# Data-Driven Public Transport Schedule Refinement: A Case Study of the Tram Network in Budapest

Balázs Varga[1a], Tamás Ormándi[a], Domokos Esztergár-Kiss[b]

[a] *Department of Control for Transportation and Vehicle Systems, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, Budapest, H-1111 Hungary*

[b] *Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, Budapest, H-1111 Hungary*

**Abstract**

In most cities, the urban rail network forms the backbone of the public transport system. Since this infrastructure is already in place, there is limited room for major structural changes. Optimizing timetables offers a cost-effective way to improve service quality.This paper presents a data-driven method for refining tram schedules using trip-planning queries as a proxy for passenger demand in Budapest. The proposed approach integrates demand modeling with timetable optimization by iteratively planning user trips with OpenTripPlanner and adjusting the General Transit Feed Specification-based timetable. The optimization relies on a Genetic Algorithm, offsetting the departure times of trams. The results show that the optimization can refine the schedules and reduce travel time. Although the relative reduction in total travel time is modest, small timetable offsets can measurably influence passenger routing outcomes. The magnitude of the intervention is small, as well. When applied at the scale of Budapest's public transport system, these modest relative changes correspond to tangible aggregate annual opportunity-cost savings. The findings underline the necessity of analyzing schedules in a network-wide context rather than optimizing isolated lines. Beyond proving the feasibility of the optimization, the methodology is both generalizable to other urban contexts and directly actionable without significant investment.

*Keywords:* Public transport, Demand planning, Timetable optimization, Genetic algorithm

[1] Corresponding author

## 1. Introduction

Efficient public transport is essential for the sustainability of urbanization. In dense urban environments, expanding the road network to accommodate more car traffic is costly and space-consuming, while increasing the capacity of the existing public transport systems is more practical. Tram and metro lines are the backbone of the public transport systems in several cities. In most cases, the urban rail network has already been developed in the last century which allows limited opportunities for rerouting or new constructions (?). As a result, cost-effective improvements are largely confined to schedule optimization or increases in speed and frequency. However, the latter options can be expensive and require infrastructure upgrades, regulatory adjustments, or additional rolling stock with significant upfront and operating costs. Furthermore, commuters are often resistant to changing their established travel habits (?), and real-time demand-responsive adjustments are generally infeasible for large-scale transit systems (?). For these reasons, the current study assumes that the public transport system is provided as it is and merely minor improvements are possible through schedule optimization.

Public-transport interventions can be distinguished by planning scope. Strategic decisions alter network topology or line structure; tactical decisions determine frequencies and timetables; and operational control reacts to realized disturbances. Timetable coordination is therefore a tactical problem when routes, stop patterns, and service frequencies are fixed (??). In the present study, the Buda tram network is treated as given: the route set, stop sequence, scheduled running times, and headways are not decision variables, and only the relative departure phase of each modeled tram line is adjusted. Within this tactical scope, fixed-headway timetable synchronization and line-offset optimization are well established. Earlier methods created coordinated departure times while retaining required service frequencies (?). ? shifted existing timetables with a genetic algorithm and systemwide transfer data, while ? formulated fixed-headway synchronization as a network optimization problem for passenger transfer waiting time. Models that account for travel-time uncertainty further show that gains in planned coordination should not automatically be interpreted as operational robustness (?). Several studies also account for passenger response to timetable changes. ? varied line offsets, supplied each revised timetable to a transit assignment model, and iterated the two calculations. **?** simultaneously optimized bus timetables and passenger path choices. These precedents establish that line offsets and timetable-passenger feedback are not new in themselves; they also show why evaluating a candidate timetable with fixed passenger paths can miss route changes induced by the schedule. A separate line of literature deals with passenger-demand estimation. Automated fare-collection and smart-card data provide large samples of realized transit use and can support trip chaining, transfer detection, and network-level origin–destination inference (**???**). However, depending on the fare system, destinations and transfer chains may require inference, while access and egress stages may not be observed directly. Trip-planner logs represent a different data source: they record prospective information requests rather than completed journeys. Route-planning query logs can reveal aggregate travel-flow patterns (**?**), and trip-planner logs have also been shown to support transit service planning (**?**). Trip-planner requests can improve short-term ridership prediction when validated against fare-collection observations (**?**). Smartphone-application logs have also been used to infer transit origin-destination matrices (**?**). Together, these studies demonstrate that trip-planner logs contain meaningful information for demand analysis and service planning. This is particularly relevant to the present study because journey planning is most useful for unfamiliar, non-routine, or uncertain trips, when travelers actively compare routes and transfers (**???**). Routine commuters who already know their route may instead use next-arrival information, which represents a different type of data. Consequently, trip-planning queries provide an intention-based demand proxy that is especially informative for journeys in which timetable-dependent route and transfer choices are actively considered. This interpretation does not require treating every query as a realized trip or assuming that app users are representative of all passengers.

Demand estimation must also be distinguished from multimodal journey evaluation. Passenger-oriented timetable analysis considers the complete journey, including access, waiting, in-vehicle time, transfers, and egress (**?**). Schedule-based assignment and path-choice models can represent how exact departure times change the feasible and attractive itineraries in mixed public-transport networks (**???**). In the present study, OpenTripPlanner is used as a deterministic, schedule-based itinerary evaluator: for every candidate tram timetable, the full planned itinerary is recomputed for each query. This captures schedule-induced changes in walking, waiting, transit, and transfer components, but it does not model capacity constraints, crowding, or probabilistic route choice.

The studies reviewed above use trip-planner or smartphone-app logs to estimate or predict demand, while the timetable-offset studies use OD matrices, passenger flows, or transfer counts. This study instead uses query-level intended trips to evaluate small, network-wide timetable offsets in the tram network of Budapest. The line-wide offset is selected as a deliberately constrained decision variable. Applying one phase shift to all modeled trips of a line preserves its headway, frequency, stop sequence, and scheduled interstop running times while limiting the dimensionality of the optimization. Each phase is selected from five candidate positions within one headway and is expressed in integer minutes; the offset should thus be interpreted as a timetable phase within the headway rather than as a delay accumulated from trip to trip. Minute-scale temporal resolution can materially affect schedule-based assignment results (**?**), while the discrete grid also avoids unsupported second-level precision and reduces the number of computationally expensive OTP evaluations. This resolution is a modeling and implementation constraint, not proof of compatibility with traffic-signal programs, vehicle blocks, crew duties, or disruption robustness, all of which require operator validation before implementation. Accordingly, this study asks: given trip-planner queries as a proxy for intended demand, can small phase changes to existing tram timetables reduce aggregate planned journey time without changing routes or service frequencies? To answer this question, a genetic algorithm selects line-level phases, each candidate timetable is represented through GTFS updates, and OTP recomputes the complete itinerary of every sampled query. The reported effects should be interpreted as changes in planned schedule-based itineraries for the evaluated time windows; realized ridership, capacity effects, and operational reliability remain outside the present analysis.

The paper is organized as follows. In Section 1, the introduction is presented. Section 2 demonstrates the dataset and the optimization algorithm along with its practical implementation. Section 3 outlines the results of the timetable optimization. Section 4 discusses the limitations and implications of the findings and sets several future research directions. Finally, Section 5 concludes the findings of the paper.

## 2. Methodology

This section presents the data used, the formulation, and implementation of the timetable optimization problem.

### *2.1. Data collection*

This case study is based in Budapest, the capital of Hungary, with 1.68 million inhabitants. The network includes four metro lines and 35 tram lines running on roughly 174 kilometers of track. The transportation system consists of more than 200 daytime bus lines along with around 40 night bus lines, and 15 trolleybus routes. In addition, several suburban rail lines connect the city to nearby towns. The city is split into two parts by the Danube River. The current study deals with 11 tram lines on the right bank of the river (Figure 1).

The trip planning data are obtained from the BudapestGO public transport application (**?**). It is the official application for the public transport of Budapest, which includes trip planning, real-time vehicle information, and ticket purchase. For this study, the data were collected for six months, between November 2024 and May 2025. The dataset contains 9,677,858 trip-planning queries submitted through the application, each containing origin and destination coordinates and a planned departure date and time. The itineraries originally returned by the BudapestGO app to users were not included in the dataset and were instead recomputed in OTP for this study. Additionally, the GTFS for the given time period is provided.

As discussed in Section 1, the trip-planning queries are used as a proxy for intended passenger demand. The scale of the sample is substantial: 1.4 million individuals used BudapestGO in August 2024, compared with 2.6 million inhabitants in the metropolitan area (**??**), and the 9,677,858 queries correspond to approximately 1.2 planning requests per 100 BKK passenger boardings over the stated observation period (about one query per 79–92 boardings) (**?**).

### *2.2. Data filtering*

Due to the sheer amount of data and the complexity of the optimization, the trips used for the optimization have to be filtered. The data was already deduplicated. First, trip-planning queries shorter than 500 meters are removed. Afterward, the data are spatially and temporally filtered to optimize solely a portion of the schedule. Approximately 66% of the planned trips

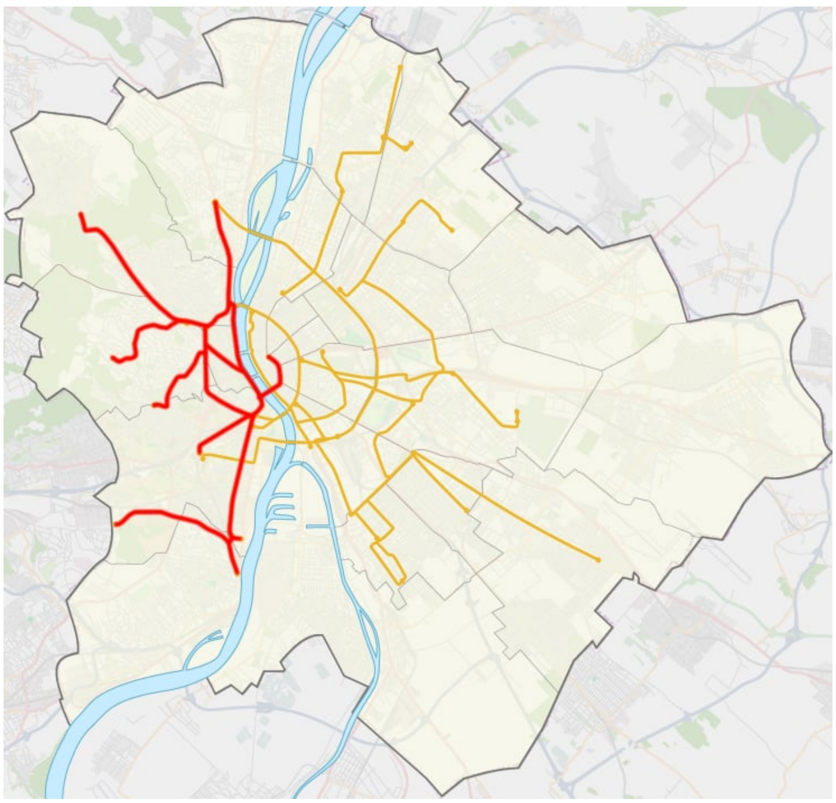


Figure 1: The tram network of Budapest (original image from ?). The tram lines manipulated in this study are highlighted in red. (GPS latitude: 47.49835, longitude: 19.04045)

in the dataset are shorter than 500 meters. It might be because the users often use the application for navigation to plan the last leg of their journeys. For this analysis, these shorter trips are filtered out. Additionally, merely the tram network of the Buda side (i.e., right bank of the Danube River) is optimized. During the evaluation periods, out of the 11 lines, solely nine run in the morning and eight during the evening (i.e., some trams run exclusively on weekends or at specific times of the day). Therefore, trip plans that originate from and end in this Buda part of the city (i.e., based on origin and destination GPS coordinates) are included in the evaluation.

In this research, one hour of Monday morning rush hour traffic and one hour of evening off-peak traffic are studied. The two periods were selected as contrasting scheduled-service conditions, not as uniform samples of the day or to estimate an all-day average effect. The same methodology can be applied to other time windows using the corresponding query demand and timetable. Precisely, trips that fall on Mondays and depart between 8:00 and 9:00 for the morning dataset and between 21:00 and 22:00 for the evening dataset are used. In addition, to verify the temporal robustness of the solution and ensure the optimized timetable does not overfit to random daily travel variations, 1/5 of the morning data is used to evaluate the optimization, while another randomly selected 1/5 serves as an unseen validation set. In the case of the evening dataset, it is 1/3-1/3. The rest of the data is not used to speed up the optimization. It is assumed that the sampled queries are representative of the trip-planning demand observed in the selected time windows. Data filtering is summarized in Figure 2. Finally, the trips used for the baselines are depicted in Figure 3.

### *2.3. Problem formulation*

The decision variable in the optimization is a set of offsets in the schedule of the trams. Their frequency and headways are left unchanged. In total, $k_m = 9$ and $k_e = 8$ tram lines are manipulated in the morning and evening scenarios, respectively. The baseline schedule is used as a reference, and the headway of each tram is partitioned into five evenly spaced temporal offsets. Offsets are given in integer minutes as tram lines are usually matched with the traffic signal program along the line. Offsetting by whole minutes, assuming 60-second or 90-second traffic light programs, the departures are offset by whole traffic light cycles not delaying them unnecessarily. For example, a headway of 15 minutes results in offsets of 0, 3, 6, 9, and 12 minutes.

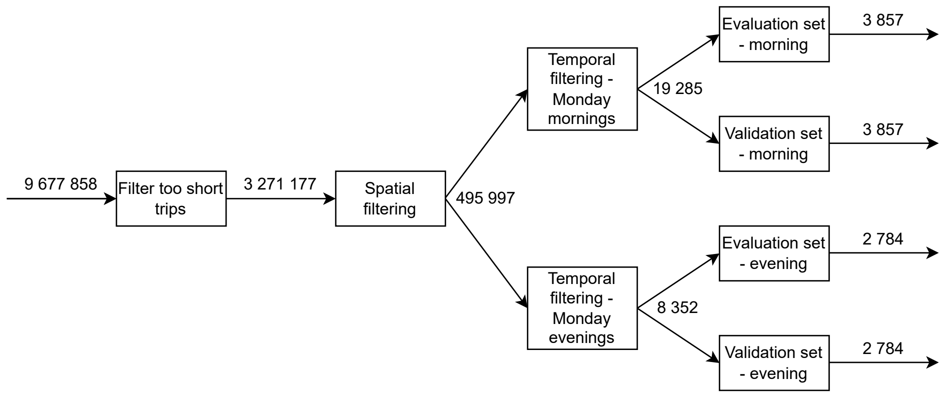
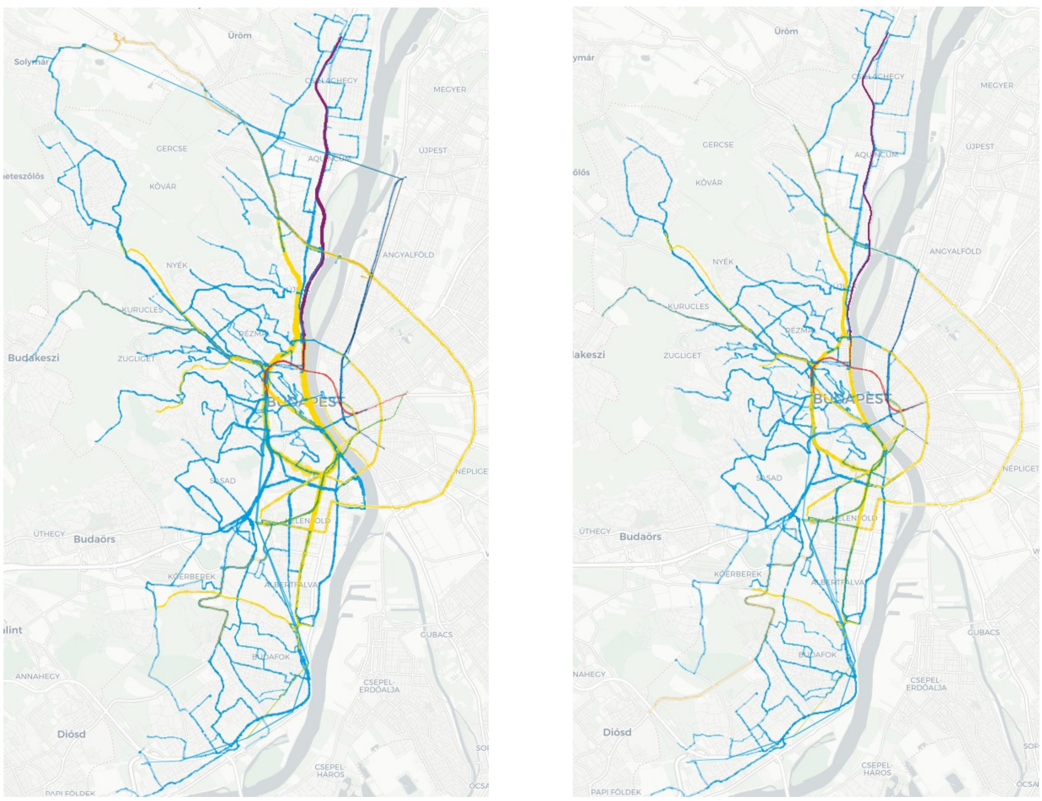



Figure 2: Filtering the relevant trips

This approach allows for systematic variation in departure times while preserving the overall service frequency. Therefore, the search space has $5^9$ and $5^8$ combinations. The objective is to minimize the total travel time $T$ for all passengers' trips $t_i$ ($T = \sum_{i=1}^{N} t_i$, where $N$ is the number of trips in the evaluation set).

To account for the trams departing before the investigated period and the users finishing their trips after it, all trams with departures plus and minus one hour of the investigated period are manipulated, too. The relationship between offsetting the tram departures and travel times is highly nonlinear. While the $5^9$ and $5^8$ combinatorial search spaces might initially appear constrained enough for an exhaustive search, this dimensionality hides the immense computational weight of the objective function. Because offsetting departures can alter the itineraries returned by OTP, evaluating a single schedule combination requires replanning thousands of individual journeys. Brute-forcing the approximately $1.95 \times 10^6$ combinations for the morning scenario would demand billions of individual trip plan calculations, rendering an exhaustive search mathematically and computationally intractable. Therefore, a metaheuristic optimization algorithm, specifically a genetic algorithm (GA), is strictly required to solve this problem efficiently (?). It is assumed that the current public transport schedule is already near-optimal with merely minor improvements possible. Therefore, the schedule is included in the initial GA population as the first individual, which is initialized with zero delay to represent the real timetable.

This baseline serves both as a reference point for comparison and as the




(a) Monday mornings - validation set

(b) Monday evenings - validation set

Figure 3: Trips used for the optimization. The yellow lines are trams, the blue lines are buses, and the others are metro or rail. Line widths are proportional to the number of the trips using that line.

starting point for optimization. The goal of the algorithm is to refine and enhance this current optimum. Mathematically, the optimization is formulated as follows. Denote the vector of tram departure offsets with $u \in \mathbb{Z}^{0+,k_m}$ or $u \in \mathbb{Z}^{0+,k_e}$ (with $u = \underline{0}$ being the baseline). Then,

$$\min_u T(u) = \sum_{i=1}^{N} t_i(u) \tag{1}$$

subject to

$$u \in S$$

where $S$ denotes all possible combinations of departure offsets.

The genetic algorithm is configured with a population size of 10 and runs until convergence. Crossover and mutation rates are set to 0.7 and 0.2, respectively. These parameter settings are consistent with commonly used ranges in the genetic algorithm literature, as reported, for example, by **?**. The GA was configured with a maximum of 100 generations and a fixed random seed of 42 to ensure reproducibility. The stopping criterion was no improvement in the best timetable for 10 consecutive generations. The workflow of the fitness evaluation for each individual is summarized in Algorithm 1.

---

**Algorithm 1:** Compute the total travel time

---

**Input:** individual, demands
**Output:** $T$: total travel time
**Data:** $H$: mapping of previous individuals to their total travel time
**if** *individual* $\in H$ **then**
    $T \leftarrow H[individual]$;
**else**
    $T \leftarrow 0$;
    set_GTFS_RT(individual);
    **for** *demand in demands* **do in parallel**
        $t_i \leftarrow$get_travel_time(demand);
        $T \leftarrow T + t_i$;
    $H[individual] \leftarrow T$;
**return** $T$;

---

Since for each individual (i.e., timetable variant), all travel demands need to be evaluated, and the trip planning is deterministic, a hashtable is defined to store the results of the individuals already calculated. The above evaluation runs until the optimization converges to a (local) optimum and a better candidate cannot be found for 10 consecutive generations. Each generation of the algorithm consists of 10 individuals, and the algorithm terminates within a few dozen iterations (depending on initialization). The optimized timetables reported below correspond to the best individuals retained over the GA runs.

### *2.4. Implementation*

The optimization runs on a dedicated Virtual Private Server (VPS) managed by a Python script. This script first initializes an MQTT (?) (i.e., a lightweight, publish-subscribe-based messaging protocol designed for resource-constrained devices and low-bandwidth, high-latency networks) client to establish the communication with the OTP instances on a designated topic. Afterward, it sets up the genetic algorithm for the optimization process.

In this research, OpenTripPlanner (OTP) version 2.7 was used as an open-source, GTFS-based itinerary evaluator (?). OTP combines GTFS timetable data with an OpenStreetMap street network to compute multimodal, schedule-based itineraries (?). Each trip planning request supplied the query origin and destination coordinates through `fromPlace` and `toPlace`, a common Monday service `date`, and the planned departure `time` at minute resolution. The requests used `mode=WALK,TRANSIT`, `numItineraries=1`, and `walkReluctance=10`, while `showIntermediateStops` and `full_elevation` were set to false. Because `arriveBy` was omitted, the requests were evaluated as depart-at searches. The first and only returned itinerary, was used in the objective. Thus, OTP recomputed the recommended schedule-based itinerary for each timetable rather than simulating probabilistic passenger choice. Vehicle capacity, denied boarding, and crowding-dependent route choice were not modeled; OTP therefore treated each scheduled service as available for boarding.

GTFS Realtime (GTFS-RT) is used to adjust tram departure times during the optimization. No additional real-time disruptions, delays, or service cancellations were considered. At initialization, the original calendar dates are mapped to a common regular-Monday service date, while the planned departure times of day are preserved. Each recorded origin-destination-time query is then submitted to OTP both for the baseline timetable and the candidate timetables. During the evaluation, for each individual in the genetic algorithm, for every tram in the relevant time interval, the GTFS stop times are offset as defined by the genome of that individual. The updated stop times are packaged as a GTFS-RT (?) feed and published to the OTP instances via the MQTT topic. OTP updates its schedules based on this real-time feed for the target date. Afterward, every trip in the evaluation dataset is planned by using OTP with the offset stop times. Then, the resulting travel times are summed. Since a different set of tram departure delays can result in varied route plans for the travelers, it is necessary to re-evaluate every trip for every individual. The designed optimization loop uses two OTP instances to improve the performance and scalability by planning demand trips in parallel. For the itinerary returned for query $i$, total travel time $d_i$, waiting time $w_i$, walking time $a_i$, and in-vehicle transit time $v_i$ were read directly from the OTP response fields `duration`, `waitingTime`, `walkTime`, and `transitTime`, respectively. All four fields are expressed in seconds, and the responses satisfy $d_i = w_i + a_i + v_i$. Records without a valid route-planning output were excluded before analysis. The values in Table 2 were obtained by summing each field across the retained queries and converting the totals to minutes. The goal of the genetic algorithm is to minimize the total travel time; the fitness value for each individual is the total travel time with a negative sign. During the operation, trip planning responses are logged for each traveler for better comprehension of the optimization results.

OTP is deployed on a server configured with an Intel Xeon Gold 63 processor, which provides access to four cores operating at 2.9 GHz. The server is equipped with 31.3 GB of RAM, and each OTP instance is allocated 10 GB of memory.

## 3. Results

The outcomes obtained for the four data subsets are outlined (Figure 2). A summary of the corresponding modified schedules across scenarios is provided in Table 1.

The key metric for assessment is the improvement in the total travel time compared to the baselines (i.e., original schedules) in the morning and evening scenarios. In both cases, the optimization converges to a better result than the baseline on the evaluation sets (see Table 2). Thus, it is confirmed that there is room for improvement in the tram schedules. To assess out-of-sample performance within each selected period, the optimized timetables are also evaluated on the corresponding validation samples.

Table 1: Tram frequencies and departure offsets from the optimization

| Tram ID | Morning frequency (min) | Evening frequency (min) | Morning departure offset (min) | Evening departure offset (min) |
|---|---|---|---|---|
| 17 | 7 | 20 | 0 | 0 |
| 19 | 15 | 20 | 0 | 12 |
| 41 | 15 | 20 | 0 | 0 |
| 47 | 7 | 15 | 4 | 9 |
| 49 | 10 | 20 | 2 | 12 |
| 56 | 10 | - | 0 | 0 |
| 56A | 10 | 20 | 0 | 0 |
| 60 | 20 | 30 | 0 | 6 |
| 61 | 15 | 20 | 0 | 9 |

According to Table 2, the total travel times decrease between 0.07% and 0.21% depending on the dataset used. The decrement is similar in magnitude for the evaluation and validation sets; therefore, there is no overfitting. To separate timing effects from route and mode changes, Table 3 classifies itineraries before and after the optimization as unchanged, changed route/transfer patterns with the same modal sequence, or changed modal sequences. In the morning samples, most of the saving occurred along unchanged itineraries and was primarily associated with reduced waiting time; in the evening samples, a small subsample of users saw substantial time savings from changed modal sequences.

Furthermore, the results are subject to a more in-depth analysis to make the root causes of the improvement transparent. The following diagrams (4) analyze the altered trips and the number of tram users on each tram. Trip planning results show that almost 50% of the travelers use trams during their trips, but according to Figure 4, merely a small portion of the trips are changed due to updating the schedule. The figure shows that the randomly sampled evaluation and validation sets exhibit similar distributions of tram line usage. This suggests that they reflect the same underlying distribution as the full dataset.

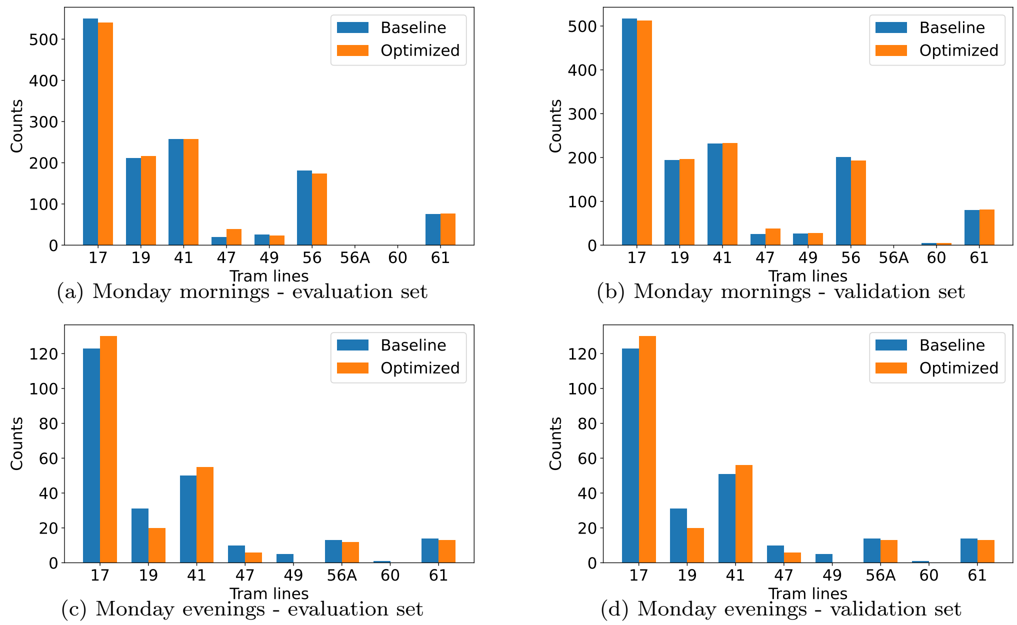
Analyzing the changed trips, trams that travel (partially) on the same route (e.g., trams 17 and 19 or trams 47 and 49) are sometimes swapped between the baseline and the iterated plans. That is because their departures are altered, and one might come earlier than the other. Although the total

Table 2: Travel time results

| | | Total waiting time (min) | Total walking time (min) | Total transit time (min) | Total travel time (min) |
|---|---|---|---|---|---|
| Morning Evaluation | Baseline | 9,683 | 36,184 | 57,675 | 103,542 |
| | Optimized | 9,652 | 36,161 | 57,655 | 103,468 |
| | Difference | −31 | −23 | −20 | −74 |
| | Rel. diff. | −0.32% | −0.06% | −0.03% | −0.07% |
| Morning Validation | Baseline | 9,977 | 36,366 | 57,682 | 104,025 |
| | Optimized | 9,935 | 36,337 | 57,676 | 103,948 |
| | Difference | −42 | −29 | −6 | −77 |
| | Rel. diff. | −0.42% | −0.08% | −0.01% | −0.07% |
| Evening Evaluation | Baseline | 2,181 | 14,134 | 10,518 | 26,832 |
| | Optimized | 2,164 | 14,176 | 10,437 | 26,776 |
| | Difference | −17 | 42 | −81 | −56 |
| | Rel. diff. | −0.78% | 0.30% | −0.77% | −0.21% |
| Evening Validation | Baseline | 2,186 | 14,170 | 10,527 | 26,883 |
| | Optimized | 2,171 | 14,211 | 10,453 | 26,835 |
| | Difference | −15 | 41 | −74 | −48 |
| | Rel. diff. | −0.69% | 0.29% | −0.70% | −0.18% |

Table 3: Itinerary decomposition of travel-time changes. Note: the reported row totals are slightly lower than the values reported in Figure 2 because some trip plans failed in OTP.

| | Unchanged itinerary | | Route/transfer change | | Mode-sequence change | |
|---|---|---|---|---|---|---|
| | $n$ (%) | Gain/loss (min) | $n$ (%) | Gain/loss (min) | $n$ (%) | Gain/loss (min) |
| Morning evaluation | 3,788 (98.65%) | −45 | 22 (0.57%) | −2 | 30 (0.78%) | −28 |
| Morning validation | 3,794 (98.83%) | −55 | 22 (0.57%) | −4 | 23 (0.60%) | −18 |
| Evening evaluation | 2,732 (98.34%) | 6 | 19 (0.68%) | 6 | 27 (0.97%) | −67 |
| Evening validation | 2,736 (98.42%) | 2 | 18 (0.65%) | 6 | 26 (0.94%) | −55 |




(a) Monday mornings - evaluation set

(b) Monday mornings - validation set

(c) Monday evenings - evaluation set

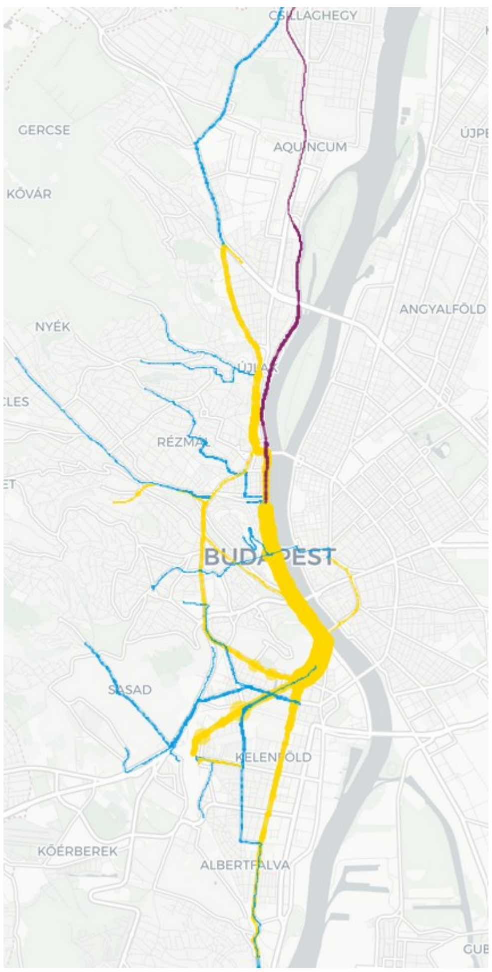
(d) Monday evenings - validation set

Figure 4: Number of trips for each tram line

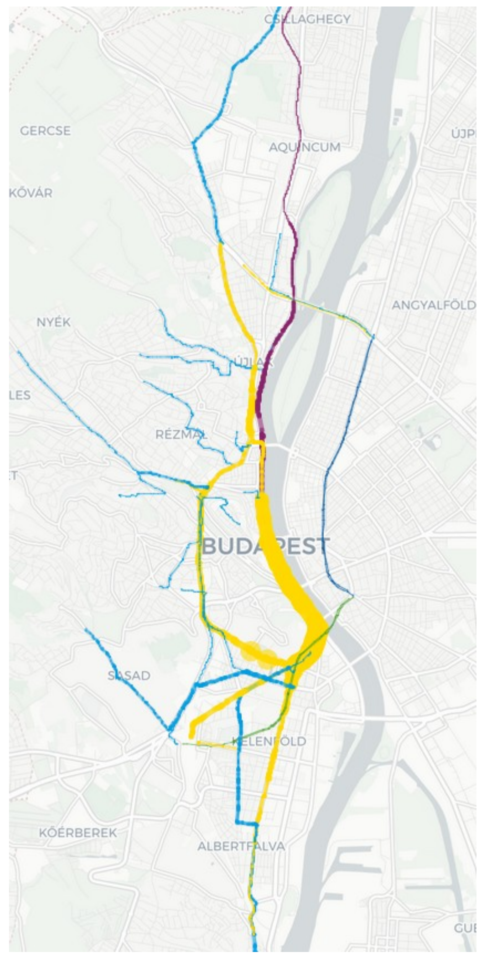
ridership of the trams does not change drastically, mode shifts happen: the algorithm naturally adjusts the schedule in a way that it moves the tram departures where there are alternative modes (i.e., bus or subway); therefore, there is no significant trade-off in travel times. In turn, it adjusts the schedule to fill in the gaps in the timetable.

In the morning scenarios, the baseline schedule leans more heavily on the tram network than in the evening scenario. In the iterated schedule, more subway and bus substitutions are introduced. Although trams 47 and 49 have a relatively small ridership according to Figure 4, altering their schedule has a broad impact. In most cases, these trips are moved to other trams, primarily to trams 17 and 19 in the iterated plans, and vice versa. The shared end station of trams 47 and 49 lies outside the study area; thus, several trips in that direction across the Danube River are filtered out.

In the evening scenarios, the baseline plans with tram line 19 are more frequent. In the optimized version, trams 17 and 41 become more used because the departure of line 19 is offset, and the end stations of all three lines in the southbound direction are the same. Therefore, the distribution of the travelers becomes more uniform. Additionally, the iterated timetable results in slightly more transfers to the buses or the subway. Offsetting the departure for tram 19 leads to the largest change in the planned trips. Some of the schedule adjustments result from synchronizing the tram departures with tram 19, with which tram 49 shares a track in the southern part of the city. Consequently, the departure offsets for tram 49 match those of tram 19. Since tram 49 shares a line with tram 47 in another part of the city, adjusting the schedule of tram 49 indirectly affects the departures of tram 47, as well.

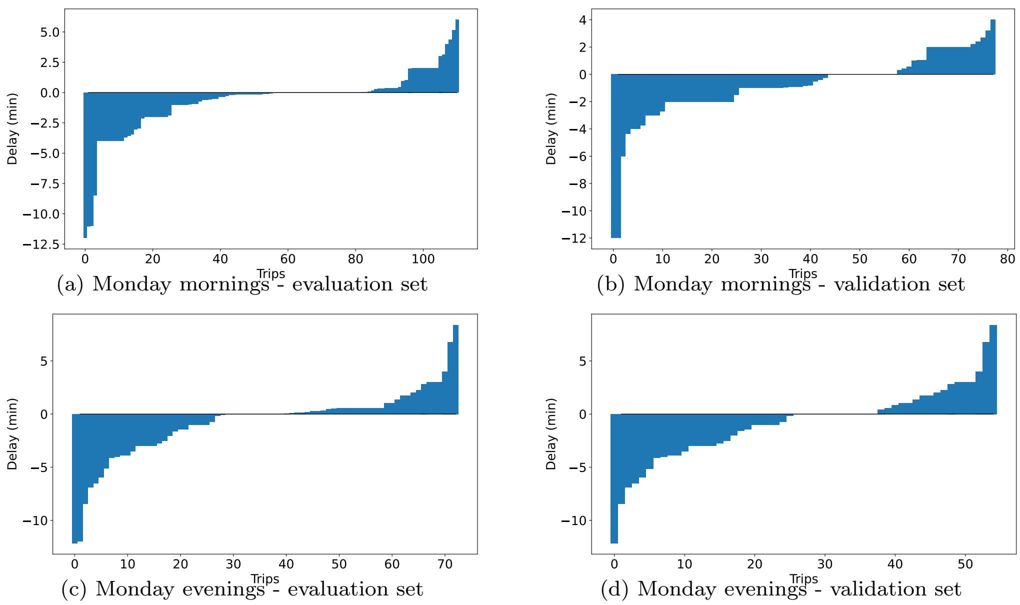



(a) Changed evening trips - validation set, baseline




(b) Changed evening trips - validation set, optimized

Figure 5: Planned trips affected by the timetable optimization in the evening validation dataset




(a) Monday mornings - evaluation set

(b) Monday mornings - validation set

(c) Monday evenings - evaluation set

(d) Monday evenings - validation set

Figure 6: Travel time changes for the altered trips

Although the relative travel time savings are merely $0.07\% - 0.21\%$, in absolute terms, it can be translated into substantial savings in opportunity cost. Small travel-time savings can have aggregate value when repeated across millions of journeys (?).

## 4. Discussion

The main result of this research is that it can make the public transport service more efficient by reducing aggregate user delays without requiring any capital or infrastructure investment. The same number of vehicles are used with the same headways, not compromising the drivers' schedule or altering their lines. Therefore, the recommendations in this study are limited operational intervention and actionable. In addition, the methodology is universal and can be adapted to any public transport context as it is not limited to a specific city. The mere requirements for the algorithm to work are a sufficiently large set of origin-destination-time demand observations relevant to the target application and a public transport network represented by a static GTFS schedule. GTFS-RT is not a requirement from the service provider's side. It is a tool used for optimization alongside OTP.

The optimization typically leads to the same result: changing the departure times for a few lines. While altering schedules inevitably creates minor travel-time penalties for a small subset of individual users (as observed in Figure 6), the algorithm dynamically displaces these departures to corridors well-served by alternative modes (i.e., bus or subway). Therefore, the negative impact on individual trips is minimized, ensuring a net-positive aggregate efficiency gain for the network without severe localized trade-offs. In turn, it adjusts the schedule to fill in the gaps of the timetable and induces a natural mode shift between public transport modes. The algorithm can uncover relationships or relevant transfers that might not immediately be obvious for public transport planners.

There are some limitations of the study. Trip-planning queries capture users actively requesting complete origin-destination itineraries rather than simply checking the next arrival, indicating a willingness to evaluate route alternatives and multimodal connections (**??**). Previous studies also show that planner logs contain useful aggregate demand information and that route planning is particularly relevant for unfamiliar, non-routine, and uncertain journeys (**??????**). With nearly 9.7 million requests (approximately 1.2 queries per 100 BKK boardings) the dataset provides a substantial demand basis for identifying timetable changes that affect planned routes and transfers. Accordingly, the queries are used as a proxy for the demand, rather than as a representative sample of all passengers. The importance of population representativeness depends on whether the transfer effects point in the same direction. When a recurring coordination problem penalizes many journeys and an offset improves those connections without producing comparable losses elsewhere in the observed query demand, the optimization identifies a clear direction of improvement even if the sample does not reproduce the complete passenger population. If an offset improves some transfers while worsening others, however, the preferred solution depends on the relative weights of the affected journeys, and representative demand becomes important for identifying the population-wide optimum. The similar improvements obtained for the evaluation and held-out validation samples indicate that the direction of the identified gains is stable within the observed query demand. Their magnitude and distribution across the complete passenger population nevertheless remain uncertain. This limitation is most relevant during the morning peak, when crowding may alter realized journeys and travel times. However, Table 3 shows that only a small share of modeled queries received a different route/transfer pattern or modal sequence, indicating limited reassignment within the evaluated sample; localized crowding effects were not assessed. Real-time disruptions, delays, and congestion were not assessed in the paper.

Furthermore, the decision to evaluate the optimization exclusively against the current operational timetable, rather than simpler theoretical heuristics, is a practical one. The existing timetable is not a naive baseline; it is a highly refined, expert-crafted schedule designed by the transit agency (BKK). Demonstrating that an automated, low-cost algorithm can uncover latent inefficiencies and consistently outperform this expert baseline proves the practical value of integrating full-journey, intent-to-travel data into the optimization loop.

Although the resulting travel-time improvements are modest, the findings should not be interpreted as evidence of large operational gains. Rather, they demonstrate that app-based demand data can be incorporated into timetable refinement workflows and can influence network-wide routing outcomes. As noted by **?**, opportunities for major structural modifications in mature urban rail systems are often limited, motivating the investigation of timetable-based interventions. The proposed framework modifies only departure times while preserving service frequencies and fleet requirements.

Beyond the limitations associated with the trip-planning query data, the analysis currently operates within a limited spatial, temporal, modal, and computational scope. Spatially, the study does not consider the whole city, merely a limited area. However, in transportation, there are no closed systems: each subset of the network is interlinked with its neighbors, e.g., if the whole city were considered, there would be interaction with its suburbs, this is a classical boundary problem. Temporally, the analysis is limited to two Monday windows selected to contrast morning peak and late-evening off-peak conditions. Similar results for the evaluation and validation samples support stability within these periods, but do not establish transferability to other hours, weekdays, weekends, seasons, or disrupted operating conditions. Broader temporal validation is therefore required before drawing system-wide conclusions. For the same reason, the case-specific percentage reductions cannot be applied directly to annual system-wide passenger time; such an extrapolation would require evidence on the share of journeys, network areas, and operating periods experiencing comparable effects. The limitation of the model is that the study focuses exclusively on the tram network, which is merely one part of the public transport network. Although the outcomes with the optimized schedule show better alignment with other modes (i.e., better transfers), a multimodal optimization would most likely yield superior results. This leads to the final limitation, which is computational in nature. Evaluating all the trips multiple times for the sake of optimization is very resource-intensive and scales poorly. Therefore, the scope of the optimization must be matched with the available computational resources.

Yet, the findings of this study show that the data-driven schedule optimization is feasible, and if sufficient computational resources are available, it is worth scaling up.

The optimization is developed in a way that it can be easily adopted by public transport service operators, without incurring any extra operational costs. The scalability problem can be overcome by involving the transportation service operators to point out problematic areas in terms of the quality of service thus appropriately selecting the scope of the optimization (i.e., which lines to optimize and in what time interval). Although the algorithm can easily be adapted to different cities or public transport networks in general, unthinkingly following the optimization results might compromise the transparency. Expert oversight and open discussion with the stakeholders are always required when making timetable alterations. Three main directions for future research are identified. First, the method- ology can be adapted to varied urban contexts, not merely to Budapest. Second, the filtering of data should be improved to consider exclusively the relevant trips and discard those not impacted by the schedule change at all. Third, with the help of trip planning, secondary metrics, such as emissions or health impacts, can be calculated, as demonstrated by **?**. Therefore, the schedule optimization problem can be formulated by taking multiple quality-of-service-related objectives into account.

## 5. Conclusions

This paper presents a data-driven tram schedule optimization method relying on a genetic algorithm and travel-demand data derived from a public transport trip-planning application. The proposed framework integrates demand modeling with timetable optimization by iteratively evaluating timetable modifications through route planning. Similar travel patterns and improvements were observed in both the evaluation and validation datasets. The optimized timetables reduced total travel times by approximately $0.07\%-0.21\%$ depending on the analyzed scenario. Although these improvements are modest, the results demonstrate that small timetable modifications can influence passenger routing outcomes. The findings further highlight that public transport schedules should be analyzed in a network-wide context, since timetable changes on one line can affect route choices and travel times throughout the network. The main contribution of this work is the demonstration of a practical framework that combines app-based demand data, timetable optimization, and multimodal route planning. While the case study is limited to a subset of the Budapest tram network and specific time periods, the results indicate that demand-informed timetable refinement is feasible and can identify opportunities for modest travel-time reductions. Additional validation on larger networks, broader temporal horizons, and under operational constraints would be required before drawing wider operational conclusions. Finally, the interpretation and implementation of timetable modifications require detailed knowledge of local travel patterns and operational conditions. Therefore, optimization results should be considered as decision-support tools to complement, rather than replace, the expertise of public transport operators.

## Acknowledgements

This work has received funding from the European Union under the Digital Europe Programme (DEP) through the project deployEMDS (Grant Agreement No. 101123520) and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

## Declaration of Interest

None.

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Microsoft Copilot in order to improve the language and clarity of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.