Optimizing passengers' boarding and alighting operations in urban mass transit

Laura Knappik
RWTH Aachen University, Chair for Data and Business Analytics
e-mail: knappik@analytics.rwth-aachen.de

Dr. Lorena S. Reyes Rubiano
RWTH Aachen University, Chair for Data and Business Analytics
e-mail: reyes@analytics.rwth-aachen.de

Prof. Dr. Sven Müller
RWTH Aachen University, Chair for Data and Business Analytics
e-mail: mueller@analytics.rwth-aachen.de

Abstract: During boarding and alighting in urban mass transit, some train doors become overcrowded while others are underutilized. This imbalance increases dwell time as trains wait for the last passengers to board or alight. We propose a nonconvex MINLP to minimize boarding and alighting times—a function of the number of boarding and alighting passengers—by optimally allocating passengers to doors. To capture the process of passengers' choosing doors, we developed a choice model with financial incentives to guide passengers to specific doors. Optimal discounts are determined within the choice model and integrated into the optimization framework to adjust passengers per door.

1 Introduction

Urban mass transit systems often struggle with high operational costs and tight schedules. A major source of inefficiency is the dwell time of vehicles at stations, which is significantly influenced by variability in boarding and alighting times (Kuipers et al., 2021). Passengers tend to favor boarding through particular doors, such as those near the station entrance or closer to their destination's exit. This preference often leads to overcrowding at certain doors while others remain underused (Oliveira et al., 2019), creating an imbalance that increases dwell time as the vehicle waits for the last passengers to board or alight. This research aims to reduce dwell times during peak hours by optimizing the allocation of passengers to doors and trips, thereby enhancing overall system efficiency (Figure 1).

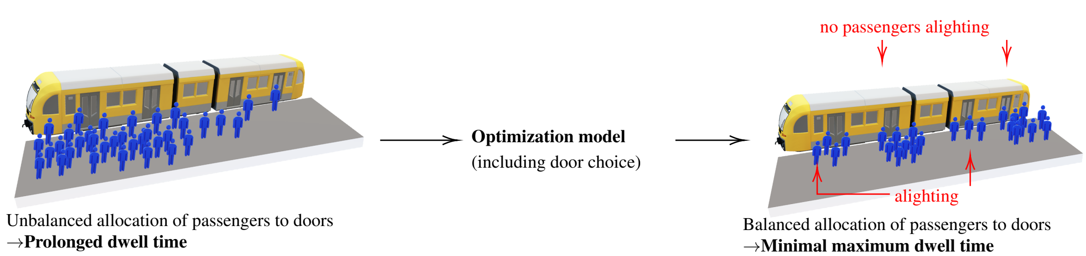


Figure 1: Overview of research objective.

To achieve this, our methodology integrates three main components:

- *Optimization of on- and off-boarding processes*: This involves developing an optimization model that minimizes dwell time by strategically allocating passengers to doors and trips.
- *Door-trip choice prediction*: This component utilizes discrete choice models to predict passengers' selection of doors and trips. The goal is to better understand and influence passenger behavior, promoting a more balanced distribution across doors and trips—e.g., encouraging passengers to align with optimization strategies.
- *Dwell time prediction*: We use a combination of pedestrian simulation models and empirical data to predict dwell times based on the number of boarding and alighting passengers.

Our integrated framework for simultaneously combining these components focuses on the operational and behavioral aspects of passenger flow management in urban mass transit systems (see Figure 2). This approach not only aims to reduce dwell times but also enhances the overall efficiency and reliability of transit operations. The integrated approach has the advantage of identifying the overall optimal solution by considering the interaction between door choice and the optimization model. In contrast, the sequential approach tends to result in suboptimal solutions, as it neglects the interdependence between these two components.

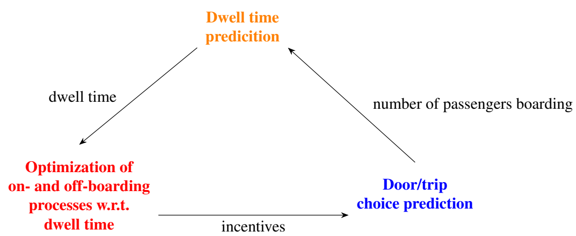


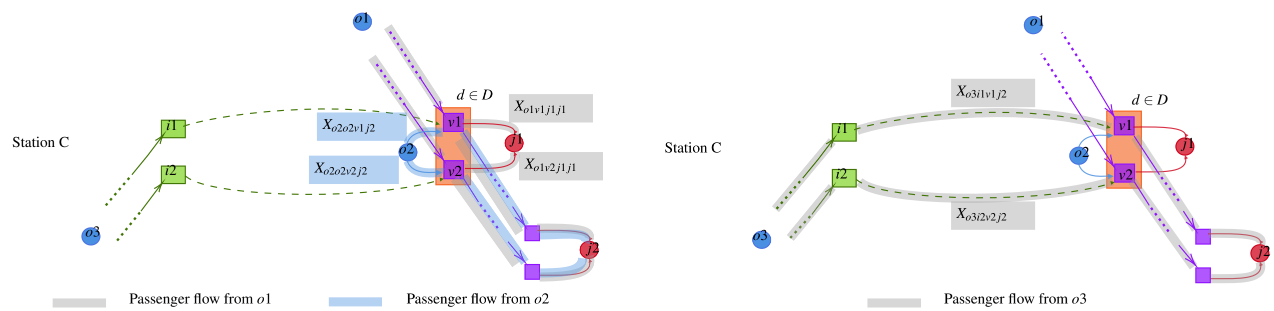
Figure 2: Overview of research objective.

2 PROBLEM DESCRIPTION

The objective is to minimize the maximum total dwell time by optimally allocating passengers to doors and trips, achieved by subtly influencing their door choices. Since boarding decisions affect dwell times during later alighting, the model considers the entire public transport network to achieve an optimal flow of passengers. The public transportation network comprises stations served by multiple lines, with each line connecting stations to others through a series of trips scheduled according to a predefined line plan and timetable for a specific operational day. The optimization is tailored to this particular day, ensuring a comprehensive and context-specific approach. Each vehicle operating within the network has a fixed number of doors for **boarding** and **alighting**. Additionally, **transfers** are possible at stations where different lines intersect, allowing passengers to reach their destination when direct trips are unavailable.

The problem is modeled as a directed network $G = (I, A, q)$, where the nodes correspond to points in time and space where passengers enter, exit, or travel along the public transport system using the lines. We define a set I, which includes origin nodes $O \subset I$ (indexed by o), destination nodes $J \subset I$ (indexed by j), and door-trip nodes $E \subset I$ (unique combinations of train doors and trips for a given trip and station, indexed by i, k, v). The operational day is divided into multiple time intervals $\{t, t+1, t+2, \ldots\}$ to monitor the times at which each line arrives at each station and thus determine the time associated with each node. To track dwell time at each station for each trip, we define a set D, representing all door-trip nodes belonging to the same station and trip. The set $E_d \subset E$ represents the subset of door nodes within the same door-trip set D, e.g., all doors belonging to the same train.

Set A represents the arcs, which indicate feasible connections between nodes regarding time and space. An arc $(i, v) \in A$ indicates that passengers can feasibly move (board or alight) from door-trip i to door-trip v. To ensure feasible flows for each origin $o \in O$, we define the set $\tilde{A}(oiv)$. We assume that passengers board and alight through the same door, as they are unlikely to move to a different door within a crowded train. Transfers between $i \in E$ and $v \in E$, are represented by the subset $F \subset A$. The parameter q represents the weights of the arcs, representing passenger flow capacities. Set $B \subset A$ represents the arcs where boarding and transfer occurs. Specifically, an arc i, v belongs to B if $(i, v) \in A$ and either $i \in O$ or $i \in E$ with $(i, v) \in F$. The total demand is defined by a time period-specific origin-destination matrix, with a predetermined path from each origin to its destination. We assume passengers choose the shortest path in terms of travel time from their origin to their destination (Müller et al., 2022). To manage capacity constraints while meeting demand, we allow passengers arriving for a trip at time period t can either board a vehicle on the same trip or wait for a subsequent trip of the same line (time periods are defined by dividing the operational day into discrete intervals, each referred to as a period. Period t represents the specific time window in which passengers arrive at a station and may board a scheduled trip departing during t or wait for a subsequent trip in a later period $(t+1, t+2, \ldots)$. The set of all periods, denoted by T, spans the entire scheduling horizon). The non-negative variable X_{oivj} indicates the number of passengers starting at the station and trip of node o and traveling along arc (i, v) to reach destination j. The four indices are essential for capturing the complexity of the system, as they enable us to account for passengers transferring between trains at various stations and times. The variable X_{oivj} tracks passenger flow in one direction (boarding or alighting) and also enables tracking counterflows that occur simultaneously through the same door-trip, where passengers board and alight at the same time. For simplification and a better understanding of passenger flow, Figure 3 illustrates the counterflows at node $v \in E_d$. The highlighted arcs in gray illustrate the variable X_{oivj}, demonstrating that passenger flows have a specific origin and destination determined by demand.



(a) Passenger flow from o_1 and o_2 boarding the train through doors $v1, v2 \in E_d$ to reach their destination.

(b) Passenger flow from o_3 transferring from $i1, i2$ boarding the train through doors $v1, v2 \in E_d$ to reach destination $j2$.

Figure 3: Simplified example of the passenger flow considering part of the network variable X_{oivj}.

This level of detail is crucial for the choice model, as it ensures the model can differentiate between passengers who remain on the train, alight, or transfer and correctly apply the probabilities of passengers choosing a specific boarding door at the transfer station. This network model allows us to represent passenger flows over time and space.

Figure 4 shows a simplified example of a network, with Line I and Line II operating in only one direction. At station C, passengers can transfer between lines without a direct connection from their origin to their destination. For instance, when traveling from station B to station A, passengers first board Line II and then transfer at station C to take Line I. Passengers who need to transfer can either take the next available trip t on Line I or wait for the following trip $t+1$. To minimize dwell times effectively, it is essential to define a function that captures the factors contributing to dwell time at each station and for each trip. This general dwell time function serves as the foundation for the optimization process. For a train at each station, the maximum dwell time τ_d, associated with each set of door-trips v at the station and time period combination $d \in D$, with $v \in E_d$ is modeled as a function of the number of passengers boarding and alighting through door v. Specifically, the relationship is expressed as:

$$P_v^1 = \sum_{o \in O} \sum_{\substack{i \in B_{iv} \\ \wedge \tilde{A}_{oiv}}} \sum_{j \in J} X_{oivj} \quad \forall v \in E_d \tag{1a}$$

$$P_v^2 = \sum_{o \in O} \sum_{\substack{j \in J \\ \tilde{A}_{ovj}}} X_{ovjj} \quad \forall v \in E_d \tag{1b}$$

$$\tau_d \geq \alpha_0 + \alpha_1 \cdot P_v^1 + \alpha_2 \cdot P_v^2 + \alpha_3 \cdot P_v^1 \cdot P_v^2 \quad \forall d \in D, v \in E_d. \tag{1}$$

The dwell time function, derived from Puong (2000), Lam et al. (1998), reflects pedestrian behavior and integrates insights from previous studies on dwell time modeling. Thereby P_v^1 is the number of boarding passengers (including transfer passengers) and P_v^2 the number of alighting passengers at trip-door node v. The α values represent the corresponding coefficients.

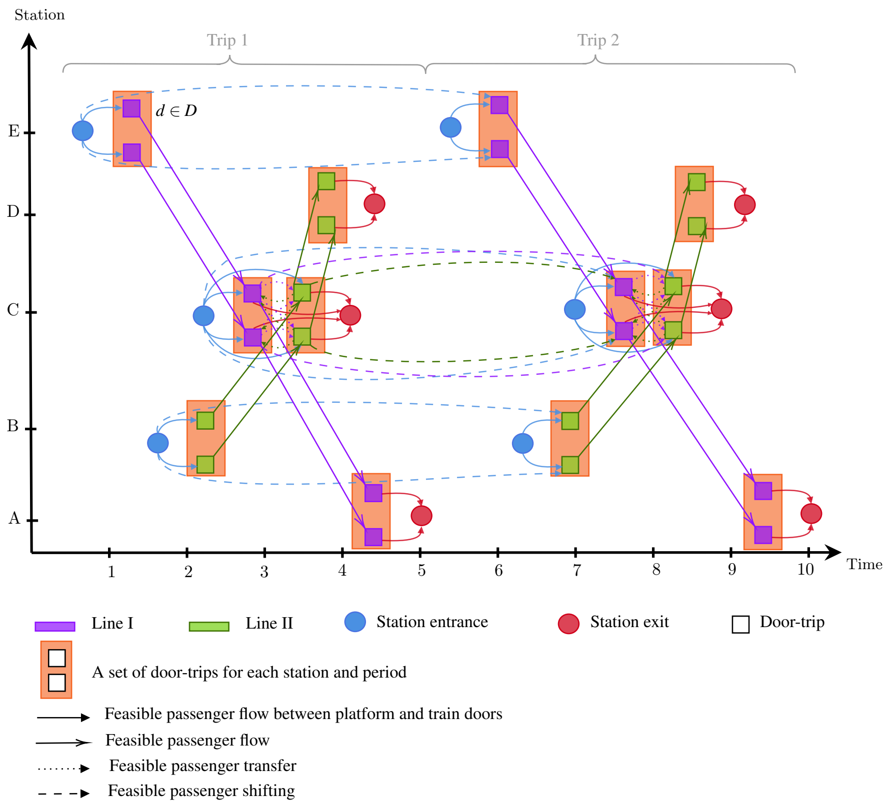


Figure 4: Public transport network visualization, showing two trips and two lines.

We proved that the counterflows of boarding and alighting passengers at the doors result in a nonconvex optimization problem. Nonconvexity presents a significant challenge, as finding the optimal values for the decision variables X_{oivj} to minimize dwell time involves solving a nonconvex optimization problem. Furthermore, these values depend on passenger behavior, particularly their choices regarding doors and trips, directly influencing dwell time. So far, this behavioral aspect has not been incorporated into the analysis.

2.1 Door choices

In this section, we explain how passenger preferences for waiting areas and boarding doors are incorporated into the previously mentioned optimization model. This approach acknowledges that passengers often have specific preferences—such as doors located near the station entrance or close to the exit at their destination (Küpper and Seyfried, 2023). To capture these behavioral factors, we integrate a choice model within the optimization framework. Inspired by Yu et al. (2021), we influence passengers' door and trip choices to encourage compliance with the optimal allocation of passengers. We use economic incentives, such as discounts, to motivate the passengers to board through specific (less crowded) doors. This implies that passengers originating from o and traveling to j receive a discount on the regular fare when they use door-trip v. Thus, the second problem we aim to solve is determining the optimal discounts to ensure that the values of X_{oivj} lead to a minimization of the maximum dwell time.

To predict passengers' choices, we assume that both the vehicle design and station platform layout are known. The platform is divided into the same number of waiting areas as the number of train doors. For example, if a train has two doors, the platform is divided into two corresponding areas. We assume that passengers choose the door for boarding that is closest to their waiting positions. This allows us to track the expected number of passengers who wait at the station and choose to board the train through a specific door-trip (see Figure 5). The choice model predicts the probabilities of a passenger selecting each feasible door.

Passengers select one door-trip for boarding from all feasible options. The remaining doors for the trip in time period t and any subsequent trips up to $|T|$ define the set of alternatives AT, $AT \subset E$ (see Figure 5). The choices are made based on the principle of utility maximization using a Multinomial Logit (MNL) framework. The probability of a passenger

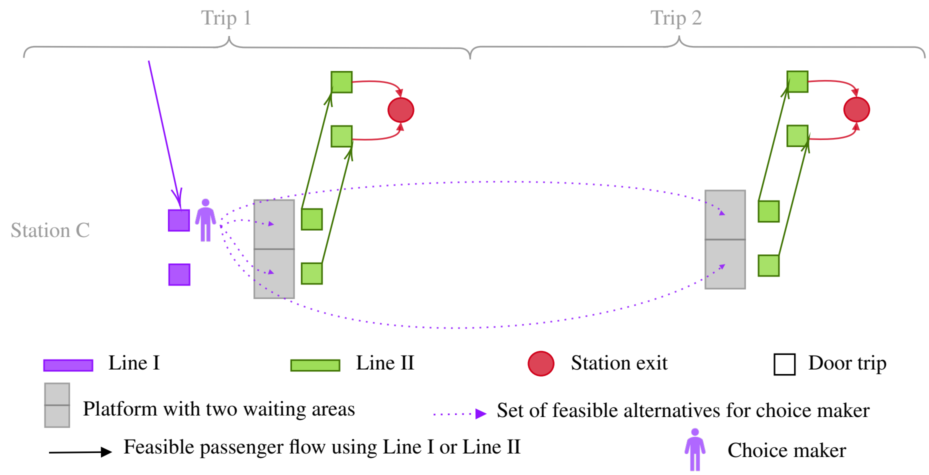


Figure 5: Illustration of passengers selecting a specific door-trip for boarding. This is an example of a passenger choosing between four feasible door-trips for boarding when changing from Line I to Line II at trip $t = 1$.

choosing a door increases with its utility. The probability of a passenger originating from o and traveling to j, boarding or transferring from node i to door-trip v, is denoted by Y_{oivj}:

$$Y_{oivj} = \frac{\sum_{l \in L} \exp\left(U_{oivjl}\right) \cdot L_{oivjl}}{\sum_{v' \in AT} \sum_{l' \in L} \exp\left(U_{oiv'jl'}\right) \cdot L_{oiv'jl'}} \qquad \forall o \in O, j \in J, i \in B(i,v), v \in E, \tilde{A}(o,i,v) \tag{2}$$

where U_{oivjl} represents the exponential utility of door-trip v, defined as $U_{oivjl} = \exp(\tilde{u}_{oivjl})$, with $\tilde{u}_{oivjl}$ denoting the systematic utility. It depends on attributes such as the distance to the entrance and the assigned price, including a potential discount. The binary decision variable L_{oivjl} determines whether a specific pricing level l is applied for the use of door-trip v. For each combination of o, i, v, j, exactly one pricing level can be selected, ensuring that probabilities for non-selected discounts are zero. The cumulative discount offered along the path from o to j is limited by an upper bound. A higher discount for a specific door-trip increases its utility, raising the probability Y_{oivj} of passengers selecting that door-trip for boarding.

This probability is used to define the number of boarding and alighting passengers for each node in the optimization model, the flow variable X_{oivj}. For the boarding passengers, the probability is multiplied by the origin-destination demand OD_{oj}, representing the total number of passengers traveling from o to j, to calculate the flow through door-trip v:

$$X_{oovj} = Y_{oovj} \cdot \text{OD}_{oj} \qquad \forall o \in O, j \in J, v \in E, A(o,v) \tag{3}$$

Similarly, the origin-destination matrix can determine the number of transfer passengers. Transfer passengers, meaning those changing trains to reach their destination, are similarly assigned their utility, price discount, and resulting probability Y_{oivjl}. This enables the model to assign different discount levels for boarding and transferring passengers while ensuring the cumulative discounts along each path from o to j remain within predefined limits. Since the number of passengers traveling from o to j is known, and the feasible paths are defined, the transfer passengers can be derived based on the number of passengers arriving at node i. For instance, if 10 passengers originating from o and traveling to j arrive at trip-door i, and $A(i,v)$ is a feasible transfer flow, it follows that they must transfer from i to door v or one of its alternatives. The resulting flow in the optimization model is:

$$X_{oivj} = \sum_{k \in I \wedge \tilde{A}(o,k,i)} Y_{oivj} \cdot X_{okij} \qquad \forall o \in O, j \in J, i \in E, v \in E, F(i,v), \tilde{A}(o,i,v) \tag{4}$$

The resulting flows are decision variables in the optimization model, directly impacting the station dwell time. Overall, the optimization process integrates the influence of pricing levels l on the choice probabilities Y_{oivj}, and the resulting flows X_{oivj} through door-trips. The flow is used to determine the maximum dwell time.

3 FUTURE WORK

As this is work in progress, there are multiple aspects to be addressed to enhance the model and its performance.

The nonlinearity and nonconvexity of the model, particularly in the multiplication of the flow variable by the probability variable Y_{oivj} in Equation 4, presents challenges in finding optimal solutions, even for small instances. While good solutions with a gap of 0.1% can be obtained within seconds, achieving an optimal solution within a reasonable runtime (e.g., 48 hours) remains infeasible, even for small instances. To overcome the nonlinearity, a linearization technique will be applied to approximate the multiplication of a continuous variable X_{oivj} and a positive variable Y_{oivj}, where $Y_{oivj} \in [0,1]$ (Asghari et al., 2022). Similar approaches will be used to approximate the multiplication of the two continuous positive flow variables X_{oivj} in the dwell time function (Equation 1). Future work will, furthermore, focus on proving the optimality gap introduced by this approximation, quantifying the trade-off between solution quality and computational efficiency. Additionally, the probability function, which is also nonlinear due to the presence of exponential terms in the utility function and the division by the sum of these exponential terms for all alternatives, contributes to the overall complexity of the model. It will be linearized using methods similar to existing choice modeling literature (Haase, 2009, Ljubić and Moreno, 2018).

The model currently relies on outdated and sparse dwell time functions from the literature. To address this, we plan to validate these functions using observational data from Frankfurt Main Station provided by Jülich Institute and simulations based on the social force model. However, the nonlinear and nonconvex nature of the function remains essential to accurately capture the counterflows at the doors.

Another step to ensure the model's scalability and applicability is to test it on multiple instances of real-world size. For this purpose, we have developed an instance generator capable of efficiently creating instances with varying numbers of stations, lines, trips, and passengers. Furthermore, the public transport network's structure will be allowed to change dynamically over time to account for variations in operational patterns, such as different lines operating during peak and off-peak hours. This dynamic adjustment will better reflect real-world transportation patterns throughout the day.

Finally, we plan to include more attributes to the utility function that influence passengers' decisions regarding which area of the platform they choose to wait in before boarding the train. These additional factors will help refine the choice model and better capture passenger preferences.

References

Asghari, M., Fathollahi-Fard, A. M., Mirzapour Al-E-Hashem, S. M. J. and Dulebenets, M. A. (2022), 'Transformation and linearization techniques in optimization: A state-of-the-art survey', *Mathematics* **10**(2), 283. https://doi.org/10.3390/math10020283.

Haase, K. (2009), Discrete location planning, Working Paper ITLS-WP-09-07, Institute of Transport and Logistics Studies, The University of Sydney, Sydney, Australia.

Kuipers, R. A., Palmqvist, C.-W., Olsson, N. O. and Hiselius, L. W. (2021), 'The passenger's influence on dwell times at station platforms: a literature review', *Transport Reviews* **41**(6), 721–741. https://doi.org/10.1080/01441647.2021.1887960.

Küpper, M. and Seyfried, A. (2023), 'Identification of social groups and waiting pedestrians at railway platforms using trajectory data', *PLOS ONE* **18**(3), e0282526. https://doi.org/10.1371/journal.pone.0282526.

Lam, W. H., Cheung, C. Y. and Poon, Y. F. (1998), 'A study of train dwelling time at the hong kong mass transit railway system', *Journal of Advanced Transportation* **32**(3), 285–295. https://doi.org/10.1002/atr.5670320303.

Ljubić, I. and Moreno, E. (2018), 'Outer approximation and submodular cuts for maximum capture facility location problems with random utilities', *European Journal of Operational Research* **266**(1), 46–56. https://doi.org/10.1016/j.ejor.2017.09.026.

Müller, S., Haase, K. and Reyes-Rubiano, L. (2022), 'Revenue maximizing tariff zone planning for public transportation', SSRN Preprint. https://doi.org/10.2139/ssrn.4281537.

Oliveira, L. C., Fox, C., Birrell, S. and Cain, R. (2019), 'Analysing passengers' behaviours when boarding trains to improve rail infrastructure and technology', *Robotics and Computer-Integrated Manufacturing* **57**, 282–291. https://doi.org/10.1016/j.rcim.2018.12.008.

Puong, A. (2000), 'Dwell time model and analysis for the mbta red line', MIT OpenCourseWare Project.

Yu, J., Kuai, C., Ji, Y. and Gao, L. (2021), 'Can monetary incentive regulate boarding passenger distribution on the metro station platform?', *Journal of Advanced Transportation* **2021**, 1–12. https://doi.org/10.1155/2021/8373703.