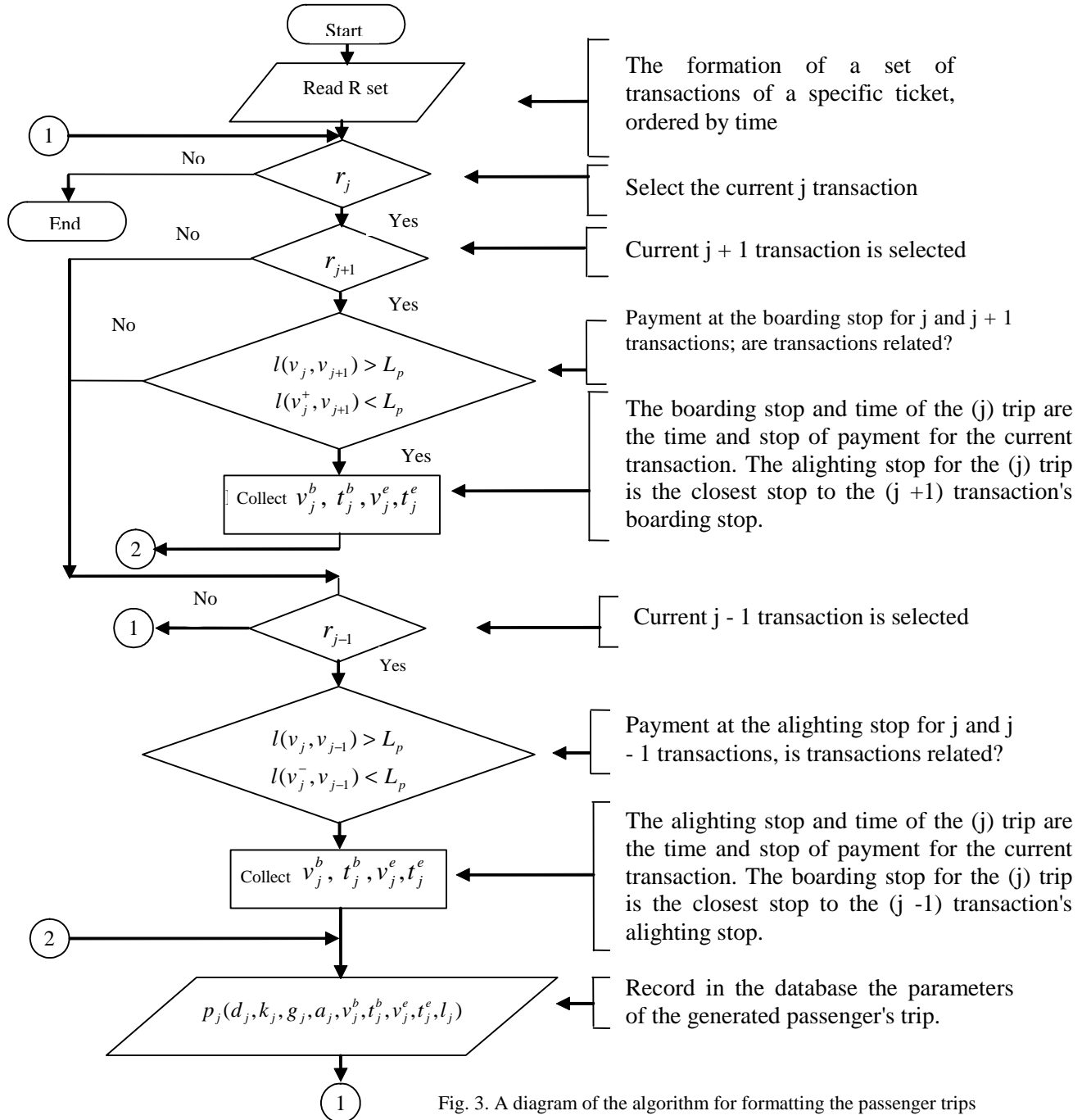

Fig. 3. A diagram of the algorithm for formatting the passenger trips

transactions (j) and (j-1):

a) the payment points in the current and previous transactions are not located in one buffering zone (the distance between them is bigger than a walking distance) $l(v_j, v_{j-1}) > L_p$;

b) in the route of current transaction, there is a stop before the payment point, which is within walking distance from the payment point of the previous transaction $l(v_j^-, v_{j-1}) < L_p$.

The alighting stop and time of the (j) trip (v_j^e, t_j^e), if these conditions are met, are the time and stop of payment for the current transaction (v_j, t_j) and the boarding stop (v_j^b) for (j) trip is selected from the set of the (j) transaction route's stops that has the shortest distance to the previous transaction payment point (j - 1). The boarding time (t_j^b) of the (j) trip can be obtained from the database of the monitoring the movement of the public transport along the routes. In this paper, it was determined using the average speed of transport vehicle on the route.

The trip distance (l_j) is calculated as the sum of the distances between the stops along the route (g_j) from the boarding stop (v_j^b) to the alighting stop (v_j^e).

Proceed to step 1.

Figure 3 presents the proposed algorithm.

VI. DISCUSSION OF RESULTS

The previous algorithm for formatting the passenger trips, as well as the algorithm for determining passenger correspondence, which was described in [10], is implemented in a computer program based on Delphi programming language with using a database management system (DBMS). In the current version of the program, it is adapted for Microsoft Access and MS SQL Server. The standard software components used in the program allow using modern DBMS like: Oracle, InterBase, DB2, etc. without significant correction.

The DBMS was chosen due to the efficiency of the relational algebra engine through the SQL language. As a result, a software package has been obtained that allows processing a large amount of data. For example, the monthly volume of non-cash payment transactions of public transportation for Krasnoyarsk city (more than 6 million transactions) is processed in MS SQL Server in 20 hours of computer operation.

The analysis of the efficiency of the proposed algorithm was carried out with the use of 6.2 million non-cash transactions (for October, 2016), provided by the Municipal Public Institution of Krasnoyarsk city (MSE) "Krasnoyarskcitytrans".

The summary of the processing results is presented in Table 1. As shown in the table, the algorithm allows

interpreting in average about 71.9% of transactions. The decrease in the percent of the formatted trips by the end of the month is due to the absence of subsequent trips in the transaction database, which were carried out in the next month, i.e. some passengers continue the sequence of their transactions after a pause interval of several days. When formatting trips for the required day, the database should contain transactions for five days later. Thus, the average percent of the formatted trips and the transfer factor are determined for the 25th day of the month (Table 1).

Figure 4 shows the data of the formatted trips' percent by days of the month. The significant decrease in the percent of the formatted trips by the end of the analyzed period can be clearly seen.
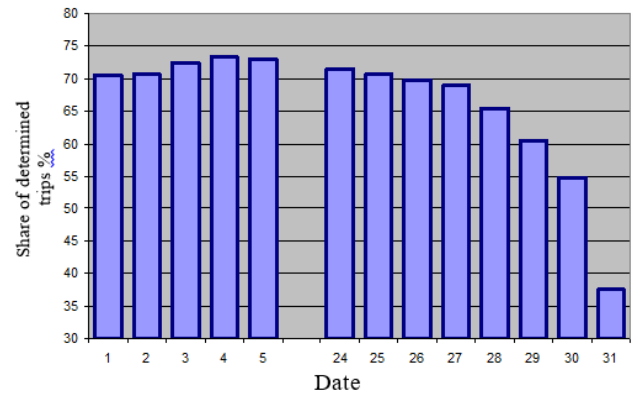

Fig. 4. Dynamics of the formatted trips' percent by days of the month

Figure 5 shows the distribution of the number of passengers' correspondences by hours of the day, which have been obtained based on the non-cash payment transactions data from the AFC system for October 2016 (the solid red line) and the results of intercept survey of passenger movement in November 2006 (the dotted line). From the graph it can be assumed that the sample of passenger correspondences, obtained from the AFC system data, corresponds to the general set of passenger correspondence obtained from the survey.
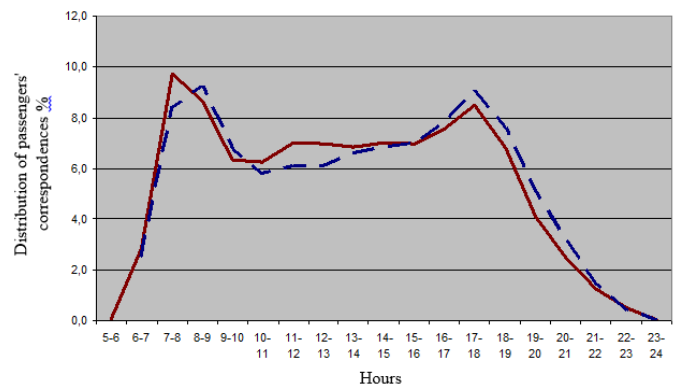

Fig. 5. Distribution of passengers' correspondences by hours of the day

TABLE I.   Summary processing results of the non-cash payment transactions in Krasnoyarsk city for October 2016

Date	Day of the Week	Transaction (thousand)	Determined Trips (thousand)	Share of Determined Trips, %	Passengers' Correspodance, (thousand)	Transfer Factor
1	6	151,0	106,4	70,5	96,3	1,10
2	7	116,2	82,1	70,7	74,3	1,11
3	1	223,9	162,2	72,4	147,2	1,10
4	2	228,2	167,3	73,3	152,3	1,10
5	3	232,6	169,9	73,0	154,5	1,10
6	4	230,1	167,2	72,7	151,9	1,10
7	5	233,3	167,9	72,0	152,5	1,10
8	6	152,5	108,6	71,2	98,8	1,10
9	7	116,1	81,9	70,6	74,5	1,10
10	1	226,6	163,8	72,3	149,2	1,10
11	2	233,4	168,9	72,4	153,7	1,10
12	3	231,7	169,6	73,2	154,3	1,10
13	4	233,5	170,9	73,2	154,9	1,10
14	5	242,6	176,4	72,7	160,0	1,10
15	6	161,1	114,7	71,2	104,0	1,10
16	7	119,4	83,9	70,2	76,0	1,10
17	1	241,5	174,2	72,1	158,2	1,10
18	2	243,2	176,5	72,5	160,3	1,10
19	3	239,1	173,2	72,4	157,5	1,10
20	4	239,6	174,9	73,0	159,0	1,10
21	5	234,1	166,1	70,9	151,2	1,10
22	6	147,5	100,8	68,3	92,1	1,09
23	7	115,6	79,3	68,6	72,1	1,10
24	1	223,1	159,3	71,4	145,3	1,10
25	2	231,0	163,3	70,7	149,1	1,10
Average for 25 days				71,9		1,10
26	3	236,3	164,9	69,8	150,4	1,10
27	4	229,0	158,3	69,1	144,8	1,09
28	5	223,9	146,5	65,5	134,0	1,09
29	6	147,4	89,2	60,5	81,1	1,10
30	7	107,1	58,7	54,8	53,6	1,10
31	1	216,5	81,3	37,5	74,1	1,10
Monthly average				69,7		1,10

To evaluate the set of passenger correspondence, the following approximation parameters as in (1) and (2) can be calculated:

1. Coefficient of non-cash payment transaction:

$$\alpha_t = R / N_t \qquad (1)$$

where R - the number of non-cash payment transactions;

N_T - the number of all transactions (cash and non-cash payments).

2. Coefficient trips formation:

$$\alpha_p = N_t^{'} / R \qquad (2)$$

where $N_t^{'}$ - the number of trips formatted from non-cash payment transactions data.

To calculate the coefficient of non-cash payment transaction, the results of a passenger movement survey or reporting data from carriers can be used.

Obviously, α_T cannot be regarded as a constant value. In the future, the coefficient is likely to grow due to the dramatic increase in the use of the smart card and non-cash payment in public transportation, which will result in continuous development of the monitoring systems and upgrading the values of the non-cash payment transaction coefficient.

VII. CONCLUSIONS AND NEXT STEPS

1. Studying the patterns of population's mobility is a golden key to solve a wide range of transport challenges and town planning problems such as: the development of a streets network, the development and optimization of transport routes and others. Currently in megapolises and large cities, there is an urgent need in the mobility of the population; therefore, studies of this kind seem topical and relevant.

To determine the mobility of the population, survey methods of questioning the population or examining passenger flows were used. Therefore, defining the transport demand by using the data from the AFC system is very effective because it significantly reduces costs of resources that are needed to study the mobility of the population.

2. Through the AFC system data, the following tasks of monitoring public transport demand can be solved:

a) Formation of passenger trips (boarding and alighting stops).

b) Determination of passenger correspondence (passengers' O-D trips).

3. None of the reviewed studies [1...12] had discussed the procedure to determine the trip's alighting and boarding stops in the situation where the AFC system type is exit-only. Nowadays such AFC systems exist at least in Russia, so to correctly use the algorithm for estimating passengers' O-D trips based on AFC system data, it is necessary to ensure that data from such AFC systems (exit-only system) are processed.

4. In majority of the studies, inferring trip's missing stop is based on the analysis of AFC data from one day only. In this case, the last trip's alighting stop is determined, based on the assumption that the passenger returns to the first trip's boarding stop of the day. In this paper, the task of determining passenger trips is based on the analysis of smart card transactions data not only for one day but also during the following periods (up to 5 days).

5. The paper presents a relational model that allows the use of the relational algebra engine through the SQL language to solve the task of formatting passenger trips based on data from both types of AFC systems (enter-only and exit-only).

The developed algorithm allows interpreting in average about 71.9% of transactions. By the end of the analyzed period, the percent of the formatted trips is decreasing, which is explained by the fact of absence of the subsequent trips in the transaction database, which were carried out the next month, i.e. some passengers continue the sequence of their transactions after a pause interval of several days.

6. The task of matching the sample set of trips, obtained from the AFC system data, with a general set of trips (for cash and non-cash transaction) should be included in subsequent studies. It is necessary to prove that the basic parameters of the sample correspond to the general one with sufficient accuracy: the distribution of trips over the city network, transfer factor, trips distances, etc. It is also necessary to consider the issue with the coefficient of non-cash payment transaction: should it be calculated on average over the city network, or is the differentiation based on the types of the public transport (bus, tram trolleybus, microbus), etc. required?

References

[1] J.J. Barry, R. Newhouser, A. Rahbee, S. Sayeda, "Origin and destination estimation in New York City with automated fare system data," Transportation Research, Record 1817, 2002, pp.183–187.

[2] J. Zhao, A. Rahbee, N. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems;" Computer-Aided Civil and Infrastructure Engineering, vol. 22, 2007, pp.376–387.

[3] M. Trépanier, N. Tranchant, R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," Journal of Intelligent Transportation Systems, vol. 11, 2007, pp.1–14.

[4] F. Devillaine, M.A. Munizaga, M. Trépanier, "Detection activities of public transport users by analyzing smart card data," Transportation Research Record: Journal of the Transportation Research Board, vol. 2276, 2012, pp.48–55.

[5] M. Munizaga, C. Palma, "Estimation of a disaggregate multimodal public transport OD matrix from passive smartcard data from Santiago, Chile," Trans-portation Research Part C, Vol. 24, 2012, pp.9-18.

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

[7] N. Nassir, A. Khani, S.G. Lee, H. Noh, M. Hickman, "Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system," Transportation Research Record: Journal of the Transportation Research Board, vol. 2263, 2011, pp.140–150.

[8] A. Cui, Bus passenger origin–destination matrix estimation using automated data collection systems master's dissertation, Massachusetts Institute of Technology, 2006.

[9] W. Wang, P. John, H.M. Nigel, "Bus passenger origin–destination estimation and travel behavior using automated data collection systems in london," Journal of Public Transportation, Vol. 14, No. 4, 2011.

[10] A. Alsger, M. Mesbah, L. Ferreira, H. Safi, "Use of smart card fare data to estimate public transport origin–destination matrix," Transportation Research Record: Journal of the Transportation Research Board, vol. 2535, 2015, pp.88–96.

[11] J.J. Barry, R. Freimer, H.L. Slavin, "Use of entry-only automatic fare collection data to estimate linked transit trips in New York City," Transp. Res. Rec. J. Transp. Res. Board, vol. 2112, 2009, pp.53–61.

[12] K. El. Mahrsi, E. Côme, L. Oukhellou, M. Verleysen, "Clustering smart card data for urban mobility analysis," IEEE Transactions on intelligent transportation systems, vol. 18, NO. 3, March 2017.