Transportation Research Record

Understanding Reverse Routing Path Choice Behavior in Congested Metro Systems

--Manuscript Draft--

Full Title:	Understanding Reverse Routing Path Choice Behavior in Congested Metro Systems
Abstract:	Passenger path choice in highly congested metro systems is affected by crowding in vehicles and denied boarding to trains. This paper deals with a special situation where passengers choose to stay in the train longer than what would be preferable under normal conditions and transfer at a station further along the line in order to travel backwards and pass the same station which has already been passed. An earlier study have identified that some passengers take advantage of this reverse routing behavior in the MTR metro system in Hong Kong, which serves as the case study for this paper. Smart card data and a passenger-to-train assignment model combining the automatic fare collection data to automatic vehicle location data are used for analyzing the underlying causes for this behavior. A detailed model for passenger path choice developed by Ma et al. (2019) is used to determine the fractions of reverse routing passengers. The case study is used to analyse the fractions and underlying causes of reverse routing. The results show that 35%-55% of passengers in the case study in the peak hour from 18:00-19:00 are reverse routing. The passengers who travel furthest after transferring have a significantly higher probability of doing reverse routing. The analysis also shows that more experienced passengers have lower journey times than less experienced passengers indicating lower probability of doing reverse routing because of their system knowledge. The results can help agencies evaluating new operational strategies and reduce the crowding levels ultimately benefitting the passengers.
Manuscript Classifications:	Public Transportation Planning and Development AP025; Transit Assignment; Public Transportation; Route Choice and Spatio-Temporal Behavior ADB30SC
Manuscript Number:	
Article Type:	Presentation and Publication
Order of Authors:	Morten Eltved, PhD Candidate
	Haris N. Koutsopoulos, Professor
	Nigel H.M. Wilson, Professor Emeritus
	Kerem Tuncel, PhD Candidate
	Zhenliang (Mike) Ma, Lecturer

Understanding Reverse Routing Path Choice Behaviour in Congested Metro Systems

Morten Eltved
Department of Management
Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
Email: morel@dtu.dk
ORCID: 0000-0003-1079-8127

Haris N. Koutsopoulos
Department of Civil and Environmental Engineering
Northeastern University, Boston, MA, 02115
Email: h.koutsopoulos@northeastern.edu

Nigel H.M. Wilson
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology, Cambridge, MA, 02139
Email: nhmw@mit.edu

Kerem Tuncel
Department of Mechanical and Industrial Engineering
Northeastern University, Boston, MA, 02115
Email: k.tuncel@northeastern.edu

Zhenliang (Mike) Ma
Department of Civil Engineering
Monash University, Clayton, VIC, Australia, 3800
email: mike.ma@monash.edu
ORCID: 0000-0002-2141-0389

Word Count: 4,925 words + 2 tables * 250 words = 5,425 words

Submitted August 1st 2019

ABSTRACT

Passenger path choice in highly congested metro systems is affected by crowding in vehicles and denied boarding to trains. This paper deals with a special situation where passengers choose to stay in the train longer than what would be preferable under normal conditions and transfer at a station further along the line in order to travel backwards and pass the same station which has already been passed.

An earlier study have identified that some passengers take advantage of this reverse routing behavior in the MTR metro system in Hong Kong, which serves as the case study for this paper. Smart card data and a passenger-to-train assignment model combining the automatic fare collection data to automatic vehicle location data are used for analyzing the underlying causes for this behavior. A detailed model for passenger path choice developed by Ma et al. (2019) is used to determine the fractions of reverse routing passengers.

The case study is used to analyse the fractions and underlying causes of reverse routing. The results show that 35%-55% of passengers in the case study in the peak hour from 18:00-19:00 are reverse routing. The passengers who travel furthest after transferring have a significantly higher probability of doing reverse routing. The analysis also shows that more experienced passengers have lower journey times than less experienced passengers indicating lower probability of doing reverse routing because of their system knowledge. The results can help agencies evaluating new operational strategies and reduce the crowding levels ultimately benefitting the passengers.

Keywords: Transit, Metro systems, Crowding, Reverse routing, Travelling backwards

INTRODUCTION

Ridership in transit systems are constrained by the capacity of the system and overcrowding in peak hours on these systems can lead to special circumstances where the passenger can gain an advantage by choosing paths, which under normal conditions would be dominated alternatives. This paper deals with the special situation where passengers choose to stay in the train longer than what would be preferable under normal conditions and transfer at a station further along the line in order to travel backwards and pass the same station which has already been passed. The principle of reverse routing is illustrated in Figure 1 below. A passenger travelling from O to D would under normal conditions transfer at station T if the transfer at station T is as convienient as the transfer at station A. Transferring at station T minimizes the in-vehicle time and in high-frequent systems almost certainly allows the passenger to board an earlier departing train than if the transfer was made at station A. However, if the passenger risks to be denied boarding several times at station T or have a higher probability of obtaining a seat by transferring at station A, the passenger might choose to stay on line X until station A and transfer here and thus perform what in this paper is defined as reverse routing.

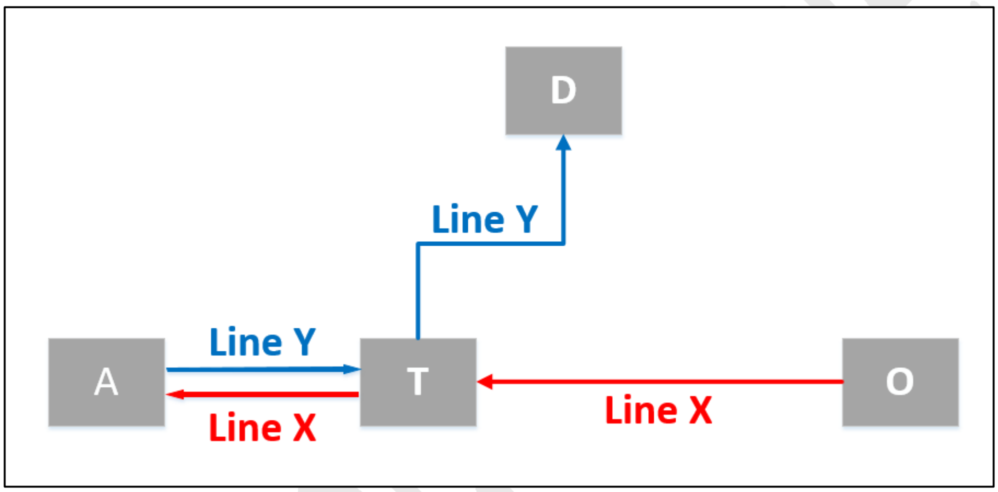

FIGURE 1 - Concept of reverse routing

It is well-known in the literature that crowding in public transport is not comfortable for passengers (see e.g. Li and Henscher, 2013, Tirachini et al., 2017, Haywood, Koning & Monchambert, 2017 or Batarce, Muñoz & Ortúzar, 2017). Under very crowded situations passengers experience the in-vehicle time up to 2.5 times the actual in-vehicle time (Batarce, Muñoz & Ortúzar, 2017), however, it is also important to note that some passengers experience crowding with a high degree of discomfort while others experience it less uncomfortable (Tirachini et al., 2017).

Given the high discomfort of in-vehicle crowding passengers also react to this in their path choice in metro systems. Kim et al., 2015 investigated the effect of crowding on passenger path choice and found that not only does crowding and the resulting increased travel time affect the path choice, but the discomfort of crowding itself does also affect the path choice. This shows that passengers do not solely base their path choice on travel time, but also take into account the comfort of the trip including well-known parameters such as waiting time, walking time and number of transfers (see e.g. Raveau et al., 2014).

Literature on unusual path choice behavior

Although many studies have focused on evaluating the cost of crowding and investigated the effect on path choice in transit systems, very few studies have dealt with unusual path choices such as reverse routing. No studies have specifically dealt with the example of reverse routing as shown in Figure 1, however a handful of studies have considered the concept of “travelling backwards”. This concept is a slightly different case than explained on Figure 1, where passengers instead board a train in the “backwards” direction of the direction they want to travel to transfer at a station further up the line and pass the origin station in the “forward” direction. This behavior has been identified in the metro systems in Singapore (Chakirov & Erath, 2011, Othman et al., 2015 and Tirachini et al., 2016) and Bejing (Li et al., 2017 and Xu et al., 2018). The behavior is typically seen at penultimate stations (stations near the start of a line), where passengers at the second or third station travel to the starting station of the line and in this way has a very high probability of obtaining a seat.

Chakirov & Erath (2011) was the first paper to verify the unusual behavior of travelling backwards based on data from Singapore. They used estimations of waiting times calculated based on the fastest possible person through the system to find that the distribution of waiting times at stations close to the starting station was bimodal. Since no denied boarding was observed at these stations they were able to justify that this bimodal distribution must then stem from passengers travelling backwards. Based on this finding, they concluded that some passengers were in this case willing to offer ten minutes of extra in-vehicle time in exchange of a seat. Othman et al. (2015) also studied the case of Singapore and focused on the development of an agent-based model to estimate effects of crowding in the metro system. They developed a simple model to replicate the empirically observed bimodal journey time distributions, which took into account the number of stations the passenger had to travel on a given line and how many stations the passenger travelled backwards. This improved their model and gave a more realistic determination of the crowding levels in the system. The final study which used Singapore as the case was Tirachini et al. (2016). They specifically used the observations of passengers travelling backwards and quantified the standing multiplier to be around 1.2 compared to being seated with the current crowding levels.

The studies in Bejing focused on analyzing the fraction of passengers travelling backwards, also focusing on cases where passengers at penultimate stations travel backwards to the first station of the line. Li et al. (2017) developed a clustering methodology to group passengers based on their journey times. By comparing the results to observed travel behavior at some stations, they were able to estimate that up to 10% of passengers travelled backwards in some OD relations in peak hours and that the proportion of passengers travelling backwards increased with the distance between origin and destination. Xu et al. (2018) refined the methodology developed in Li et al. (2017) and developed a clustering methodology determining if the passenger travelled backwards or not. They were able to identify specific stations on a specific line, where up to 10% of passengers travel backwards.

This paper extends the knowledge on unusual path choices and investigates the underlying causes for why passengers are doing reverse routing and uses a detailed path choice model to estimate the fraction of reverse routing passengers in the MTR metro system in Hong Kong. The next section presents the methodology used to analyse the problem. This is followed by the case study and the results of the analyses and finally a conclusion wrap up the findings.

METHODOLOGY

The methodology builds on two available data sources: automatic fare collection (AFC) data with tap-in and tap-out information and automatic vehicle location (AVL) data with departure and arrival times to stations. Passengers in closed metro systems only tap-in at the origin and tap-out at the destination and no information on the transfer stations is recorded in the AFC data. The idea for analyzing the potential factors affecting the reverse routing behavior is therefore to benefit of passenger-to-train assignment models to identify which trains passengers have boarded and in this way eliminating some uncertainty in the journey times from tap-in to tap-out time. As this paper is only focusing on reverse routing cases, where passengers choose to transfer at a station further from the station where passengers would mostly transfer at under normal conditions, the aim of the passenger-to-train assignment is to determine the specific train that the passenger boarded. For the first part of the analysis the focus is solely on determining the train that passengers boarded on the second leg of their trip. Returning to the sketch in Figure 1 this means, that the time which is analysed is the time from tap-in at station O to when the passenger leaves station T. The time spent on line Y between station T and D can be eliminated from the journey time since the passenger is assigned to a specific train where the departure time from station T is known. Below the passenger-to-train assignment methodology is described in further detail.

Passenger-to train assignment

The passenger to train assignment utilizes the egress time on the destination station. It is assumed that each group of passengers who boarded different trains have the same egress time distribution. Thus the egress times of passengers tapping-out in different time intervals is assumed to be generated from the same distribution that is specific to the destination platform. Based on this assumption, a sample of egress times can be acquired by looking at the passengers who have single feasible trains in their feasible itineraries and tapped-in on the same line. However, this is a biased sample since the passengers who have a single feasible train are conditioned to have an egress time that is smaller than the headway between their boarded train and the next train. Therefore, some correction for this bias is necessary. This correction is made using a truncated distribution to represent the observed egress times (Zhu, Koutsopoulos & Wilson, 2017). Given that the headway of each passenger serves as the upper bound for the egress time, the egress time distribution can be written as a truncated random variable using the following equation;

$$f(t^e|t^e < H) = \frac{g(t^e)}{F(H)} \tag{1}$$

where t^e is the egress time, H is the headway, f(x) is the probability density function associated with the egress time and F(x) is the cumulative distribution function associated with the egress time. Also, g(x) = f(x) for all x<H and g(x) = 0 for other values. Using this formulation, any continuous probability distribution can be fitted to the observed egress times using the following likelihood function;

$$L = \prod_i f(t_i^e|t_i^e < H_i) \tag{2}$$

where t_i^e is the egress time for i[th] passenger. Based on the corrected egress time distribution, we can evaluate all the possible egress time values that can be associated with a passenger. Then, it is trivial to assign each passenger to the train with the highest probability within her feasible train set. A posterior probability can be calculated for each passenger and each feasible train using the possible egress times;

$$P_{ij} = \frac{f(t_{ij}^e)}{\sum_k f(t_{ik}^e)} \tag{3}$$

where P_{ij} is the probability of passenger i boarding train j and t_{ij}^e is the egress time associated with passenger i, if that passenger boarded train j. Thus $f(t_{ij}^e)$ is the pdf value of observing that egress time value. For the purposes of this study, lognormal distribution is used to represent the egress time distribution since it is known to be used to represent walking times (Zhu, Koutsopoulos & Wilson, 2017). In Figure 2 an example of the passenger-to-train assignment is shown. The egress time distribution is modelled in a previous step and based on this distribution the most likely train the passenger boarded on line Y is train Y3. With this knowledge the departure time from station T can be found using the AVL data and the time from tap-in to departure from station T denoted τ is defined by:

$$\tau = t_T - t_O \tag{4}$$

The time τ does not explain whether the passenger transferred at station T or station A, but can give an indication whether some passengers spend more time than others given that they departed with the same train on line Y.

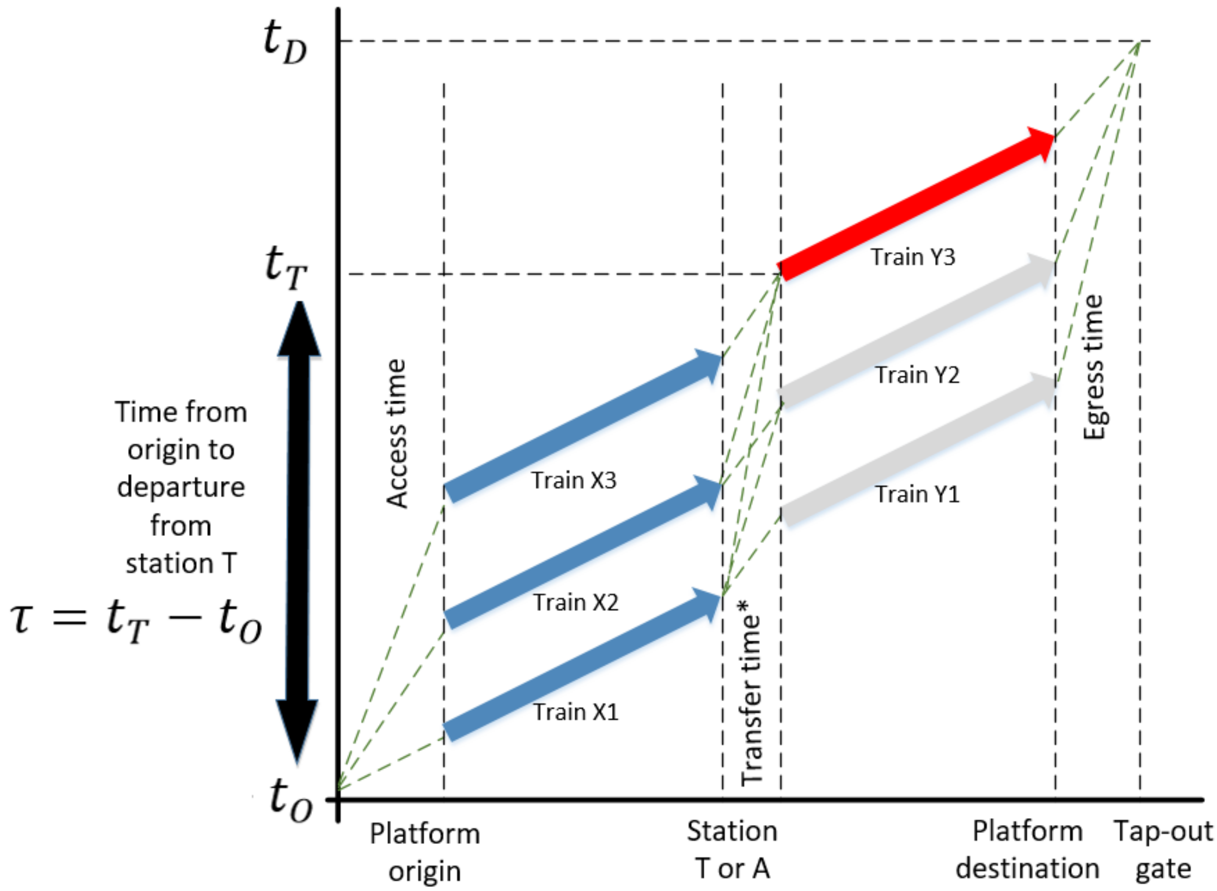

FIGURE 2 - Passenger to train assignment

Methodology for determining fraction of reverse routing passengers

The above mentioned methodology will be used analyse the journey times from tap-in to departure from the transfer station which all passengers have to depart from no matter if they are reverse routing or not. However, the analysis of journey times only indicates probable causes for reverse routing and a detailed model developed in Ma et al. (2019) is used to calculate the fractions of passengers who are reverse routing. The model is an extension to the passenger-to-train assignment model in Zhu, Koutsopoulos & Wilson (2017). The model builds on several submodels; first the denied boarding rates at each station is calculated and walking time distributions are defined using the methods in Zhu, Koutsopoulos & Wilson (2017). Then a multidimensional optimization model is used to determine the route choice parameters for passenger path choice. The route choice parameters included in the model are in-vehicle time, out of vehicle time (access, egress, transfer walking time and waiting time at uncongested periods), number of transfers and waiting time spent due to denied boarding.

CASE STUDY

The case study for this paper is the MTR metro system in Hong Kong. The system has almost 5 million daily passengers (MTR, 2019) and some sections experience severe congestion in peak hours. The specific case study is concerning two major stations in the central part of Hong Kong, station 1 and 2 on Figure 3 below. Passengers travelling from stations 27-30 on the blue line must transfer at either station 1 or 2 to get to stations 3-17 on the red line.

FIGURE 3 - Case study network

In a survey carried out in 2012 and analysed in Li (2014) approximately 30,000 passengers in the MTR system were asked about their route choice and the survey revealed that around 8% of the passengers in the evening peak period going from the blue line to the red line transferred at station 1, whereas all passengers outside the peak period transferred at station 2. This indicates that denied boarding and crowding leads to a different behavior for some of the passengers. Since 2012 the number of passengers in the system has increased with 14% until 2017 where another line also opened with terminal station at station 2 adding more congestion to this already crowded station (MTR, 2019). The transfer at station 2 is a cross-platform transfer whereas the passengers at station 1 have to walk up one level to transfer to northbound departures on the red line. The additional train travel time going to station 1 is around 3 minutes (1.5 minutes in each direction).

Data description and passenger-to-train assignment

To analyse the factors influencing reverse routing behavior data from three weekdays, 21st-23rd of March 2017 (Tuesday-Thursday), was used. A decision was made to limit the sample to adult passengers, as other passenger groups, such as pensioners, might have more heterogenous travel behavior. The passenger-to-train assignment model was used to assign passengers to a specific train on the red line and outliers were removed which had a journey time (τ) outside two standard deviations of the mean from a given origin and a specific train departure. Table 1 shows the number of passengers assigned with the probability of the most likely train on the red line with departure from station 2 between 17:30-19:30. In total 20,675 trips were included in the analysis and 70% of these trips could be assigned with a probability higher than 95%. Following the passenger-to-train assignment only the passengers with an assignment probability of more than 95% to the most likely train is included in the analysis. For these passengers the journey time τ is calculated based on the tap-in time at the origin station and departure from station 2 of the assigned train.

TABLE 1 – Statistics of probability of the most likely train

Probability of most likely train	Number of pax	Share of pax
<50 %	5	0.0%
50-75%	3,061	14.8%
75-90%	4,125	20.0%
90-95%	3,423	16.6%
95-99.9%	14,819	71.7%
100%	11,617	56.2%
Total	37,050	100%

Given the journey times τ it is possible to analyse several factors leading to different behaviors in terms of reverse routing using a multiple linear regression model. The dependent variable is the journey time τ, which is explained by the following function:

$$\tau \sim \beta_{base} + \boldsymbol{\beta_d d} + \boldsymbol{\beta_o o} + \boldsymbol{\beta_k k} + \boldsymbol{\beta_x x} \tag{5}$$

, where d are the destinations, o are the origins, k is a specific 30-minute timeinterval and x is the travel experience.

In previous studies on reverse routing in metro systems one of the clear tendencies have been that passengers travelling furthest are more likely to do reverse routing (e.g. Tiranchini et al. 2017, Othman et al., 2015 and Li et al., 2017). This hypothesis is tested by using the destination stations as explanatory variables for the journey time from tap-in until departure from station 2. The origins are included in the model, as passengers naturally have higher journey times for trips with further distance to the transfer stations and the variable for timeintervals is included to explain the extra travel time imposed from crowding in the peak hours.

The variables regarding passenger experience are included as Kim et al. (2014) showed that passengers with more experience choose a specific metro car to minimize the walking distance at the destination station. In the case of reverse routing a hypothesis is that passengers with more experience have lower journey times as they are able to observe the current conditions and choose whether to transfer at the normal transfer station (1) or do reverse routing using station 2. The passengers experience is given based on the number of times the passenger have travelled from the blue line to the red line from 17:00-20:00 in March 2017. By testing different specifications of this variable it was found that intervals of <5, 5-9, 10-19 and >=20 trips in a month was giving the best fit of the model.

RESULTS

This section presents the results of the analysis of journey times from the origin to departure from station 2. First the hypothesis of the journey times being dependent on the destination station and the passenger experience is explored visually followed by the multiple linear regression model explaining the journey times. Then the results of applying the model developed in Ma et al. (2019) is shown with the fractions of reverse routing passengers.

Histograms of journey times dependent on explanatory factors

A simple histogram of the mean journey times τ shown in Figure 4 reveal that the journey times from a given origin to the departure from station 2 is increasing the further the passenger has to go on the red line. The journey time for passengers going to station 3 is higher than for passengers going to station 4 and 5, but never higher than the time for passengers going to station 17.

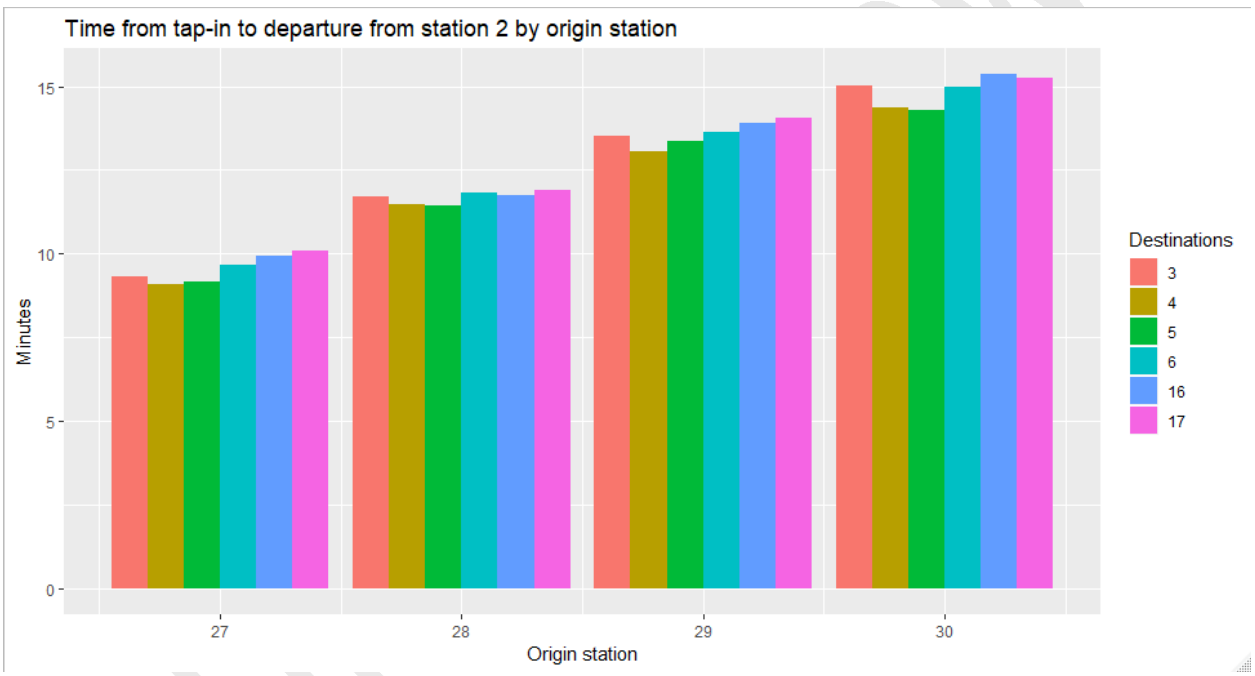

FIGURE 4 - Histogram of mean journey times from tap-in to departure from station 2 by origin and destination

When considering the other factor, passenger experience, Figure 4 clearly shows that passengers with more experience have lower journey times compared to less experienced passengers on this specific path. It is also clear that passengers with 20 or more trips during March 2017 on this path, i.e. mostly commuters, have a much lower travel time. Although these histograms does not provide any statistical test of significance of passenger experience, it seems that the factors influence the journey times.

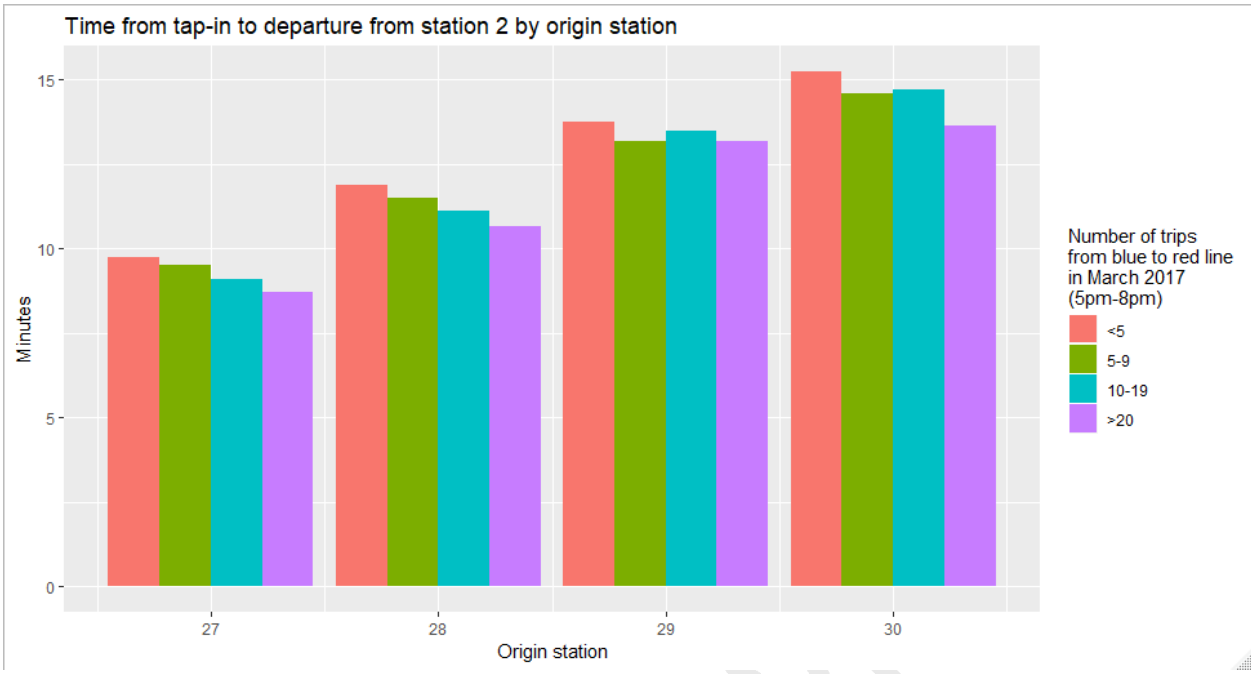

FIGURE 5 - Histogram of mean journey times from tap-in to departure from station 2 by traveler experience

Statistical model for journey times indicating potential reverse routing

Table 2 below shows the final model explaining the journey time τ from origin to departure from station 2. The intercept of 8.36 minutes explains the journey time from station 27 to destination station 3 (closest destination to transfer) for a passenger in the interval 17.30-18.00 and with less than five trips on this route in March 2017. The train travel times including dwell time is approximately 2 minutes between two consecutive stations, which is also reflected in the estimates for the origins. However, the estimate for station 30 is only around 1.25 minutes higher than for station 29, but this can be explained by a much shorter access walking distance from tap-in gates to platform at station 30 compared to the other stations.

The variables indicating the different time periods show significant differences between the four 30-minute intervals. As expected, the most congested time period is from 18.30-19.00, where passengers spend three minutes extra compared to the reference level from 17.30-18.00. The estimation of the differences between time intervals does not add information on whether reverse routing is more likely in a given time period, but helps correcting for the extra congestion in the system, so that the parameters for destinations and travel experience is not affected by the additional congestion.

Given that the journey time τ is explaining the time from tap-in to departure from station 2 there should intuitively not be any difference in the journey times for passengers going to different destinations. However, as it is seen in Table 2 most of the estimates for the destinations are significantly different from the reference level of destination 3. The estimate for station 5 is not significantly different from station 3 and the estimate for station 4 is only slightly lower for station 3 with a difference of 15 seconds. For passengers going to station 6 the journey time is approximately 20 seconds longer, for station 16 the journey times are around half a minute higher than station 3 and the journey times for passengers going to station 17 is almost one minute higher than station 3. This indicates, that passengers travelling further after transferring has a higher probability of doing reverse routing, possibly because of a higher preference of obtaining a seat, as also found in Tiranchini et al. (2017).

When investigating the parameters for travel experience there is a clear tendency, similar to the histogram, that passengers with more experience have lower journey times. This is in line with the findings in Kim et al. (2014), where passengers with more experience chose metro cars minimizing walking distance and thereby their journey time. A very experienced passenger on the route from the blue line to the red line saves more than one minute compared to unexperienced passengers. As only data from March 2017 was available for the analysis, it was not possible to check for whether passengers also travelled many times in other months and thereby could be determined as commuters, but more elaborate clusterings of different passenger groups could give more insights to which types of passengers minimize their journey times the most. Given the lower journey times for experienced passengers, this also indicates that it is mostly unexperienced passengers who do reverse routing, as transferring at station 1 is in almost all cases slower than transferring at station 1 under normal peak hour conditions.

TABLE 2 - Results of Multiple Linear Regression Model for Journey Time τ
Sig. levels: *0.05, **0.01, ***0.001

Parameter	Estimate	T-value
Base (intercept)	8.36	105.80***
Origin Station 27	-	Ref. Level
Origin Station 28	2.10	35.02***
Origin Station 29	4.05	31.68***
Origin Station 30	5.31	57.71***
Destination Station 3	-	Ref. Level
Destination Station 4	-0.26	-3.20**
Destination Station 5	-0.10	0.28
Destination Station 6	0.29	3.55***
Destination Station 16	0.46	4.81***
Destination Station 17	0.79	8.04***
Time interval 17.30-18.00	-	Ref. Level
Time interval 18.00-18.30	1.26	15.95***
Time interval 18.30-19.00	3.00	38.72***
Time interval 19.00-19.30	0.56	6.90***
Less than 5 trips in month	-	Ref. level
5-9 trips in month	-0.47	-6.16***
10-19 trips in month	-0.73	-10.01***
More than 20 trips in month	-1.16	-9.70***
Number of observations		**26,436**
Adj. R-Squared		**0.19**

Determination of fraction of reverse routing passengers

Using the model developed in Ma et al. (2019) on the case study reveals that passengers to a large extent do reverse routing in the peak hour. Figure 6 shows the reverse routing fractions for passengers from station 28, which is the largest of the considered origins, to stations 3-6 on March 16th[1]. The results of the model show that in peak shoulder periods less than 10% of passengers do reverse routing transferring at station 1 and there seems to be no clear dependence on the fractions for the different destinations. In the period of 18:00-18:30 the fraction of reverse routing passengers is between 35%-51% and in the peak of the evening peak period between 18:30-19:00 the fraction is even higher between 43%-54%. The results for these periods also indicate that the fractions is positively correlated with the distance to the destination. These results show that reverse routing is a behavior extensively utilized by passengers in the MTR metro network. From an operational perspective the large number of reverse routing passengers who use station 1 might be a benefit for the operations, as dwell time on the terminal station 1 is not as critical as on station 2. The high number of reverse routing passengers therefore lowers the dwell time at station 1 as less passenger have to board at this station to fill up the train. Further analysis should identify the magnitude of this possible gain for the operations.

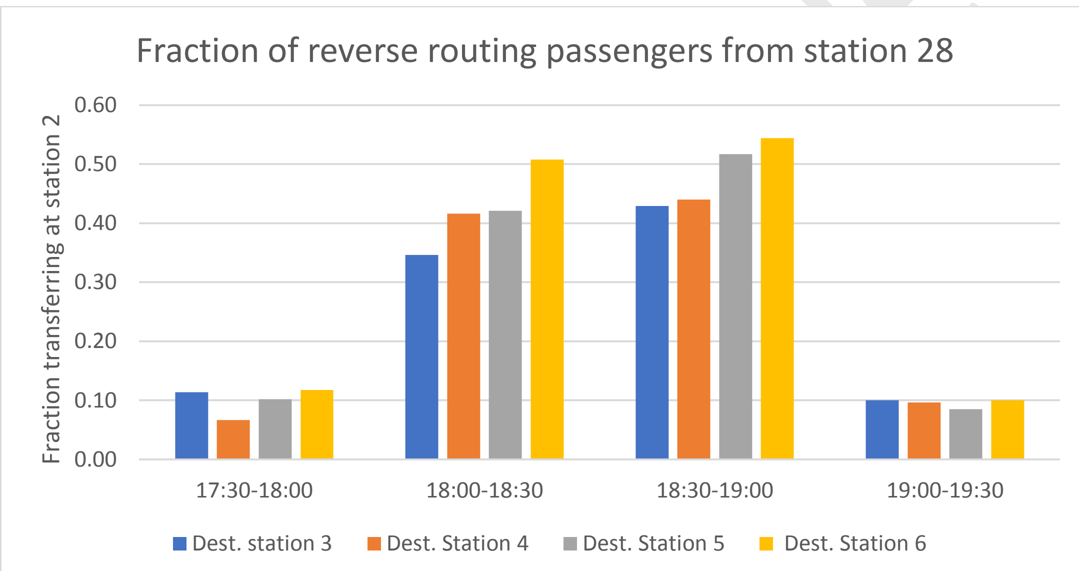

FIGURE 6 – Reverse routing fractions for passengers from station 28 on March 16th 2017

CONCLUSION

This paper deals with the special path choice named reverse routing, where passengers choose to stay longer on a train and transfer at a station further down the line before travelling backwards and passing the station where the transfer would under normal conditions take place. By applying a passenger-to-train assignment model to the second leg of the trip, it is possible to calculate the time from tap-in to departure from the transfer station, which would be used under normal conditions. These times are analysed and a model combining multiple data sources and optimizing the route choice parameters of passengers is used for calculating the fractions of passengers doing reverse routing.

[1] Due to time limitations for obtaining walking speed distributions and other distributions needed for the analysis March 16th was used for the analysis and stations 16 and 17 was not included in the analysis.

The results show that passengers who travel further on the second leg of the trip spends significantly more time from tap-in to departure from the normal transfer station when correcting for the time from the origin and the longer travel times during different time intervals in the evening peak period. The extra time spent is to a large extent the effect of reverse routing passengers. The results show that less than 10% of passengers in peak shoulder periods are reverse routing, while these fractions are up to 35%-54% in the peak hour from 18:00-19:00.

The analysis also shows that passengers who travel more during a month have significantly lower travel times compared to passengers who travel less frequently. This indicates that passengers to a high degree make decisions based on their previous knowledge and that passengers with high travel experience can assess the current denied boarding conditions and choose whether to do reverse routing or not. Future work should include clustering of passengers to determine the probability of reverse routing for different passenger segments. The revelation of the high number of reverse routing passengers should also be used to assess the impact of different operational strategies, which might lower the fraction of up to 50% of passengers doing reverse routing in the evening peak hour.

ACKNOWLEDGMENTS
MTR is thanked for their support and for providing automatic fare collection and automatic vehicle location data for this research project.

AUTHOR CONTRIBUTIONS
The authors confirm contribution to the paper as follows: all authors contributed to the study conception and design. K. Tuncel, M. Eltved and Z. Ma prepared the analysis in the paper and all authors contributed to the interpretation of the results. M. Eltved prepared the draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

Batarce, M., Muñoz, J.C., De, J., Ortúzar, D. Valuing crowding in public transport: Implications for cost- benefit analysis. Transportation Research Part A, 2016. 91: 358–378. https://doi.org/10.1016/j.tra.2016.06.025

Chakirov, A., Erath, A. Use of Public Transport Smart Card Fare Payment Data for Travel Behaviour Analysis in Singapore. 16th International Conference of Hong Kong Society for Transportation Studies, Hong Kong, 2011. https://doi.org/10.3929/ethz-a-006742061

Haywood, L., Koning, M., Monchambert, G. Crowding in public transport: Who cares and why? Transportation Research Part A, 2017. 100: 215–227. https://doi.org/10.1016/j.tra.2017.04.022

Kim, K.M., Hong, S.P., Ko, S.J., Kim, D. Does crowding affect the path choice of metro passengers? Transportation Research Part A, 2015. 77: 292–304. https://doi.org/10.1016/j.tra.2015.04.023

Kim, H., Kwon, S., Wu, S.K., Sohn, K. Why do passengers choose a specific car of a metro train during the morning peak hours? Transportation Research Part A, 2014. 61: 249–258. https://doi.org/10.1016/j.tra.2014.02.015

Li, W. Route and Transfer Station Choice Modeling in the MTR System. Report, Boston, 2014

Li, Z., Hensher, D. Crowding in Public Transport: A Review of Objective and Subjective Measures. Journal of Public Transportation, 2013. 16: 107–134. https://doi.org/10.5038/2375-0901.16.2.6

Li, Y.-N., Zhou, F., Hong, L., Zhu, W. Empirical Analysis of Failing to Board and Traveling Backward in an Overcrowded Urban Rail Transit System. 17th COTA International Conference of Transportation Professionals, 2017, pp. 1979–1989.

Ma, Z, Koutsopoulos, H.N, Zhu, Y, Chen, Y. Estimating Passengers' Path Choice Using Smart Card Data in Urban Rail Systems. TransitData 2019, Paris, France, 2019.

MTR. Annual Report 2018, 2019. https://www.mtr.com.hk/en/corporate/investor/2018frpt.html

Othman, N. Bin, Legara, E.F., Selvam, V., Monterola, C. A data-driven agent-based model of congestion and scaling dynamics of rapid transit systems. Journal of Computational Science, 2015. 10: 338–350. https://doi.org/10.1016/j.jocs.2015.03.006

Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H.M. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. Transportation Research Part A, 2014. 66: 185–195. https://doi.org/10.1016/j.tra.2014.05.010

Tirachini, A., Sun, L., Erath, A., Chakirov, A. Valuation of sitting and standing in metro trains using revealed preferences. Transport Policy, 2016. 47: 94–104. https://doi.org/10.1016/j.tranpol.2015.12.004

Xu, R., Li, Y., Zhu, W., Li, S. Empirical Analysis of Traveling Backwards and Passenger Flows Reassignment on a Metro Network with Automatic Fare Collection (AFC) Data and Train Diagram. Transportation Research Record, 2018. 2672: 1–13. https://doi.org/10.1177/0361198118781395

Zhu, Y., Koutsopoulos, H.N., Wilson, N.H.M. A probabilistic Passenger-to-Train Assignment Model based on automated data. Transportation Research Part B, 2017. 104: 1–21. https://doi.org/10.1016/j.trb.2017.04.012