Policy Analytics in Public School Operations

Dimitris Bertsimas
MIT Sloan School of Management, dbertsim@mit.edu

Arthur Delarue
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, arthur.delarue@isye.gatech.edu

Getting students to the right school at the right time can pose a challenge for school districts in the United States, which must balance educational objectives with operational ones, often on a shoestring budget. Examples of such operational challenges include deciding which students should attend, how they should travel to school, and what time classes should start. From an optimizer's perspective, these decision problems are difficult to solve in isolation, and present a formidable challenge to solve together. In this paper, we develop an optimization-based approach to three key problems in school operations: school assignment, school bus routing, and school start time selection. Our methodology is comprehensive, flexible enough to accommodate a variety of problem specifics, and relies on a tractable decomposition approach. In particular, it comprises a new algorithm for jointly scheduling school buses and selecting school start times, that leverages a simplifying assumption of fixed route arrival times, and a post-improvement heuristic to jointly optimize assignment, bus routing and scheduling. We evaluate our methodology on simulated and real data from Boston Public Schools, with the case study of a summer program for special education students. Using summer 2019 data, we find that replacing the actual student-to-school assignment with our method could lead to total cost savings of up to 8%. A simplified version of our assignment algorithm was used by the district in the summer of 2021 to analyze the cost tradeoffs between several scenarios and ultimately select and assign students to schools for the summer.

Key words: optimization, education, transportation, school bus scheduling, modeling, public sector

---

1. Introduction

Public schools in the United States play a significant role in ensuring all children have equitable access to education. Yet before students even step through the school doors, administrators must solve a wide array of operational challenges to ensure their school district is able to deliver on its educational mission. In particular, districts must decide which school students will attend, how they will get there, and what time school should start.

While simple to state, these questions can prove difficult to answer in practice. While public-school students largely attended the nearest school for the better part of the twentieth century, increasingly widespread school choice programs allow students to attend schools further from their home. Correspondingly, they raise important questions around the fair allocation of school seats (Abdulkadiroglu and Sönmez 2003). Farther home-to-school distances, coupled with the burgeoning population of major US cities, make managing local bus fleets — which collectively add up to half a million yellow school buses nationwide — a challenge. The average U.S. school district spends about \$1,000 per transported student per year for a total nationwide expenditure of \$25 billion, and large districts such as Boston with many special education students spend as much as \$5,000 per transported student (NCES 2021, Boston Public Schools 2018). To keep costs down, many districts opt to stagger the start and end times of different schools, allowing buses to serve multiple schools in succession instead of remaining idle for most of the day. This strategy can create large savings, but also increases the difficulty of finding a cost-efficient schedule that works for students, teachers and staff (Bertsimas et al. 2019).

These operational challenges are the source of major headaches for school districts across the US. Suboptimal solutions can divert precious funds from classrooms, which ultimately can negatively impact the quality of education students receive. School assignment, school transportation, and school start times are interconnected problems; it is complex enough for practitioners to solve them in isolation, much less to unravel their interplay. Yet understanding how these decisions interact is a key priority for administrators when considering strategic policy changes such as school or transportation eligibility. The goal of this paper is to develop optimization models for key problems in public school operations, and to show how these models can be applied in practice to inform policy decisions.

Our work is motivated by a real-world case study at Boston Public Schools. Each summer, the district runs a 5-week program called "Extended School Year" (ESY) for students with special needs, designed to prevent summer learning loss through continued academic programming. From an operational perspective, the ESY program presents a microcosm of the challenges faced by the school district during the normal school year, from assignment to transportation and start times. As a result, it constitutes an informative case study for our optimization models.

1.1. Related Work

School assignment. The problem of assigning students to schools has received significant attention in the operations research literature. In the 1970s, the push to desegregate US public schools motivated Franklin and Koenigsberg (1973) and Liggett (1973) to redraw the attendance boundaries of schools with optimization algorithms assigning census blocks to particular schools. The objective is minimizing distance (or squared distance, out of compactness concerns) subject to capacity and racial balance constraints. More recently, Caro et al. (2004) propose an optimization model for creating attendance zones for schools, explicitly modeling various desirable properties including contiguity, grade-balance, and distance from a previous solution.

A related line of work concerns assigning students to schools while simultaneously deciding which schools to open and close (capacitated median problem). Teixeira and Antunes (2008) propose a model to decide which grades should be hosted at each school, while Araya et al. (2012) develop a multiobjective model to select schools in rural Chile. Delmelle et al. (2014) formulate a multi-period problem to model the impact of long-term demographic changes on enrollment. In this work, we refer to the related areas of opening and closing schools and assigning students as *classroom operations*.

School transportation. School bus routing is perhaps the most active area of research in school operations (see the reviews by Park and Kim (2010) and Ellegood et al. (2020) for details). The goal is to construct bus routes which transport students from their home (or potentially a nearby bus stop) to school in the morning and back in the afternoon, while satisying a number of individual and system-wide constraints. The problem is typically decomposed into three steps: bus stop assignment (clustering nearby students at a single pickup and dropoff location), bus routing (connecting stops into routes), and bus scheduling (assigning routes or sequences of routes to a bus).

Of these three subproblems, stop assignment is considered straightforward in isolation; Schittekat et al. (2013) develop a local search approach to jointly assign students to stops and connect stops into routes. Given stops, the bus routing problem is essentially a capacitated vehicle routing problem, with bounds on vehicle capacity and maximum riding time. The Clarke-Wright savings heuristic is a popular starting point, and Levin and Boyles (2016) show that it can already outperform the manual approaches used by many school districts. High-quality solutions can be obtained quickly with vehicle routing heuristics (Cordeau et al. 2007) or metaheuristics (Chen et al. 2015).

Once routes have been computed, the school bus scheduling problem aims to assign each bus a route or succession of routes to serve, such that the bus reaches each school either at a specified time or within a specified time window. Swersey and Ballard (1984) propose a discrete-time integer optimization formulation of the problem. Desrosiers et al. (1986) introduce an additional degree of freedom by allowing school start times to vary along with bus schedules, and adopt an alternating minimization approach. Fügenschuh (2009) applies a branch-and-cut approach to a slightly different setting which allows transfers between buses. Most recently, Zeng et al. (2022) propose an integer programming reformulation of the school bus scheduling and start time selection problem under a constant route transition time assumption, which naturally yields a randomized rounding algorithm with guaranteed bounds.

While necessary for tractability, it is clear that considering the routing and scheduling subproblems separately can lead to suboptimality. A few approaches seek to bridge the divide. Braca et al. (1997) develop a constructive heuristic that inserts stops (and the corresponding school, if necessary) into a route until no more stops can be inserted. Shafahi et al. (2018) take into account other schools' locations and bell times when constructing routes for each school. Bertsimas et al. (2019) develop an approach called bi-objective routing decomposition (BiRD) in which several different sets of routes are constructed for each route, and the scheduling step jointly selects the best set for each school while constructing bus schedules.

Because school bus routing research is largely driven by practice, many different problem variations exist in the literature, including inter-bus transfers (Fügenschuh 2009) and mixed loads, i.e., co-riders from different schools (Park et al. 2012). Some works focus on transportation for special education, which can be characterized by longer routes (Russell and Morrel 1986) and heterogeneous fleets (Caceres et al. 2019). In addition, because school bus routes are generally fixed and cannot adjust to traffic, school bus routing problems are generally treated as deterministic. Recent work by Caceres et al. (2017) explores a setting in which student demand and travel times are uncertain.

Finally, we note that despite the abundance of work on both school assignment and school transportation, few studies examine the two in tandem. Mandujano et al. (2012)

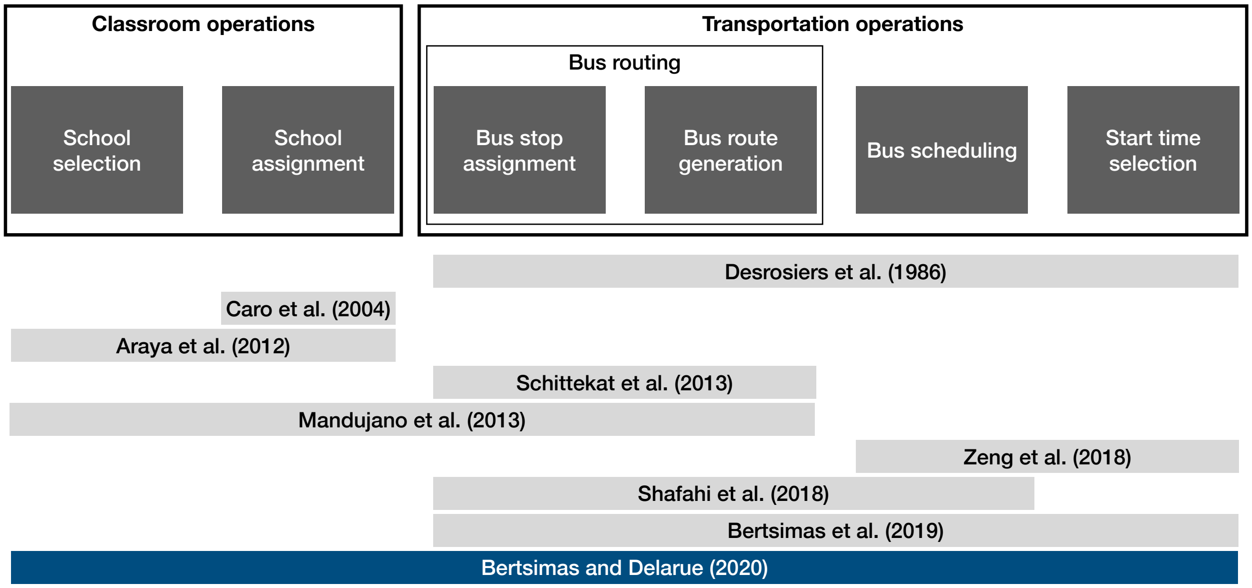
design a two-step assignment-routing procedure, assuming a hub-spoke structure on routes. Meanwhile, Kamali et al. (2013) consider an optimization formulation of the school assignment problem, followed by a greedy routing heuristic. The approach scales to instances with up to three schools and four buses.

From optimization to policy. Policy decisions made by school administrators can have far-reaching implications, and can thus benefit from analytical support. For instance, Ellegood et al. (2015) develop a continuous approximation model to quantify the benefits of mixed-load routing in different school districts. Another avenue to impact decision-making is the creation of accessible software support tools (Caro et al. 2004, Chu et al. 2020).

Addressing challenges in school operations also means wrestling with significant externalities. For instance, adjusting school start times raises questions of equity between families (Banerjee and Smilowitz 2019), as well as public health questions. Indeed, a growing body of medical literature links too-early school starts for teenagers to health and academic issues (Carrell et al. 2011). On the assignment side, the last two decades have seen the emergence of *school choice* programs, in which districts allow families to express preferences between many schools beyond the nearest ones. These programs are the direct result of a vast literature in mechanism design (Abdulkadiroglu and Sönmez 2003), in which the aim is to fairly and efficiently allocate students to schools based on their preferences. Though beyond the scope of this paper, the impact of choice mechanisms on school operations, particularly transportation, is an exciting direction for operations research. As evidence, we cite the work of Shi (2015), in which optimization is used to decide which schools families can rank based on location.

We note that this paper does not explicitly model externalities such as family preferences in the school operations problems we consider—however, our models allow us to simulate many scenarios and quantify the costs and benefits of various policies, providing a tool for policy makers to weigh operational considerations with external concerns. The potential impact of this direction of research is significant: though which school students attend has a disproportionate impact on districts' transportation spending, in most cases decision making is not coordinated across these two areas. Yet the potential benefits are huge, particularly in school districts with generous choice programs that give the students many school options in which to enroll (e.g., under "districtwide" choice programs, students could attend any school in the district). The current reality of school choice programs means

Figure 1 Taxonomy of problems in school operations.



Note. Problems are listed from left to right in the order they are typically considered. Light horizontal bars indicate the problems considered (jointly or separately) by selected studies.

that a purely optimization-based approach is unlikely to be directly and fully implemented by a district. However, it can serve as a baseline against which to evaluate current policies and provide insights to inform new policies.

1.2. Contributions

This paper presents an operations research approach to three problems in school operations, developing a methodology to assign students to schools, construct school bus routes, and schedule school start times. Our approach is:

- **Comprehensive:** given a list of schools and students, and some problem parameters, our method produces a solution in which each student is assigned to a school and a particular bus route, school start times are scheduled, and each bus is assigned an itinerary consisting of one or more consecutive routes. While the problem is initially divided into three main subproblems, we introduce a post-improvement heuristic which jointly optimizes assignment, routing and scheduling, improving the desired objective by up to 20%. To our knowledge, no existing study even considers, let alone optimizes, all these problems simultaneously (see Fig. 1).
- **Flexible:** our models can accommodate a variety of practical constraints, particularly on the transportation side. For instance, we can accommodate a heterogeneous bus fleet,

with restricted student-bus compatibility, as well as time-varying travel times. Our models also allow us to consider multiple objectives, including school and classroom costs on the assignment side, and fleet size, driving distance and driving time on the routing side (contrasting with the focus on total number of buses in most other studies).

- **Scalable:** our approach relies on a three-step decomposition of the overall problem. Students are assigned to school in the first step, bus routes are constructed for each school at the second step, and school start times and bus schedules are jointly selected in the third step. On a case study of medium size (10 schools, 3500 students), we can find high-quality solutions in under an hour.

We highlight two novel algorithms of independent interest in our decomposition approach. The first jointly targets the school bus scheduling and start time selection problems, building on earlier work by Bertsimas et al. (2019). As in that work, we enforce a single arrival time (instead of an arrival time window) for all routes of a particular school and formulate the school bus scheduling problem usign a network flow formulation. We extend the earlier work in three important ways: first, we modify the flow graph substantially so that school bus schedules and school start and end times can be optimized jointly instead of separately; second, we show how the formulation can admit more complex objectives than the number of buses, including time and distance; third, we provide a theoretical analysis of the core underlying assumption (collapsed time windows) in a simplified but realistic setting, finding in particular that this assumption comes at no cost when start and end times are organized into two groups, or "tiers".

The second novel algorithm is a post-improvement heuristic which jointly optimizes school assignment, school bus routing, and school bus scheduling. Like most improvement heuristics, it works by partially destroying the input school bus schedule, then reassembling a solution with lower total cost. A key innovation is that the re-assembling step can modify not only which bus students ride, but also which schools they attend. Crucially for tractability, we are able to formulate the re-assembling step as an integer network flow problem, which allows greater flexibility in defining local search neighborhoods. Evaluated on synthetic data, the method yields an improvement of up to 20% over the decomposition approach.

We also evaluate our integrated optimization approach (or "pipeline") on a real-world case study from Boston Public Schools, a yearly five-week summer program targeted at students with special needs. Our model is flexible enough to accommodate a range of problem specifics (e.g. how to account for students in wheelchairs during school bus routing). We show that our optimization method for assignment alone could reduce total program cost by up to 8%. We then demonstrate our models' ability to describe and explore policy trade-offs, including reducing student travel times, restricting building utilization, and changing the set of start times schools can choose from. Finally, we describe how a simplified version of our model was used by the district to plan for the summer of 2021.

We detail our optimization algorithms in Section 2. In Section 3, we study the implications of arrival time windows in school bus scheduling and start time selection. We evaluate the performance of the assignment-routing post-improvement heuristic in Section 4, and explore the implications of our models on data from Boston Public Schools in Section 5.

2. An optimization pipeline

In this section, we describe the methodology we use to solve the operational challenges faced by a school district, as motivated by the ESY program at Boston Public Schools. We first give an overview of the problem and our decomposition approach. We then detail our approach to school assignment, followed by school bus routing, before discussing our algorithm for school bus scheduling and start time selection.

2.1. Problem overview

The decisions faced by the school district are straightforward: assign each student a school, assign each school a start time, construct bus routes covering each school's students, and decide which bus will serve which route. In doing this, the district has two principal objectives in mind. The first is financial cost. The second is student travel time. These objectives are partially aligned—a bus driving less can reduce student travel time while also saving fuel costs. However, the fixed cost associated with operating a school bus means that it is financially advantageous to operate buses closer to capacity.

The problem described here is too large to formulate directly. In the school bus routing literature, decompositions are often used to manage tractability. We employ a similar approach here, decomposing the problem into three main stages, each of which we formulate as an optimization problem. We first describe individual algorithms for each stage, then introduce a post-improvement heuristic that combines assignment, routing, and scheduling. We provide a notation guide for the reader in Table 1.

Table 1 Summary of notation. Note that notation follows morning conventions (see Section 2.3 for details)

Symbol	Description
School assignment	
$\mathcal{S}$	Set of all schools that can host students
$\mathcal{I}$	Set of all students that must be assigned a school
$\mathcal{P}$	Set of all programs (classroom types) with which students can be associated
$\mathcal{P}_s$	Subset of programs that can be held at school $s \in \mathcal{S}$
$\mathcal{I}_p$	Subset of students associated with program p
K_p	Maximum number of students in a classroom for program type p
Y_s	Maximum number of classrooms available at school s
Bus stop assignment	
$\mathcal{I}_s$	Subset of students assigned to school s
$\hat{\mathcal{I}}_s$	Subset of students assigned to school s who need to be assigned a stop
$\mathcal{L}$	Set of all allowed bus stop locations
$\mathcal{L}_i$	Subset of stop locations within walking distance for student $i \in \mathcal{I}$
Bus routing	
d	Bus depot
$\mathcal{H}_s$	Set of active bus stops for school $s \in \mathcal{S}$
$\mathcal{H}$	Set of all active bus stops
$\mathcal{I}_h$	Subset of students assigned to stop $h \in \mathcal{H}$
$t^{\text{drive}}_{\alpha,\beta}$	Driving time from location α to location β
t^{stop}_h	Pickup time at stop $h \in \mathcal{H}$
t^{school}_s	Dropoff time at school $s \in \mathcal{S}$
$\mathcal{B}$	Set of bus types
Q_b	Maximum number of seats on bus of type b
$\mathcal{R}_s$	Set of all feasible routes for school $s \in \mathcal{S}$
$\mathcal{B}_r$	Set of bus types that can serve route r
θ^r_h	Travel time from stop h to school along route r
Bus scheduling and start time selection	
$\mathcal{T}_s$	Set of start times allowed for school $s \in \mathcal{S}$
$\mathcal{R}^{\text{AM}}_s(t)$	Set of morning routes for school s if starting at time $t \in \mathcal{T}_s$
$\mathcal{R}^{\text{PM}}_s(t)$	Set of afternoon routes for school s if starting at time $t \in \mathcal{T}_s$
t^{service}_r	Total time to serve route r, from first pickup to last dropoff
$\mathcal{V}$	Vertices of the bus scheduling graph
$\mathcal{E}$	Edges of the bus scheduling graph

2.2. School assignment

Formulation. Let $\mathcal{S}$ designate the set of schools, and let $\mathcal{I}$ designate the set of students. Each student is associated with a program p. Let $\mathcal{P}$ designate the set of all programs, and let $\mathcal{P}_s$ designate the set of programs that can be held at school s. We also designate as $\mathcal{I}_p \subseteq \mathcal{I}$ the subset of students associated with program p. Each program p is assigned with a classroom capacity K_p, and each school s has a maximum number of classrooms Y_s.

We define the binary decision variable x_{is}, which takes the value 1 if student $i \in \mathcal{I}$ attends school $s \in \mathcal{S}$, and 0, otherwise. We also define the nonnegative integer variables y_{ps} that indicate the number of classrooms staffed at school $s \in \mathcal{S}$ for program $p \in \mathcal{P}$. Then the basic formulation of the assignment problem can be written as:

$$
\begin{aligned}
\min \quad & \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}} d_{is}x_{is} + \gamma \sum_{s\in\mathcal{S}}\sum_{p\in\mathcal{P}} y_{ps} && && \text{(1a)}\\
\text{s.t.} \quad & \sum_{s\in\mathcal{S}} x_{is} = 1 && \forall i \in \mathcal{I} && \text{(1b)}\\
& \sum_{i\in\mathcal{I}_p} x_{is} \le K_p y_{ps} && \forall p \in \mathcal{P}, s \in \mathcal{S} && \text{(1c)}\\
& \sum_{p\in\mathcal{P}} y_{ps} \le Y_s && \forall s \in \mathcal{S} && \text{(1d)}\\
& x_{is} \in \{0,1\} && \forall s \in \mathcal{S}, i \in \mathcal{I} && \text{(1e)}\\
& y_{ps} \in \mathbb{Z}_+ && \forall s \in \mathcal{S}, p \in \mathcal{P}. && \text{(1f)}
\end{aligned}
$$

Constraint (1b) ensures that each student is assigned to exactly one school. Constraint (1c) relates the number of students of program p assigned to school s to the number of classrooms that need to be staffed for program p at school s. Constraint (1d) imposes a limit on the number of classrooms that can be staffed in each school building. The objective (1a) trades off the total student-to-school distance with the cost of staffing each classroom via the hyperparameter γ. Ideally, we would replace the first term in the objective with the school bus routing cost corresponding to the assigment represented by the $\boldsymbol{x}$ variables, but representing this function explicitly requires modeling the downstream routing problem in its entirety, which is intractable.

Formulation (1) models the key decisions of the school assignment problem. It can be extended to take into account other desired problem specifics.

- **Student and program restrictions:** Some schools may not be able to host certain programs due to a lack of appropriate facilities or equipment. Additionally, students may not be allowed to attend any school (for example, bathroom fixtures at elementary schools cannot be used by high school students). In both cases, the relevant y or x variables can be fixed to 0.

- **Student cohorts:** A particular subset of students $\bar{\mathcal{I}} \subseteq \mathcal{I}$ (e.g., siblings) may need to be assigned to the same school. In this case, we can pick one student $i_0 \in \bar{\mathcal{I}}$, and impose $x_{i_0 s} = x_{is}$ for each school s and each student $i \in \bar{\mathcal{I}}$.

- **Program-specific classroom costs:** Each program may have slightly different staffing and facility needs, leading to a different contribution to the objective. We can easily modify the objective to include weights μ_p capturing the relative cost of classrooms for different programs.

School selection. Sometimes, the set of schools to open is not fixed. For a summer program like BPS's Extended School Year, the district must sometimes decide which school sites to utilize, turning the assignment problem into a facility location problem Teixeira and Antunes (2008). We can adjust formulation (1) to take into account this additional decision, by introducing a new binary variable z_s for each school s, taking the value 1 if school s is open, and 0 otherwise, and adding the constraints:

$$x_{is} \le z_s \qquad \forall s \in \mathcal{S}, i \in \mathcal{I} \tag{2a}$$

$$\sum_{p \in \mathcal{P}} y_{ps} \le Y_s z_s \qquad \forall s \in \mathcal{S}. \tag{2b}$$

Note that the inclusion of the latter constraint renders constraint (1d) superfluous.

Even if the set of schools is fixed, the district may decide to open some schools, but only for a subset of programs or students—for example, opting to open a school for only elementary school, middle school or high school students (denoted as $\mathcal{I}_{\text{ES}}, \mathcal{I}_{\text{MS}}, \mathcal{I}_{\text{HS}} \subseteq \mathcal{I}$). In this case, we can define binary decision variables z_s^{ES} (respectively z_s^{MS}, z_s^{HS}), taking the value 1 if school s is open for elementary school (respectively middle school, high school) students, and 0 otherwise. We then add the constraints

$$x_{is} \le z_s^{\text{ES}} \qquad \forall i \in \mathcal{I}_{\text{ES}}, \tag{2c}$$

(respectively for $\mathcal{I}_{\text{MS}}$ and $\mathcal{I}_{\text{HS}}$), allowing us to introduce rules restricting the co-location of different age groups in the same building, e.g.

$$z_s^{\text{ES}} + z_s^{\text{HS}} \le 1 \qquad \forall s \in \mathcal{S}. \tag{2d}$$

We note that formulation (1) uses a weighted-sum-of-objectives technique to trade off student-to-school distance with the number of classrooms. An almost equivalent method to study this tradeoff is a goal programming approach, where we minimize the number of classrooms (more generally, the total classroom cost) subject to the student-to-school distance not exceeding a parametric threshold D.

Regardless of the modeling choices made at the assignment step, the output is always the same: a set of students $\mathcal{I}_s \subseteq \mathcal{I}$ assigned to each school $s \in \mathcal{S}$.

2.3. Bus routing

Once students are assigned to a school, they need to be provided a means to get there. Each student that is eligible for transportation must be assigned a bus stop, then these stops must be visited in sequence by a bus. A common assumption in the school bus literature, which usually holds in practice, is that buses can only carry students from the same school. This fact suggests a natural decomposition of the school bus routing problem: at the *routing* step, stops for each school are connected into routes; at the *scheduling* step, each bus is assigned a set of routes for different schools, to be served in succession. The advantage of this decomposition is that each school's routing problem can be solved in isolation. Each routing subproblem is much smaller than the overall school bus routing problem, and these subproblems can easily be solved in parallel.

From students to stops. While the assignment step of our pipeline deals with students, school bus routing considers *stops*, which are associated with one or more students. As a preprocessing step to the routing problem, students need to be assigned to bus stops. One approach is simply to consider that each student will be picked up outside their home, i.e., there is a one-to-one mapping between students and stops. In practice, many students with special needs are assigned their own bus stop. However, treating every student in this way can be inefficient, especially when many students live close by.

We therefore adopt the following approach to map students to stops. For students who need their own stop, called *door-to-door* students, we simply create a stop at their address and assign them to it—note that the stop may still have multiple students, e.g., siblings living at the same address. For the remaining students, denoted $\hat{\mathcal{I}}_s \subset \mathcal{I}_s$, we consider the set $\mathcal{L}$ of stop locations historically used by the district. Each student $i \in \hat{\mathcal{I}}_s$ is associated with a subset $\mathcal{L}_i \subseteq \mathcal{L}$ of locations within walking distance of their home. Then we solve the following hitting set problem for each school to minimize the number of stops:

$$
\begin{aligned}
\min \quad & \sum_{\ell \in \mathcal{L}} v_\ell && && \text{(3a)} \\
\text{s.t.} \quad & \sum_{\ell \in \mathcal{L}_i} w_{i\ell} = 1 && \forall i \in \hat{\mathcal{I}}_s && \text{(3b)} \\
& w_{i\ell} \le v_\ell && \forall i \in \hat{\mathcal{I}}_s, \ell \in \mathcal{L}_i && \text{(3c)} \\
& v_\ell \in \{0,1\} && \forall \ell \in \mathcal{L} && \text{(3d)} \\
& w_{i\ell} \in \{0,1\} && \forall i \in \hat{\mathcal{I}}_s, \ell \in \mathcal{L}_i. && \text{(3e)}
\end{aligned}
$$

The binary decision variable $w_{i\ell}$ takes the value 1 if student i is assigned to location ℓ, and 0 otherwise, while the binary decision variable v_ℓ indicates whether location ℓ has any assigned students. The objective (3a) is to minimize the number of locations used, while constraint (3b) enforces that each student is assigned exactly one stop and constraint (3c) ensures that location ℓ is marked as used if any student is assigned to it. At the end of this preprocessing step, we obtain a set of stops $\mathcal{H}_s$ for each school $s \in \mathcal{S}$. Each stop $h \in \mathcal{H}_s$ is associated with a single location, and with a subset of the students assigned to school s, which we denote $\mathcal{I}_h \subseteq \mathcal{I}_s$ in a slight abuse of notation. We write the set of all stops as $\mathcal{H} = \cup_{s\in\mathcal{S}} \mathcal{H}_s$.

From stops to routes. The key outcome of the bus routing subproblem is the creation of sequences of stops called *routes*, or sometimes *trips*. We first focus on the *morning* routing problem, in which students are picked up at home and dropped off at school. Formally, we define a route r as a vector of $|r|$ stops; each route may contain a different number of stops, and must terminate at the school it serves (which is not included in the list of stops). The k-th stop in a route r is denoted as $h_k(r)$, and in another slight abuse of notation, the set of stops covered by route r is denoted as $\mathcal{H}_r = \{h \in \mathcal{H} : \exists k \in \{1,\ldots,|r|\}, h_k(r) = h\} \subseteq \mathcal{H}_s$.

For a particular school $s \in \mathcal{S}$, we first assume that the start time of each school s is fixed, and that all routes must arrive at school at the same time. The bus routing problem then becomes a variant of the capacitated vehicle routing problem, where the total number of students on each route is bounded by the bus capacity, and each route's length is bounded by the maximum riding time, a hard constraint determined by the district.

For any two locations $\alpha, \beta \in \mathcal{S} \cup \mathcal{H} \cup \{d\}$, where d designates the (assumed unique) bus depot, we define $t^{\text{drive}}_{\alpha,\beta}$ as the time necessary to drive from α to β. For simplicity, we assume this time is constant, but our models are flexible to driving times varying based on time of day. We further define t^{stop}_h as the time necessary to pick students up at stop $h \in \mathcal{H}$, and t^{school}_s as the time necessary to drop students off at school $s \in \mathcal{S}$. We note that in the afternoon, students are picked up from school and dropped off at home, but we use the morning conventions in conceptualizing travel times. Calling T the maximum time that students can spend on the bus, a route r for school s is time-feasible if it verifies

$$\sum_{k=1}^{|r|-1} \left(t^{\text{drive}}_{h_k(r),h_{k+1}(r)} + t^{\text{stop}}_{h_{k+1}(r)} \right) + t^{\text{drive}}_{h_{|r|}(r),s} \le T, \tag{4a}$$

and capacity-feasible if it verifies

$$\sum_{h\in\mathcal{H}_r} |\mathcal{I}_h| \leq Q, \tag{4b}$$

where Q designates the number of seats on the bus. Note that we assume here that all buses are identical. If the fleet comprises several vehicles, we define $\mathcal{B}$ as the set of bus *types*, and assume all buses of the same type have the same capacity Q_b. A route r can then be associated with a set of comptible bus types $\mathcal{B}_r = \{b\in\mathcal{B}|\sum_{h\in\mathcal{H}_r}|\mathcal{I}_h|\leq Q_b\}$, and we say r is capacity-feasible if $\mathcal{B}_r \neq \emptyset$, i.e., if there is at least one bus type that can serve route r.

Given the set of stops $\mathcal{H}_s$, we call $\mathcal{R}_s$ the (possibly very large) set of all feasible routes for school s. In another slight abuse of notation, we further denote the set of all routes visiting stop $h\in\mathcal{H}_s$ as $\mathcal{R}_h$. We can then write the bus routing problem using a set cover formulation:

$$\min \quad \sum_{r\in\mathcal{R}_s} c_r u_r \tag{5a}$$

$$\text{s.t.} \quad \sum_{r\in\mathcal{R}_h} u_r \geq 1 \qquad \forall h\in\mathcal{H}_s \tag{5b}$$

$$u_r \in \{0,1\} \qquad \forall r\in\mathcal{R}_s. \tag{5c}$$

The parameter c_r designates the cost associated with route r, and the binary decision variable u_r takes the value 1 if route r is selected, and 0 otherwise. Constraint (5b) ensures that the set of routes selected visits every stop.

Following Bertsimas et al. (2019), we define the cost of a route as

$$c_r = \lambda + \sum_{h\in\mathcal{H}_r} |\mathcal{I}_h|\theta_h^r, \tag{6}$$

where θ_h^r designates the travel time from stop h to the school on trip r, i.e., if $h = h_k(r)$, we can write

$$\theta_h^r = \sum_{j=k}^{|r|-1}\left(t^{\text{drive}}_{h_j(r),h_{j+1}(r)} + t^{\text{stop}}_{h_{j+1}(r)}\right) + t^{\text{drive}}_{h_{|r|}(r),s}.$$

The parameter λ trades off the number of buses with the total student travel time.

Formulation (5) is simple to state, but difficult to solve, mostly because the set of feasible routes $\mathcal{R}_s$ is potentially very large. Bertsimas et al. (2019) overcome this challenge by solving problem (5) over a heuristically generated restricted set of routes $\bar{\mathcal{R}}_s \subset \mathcal{R}_s$, with

$|\bar{\mathcal{R}}_s| \ll |\mathcal{R}_s|$. We adopt a different approach, common in vehicle routing settings, in which we generate a greedy solution, then improve it using local search. An initial solution is generated using a randomized greedy heuristic, then improved using a specialized variant of k-OPT. Given a set of routes represented as a directed graph over stops and the school, we delete k edges, yielding a set of *route fragments* that may or may not include the school. We then find the minimum-cost way to recombine these fragments into routes, as detailed in Algorithm 1 in Appendix A.

Throughout this section, we have described the routing problem from the morning perspective, in which students are picked up at home and dropped off at school. However, as previously discussed, routes must also be constructed in the afternoon, picking up students at school and dropping them off at home. Fortunately, the bus routing subproblem in the morning is a mirror image of the bus routing subproblem in the afternoon, and both can be solved in the same way. For example, given a morning route r, where $h_1(r)$ is the first stop visited, and $h_{|r|}(r)$ the last stop visited before reaching the school, we can obtain an afternoon route by reversing the order of stops, so that $h_{|r|}(r)$ is the first stop visited after leaving school, and $h_1(r)$ is the last stop visited. Feasibility with respect to bus capacity remains unchanged, and time-feasibility now means verifying

$$t^{\text{drive}}_{s,h_{|r|}(r)} + \sum_{k=1}^{|r|-1} \left(t^{\text{drive}}_{h_{k+1}(r),h_k(r)} + t^{\text{stop}}_{h_{k+1}(r)} \right) \leq T. \tag{7}$$

Notice that expression (7) can be obtained from expression (4a) by applying the one-to-one mapping $t^{\text{drive}}_{\alpha,\beta} \to t^{\text{drive}}_{\beta,\alpha}$. Though the routes for the morning and afternoon may differ, because driving times are not necessarily symmetric and may depend on time of day, performing this substitution allows us to consider the afternoon problem using the methodology developed above for the morning problem.

2.4. Bus scheduling and start time selection

The methods provided in the previous section allow us to construct routes for a particular school at a particular time. Two consequential decisions remain: the first is to determine which routes will be served in succession by the same bus, both in the morning and in the afternoon; the second is to select each school's start time.

Integer network flow formulation. We propose to make these decisions jointly. For each school, let $\mathcal{T}_s$ designate the set of possible start times, and let $\mathcal{R}_s^{\text{AM}}(t)$ (respectively $\mathcal{R}_s^{\text{PM}}(t)$) designate the set of routes associated for school s for start time $t \in \mathcal{T}_s$ in the morning (respectively afternoon). Note that if we assume morning driving conditions (travel times, etc.) are the same at times t_1 and t_2, then we can have $\mathcal{R}_s^{\text{AM}}(t_1) = \mathcal{R}_s^{\text{AM}}(t_2)$, but our model does not rely on this assumption. We note that we write the school associated with route r as s_r.

To solve the scheduling and start time selection problem, we consider the following network flow formulation. We first assume for simplicity that all buses in the fleet are identical (single bus type), and originate from a single bus depot d. Two morning routes can be served by the same bus if there is enough time for the bus to serve the first route, drop off students at school, then travel to the beginning of the second route, and serve the second route before the scheduled arrival time at school. More formally, we say that $r_1 \in \mathcal{R}_{s_1}^{\text{AM}}(t_1)$ and $r_2 \in \mathcal{R}_{s_2}^{\text{AM}}(t_2)$ are (t_1, t_2)-*compatible*, denoted as $(t_1, r_1) \bowtie (t_2, r_2)$, if and only if

$$t_1 + t_{s_1, h_1(r_2)}^{\text{drive}} + t_{r_2}^{\text{service}} + t_{s_2}^{\text{school}} \leq t_2,$$

where t_r^{service} is the total time required to serve a route r, given in the morning by

$$t_r^{\text{service}} = t_{h_1(k)}^{\text{stop}} + \sum_{k=1}^{|r|-1} \left(t_{h_k(r), h_{k+1}(r)}^{\text{drive}} + t_{h_{k+1}(r)}^{\text{stop}} \right) + t_{h_{|r|}(r), s}^{\text{drive}}.$$

Correspondingly, we say that $r_1 \in \mathcal{R}_{s_1}^{\text{PM}}(t_1)$ and $r_2 \in \mathcal{R}_{s_2}^{\text{PM}}(t_2)$ are (t_1, t_2)-*compatible* if and only if

$$t_1 + t_{s_1}^{\text{school}} + t_{r_1}^{\text{service}} + t_{h_1(r_1), s_2}^{\text{drive}} \leq t_2.$$

Then we construct a directed graph $G = (\mathcal{V}, \mathcal{E})$. The vertex set can be written as $\mathcal{V} = \mathcal{V}_{\text{depot}} \cup \mathcal{V}_{\text{routes}}$, where $\mathcal{V}_{\text{depot}}$ contains exactly two nodes, which we label v_{AM} and v_{PM}, representing the depot at the beginning of the morning, and at the beginning of the afternoon. The remaining vertices are defined as $\mathcal{V}_{\text{routes}} = \cup_{a \in \{\text{AM}, \text{PM}\}} \mathcal{V}_a$ with

$$\mathcal{V}_a = \bigcup_{s \in \mathcal{S}} \left\{ (t, r) \mid t \in \mathcal{T}_s, r \in \mathcal{R}_s^a(t) \right\}.$$

In other words, we create one node per route, for each possible start time and school, both in the morning and afternoon. We then define the edge set as

$$\mathcal{E} = \mathcal{E}_{\text{depot} \to \text{route}} \cup \mathcal{E}_{\text{route} \to \text{depot}} \cup \mathcal{E}_{\text{route} \to \text{route}} \cup \mathcal{E}_{\text{depot} \to \text{depot}},$$

where

$$\mathcal{E}_{\text{depot}\to\text{route}} = \{(v_{\text{AM}}, v) | v \in \mathcal{V}_{\text{AM}}\} \cup \{(v_{\text{PM}}, v) | v \in \mathcal{V}_{\text{PM}}\}, \tag{8a}$$

$$\mathcal{E}_{\text{route}\to\text{depot}} = \{(v, v_{\text{PM}}) | v \in \mathcal{V}_{\text{AM}}\} \cup \{(v, v_{\text{AM}}) | v \in \mathcal{V}_{\text{PM}}\}, \tag{8b}$$

$$\mathcal{E}_{\text{depot}\to\text{depot}} = \{(v_{\text{AM}}, v_{\text{PM}}), (v_{\text{PM}}, v_{\text{AM}})\}, \tag{8c}$$

$$\mathcal{E}_{\text{route}\to\text{route}} = \{(v := (t, r), v' := (t', r')) | (t, r) \bowtie (t', r')\}. \tag{8d}$$

In other words, we include an edge connecting the depot to and from every route (Equations (8a) and (8b)), edges between the morning and afternoon depot nodes (Equation (8c)), and an edge between every pair of compatible routes, i.e., routes that can be served by the same bus (Equation (8d)). For a given vertex v, we denote incoming edges as $\mathcal{E}_{\text{in}}(v) \subseteq \mathcal{E}$, and outgoing edges as $\mathcal{E}_{\text{out}}(v) \subseteq \mathcal{E}$.

Scheduling buses and selecting start times can then be formulated as an integer network flow problem on the graph G. We associate each edge $(v, v') \in \mathcal{E}$ with a nonnegative integer flow variable $f_{v,v'}$; for two route nodes $v = (t, r)$ and $v' = (t', r')$, $f_{v,v'}$ takes the value 1 if r and r' are . Additionally, we define binary decision variables $g_{s,t}$ for each school s and potential start time $t \in \mathcal{T}_s$, taking the value 1 if school s is assigned start time t, and 0 otherwise. The problem can then be formulated as follows:

$$\min \sum_{(v,v')\in\mathcal{E}} C_{v,v'} f_{v,v'} \tag{9a}$$

$$\text{s.t.} \sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} f_{v',v} = \sum_{(v,v')\in\mathcal{E}_{\text{out}}(v)} f_{v,v'} \qquad \forall v \in \mathcal{V} \tag{9b}$$

$$\sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} f_{v',v} = g_{s_r,t} \qquad \forall v := (t, r) \in \mathcal{V}_{\text{AM}} \tag{9c}$$

$$\sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} f_{v',v} = g_{s_r,t} \qquad \forall v := (t, r) \in \mathcal{V}_{\text{PM}} \tag{9d}$$

$$\sum_{t\in\mathcal{T}_s} g_{s,t} = 1 \qquad \forall s \in \mathcal{S} \tag{9e}$$

$$g_{s,t} \in \{0, 1\} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s \tag{9f}$$

$$f_{v,v'} \in \mathbb{Z}_{\geq 0} \qquad \forall (v, v') \in \mathcal{E}. \tag{9g}$$

Constraint (9b) enforces flow conservation at every node in the network. Constraints (9c) in the morning and (9d) in the afternoon force the flow through a node to equal one if the start time associated with the node is selected, and 0 otherwise. Finally, constraint (9e) enforces that each school is assigned exactly one start time. The objective (9a) assigns a cost $C_{v,v'}$ to each selected edge, and can therefore encompass many different potential costs. For instance, the total number of buses in use can be minimized by setting $C_{v,v'} = 1$ if $v = v_{\text{AM}}$, and 0 otherwise. The total driving time can be minimized by defining costs for morning routes as

$$C_{v,v'} = \begin{cases} t^{\text{drive}}_{d,h_1(r')} + t^{\text{service}}_{r'} + t^{\text{school}}_{s_{r'}} & \text{if } v = v_{\text{AM}}, v' = (t', r') \in \mathcal{V}_{\text{AM}}, \\ t' - t & \text{if } v = (t, r), v' = (t', r') \in \mathcal{V}_{\text{AM}}, (v, v') \in \mathcal{E}_{\text{route}\to\text{route}} \\ t^{\text{drive}}_{s_r,d} & \text{if } v = (t, r) \in \mathcal{V}_{\text{AM}}, v' = v_{\text{PM}}, \end{cases}$$

and similarly for afternoon routes. The total driving distance can be minimized with a similar definition of edge costs.

When school districts choose start times, they may consider many external considerations, such as the opinions of families who are not eligible for transportation. Formulation (9) can be adjusted to take such external factors into account, either by restricting the set of allowed start times $\mathcal{T}_s$ for school s, or by augmenting the objective with an additional term $\sum_{s\in\mathcal{S}} \sum_{t\in\mathcal{T}_s} m_{s,t} g_{s,t}$, where $m_{s,t}$ designates the cost of assigning time t to school s.

Bus types and school scenarios. The network flow formulation presented above assumes that all buses are of the same type. This assumption was introduced for clarity, but it is unrealistic in practice. We now show how to adjust formulation (9) to take this fact into account.

As described in Section 2.3, it is more realistic to consider that the school bus fleet is heterogenous, with a set of bus types denoted as $\mathcal{B}$, and that each route can be served by a subset of bus types $\mathcal{B}_r$. In this case, we modify the scheduling graph $G = (\mathcal{V}, \mathcal{E})$ as follows:

- Instead of a single depot node for the morning and afternoon, we create one such node for each bus type, i.e., $\mathcal{V} = \cup_{b\in\mathcal{B}} \{v^b_{\text{AM}}, v^b_{\text{PM}}\}$.
- Instead of a single node for each potential start time t and route r, we create one node per bus type that can serve route r, i.e.

$$\mathcal{V}_a = \bigcup_{s\in\mathcal{S}} \left\{(t, r, b) \mid t \in \mathcal{T}_s, r \in \mathcal{R}^a_s(t), b \in \mathcal{B}_r\right\}.$$

- The edge set $\mathcal{E}$ is constructed in the same way, but only nodes associated with the same bus type b can be connected.

For each set of nodes associated with a single time t and route r, denoted as $\mathcal{V}_{t,r} = \{(t,r,b) | b \in \mathcal{B}_r\}$, we modify constraints (9c) and (9d) as follows:

$$\sum_{v \in \mathcal{V}_{t,r}} \sum_{(v',v) \in \mathcal{E}_{\text{in}}(v)} f_{v',v} = g_{s,t} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s, r \in \mathcal{R}_s^{\text{AM}}(t), \tag{10a}$$

$$\sum_{v \in \mathcal{V}_{t,r}} \sum_{(v',v) \in \mathcal{E}_{\text{in}}(v)} f_{v',v} = g_{s,t} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s, r \in \mathcal{R}_s^{\text{PM}}(t). \tag{10b}$$

These updated constraints enforce that the flow through all nodes associated with a particular time and route is 1 if that time is selected, and 0 otherwise. The particular node (t,r,b) traversed by a unit of flow indicates which bus type $b \in \mathcal{B}_r$ is assigned to route r.

Finally, we notice that the decomposition of the school bus routing problem into routing and scheduling subproblems may lead to suboptimality. Indeed, it is not only of interest to find optimized routes for each school, but also to find routes that can readily be connected at the scheduling stage. A technique called bi-objective routing decomposition (BiRD) was introduced by Bertsimas et al. (2019), in which not one, but several sets of routes are generated for each school (for example, a set of long routes, medium routes and short routes).

In this case, instead of one set of morning routes for each school and time $\mathcal{R}_s^{\text{AM}}(t)$, we are given n sets of morning routes $\{\mathcal{R}_s^{\text{AM},j}(t)\}_{j=1}^n$, obtained for example by choosing n values for the parameter λ in (6). We create the scheduling graph G as before, with one node per time, route pair in each of the n provided route sets. We then modify formulation (9) to include selection variables $\omega_{s,j,t}^{\text{AM}}$ (respectively $\omega_{s,j}^{\text{PM}}$), taking the value 1 if the j-th set of routes is selected in the morning (respectively afternoon) for school s at time t.

We then replace constraints (9c) and (9d) with the following,

$$\sum_{(v',v) \in \mathcal{E}_{\text{in}}(v)} f_{v',v} = \omega_{s,j,t}^{\text{AM}} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s, r \in \mathcal{R}_s^{\text{AM},j}(t), 1 \le j \le n \tag{11a}$$

$$\sum_{(v',v) \in \mathcal{E}_{\text{in}}(v)} f_{v',v} = \omega_{s,j,t}^{\text{PM}} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s, r \in \mathcal{R}_s^{\text{PM},j}(t), 1 \le j \le n \tag{11b}$$

$$\sum_{j=1}^{n} \omega_{s,j,t}^{\text{AM}} = g_{s,t} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s, 1 \le j \le n \tag{11c}$$

$$\sum_{j=1}^{n} \omega_{s,j,t}^{\text{PM}} = g_{s,t} \qquad \forall s \in \mathcal{S}, t \in \mathcal{T}_s, 1 \le j \le n, \tag{11d}$$

where constraints (11a) and (11b) enforce that selecting the j-th route set $\mathcal{R}_s^{\mathrm{AM},j}(t)$ for school s at time t means selecting each route $r \in \mathcal{R}_s^{\mathrm{AM},j}(t)$, and constraints (11c) and (11d) ensure that a route set for school s is only selected at a particular time t if school s does indeed start at time t.

2.5. Joint assignment and routing

So far, our decomposition-based approach to school operations has treated school assignment and routing as separate steps. While there is already value in simply considering these related optimization problems together to expose interplay between their various objectives, from an optimization perspective solving them separately is likely to result in suboptimal solutions. At the same time, a key advantage of the decomposition approach is its tractability. In this section, we discuss a post-improvement heuristic to jointly adjust assignment and routing, starting from a solution produced by the decomposition approach described in the previous sections. Similarly to the bus scheduling and start time selection solution approach, the heuristic involves solving an integer network flow problem at each step. The crucial distinction is that we allow "fragments" of routes (and correspondingly, the students they serve) to be assigned to another school if it improves overall cost. Because student assignments must remain consistent across the morning and afternoon, we further assume that the afternoon routes for each school are an exact reversal of the morning routes (though they may have different travel times due to varying traffic conditions). Such "mirroring" of routes is common in practice as it provides a more consistent transportation experience for students.

This post-improvement heuristic is an iterative algorithm, with each iteration consists of three steps. At the first step, we randomly select a subset of individual school routes. For each route, we then randomly select a stop, and split the route into two fragments, the first containing all the stops preceding and including the selected stop, and the second containing all the subsequent stops and the school. Note that the randomly selected stop may be the last one in the route, in which case the second fragment will be functionally empty (only contains the school). For school s, we denote by $\mathcal{F}_s^{\mathrm{cplt}}$ the set of route fragments for school s that contain the school ("complete" fragments), and by $\mathcal{F}_s^{\mathrm{icplt}}$ the set of route fragments for school s that do not contain the school s ("incomplete" fragments). Note that deleting the early part of a route only decreases occupancy and travel time, so each route fragment in $\mathcal{F}_s^{\text{cplt}}$ is actually a feasible route for school s (hence referred to as a "complete" fragment).

Given these route fragments, we then construct a directed flow graph $G' = (\mathcal{V}', \mathcal{E}')$ over the route fragments. In the interest of notational simplicity, we assume that (a) all buses of are of the same type, (b) start times t_s for each school remain fixed, (c) the set of possible start times $\mathcal{T}$ is the same for all schools. The approach can easily be modified to remove any or all of these restrictions. We also omit the "prime" modifiers on the graph, with the understanding that the graph described in this section is distinct from the one presented in the previous section. We describe the graph by specifying the vertices, edges, and edge cost function. The vertex set can be written as $\mathcal{V} = \mathcal{V}_{\text{depot}} \cup \mathcal{V}_{\text{schools}} \cup \mathcal{V}_{\text{fragments}}$, where $\mathcal{V}_{\text{depot}} = \{v_{\text{AM}}, v_{\text{PM}}\}$ contains the two nodes representing the bus depot in the morning and afternoon, and $\mathcal{V}_{\text{schools}} = \{v_{s,a}\}_{s\in\mathcal{S}, a\in\{\text{AM,PM}\}}$ contains two nodes per school (morning and afternoon). The remaining vertices are defined as $\mathcal{V}_{\text{fragments}} = \cup_{s\in\mathcal{S}, a\in\{\text{AM,PM}\}} \mathcal{V}_{\text{fragments}}^{s,a}$, with

$$\mathcal{V}_{\text{fragments}}^{s,a} = \{(f, a) \mid f \in \mathcal{F}_s^{\text{cplt}}\} \cup \{(t, f, a, \omega) \mid t \in \mathcal{T}, f \in \mathcal{F}_s^{\text{icplt}}, \omega \in \{0, 1\}\}.$$

In other words, we create one node for each complete fragment in the morning and afternoon. For each incomplete fragment, we create two nodes (indexed by ω) in the morning and afternoon, for each possible start time t. Nodes with $\omega = 1$ (resp. 0) are called *terminal* (resp. nonterminal). Intuitively, terminal nodes can only have incoming edges from the depot (new bus), while nonterminal nodes can only have incoming edges from other schools (re-used bus). For ease of notation, we denote by $\mathcal{V}_{s,a}^{\text{cplt}}$ the set of nodes associated with complete fragments for school s, and $\mathcal{V}_{s,a,t,\omega}^{\text{icplt}}$ the set of terminal or nonterminal nodes associated with incomplete fragments for school s at time t. Let f_v denote the fragment associated with node v.

With these vertices, we can then define the edge set for the morning nodes (with the understanding that the afternoon part of the graph is a mirror image). The structure of the network is similar to the one presented in the previous section, with the key difference that nodes associated with incomplete route fragments can be connected to nodes associated with any school, not simply the fragment's original school. This mechanism allows route recombination across schools, which modifies the underlying student-to-school assignment to improve the routing objective. In turn, it requires the creation of nodes associated with incomplete route fragments, as well as nodes associated with schools (since an incomplete route fragment could connect directly to a school to create a new route).

In the previous section, edges could only be created between pairs of routes if they were time-compatible. The addition of school nodes and incomplete fragments means time-compatibility means time-compatibility must be evaluated differently across different node pairs. As before, we write that a school $s \in \mathcal{S}$ is time-compatible with a complete route fragment $f \in \mathcal{F}_{s'}^{\text{cplt}}$, denoted as $s \bowtie f$, if and only if $t_s + t_{s,h_1(f)}^{\text{drive}} + t_f^{\text{service}} + t_{s'}^{\text{school}} \leq t_{s'}$. Additionally, we write that an incomplete route fragment $f \in \mathcal{F}_s^{\text{icplt}}$ is compatible with a complete route fragment $f' \in \mathcal{F}_{s'}^{\text{cplt}}$, denoted as $f \bowtie f'$ in a slight abuse of notation, if and only if (i) the total number of students carried on both fragments does not exceed the bus capacity, and (ii) the total length of the newly formed route $t_f^{\text{service}} + t_{h_{|f|}(f),h_1(f')}^{\text{drive}} + t_{f'}^{\text{service}}$ does not exceed the maximum route time T. We then write that an incomplete route fragment $f \in \mathcal{F}_s^{\text{icplt}}$ is time-compatible with a school s', denoted by $f \bowtie s'$, if and only if the total length of the newly created route $t_f^{\text{service}} + t_{h_{|f|}(f),s'}^{\text{drive}}$ does not exceed the maximum route length T. Checking time-compatibility from schools to incomplete fragments requires information from more than two nodes in the graph, so this constraint cannot be imposed directly on the edges.

We can therefore write the (morning) edge set as

$$\mathcal{E}_{\text{AM}} = \{(v_{\text{AM}}, v) \mid v \in \mathcal{V}_{s,\text{AM}}^{\text{cplt}} \cup \mathcal{V}_{s,\text{AM},t,1}^{\text{icplt}}, s \in \mathcal{S}, t \in \mathcal{T}\} \tag{12a}$$

$$\cup \{(v, v_{\text{PM}}) \mid v \in \mathcal{V}_{\text{schools}}\} \tag{12b}$$

$$\cup \{(v_{\text{AM}}, v_{\text{PM}}), (v_{\text{PM}}, v_{\text{PM}})\} \tag{12c}$$

$$\cup \{(v, v_{s,\text{AM}}) \mid v \in \mathcal{V}_{s,\text{AM}}^{\text{cplt}}, s \in \mathcal{S}\} \tag{12d}$$

$$\cup \{(v, v_{s',\text{AM}}) \mid v \in \mathcal{V}_{s,\text{AM},t,\omega}^{\text{icplt}}, \omega \in \{0,1\}, s \in \mathcal{S}, s' \in \mathcal{S}, t = t_{s'}, f_v \bowtie s\} \tag{12e}$$

$$\cup \{(v, v') \mid v \in \mathcal{V}_{s,\text{AM},t,\omega}^{\text{icplt}}, v' \in \mathcal{V}_{s',\text{AM}}^{\text{cplt}}, \omega \in \{0,1\}, s \in \mathcal{S}, s' \in \mathcal{S}, t = t_{s'}, f_v \bowtie f_{v'}\} \tag{12f}$$

$$\cup \{(v_{s,\text{AM}}, v') \mid v' \in \mathcal{V}_{s',\text{AM}}^{\text{cplt}}, s \in \mathcal{S}, s' \in \mathcal{S}, s \bowtie f_{v'}\} \tag{12g}$$

$$\cup \{(v_{s,\text{AM}}, v') \mid v' \in \mathcal{V}_{s',\text{AM},t,0}^{\text{icplt}}, s \in \mathcal{S}, s' \in \mathcal{S}, t_s < t\}, \tag{12h}$$

where we include an edge (12a) from the depot to each complete or terminal fragment; an edge (12b) from each school to the depot; edges (12c) between the morning and afternoon states of the depot; an edge (12d) from each complete fragment to the corresponding school; an edge (12e) from each incomplete fragment to each time-compatible school; an edge (12f) from each incomplete fragment to each time-compatible complete fragment; an edge (12g) from each school to each time-compatible complete fragment; and finally an edge (12e) from each school to each incomplete fragment. The afternoon edges $\mathcal{E}_{\text{PM}}$ form a mirror image of the morning edges.

As in the previous section, we can then formulate the problem of reconnecting the route fragments into a feasible bus schedule as a network flow problem, with key variable $u_{v,v'}$ indicating the amount of flow from node v to node v':

$$\min \quad \sum_{(v,v')\in\mathcal{E}} C'_{v,v'} u_{v,v'} \tag{13a}$$

$$\text{s.t.} \quad \sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} u_{v',v} = \sum_{(v,v')\in\mathcal{E}_{\text{out}}(v)} u_{v,v'} \qquad \forall v \in \mathcal{V} \tag{13b}$$

$$\sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} u_{v',v} = 1 \qquad \forall v \in \mathcal{V}^{\text{cplt}} \tag{13c}$$

$$\sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} u_{v',v} = m_{t,f,a,\omega} \qquad \forall v := (t,f,a,\omega) \in \mathcal{V}^{\text{icplt}} \tag{13d}$$

$$\sum_{\substack{t\in\mathcal{T}\\ \omega\in\{0,1\}}} m_{t,f,a,\omega} = 1 \qquad \forall f \in \mathcal{F}^{\text{icplt}}, a \in \{\text{AM}, \text{PM}\} \tag{13e}$$

$$\sum_{\substack{v\in\mathcal{V}_{f,\text{AM}}\\ v'\in\mathcal{V}_{f',\text{AM}}}} u_{v,v'} = \sum_{\substack{w\in\mathcal{V}_{f,\text{PM}}\\ w'\in\mathcal{V}_{f',\text{PM}}}} u_{w',w} \qquad \forall f \in \mathcal{F}^{\text{icplt}}, f' \in \mathcal{F}^{\text{cplt}} \tag{13f}$$

$$\sum_{(v',v)\in\mathcal{E}_{\text{in}}(v)} \tau_{v',v} u_{v',v} + \sum_{(v,v')\in\mathcal{E}_{\text{out}}(v)} \tau_{v,v'} u_{v,v'} \le t - t_{v'} \qquad \forall v := (t,f,a,0) \in \mathcal{V}^{\text{icplt}} \tag{13g}$$

$$m_{t,f,a,\omega} \in \{0,1\} \qquad \forall v := (t,f,a,\omega) \in \mathcal{V}^{\text{icplt}} \tag{13h}$$

$$u_{v,v'} \in \mathbb{Z}_{\ge 0} \qquad \forall (v,v') \in \mathcal{E}, \tag{13i}$$

where $\mathcal{V}^{\text{cplt}} = \cup_{s\in\mathcal{S},a\in\{\text{AM},\text{PM}\}} \mathcal{V}^{\text{cplt}}_{s,a}$ designates the set of nodes associated with a complete fragment, and analogously $\mathcal{V}^{\text{icplt}} = \cup_{s\in\mathcal{S},a\in\{\text{AM},\text{PM}\},t\in\mathcal{T},\omega\in\{0,1\}} \mathcal{V}^{\text{icplt}}_{s,a,t,\omega}$ designates the set of nodes associated with an incomplete fragment. Additionally, $\mathcal{V}_{f,a}$ designates the set of vertices associated with fragment f at time of day a. Finally, $\mathcal{F}^{\text{icplt}} = \cup_{s\in\mathcal{S}} \mathcal{F}^{\text{icplt}}_s$ designates the set of all incomplete fragments, and similarly $\mathcal{F}^{\text{cplt}} = \cup_{s\in\mathcal{S}} \mathcal{F}^{\text{cplt}}_s$ the set of all complete fragments. For a school node $v = v_s$, we denote by t_v the start time t_s of the associated school.

In addition to the flow variables $u_{v,v'}$, the binary variables $m_{t,f,a,\omega}$ ensure exactly one node associated with an incomplete fragment is selected (constraint (13e)). Constraints (13b)–(13d) impose flow conservation and make sure nodes associated with a fragment are visited (analogous to constraints (9b)–(9d)).

Formulation (13) also introduces two significantly new constraints: first, (13f) ensures that if two fragments are combined into a route in the morning, the same two fragments must also be combined into a route in the afternoon. Combining a complete fragment and an incomplete fragment may involve changing the school a student attends, and a student cannot be dropped off at one school in the morning and picked up at another in the afternoon. Second, constraint (13g) imposes time-feasibility for bus itineraries that visit a school and incomplete fragment in succession. In the morning, we can select

$$\tau_{v,v'} = \begin{cases} t^{\text{drive}}_{s,h_1(f)} + t^{\text{service}}_{f}, & \text{if } v := v_s \in \mathcal{V}_{\text{schools}},\ v' := (t,f,\text{AM},0) \in \mathcal{V}^{\text{icplt}}, \\ t^{\text{drive}}_{h_{|f|}(f),h_1(f')} + t^{\text{service}}_{f'} + t^{\text{school}}_{s'}, & \text{if } v := (t,f,\text{AM},0) \in \mathcal{V}^{icplt}, v' := (f',\text{AM}) \in \mathcal{V}^{\text{cplt}}_{s'}, \\ t^{\text{drive}}_{h_{|f|}(f),s'} + t^{\text{school}}_{s'}, & \text{if } v := (t,f,\text{AM},0) \in \mathcal{V}^{icplt}, v' := v_{s'} \in \mathcal{V}_{\text{schools}}, \\ 0, & \text{otherwise,} \end{cases}$$

to ensure that any constructed route involving an incomplete fragment preceded by a school node with a start time t_1 and followed by another school node or a complete fragment with start time t_2 will satisfy (t_1,t_2)-compatibility. It is because of this constraint that we introduced the somewhat unintuitive duplication of incomplete fragments into "terminal" and "nonterminal" nodes.

As with the previous network flow formulation, an advantage of this approach is the ability to model multiple objectives. We describe examples of potential cost functions in Appendix B.

3. Time windows and optimality gaps

The formulations presented in the previous section together form an optimization pipeline, which allows a school district to explore new operational policies at the intersection of assignment, transportation, and start times. We note that some of the assumptions we make are more restrictive than those considered in the literature. In particular, some school bus routing works consider a slightly more general setting, where routes need not arrive at school at exactly the same time, but rather within a certain time window. We choose instead to consider that all routes must arrive at the same time. In this section, we identify a special case in which the two approaches are equivalent.

3.1. Bus scheduling with time windows

The question of time windows has sparked some debate in the school bus routing literature. Some approaches, such as those of Shafahi et al. (2018) and Bertsimas et al. (2019), favor the assumption that all school bus routes for a particular school must arrive at the same fixed "anchor time". The advantage of this approach is that it is easy to reserve a time buffer (of, say, 10 minutes) between the routes' anchor time and the school's start time, as a simple, easily implementable way to account for delays caused by day-to-day variance in traffic. Other authors, including Swersey and Ballard (1984) and Zeng et al. (2022) favor time windows, citing an increased potential for connecting routes arising from this greater flexibility. In practice, school districts may use either framework.

Consider a simplified version of the bus scheduling and start time selection problem described in Section 2, in which we ignore afternoon routing, assume the route set $\mathcal{R}_s$ for each school s does not depend on the school's selected start time $t_s \in \mathcal{T}_s$, and reduce the bus fleet to a single type of vehicle. For ease of notation, we refer to the collection of all routes as $\mathcal{R} = \cup_{s\in\mathcal{S}}\mathcal{R}_s$. As previously, we consider three potential objectives: the total number of buses (denoted by z_1), the total driving distance (denoted by z_2), and the total driving time (denoted by z_3). Our general objective z is a convex combination of these three possibilities.

Given the time window length Δt_s for each school s, implying that routes are allowed to arrive at school within the interval $[t_s - \Delta t_s, t_s]$ for a start time of t_s, we can formulate the problem as follows:

$$\min_{\boldsymbol{t}\in\prod_{s\in\mathcal{S}}\mathcal{T}_s} \quad \min_{\substack{\boldsymbol{\tau}\in\mathbb{R}^{|\mathcal{R}|}\\ \boldsymbol{w}\in\{0,1\}^{|\mathcal{R}|^2}}} \quad z := \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 \tag{14a}$$

$$\text{s.t.} \quad \sum_{r'} w_{r',r} \le 1 \qquad \forall r \in \mathcal{R} \tag{14b}$$

$$\sum_{r'} w_{r,r'} \le 1 \qquad \forall r \in \mathcal{R} \tag{14c}$$

$$\tau_{r'} \ge \tau_r + T(r,r') - M(1 - w_{r,r'}) \qquad \forall r, r' \in \mathcal{R} \tag{14d}$$

$$t_s - \Delta t_s \le \tau_r \le t_s \qquad \forall s \in \mathcal{S}, r \in \mathcal{R}_s. \tag{14e}$$

The discrete decision variable t_s represents the selected start time for school $s \in \mathcal{S}$, while the continuous decision variable τ_r designates the arrival time of route $r \in \mathcal{R}$. The binary decision variable $w_{r,r'}$ takes the value 1 if route r' is served immediately after route r by the same bus, and 0 otherwise. Constraints (14b) and constraints (14c) enforce that a route is immediately preceded or followed by at most one other route in a bus schedule. We define $T(r,r')$ to be the total time necessary to travel from the end of route r to the beginning of route r', and then serve the entire route r', meaning constraint (14d) ensures that route r' can only immediately follow route r if their respective arrival times are far enough apart (time feasibility). Finally, constraint (14e) ensures that the arrival time for each route occurs within the specified school time window.

We now write the objectives of interest as a function of our decision variables. Abusing notation slightly, we define $T(0,r)$ as the time necessary to travel from the depot to the start of route r, then serve the entirety of route r, while $T(r,0)$ indicates the time necessary to travel from the end of route r to the depot. Analogously, we define the distances $D(r,r')$, $D(0,r)$ and $D(r,0)$. As shorthand, we write $\bar{T}(r) = T(0,r)$, $\underline{T}(r) = T(r,0)$, $\bar{D}(r) = D(0,r)$, and $\underline{D}(r) = D(r,0)$. We can then write the following:

$$z_1 := \sum_{r\in\mathcal{R}} \left(1 - \sum_{r'\in\mathcal{R}} w_{r,r'}\right) = |\mathcal{R}| - \sum_{r,r'\in\mathcal{R}} w_{r,r'}, \tag{15a}$$

$$z_2 := \sum_{r\in\mathcal{R}} \bar{D}(r) + \underline{D}(r) + \sum_{r,r'} w_{r,r'} \left(D(r,r') - \bar{D}(r') - \underline{D}(r)\right), \tag{15b}$$

$$z_3 := \sum_{r\in\mathcal{R}} \bar{T}(r) + \underline{T}(r) + \sum_{r,r'} w_{r,r'} \left(\tau_{r'} - \tau_r - \bar{T}(r') - \underline{T}(r)\right). \tag{15c}$$

Note that these expressions offer a clear interpretation: for each pair of routes r and r' served in succession, the total number of buses decreases by one. Similarly, the total travel distance changes by $D(r,r') - \bar{D}(r') - \underline{D}(r)$ (which is negative if distances follow the triangle inequality), by eliminating a visit to the depot between routes r and r'. Finally, the total bus service time changes by $\tau_{r'} - \tau_r - \bar{T}(r') - \underline{T}(r)$, replacing driving time to and from the depot with the difference between the arrival times of the consecutive routes r and r'.

Notice that z_3 is nonlinear, as it involves a product of decision variables. This difficulty can be overcome by introducing (many!) additional variables to model the product between $w_{r,r'}$ and τ_r, but this remains an unwelcome feature of the bus scheduling problem with time windows. We note that using formulation (14), a particular bus scheduling and start time selection problem is uniquely determined by the tuple $(\{\mathcal{R}_s, \mathcal{T}_s, \Delta t_s\}_{s\in\mathcal{S}}, T(\cdot,\cdot), D(\cdot,\cdot), \boldsymbol{\alpha})$.

Leaving aside the $\mathcal{O}(|\mathcal{T}_s|)$ variables required to model the discrete start time t_s (e.g., as $t_s = \sum_{t\in\mathcal{T}_s} tg_{s,t}$, with $g_{s,t}\in\{0,1\}$), formulation (14) comprises $\mathcal{O}(|\mathcal{R}|^2)$ discrete variables and $\mathcal{O}(|\mathcal{R}|)$ continuous variables, as well as $\mathcal{O}(|\mathcal{R}|^2)$ constraints. Fixing route arrival times, i.e., imposing $\Delta t_s = 0$ for each school $s\in\mathcal{S}$, eliminates all continuous variables in the inner problem, as well as constraints of type (14e). It is then easy to write constraint (14d) as:

$$t_{s'}w_{r,r'} \geq (t_s + T(r,r'))w_{r,r'} \qquad \forall s,s'\in\mathcal{S}, r\in\mathcal{R}_s, r'\in\mathcal{R}_{s'}, \tag{16a}$$

or equivalently

$$\sum_{t'\in\mathcal{T}_{s'}} t'g_{s',t'}w_{r,r'} \geq \sum_{t\in\mathcal{T}_s} tg_{s,t}w_{r,r'} + T(r,r')w_{r,r'} \qquad \forall s,s'\in\mathcal{S}, r\in\mathcal{R}_s, r'\in\mathcal{R}_{s'}, \tag{16b}$$

$$\sum_{t\in\mathcal{T}_s} g_{s,t} = 1 \qquad \forall s\in\mathcal{S}, \tag{16c}$$

which can be expressed as a pure binary optimization problem if we introduce new variables to represent the product of $\boldsymbol{w}$ and $\boldsymbol{g}$. The resulting problem has $\mathcal{O}\left((\sum_{s\in\mathcal{S}}|\mathcal{R}_s||\mathcal{T}_s|)^2\right)$ variables, and $\mathcal{O}\left(|\mathcal{R}|^2|\mathcal{T}|^2\right)$ variables if $\mathcal{T}_s = \mathcal{T}\;\forall s\in\mathcal{S}$ (same possible times for all schools). This is a larger number of variables, but a much stronger formulation due to the elimination of the big-M constraints. Furthermore, the objective z_3 can now be written as:

$$z_3 = \sum_{r\in\mathcal{R}} \bar{T}(r) + \underline{T}(r) + \sum_{r,r'} w_{r,r'}\left(\sum_{t\in\mathcal{T}_{s'}} tg_{s',t} - \sum_{t\in\mathcal{T}_s} tg_{s,t} - \bar{T}(r') - \underline{T}(r)\right),$$

where s is the school assocaited with route r and s' the school associated with route r'. Note that after introducing new product variables of $\boldsymbol{w}$ and $\boldsymbol{g}$, z_3 becomes linear. Indeed, formulation (9) was obtained via this method, in the more general case of a heterogeneous fleet, and after pruning pairs of routes that cannot possibly be served in succession.

Thus, by eliminating time windows, we can go from a weak, nonlinear mixed-integer formulation to a strong, linear pure-integer formulation. Furthermore, it is clear that any schedule computed assuming with $\Delta t_s = 0$ will remain feasible (but perhaps suboptimal) for larger values of Δt_s. We now consider a special case of practical interest, in which this suboptimality can be eliminated.

3.2. Two-tier problems

In practice, many school district schedule start times in *tiers*, meaning that schools can only start at a few possible times. For example, Boston Public Schools operates a three-tier schedule during the normal school year: start times are either 7:30, 8:30 or 9:30. In the summer, the ESY program operates a two-tier schedule, with schools starting at either 8:00 or 9:30 in the summer of 2019. Two-tier systems (and their perhaps less common three-tier cousins) are valued for their simplicity, particularly in the case when school bus routing is a manual or semi-manual process.

Definition 1. Consider the problem of scheduling school buses and selecting start times for a set of schools $\mathcal{S}$. The problem is *two-tier* if it verifies the following conditions: (1) only two start times are allowed for each school, i.e. $\mathcal{T}_s = \{t_1, t_2\}, \forall s \in \mathcal{S}$ ($t_1 < t_2$ WLOG); (2) all bus routes for school s must arrive within the time window $[t_s - \Delta t_s, t_s]$; (3) the early and late time windows do not overlap, i.e., $t_1 < t_2 - \Delta t$; (4) the largest time window is shorter than the shortest route, in other words, no bus can serve two consecutive routes for the same tier.

Two-tier scheduling problems are highly structured—a key reason they are often used in practice. We next prove that given a two-tier scheduling problem with time windows, we can always reduce it to an equivalent two-tier scheduling problem without time windows.

Proposition 1. *Consider a two-tier bus scheduling and start time selection problem* $(P1) \coloneqq (\{\mathcal{R}_s, \mathcal{T}_s \coloneqq \{t_1, t_2\}, \{\Delta t_s\}_{s\in\mathcal{S}}, T(\cdot,\cdot), D(\cdot,\cdot), \boldsymbol{\alpha})$. *We can define a related problem* $(P2) \coloneqq (\{\mathcal{R}_s, \{t_1 - \Delta t_s, t_2\}, 0\}_{s\in\mathcal{S}}, T(\cdot,\cdot), D(\cdot,\cdot), \boldsymbol{\alpha})$, *with the modified objective* $\hat{z} = \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 \hat{z}_3$, *where*

$$\hat{z}_3 = \sum_{r\in\mathcal{R}} \bar{T}(r) + \underline{T}(r) + \sum_{r,r'\in\mathcal{R}} w_{r,r'} \left(\max[t_2 - \Delta t_{s'} - t_1, T(r,r')] - \bar{T}(r') - \underline{T}(r)\right),$$

such that $(P1)$ *and* $(P2)$ *have the same optimal objective, and there is a polynomial-time algorithm to compute an optimal solution of* $(P1)$ *given an optimal solution of* $(P2)$.

Proposition 1 means that in a two-tier scheduling problem, we can fix route arrival times at the earliest part of the first tier and the latest part of the second tier, and still find the optimal solution of the original scheduling problem with time windows. The intuition behind this result is that spreading out routes further than necessary does not affect the total number of buses, or the total distance traveled, and only affects the total driving time in a way that can be computed in polynomial time (by solving a linear program). We defer all proofs to Appendix C. Note that while the result assumes that all buses are of the same size, the proof does not require it. Indeed, it is possible to show the same result for an arbitrary fleet composition, though at the cost of unwieldy notation.

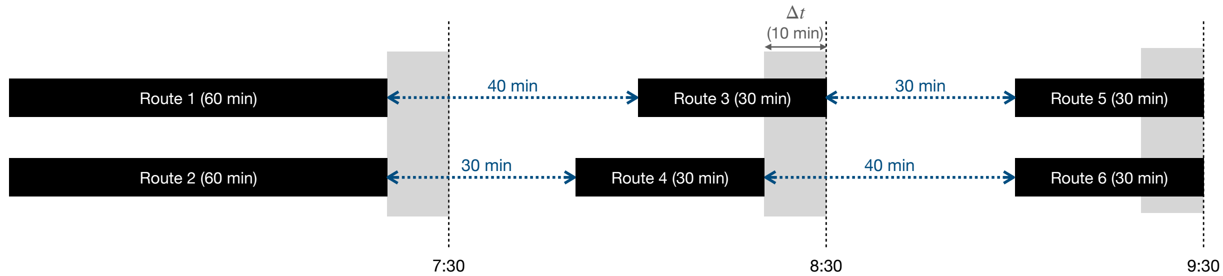
3.3. Three-tier problems

The ability to do away with time windows entirely is an unexpectedly strong result which bodes well for a school bus scheduling approach without time windows, such as the one presented in the previous section. However, it is natural to wonder whether the same reasoning holds in a setting with more tiers - for example, larger school districts may have three or even four tiers. In the case of a three-tier system, we provide a simple counter-example to show Proposition 1 does not hold, then describe a worst-case upper bound on the cost of removing time windows. We first define the notion of a three-tier system.

Definition 2. Consider the problem of scheduling school buses and selecting start times for a set of schools $\mathcal{S}$. The problem is *three-tier* if it verifies the following conditions: (1) only three start times are allowed for each school, i.e. $\mathcal{T}_s = \{t_1, t_2, t_3\}, \forall s \in \mathcal{S}$ ($t_1 < t_2 < t_3$ WLOG); (2) all bus routes for school s must arrive within the time window $[t_s - \Delta t_s, t_s]$; (3) the time windows for each tier do not overlap, i.e., $t_1 < t_2 - \Delta t < t_2 < t_3 - \Delta t$; (4) the largest time window is shorter than the shortest route, in other words, no bus can serve two consecutive routes for the same tier; (5) any pair of routes where the first arrives at a school within the time window $[t_1, t_1 - \Delta t]$ and the second arrives at a school within the time window $[t_3, t_3 - \Delta t]$ can be served by the same bus (skipping a tier implies time feasibility).

We note that a three-tier system is defined analogouly to a two-tier system, with the additional condition that routes in the first and third tier can always be served consecutively. This is a practical assumption: a typical three tier-system may have separations of up to an hour between tiers, so there is considerable time for buses to connect the earliest and latest routes. Indeed, the main value of a three-tier system is the ability to re-use buses more than twice.

Figure 2 Three-tier counter-example for Proposition 1



Note. This diagram presents the optimal solution to a bus scheduling and start time problem with three schools, each with two routes. The transition times and route lengths are also indicated. Collapsing time windows imposes that routes 3 and 4 must arrive at school 2 at the same time, which will make one of the two three-route itineraries infeasible.

Remark 1. In a three-tier problem, collapsing school arrival time windows to a single point may deteriorate the objective function. Consider the simple setting where $\alpha_1 = 1$ and $\alpha_2 = \alpha_3 = 0$ (the only objective is to minimize the number of buses). Consider a problem instance with 3 schools, possible start times of 7:30, 8:30, and 9:30, and two routes per school. We describe the optimal solution of the bus scheduling and start time problem with time windows in Figure 2, indicating both the length of each route and the length of the optimal transitions. It is straightforward to see that collapsing the time window for the middle school forces routes 3 and 4 to start at the same time, which will always make at least one three-route itinerary (and potentially both) infeasible, requiring the addition of at least one additional bus and deteriorating the optimal objective (assuming the lengths of route 1 and route 2, along with long transition times from the first school to the routes of the third school, make a start time rotation between schools impossible).

Therefore, in a three-tier problem, we cannot expect a result as strong as Proposition 1 to hold. However, we can still bound the objective deterioration from collapsing time windows.

Proposition 2. *Consider a three-tier two-tier bus scheduling and start time selection problem* $(P) \coloneqq (\{\mathcal{R}_s, \mathcal{T}_s \coloneqq \{t_1, t_2, t_3\}, \{\Delta t\}_{s \in \mathcal{S}}, T(\cdot,\cdot), D(\cdot,\cdot), \boldsymbol{\alpha} = (1,0,0))$, *with optimal objective* z^*. *Denote the collapsed version of the problem* $(P') \coloneqq (\{\mathcal{R}_s, \mathcal{T}_s \coloneqq \{t_1 - \Delta t, t_2, t_3\}, \{\Delta t\}_{s \in \mathcal{S}}, T(\cdot,\cdot), D(\cdot,\cdot), \boldsymbol{\alpha} = (1,0,0))$, *with optimal objective* z'^*. *Then*

$$z'^* \leq z^* + N^*_{2,3} - N^*_1 \leq 2z^*,$$

where $N^*_{2,3}$ is the optimal (under z^*) number of buses serving consecutive routes for schools in the second and third tiers, and N^*_1 is the optimal number of buses only serving a route for a first-tier school.

The result in Proposition 2 is considerably weaker than Proposition 1. However, it holds in general, with minimal assumptions on the underlying school bus routes. Additional computation can yield tighter bounds. For example, the term $N^*_{2,3} - N^*_1$ can be tightened to $\min(N^*_{2,3} - N^*_1, N^*_{1,2} - N^*_3)$ (second term defined analogously to the first) by solving the collapsed problem twice, once with a second-tier bell time of t_2 and once with a second-tier bell time of $t_2 - \Delta t$. In practical settings, the bound may be significantly improved. For example, collapsing the second-tier bell times may not preclude the connection of a second-tier route r with a third-tier route r', for example if $t_3 - t_2 \geq T(r, r')$. And changing school start times may allow further bus re-use.

The results presented in this section provide motivation for an approach without time windows, such as the one presented in this paper. Indeed, as we saw in Section 2, removing time windows yields a formulation with a network flow structure, both tractable and flexible enough to include heterogeneous fleet modeling as well as multiple time-dependent routing options for each school.

4. Numerical experiments

The post-improvement heuristic for joint assignment, routing and scheduling presented in Section 2.5 is a novel contribution of this work, and in this section we evaluate its performance numerically.

We consider a realistic synthetic setting, using public data from the Boston Public Schools Transportation Challenge (Boston Public Schools 2017). The dataset includes 22,420 students, assigned to 134 schools. For the purposes of our analysis, we ignore the assignment of students to schools, and assume any student can be assigned to any school. We also randomly subsample two smaller datasets: a “Small” dataset with 5,000 students and 20 schools, and a “Medium” dataset with 10,000 students and 50 schools. The full dataset is referred to as the “Large” dataset. We consider a fleet with a single bus type with capacity 70. We define travel times using Euclidean distance and a fixed speed of 8 miles per hour.

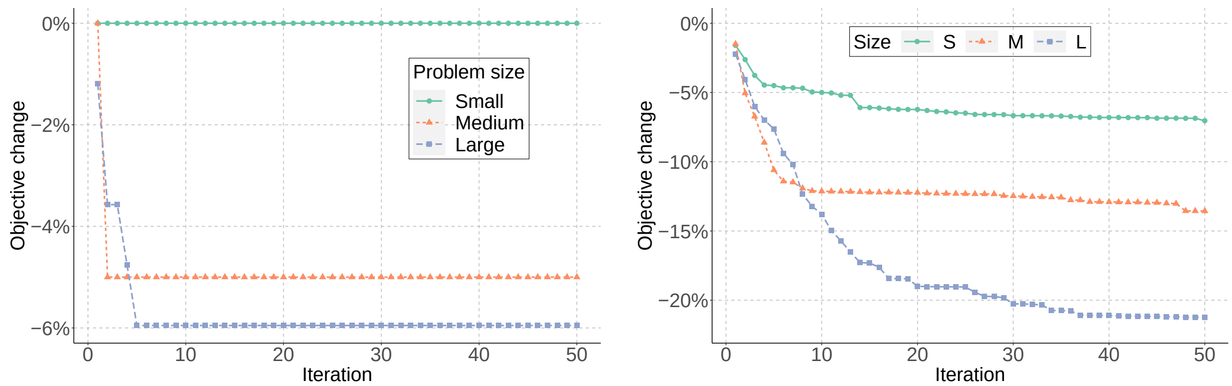
We assume each school is surrounded by a "walk zone" of radius 1 mile. Students inside the walk zone of their assigned school do not receive transportation service, except for students designated as "door-to-door", who receive transportation regardless of their distance to school. We assume bus stops cannot accommodate more than 30 students, and that visiting a bus stop requires 60s, plus 5s per student. As required by joint assignment-routing, we "mirror" morning and afternoon routes. For each of the three problem instance, we consider two potential objectives: the first is simply the total number of buses; the second is a weighted combination of the number of buses, total travel time, and total travel distance, with weights determined using parameters from the ESY case study in the following section.

Figure 3 shows the performance of the post-improvement heuristic for the three problem instances and two choices of objective. When the objective is purely the number of buses (as in Figure 3a), the algorithm struggles to make improvements, with no improvement recorded for the small instance, and about 5% improvement for the medium and large instance. When the objective is a weighted combination of buses, travel time, and distance, the heuristic provides considerably more value, with improvements of about 7% on the smaller instance and 20% on the larger instance. On the larger instance, the algorithm is able to efficiently recombine routes from different schools to visit a single school, reducing the number of schools that require transporting students (there could still be students in the walk zone of these schools).

We conjecture that the worse performance on the number of buses objective can be explained by the discrete nature of the objective, which leads to many solutions with exactly the same objective, which makes it harder for the algorithm to find locally-improving solutions. Figure 4a supports this theory, as it depicts the number of buses at each iteration, in the cases when the objective is being optimized is *not* the number of buses but instead the weighted objective. We find that including secondary objectives considerably improves the performance on the primary objective, not only in relative but also in absolute terms. For all three instances, the absolute number of buses after 50 iterations is lower when optimizing the weighted objective than when optimizing the number of buses directly.

We also seek to understand the impact of the neighborhood size (determined by the number of single-routes "broken up" by deleting a single edge between stops) in our local improvement heuristic. In Figure 4b, we study the effect of neighborhood size on runtime

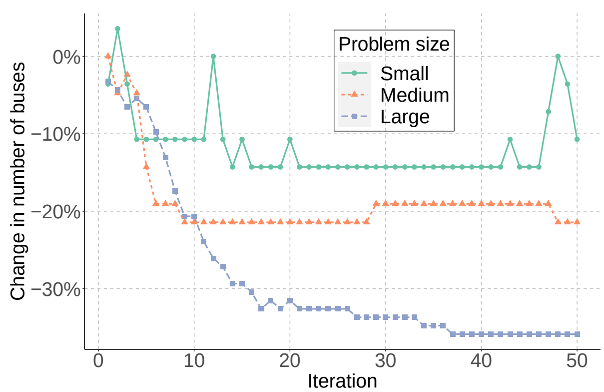
Figure 3 Performance of post-improvement heuristic



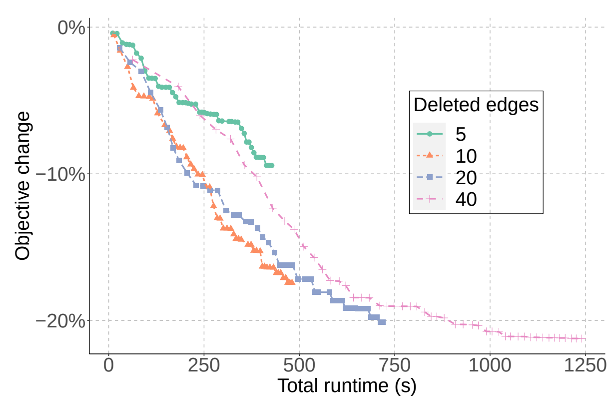
(a) Minimizing number of buses (b) Minimizing weighted objective

Note. Performance is evaluated over 50 iterations; each iteration breaks up 40 single-school routes and optimally re-assembles the fragments. The heuristic performs much better on larger problems; when minimizing the number of buses, it is more prone to getting "stuck" in local minima than when minimizing the weighted objective.

Figure 4 Secondary objectives and convergence



(a) Bus count when minimizing weighted objective



(b) Convergence and neighborhood size

Note. The left figure plots the number of buses as a function of the iteration number, when the objective being optimized is a weighted sum of number of buses, travel time, and distance. We observe significant improvement compared to Figure 3a, which optimizes the number of buses directly, suggesting the local improvement heuristic works better on a less discrete objective. On the right, we describe solution quality on the large problem instance over 50 iterations, for different numbers of single-school routes broken up at each iteration ("deleted edges" denoting the neighborhood size). The X axis depicts the cumulative runtime of the heuristic: smaller neighborhoods make less progress per iteration, while larger neighborhoods require more time per step.

and solution quality on the large instance with the weighted objective. We observe a clear tradeoff between smaller neighborhoods, which allow for more iterations with less progress per iteration, and larger neighborhoods, which make more progress per iteration but require solving a larger network flow problem at each step. Overall, the results suggest that the algorithm performs well, particularly on larger instances where there is more opportunity for optimization.

5. Case study

We now evaluate our optimization models on data from the Extended School Year (ESY) program at Boston Public Schools.

5.1. Problem setup

In the summer of 2019, the ESY program at Boston Public Schools enrolled 3,758 students across 15 different programs. These programs were held at 11 different schools: one high school, one middle school, seven elementary schools, and two all-grade sites. These school buildings vary in size, with 5 classrooms in the smallest one, and 93 in the biggest.

In our study, staffing a school for five weeks has a fixed administrative staffing cost of \$21,000 (including a site coordinator, an administrative assistant, and a nurse), plus a cost per classroom. Classrooms for the ABA (Applied Behavior Analysis) program are staffed by one teacher, incurring a cost of \$7,000, while all other classrooms are staffed by one teacher and one paraprofessional educator, incurring a cost of \$10,500. At the assignment step, we define the student-to-school distance d_{is} as the Euclidean distance between student i and school s. We refer to the costs incurred by the district to operate and staff schools and classrooms as *educational costs* or *classroom costs*.

Students are eligible for bus transportation if their Individualized Education Program (IEP) requires it or if they live further than a specified distance (as the crow flies) from school (1 mile for elementary school students, 1.5 miles for middle school students, and 2 miles for high school students). The majority of students are eligible for transportation, and most eligible students require door-to-door service, meaning they must be picked up at home and cannot be clustered into bus stops. The remaining students must be assigned a stop from one of 1,800 pre-approved locations in the city, as long as they need not walk more than 600m (approximately a third of a mile). Students who do not live close to any bus stop are treated as door-to-door students.

The Boston Public Schools fleet comports four types of buses: in our model, a full bus can hold up to 65 students, a half bus can hold up to 30, a mini bus can hold up to 10, and a wheelchair bus can hold up to 12 students, including up to three students in wheelchairs. Following Boston Public Schools' transportation assumptions, we assume that visiting a bus stop requires 60s, plus 5s per student. Each student in a wheelchair requires another 240s (4 min) to pick up or drop off, giving time to operate the specialized wheelchair lift and securing the wheelchair inside the bus. We note that students in wheelchairs can only be picked up by wheelchair buses.

We develop a more realistic travel time model than one solely based on latitude and longitude. Using data from OpenStreetMap, we create a network of the city of Boston where each node represents an intersection, and each edge represents a road. The complete network comprises 188,002 nodes and 453,703 edges. We then project each bus stop and school to the nearest road in our network, and compute the driving time between any two locations by assuming buses travel along the shortest path at a constant speed of 10 miles per hour — a conservative estimate of the average school bus velocity.

Finally, we describe the structure of our transportation costs. In contrast to many school bus applications, the objective is not simply to minimize the number of buses. Indeed, because the ESY program is held over a short period of time in the summer, the fixed cost of owning and maintaining a bus is considered sunk, as is the total cost of driver benefits such as health insurance. Instead, the cost is driven by driver hours and distance driven. The hourly salary of a driver is \$27, and each mile driven incurs a fuel and parts cost of \$0.84. There remains a fixed cost associated with each bus, because drivers must conduct a 15-minute vehicle inspection at the beginning and end of each shift in both the morning and afternoon. The ESY program runs for 25 days (5 weeks) so we multiply the computed daily costs by 25 to obtain total transportation costs for the whole program. We refer to the costs incurred by the district to operate school buses as *transportation costs* or *bus costs*.

All computations are realized in Julia (Bezanson et al. 2017), with optimization problems formulated using JuMP (Dunning et al. 2017) and solved using Gurobi (Gurobi Optimization, Inc. 2016). Computational experiments were executed in parallel on a computing server, with a single-core machine (8GB RAM) tasked with solving the assignment, routing and scheduling problems for each parameter configuration. Computations in this section do not make use of the post-improvement heuristic described in Section 2.5.

5.2. ESY-specific modeling

Modeling attendance. A particularity of the ESY program is that it is not required. Students' IEP gives them access to the program, but families may decide that this time is better spent in other activities. Students may also attend sporadically for a variety of reasons. The district must therefore make plans keeping in mind that students are not all equally likely to attend the program.

We adjust the model in the following way. For each student i, we model i's attendance on any given day using a Bernoulli random variable A_i. Let $\rho_i = \mathbb{P}(A_i = 1)$. We then replace numbers of students in capacity constraints with expected numbers. Specifically, constraint (1d) in the assignment stage becomes:

$$\mathbb{E}_{\boldsymbol{A}}\left[\sum_{i\in\mathcal{I}_p} A_i x_{is}\right] = \sum_{i\in\mathcal{I}_p} \rho_i x_{is} \leq K_p y_{ps} \forall p \in \mathcal{P}.$$

Similarly, the bus capacity constraint (4b) in the routing problem becomes

$$\mathbb{E}_{\boldsymbol{A}}\left[\sum_{h\in\mathcal{H}_r}\sum_{i\in\mathcal{I}_h} A_i\right] = \sum_{h\in\mathcal{H}_r}\sum_{i\in\mathcal{I}_h} \rho_i \leq Q,$$

In practice, students may request transportation service, but may choose to travel to the school using their own means, so we could refine the model to include a second transportation usage probability $\rho'_i < \rho_i$. This is a conservative approach: even though we only reserve a small amount of space on a bus for a student with high no-show likelihood, we still plan for a bus to travel to and from this student's location, which is potentially wasteful. The no-show problem has attracted more interest in recent years, with Caceres et al. (2017) proposing a more flexible approach to manage demand stochasticity. However, for special education students, districts may not have much leeway, as they are often legally required to provide transportation whether or not the student chooses to use it. Resolving the tension between system efficiency and legal requirements is a potentially interesting policy problem beyond the scope of this paper.

We note that our probability weighting approach is not robust. We employ it because it closely matches Boston Public Schools' regular process. In practice, uncertainty is handled in two ways: first, the attendance probabilities tend to be overestimated, so as to plan for a larger number of students than the expected value. Second, the classroom capacities K_c

can be adjusted in an ad hoc way on days when student attendance is much higher than expected. Incorporating a data-driven attendance model would be an interesting extension of this work.

Reducing site occupancy. Another particularity of the ESY problem is that it occurs over the summer. As a result, there are more restrictions on building usage, as facilities staff perform much-needed renovations on schol buildings. The ESY program handles these restrictions by penalizing the use of classrooms that would bring building usage over 50% utilization in terms of a number of classrooms.

We can model this effect in formulation (1) by separating each school's classrooms into "regular-use" and "extended-use" groups of equal size, adjusting the meaning of the decision variables y_{ps} to indicate the number of regular-use classrooms opened for program p at school s, and introducing additional variables y'_{ps} to indicate the number of extended-use classrooms opened for program p at school s. We then replace constraint (1d) with

$$\sum_{p\in\mathcal{P}} y_{ps} \leq \left\lceil \frac{Y_s}{2} \right\rceil \qquad \forall s \in \mathcal{S}, \tag{17a}$$

$$\sum_{p\in\mathcal{P}} y_{ps} + y'_{ps} \leq Y_s \qquad \forall s \in \mathcal{S}, \tag{17b}$$

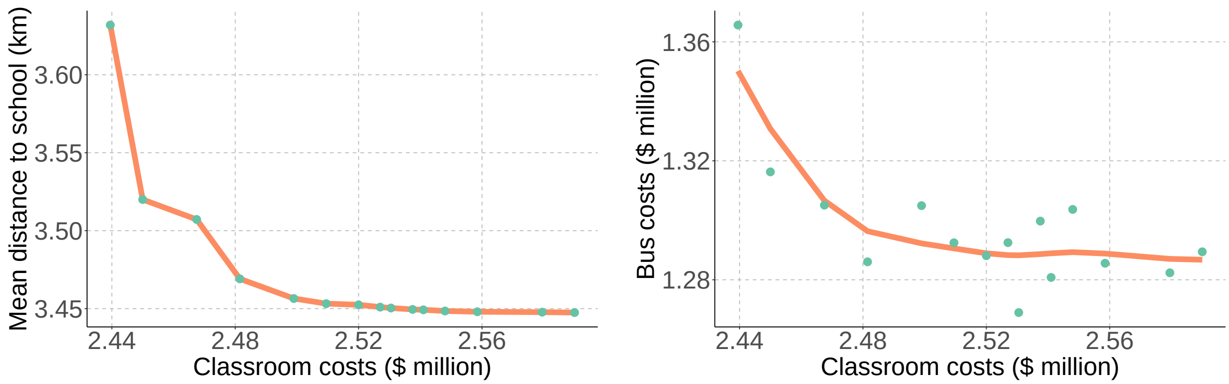
and add the term $\sum_{s\in\mathcal{S}}\sum_{p\in\mathcal{P}_s}\gamma' y'_{ps}$ to the objective (1a), where the parameter $\gamma' > \gamma$ quantifies the cost of using extended-use classrooms. Alternatively, we can use a goal-programming approach and restrict the number of extended-use classrooms below a certain threshold, e.g. $\sum_{s\in\mathcal{S}}\sum_{p\in\mathcal{P}_s} y'_{ps} \leq Y_{\text{extended}}$.

Cohorts. We also incorporate a cohort constraint, which keeps certain groups of students (e.g. students in certain programs attending the same school during the school year) together when assigning them to ESY sites. The objective of this constraint is to minimize changes in the student experience.

5.3. The edge of optimization

Our first goal in this case study is to quantify the benefit of using an optimization-based solution in operationalizing the ESY program at Boston Public Schools. To this end, we solve the assignment step (goal programming version) for a range of values of the student-to-school distance threshold D, exploring the Pareto frontier between distance to school and

Figure 5 Study of the core assignment tradeoff between transportation and classroom costs.



(a) Pareto front between distance and classroom cost

(b) Relation between routing and classroom costs

Note. As classroom costs decrease, distance to school increases, slowly at first, then sharply. Correspondingly, an increase in distance leads to an increase in total routing cost. In both cases, the points represent individual solutions, with the line indicating a cubic smoothing spline through the points.

total classroom cost. Because the problem is discrete and there are often many assignments with the same total number of classrooms (and because the optimality gaps relative to the best solver lower bound we obtain are typically in the 1-2% range at the 10-minute limit), not every solution produced by varying D is Pareto-optimal. We eliminate dominated solutions as a postprocessing step, and obtain the Pareto frontier in Figure 5a.

Surprisingly, it is possible to substantially reduce the number of classrooms in use with little to no impact on the average student's distance to school. The true metric of interest, however, is not distance, but total transportation costs incurred by the district. Therefore, for each generated assignment, we solve the school bus routing and start time selection problem using the optimization pipeline described in Section 2. We generate routes using 400 iterations of 2-OPT, for two distinct values of the routing tradeoff parameter λ ($5 \cdot 10^4$, 10^6) in both the morning and afternoon. We then solve the scheduling and start time selection problem with a 10-minute time limit, typically attaining an optimality gap of 1% for the subproblem. The resulting tradeoff curve between routing and classroom costs is shown in Figure 5b.

Figure 5 suggests that student-to-school distance is a reasonable proxy for routing costs at the assignment step (further corroborated by the correlation of 0.91 between total routing cost and student-to-school distance). However, we note that solutions with very similar student-to-school distance can vary in the resulting routing costs. This variation can be explained both by the imperfection of distance as a proxy for routing cost, and by the imperfection of our routing optimization process, including both our use of a randomized local search heuristic to construct routes, suboptimality resulting from decomposition, and the $\sim 1\%$ optimality gap when scheduling buses.

As a comparison point for our optimization-based solutions, we can compute the classroom and routing costs for the actual student-to-school assignment used by the ESY 2019 program. We find that this assignment leads to classroom costs of \$2.69 million and an average student to school distance of 3.93 km, significantly less efficient than the optimization-based solutions. Using the same routing parameters, we estimate a total routing cost of \$1.42 million, for a total estimated cost of \$4.11 million. In contrast, the minimum-cost optimization-based solution from Figure 5 has a total cost of \$3.77 million, an 8% improvement over the ground-truth baseline. We note that in all solutions, we capped the number of extended-use classrooms at 45—the number of extended-use classrooms utilized in the ground-truth baseline.

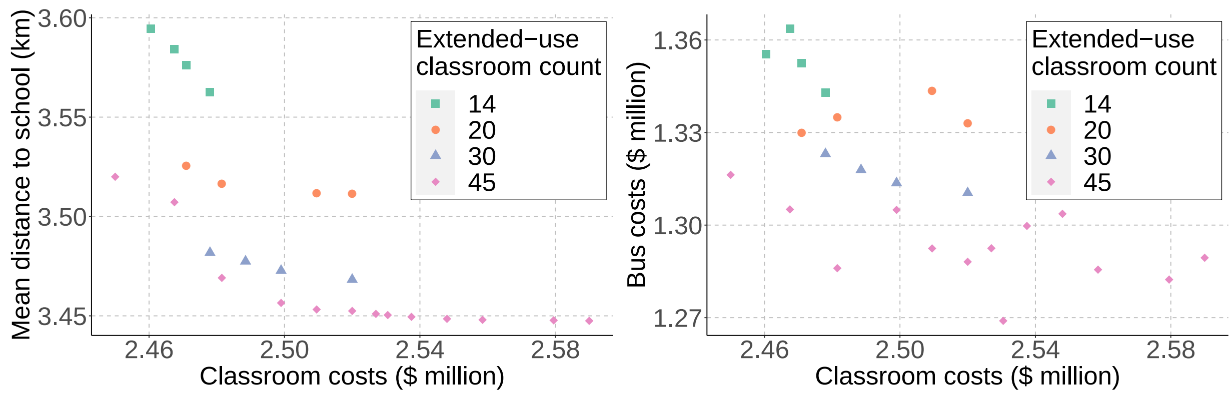
5.4. Policy tradeoffs

Beyond producing actionable, efficient solutions, our optimization model can be used to inform decision-makers of the tradeoffs of various policies.

Extended-use classrooms. For instance, it is of interest to understand the impact of restricting the usage of extended-use classrooms described in Section 5.2. Keeping site utilization below 50% is a soft constraint for the district, which can be re-evaluated based on the cost implications. Fixing the number of used extended-use classrooms to different values, we can use our model to understand how classroom and bus costs are affected. We again adopt a goal-programming approach for the assignment stage, minimizing classroom cost subject to constraints on the student-to-school distance and the number of extended-use classrooms available.

Results are presented in Figure 6, using the routing parameters defined in Section 5.1. As expected, restricting the number of extended-use classrooms affects overall efficiency: leading to a rise in student-to-school distance (and correspondingly, bus costs) for fixed classroom costs. Furthermore, fewer extended-use classrooms also imply a more constrained problem, with a narrower range of potential solutions. Indeed, the problem becomes infeasible if fewer than 10 extended-use classrooms are available.

Figure 6 Tradeoff between transportation and classroom costs for various restrictions on extended-use classrooms.

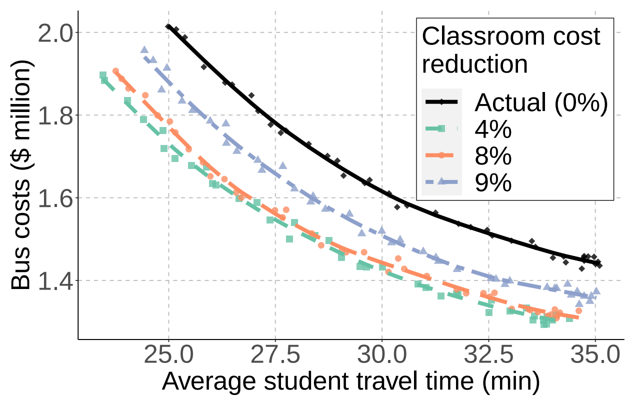


(a) Pareto frontier between student-to-school distance and total classroom cost

(b) Relationship between total routing cost and total classroom cost

Note. As fewer extended-use classrooms are available, the frontier between classroom and transportation costs shifts up and to the right. Each point represents an individual solution.

Figure 7 Tradeoff curve between average student travel time and total bus costs.



Note. Each curve corresponds to a different assignment with different classroom costs.

Student travel times. Another objective which presents an interesting tradeoff with cost is student travel time. In general, a solution in which students spend less time on the bus is preferable, particularly for special education students who may have medical needs which preclude long bus trips. We can vary the parameter λ at the route generation stage of our algorithm to create routes with different average travel times. We choose values ranging between $3 \cdot 10^3$ and $3 \cdot 10^4$.

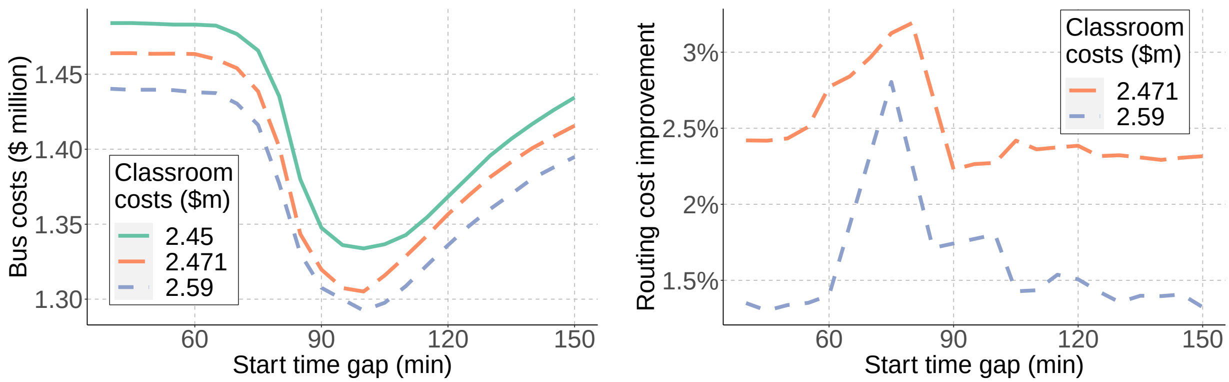
Figure 7 shows the tradeoff between transportation costs and average student travel time, as computed by the routing part of our optimization algorithm. As expected, as travel time decreases, total bus costs increase. In keeping with the results in Figure 5, as classroom costs decrease, total routing costs increase, and the efficiency frontier between travel time and bus costs shifts up. The routing solution computed by our algorithm for the actual 2019 assignment has a routing cost of \$1.42 million, for an average student travel time exceeding 35 minutes. Figure 7 thus suggests that our optimization approach could allow the district to simultaneously reduce classroom costs by 4% and average student travel time by up to 5 minutes.

Start time tiers. In 2019, Boston Public Schools decided that ESY schools would start at either 8:00 AM or 9:30 AM. The 90-minute separation between tiers contrasted with the 60-minute separation between tiers during the normal school year, and was designed to increase connectivity between bus routes for different schools, and thereby reduce transportation costs. A benefit of our analytics engine is that it can compute transportation cost estimates for different start time options and evaluate a good value for the gap between early and late schools.

We present this analysis in Figure 8. In particular, Figure 8a shows the effect of the gap, in minutes, between the earlier allowed start time (fixed at 8:00 AM) and the later allowed start time for each school, for different solutions with different classroom costs. In all cases, a small gap means routes are hard to connect between schools. As the gap increases, connecting routes becomes substantially easier and routing costs decrease. If the gap keeps increasing, however, buses must drive farther or remain idle longer on deadhead trips connecting routes, and transportation costs start increasing once more. Interestingly, in the summer of 2018, the district had a 120-minute gap between ESY start times. Our results suggest that both 90 minutes and 120 minutes are approximately equivalent in terms of routing cost, though this result is obviously heavily dependent on our travel time assumptions.

One technique introduced in Section 2 to keep costs down is the introduction of multiple sets of routes (for example, with varying emphasis on the number of routes versus the average student travel time) for each school, and allowing the optimal set of routes to be selected at the scheduling step. In Figure 8b, we show the improvement resulting from using four sets of routes per school over using just one set of routes per school (only one route set per school is used in Figure 8a). We see there is a small advantage from using multiple route sets per schools, particularly when the gap between start times is between

Figure 8 Effect of start time options on transportation cost.



(a) Effect of gap between start times on bus costs, for different computed school assignments.

(b) Cost improvement from using multiple routing scenarios per school for different start time gaps.

Note. For all three student-to-school assignments considered, the left panel shows there is a "sweet spot" for the gap between early and late start times—too short and routes are hard to connect, too long and buses spend too much time idle or on deadhead trips. The right panel shows there is a small edge from considering multiple sets of routes, or scenarios, for each school. This edge is bigger when classroom costs are lower (and students are consequently a bit further from school). We also notice that this edge is highest in the 60- to 90-minute window, just when we start reaping the highest rewards from increased route connectivity. The solid line is not depicted on the right panel because the scheduling problems were more difficult to solve with more route sets, leading to suboptimality.

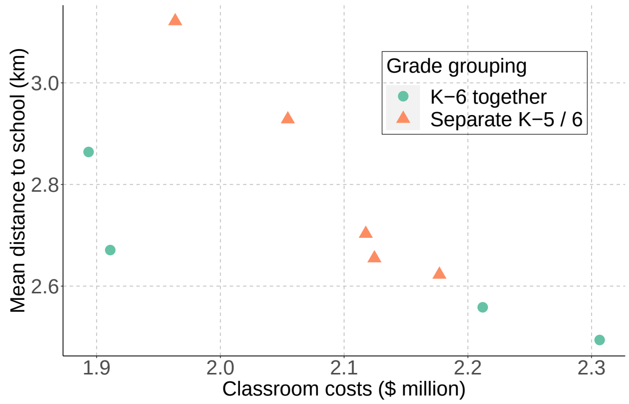
60 and 90 minutes, a zone where each additional potential route connection can potentially lead to a substantial cost reduction.

5.5. Implementation

Motivated by the case study in this section, Boston Public Schools decided to partially implement the methods described in this paper to assign students to schools for the Summer 2021 Extended School Year program. For implementation, we produced a simplified version of our optimization code, implemented in R, to be run by administrative staff at the district. Development was driven by three major concerns: portability (ensuring the code was simple enough to run with minimal software installations and prior knowledge), flexibility (ensuring the models allowed the district to experiment with many proposals), and tractability (producing a feasible solution in seconds using open-source tools).

In keeping with these priorities, we made the following simplifications: we formulated only the school assignment problem to minimize average distance to school, and allowed

Figure 9 Tradeoff curve between ESY 2021 scenarios.



Note. Each point represents a solution considered by the district. For clarity, we only show Pareto-optimal solutions according to the two plotted objectives, though many other solutions were considered using different site parameters.

the practitioners to specify site and capacity constraints to explore multiple scenarios using a goal programming approach, and allowed the computation of classroom costs after the fact. Practitioners were able to explore many scenarios, and take into account additional considerations such as idiosyncratic differences between functionally equivalent school sites. For example, the district could directly compare two different assignment policies, one which grouped 6th graders with K-5 students at elementary sites, and another which grouped 6th graders in separate middle school sites. We present a summary of Pareto-optimal solutions produced by the district in Figure 9.

Though this implementation falls short of using all the modeling tools outlined in this paper, we believe it is a good example of the practical benefits of a holistic optimization approach. Practitioners at Boston Public Schools responded positively to the flexibility this tool provides in assessing assignment options with many tradeoffs.

6. Conclusion

In this paper, we provide a sequence of optimization algorithms to address key problems in US public school operations. In particular, we develop a new network flow formulation for the bus scheduling and start time selection subproblem that can accommodate problem specifics such as a heterogeneous fleet and time-varying travel times. We apply our methods in a case study with Boston Public Schools, yielding potential cost savings of up to 8%. We further show that our methods allow policy makers to explore a variety of policy tradeoffs.

Acknowledgments

We are grateful to the Boston Public Schools team for their continued support, including Zachary Houston and Porsche Paulding from the Office of Special Education, and Shanda Williams, Emanuel Zankerkia, and Angela Zhang from the Office of Transportation. We also thank undergraduate research assistant Ashley Wang for leading the development of an interactive interface to display and visually inspect particular student assignments. Finally, we thank Julia Yan for helpful comments on early drafts, particularly regarding Figure 1.

References

Abdulkadiroglu A, Sönmez T (2003) School choice: A mechanism design approach. *American Economic Review* 93(3):729–747.

Araya F, Dell R, Donoso P, Marianov V, Martínez F, Weintraub A (2012) Optimizing location and size of rural schools in Chile. *International Transactions in Operational Research* 19(5):695–710.

Banerjee D, Smilowitz K (2019) Incorporating equity into the school bus scheduling problem. *Transportation Research Part E: Logistics and Transportation Review* 131(October):228–246.

Bertsimas D, Delarue A, Martin S (2019) Optimizing schools' start time and bus routes. *Proceedings of the National Academy of Sciences of the United States of America* 116(13):5943–5948.

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. *SIAM review* 59(1):65–98.

Boston Public Schools (2017) Transportation challenge data. `https://www.bostonpublicschools.org/transportationchallenge`, accessed 2022-02-15.

Boston Public Schools (2018) Fast facts. `https://www.bostonpublicschools.org/cms/lib/MA01906464/Centricity/Domain/187/BPS%20at%20a%20Glance%2019_final.pdf`, accessed 2021-03-19.

Braca J, Bramel J, Posner B, Simchi-Levi D (1997) A computerized approach to the New York City school bus routing problem. *IIE Transactions* 29(8):693–702.

Caceres H, Batta R, He Q (2017) School bus routing with stochastic demand and duration constraints. *Transportation Science* 51(4):1349–1364.

Caceres H, Batta R, He Q (2019) Special need students school bus routing: Consideration for mixed load and heterogeneous fleet. *Socio-Economic Planning Sciences* 65(December 2017):10–19.

Caro F, Shirabe T, Guignard M, Weintraub A (2004) School redistricting: Eembedding GIS tools with integer programming. *Journal of the Operational Research Society* 55(8):836–849.

Carrell SE, Maghakian T, West JE (2011) A's from Zzzz's? The causal effect of school start time on the academic achievement of adolescents. *American Economic Journal: Economic Policy* 3(3):62–81.

Chen X, Kong Y, Dang L, Hou Y, Ye X (2015) Exact and metaheuristic approaches for a bi-objective school bus scheduling problem. *PLoS ONE* 10(7):1–20.

Chu A, Keskinocak P, Villarreal MC (2020) Empowering Denver Public Schools to Optimize School Bus Operations. *INFORMS Journal on Applied Analytics* 50(5).

Cordeau JF, Laporte G, Savelsbergh MW, Vigo D (2007) Vehicle Routing. *Handbooks in Operations Research and Management Science*, volume 14, 367–428.

Delmelle EM, Thill JC, Peeters D, Thomas I (2014) A multi-period capacitated school location problem with modular equipment and closest assignment considerations. *Journal of Geographical Systems* 16(3):263–286.

Desrosiers J, Ferland JA, Rousseau JM, Lapalme G, Chapleau L (1986) TRANSCOL: a multi-period school bus routing and scheduling system. *TIMS Studies in the Management Sciences* 22:47–71.

Dunning I, Huchette J, Lubin M (2017) Jump: A modeling language for mathematical optimization. *SIAM Review* 59(2):295–320.

Ellegood WA, Campbell JF, North J (2015) Continuous approximation models for mixed load school bus routing. *Transportation Research Part B: Methodological* 77:182–198.

Ellegood WA, Solomon S, North J, Campbell JF (2020) School bus routing problem: Contemporary trends and research directions. *Omega (United Kingdom)* 95(xxxx).

Franklin AD, Koenigsberg E (1973) Computed School Assignments in a Large District. *Operations Research* 21(2):401–659.

Fügenschuh A (2009) Solving a school bus scheduling problem with integer programming. *European Journal of Operational Research* 193(3):867–884.

Gurobi Optimization, Inc (2016) Gurobi optimizer reference manual.

Kamali B, Mason SJ, Pohl EA (2013) An analysis of special needs student busing. *Journal of Public Transportation* 16(1):21–45.

Levin MW, Boyles SD (2016) Practice summary: Improving bus routing for KIPP charter schools. *Interfaces* 46(2):196–199.

Liggett RS (1973) The Application of an Implicit Enumeration Algorithm to the School Desegregation Problem. *Management Science* 20(2):159–168.

Mandujano P, Giesen R, Ferrer JC (2012) Model for optimization of locations of schools and student transportation in rural areas. *Transportation Research Record* (2283):74–80.

NCES (2021) Transportation fast facts. https://nces.ed.gov/fastfacts/display.asp?id=67.

Park J, Kim BI (2010) The school bus routing problem: A review. *European Journal of Operational Research* 202(2):311–319.

Park J, Tae H, Kim BI (2012) A post-improvement procedure for the mixed load school bus routing problem. *European Journal of Operational Research* 217(1):204–213.

Russell A, Morrel B (1986) Routing Special-Education School Buses. *Interfaces* 1986(October):56–64.

Schittekat P, Kinable J, Sörensen K, Sevaux M, Spieksma F, Springael J (2013) A metaheuristic for the school bus routing problem with bus stop selection. *European Journal of Operational Research* 229(2):518–528.

Shafahi A, Wang Z, Haghani A (2018) SpeedRoute: Fast, efficient solutions for school bus routing problems. *Transportation Research Part B: Methodological* 117:473–493.

Shi P (2015) Guiding school-choice reform through novel applications of operations research. *Interfaces* 45(2):117–132.

Swersey AJ, Ballard W (1984) Scheduling School Buses. *Management Science* 30(7):844–853.

Teixeira JC, Antunes AP (2008) A hierarchical location model for public facility planning. *European Journal of Operational Research* 185(1):92–104.

Zeng L, Chopra S, Smilowitz K (2022) A bounded formulation for the school bus scheduling problem. *Transportation Science* .

Appendix A: Pseudocode of single-school routing heuristic

See Algorithm 1 below.

Appendix B: Modeling costs in the joint assignment-routing step

In Section 2.5, we presented a network flow-based heuristic to jointly optimize student-to-school assignment and school bus routing and scheduling. The presented formulation is highly flexible and can model diverse objectives. We describe examples in this appendix, focusing on the morning case, with the understanding that the afternoon part of the graph is a mirror image of the morning.

Much as in the school bus routing and start time selection formulation, we can minimize the total number of buses by setting $C'_{v,v'} = 1$ if $v = v_{\mathrm{AM}}$, and 0 otherwise. Additionally, we can minimize total travel time by setting:

$$C'_{v,v'} = \begin{cases} t^{\text{drive}}_{d,h_1(f')} + t^{\text{service}}_{f'} + t^{\text{school}}_{s'}, & \text{if } v = v_{\mathrm{AM}}, v' \in \mathcal{V}^{\text{cplt}}_{s'}, \\ t^{\text{drive}}_{d,h_1(f')} + t^{\text{service}}_{f'}, & \text{if } v = v_{\mathrm{AM}}, v' \in \mathcal{V}^{\text{icplt}}, \\ t^{\text{drive}}_{s,d}, & \text{if } v' = v_{\mathrm{PM}}, \\ t_{s'} - t_s, & \text{if } v := v_s \in \mathcal{V}_{\text{schools}},\ v' \in \mathcal{V}^{\text{cplt}}_{s'}, \\ t' - t_s, & \text{if } v := v_s \in \mathcal{V}_{\text{schools}},\ v' = (t', f', \mathrm{AM}, 0) \in \mathcal{V}^{\text{icplt}} \\ t^{\text{drive}}_{h_{|f|}(f),h_1(f')} + t^{\text{service}}_{f'} + t^{\text{school}}_{s'}, & \text{if } v := (t, f, \mathrm{AM}, 1) \in \mathcal{V}^{icplt}, v' := (f', \mathrm{AM}) \in \mathcal{V}^{\text{cplt}}_{s'}, \\ t^{\text{drive}}_{h_{|f|}(f),s'} + t^{\text{school}}_{s'}, & \text{if } v := (t, f, \mathrm{AM}, 1) \in \mathcal{V}^{icplt}, v' := v_{s'} \in \mathcal{V}_{\text{schools}}, \\ 0, & \text{otherwise.} \end{cases}$$

The cost function can be defined analogously for the afternoon part of the graph, and can be adapted to minimize travel distance instead of travel time.

Appendix C: Proofs

Proof of Proposition 1 We consider formulation (14) in the two-tier case. For each school s, let the binary variable g_s take the value 1 if school s starts at time t_1, and 0 if school s starts at time

Algorithm 1 k-OPT local search heuristic for the school bus routing problem.
Input: a set of routes $\mathcal{R}_s$, the number of edges to delete k, and the number of local search iterations N. The output is an updated set of routes $\mathcal{R}_s$ with lower total cost. Our implementation of k-OPT also allows reversing the direction of route fragments that do not contain a school at the recombination step, but we omit this detail for clarity.

function k-OPT($\mathcal{R}_s$, k, N)
  for $i = 1$ to N **do**
    $c^{\text{opt}} = \sum_{r \in \mathcal{R}_s} c_r$ ▷ Initialize optimal cost
    $\mathcal{F} \leftarrow$ FragmentRoutes($\mathcal{R}_s$, k) ▷ A route fragment is a tuple $(\boldsymbol{h}, \psi)$, where $\boldsymbol{h}$ is a vector of stops, and ψ is a boolean indicating whether the fragment ends at school
    for $\hat{\mathcal{R}}_s$ in PossibleRoutes($\mathcal{F}$) **do**
      if $\sum_{r \in \hat{\mathcal{R}}_s} c_r \leq c^{\text{opt}}$ **then**
        $\mathcal{R}_s \leftarrow \hat{\mathcal{R}}_s$; $c^{\text{opt}} \leftarrow \sum_{r \in \hat{\mathcal{R}}_s} c_r$
  return $\mathcal{R}_s$
function FragmentRoutes($\mathcal{R}_s$, k)
  $\mathcal{F} \leftarrow \emptyset$
  Sample $a_1, \ldots, a_k$ uniformly from $\left\{1, 2, \ldots, \sum_{r \in \mathcal{R}_s} |r|\right\}$ ▷ Route breakpoints
  $\bar{a} \leftarrow 0$ ▷ Counter
  for $r \in \mathcal{R}_s$ **do**
    $i_0 \leftarrow 0$
    for $i = 1$ to $|r|$ **do**
      $\bar{a} \leftarrow \bar{a} + 1$
      if $\bar{a} \in \{a_1, \ldots, a_k\}$ **then**
        $\mathcal{F} \leftarrow \mathcal{F} \cup \{(\boldsymbol{h} := (h_{i_0+1}(r), \ldots, h_i(r)), \psi := \texttt{false})\}$
        $i_0 \leftarrow i$
    $\mathcal{F} \leftarrow \mathcal{F} \cup \{(\boldsymbol{h} := (h_{i_0+1}(r), \ldots, h_{|r|}(r)), \psi := \texttt{true})\}$ ▷ Remainder of route
  return $\mathcal{F}$
function NextFragments($\mathcal{F}$, $f := (\boldsymbol{h}, \psi)$)
  if $\psi = \texttt{true}$ **then**
    return $\{(\emptyset, \texttt{true})\}$ ▷ Empty fragment only contains school
  else
    return $\mathcal{F} \backslash f \cup \{(\emptyset, \texttt{true})\}$ ▷ All other route fragments, including the empty one
function PossibleRoutes($\mathcal{F}$)
  $\mathcal{F} := \{f_1, \ldots, f_n\}$ ▷ Number the fragments for clarity ($n = |\mathcal{F}|$)
  $\mathcal{P}_s(\mathcal{R}) \leftarrow \emptyset$ ▷ Initialize set of route sets
  for $\{f_{j_1}, \ldots, f_{j_n}\}$ in NextFragments($\mathcal{F}, f_1$) $\times \ldots \times$ NextFragments($\mathcal{F}, f_n$) **do**
    Consider the directed graph $G = (V, E)$ with $V = \mathcal{F} \cup \{(\emptyset, \texttt{true})\}$ and $E = \{(f_i, f_{j_i})\}_{i=1}^n$
    if G is acyclic and the maximum in-degree over $\mathcal{F} = V \backslash \{(\emptyset, \texttt{true})\}$ is 1 **then**
      $\hat{\mathcal{R}}_s \leftarrow \emptyset$ ▷ Initialize candidate route set
      for every path $(f_{\ell_0}, \ldots, f_{\ell_p})$ in G where f_{ℓ_0} has in-degree 0 and $f_{\ell_p} = (\emptyset, \texttt{true})$ **do**
        Construct a route r by concatenating the stop lists $\boldsymbol{h}_{\ell_0}, \ldots, \boldsymbol{h}_{\ell_p}$ in order
        $\hat{\mathcal{R}}_s \leftarrow \hat{\mathcal{R}}_s \cup \{r\}$
      if every route $r \in \hat{\mathcal{R}}_s$ is feasible (capacity and maximum riding time) **then**
        $\mathcal{P}_s(\mathcal{R}) \leftarrow \mathcal{P}_s(\mathcal{R}) \cup \hat{\mathcal{R}}_s$
  return $\mathcal{P}_s(\mathcal{R})$

t_2. Then we can formulate $(P1)$ as

$$\begin{aligned}
\min_{\boldsymbol{g}\in\{0,1\}^{|\mathcal{S}|}} \quad \min_{\substack{\boldsymbol{\tau}\in\mathbb{R}^{|\mathcal{R}|}\\ \boldsymbol{w}\in\{0,1\}^{|\mathcal{R}|^2}}} \quad & z && && \text{(P1a)}\\
\text{s.t.} \quad & \sum_{r'} w_{r',r} \le 1 && \forall r\in\mathcal{R} && \text{(P1b)}\\
& \sum_{r'} w_{r,r'} \le 1 && \forall r\in\mathcal{R} && \text{(P1c)}\\
& \tau_{r'} \ge \tau_r + T(r,r') - M(1-w_{r,r'}) && \forall r,r'\in\mathcal{R} && \text{(P1d)}\\
& t_1 g_s + t_2(1-g_s) - \Delta t_s \le \tau_r \le t_1 g_s + t_2(1-g_s) && \forall s\in\mathcal{S}, r\in\mathcal{R}_s, && \text{(P1e)}
\end{aligned}$$

where we recall that $t_s = t_1 g_s + t_2(1-g_s)$, and $(P2)$ as

$$\begin{aligned}
\min_{\boldsymbol{g}\in\{0,1\}^{|\mathcal{S}|}} \quad \min_{\boldsymbol{w}\in\{0,1\}^{|\mathcal{R}|^2}} \quad & \hat{z} && && \text{(P2a)}\\
\text{s.t.} \quad & \sum_{r'} w_{r',r} \le 1 && \forall r\in\mathcal{R} && \text{(P2b)}\\
& \sum_{r'} w_{r,r'} \le 1 && \forall r\in\mathcal{R} && \text{(P2c)}\\
& (t_1-\Delta t_{s'})g_{s'} + t_2(1-g_{s'}) \ge && && \\
& (t_1-\Delta t_s)g_s + t_2(1-g_s) + T(r,r') - M(1-w_{r,r'}) && \forall r,r'\in\mathcal{R}. && \text{(P2d)}
\end{aligned}$$

As a result of the two-tier property, if $w_{r,r'}=1$, route r must be associated with an early school, and route r' with a late school, i.e., $g_s=1$ and $g_{s'}=0$. Thus constraint (P2d) simplifies to

$$t_2 \ge t_1 - \Delta t_s + T(r,r') - M(1-w_{r,r'}) \qquad \forall r,r'\in\mathcal{R}. \tag{P2e}$$

Let $(\boldsymbol{g}^*,\boldsymbol{w}^*)$ (equivalently $(\boldsymbol{t}^*,\boldsymbol{w}^*)$) designate an optimal solution of $(P2)$. Consider the optimization problem $(P3)$, obtained by fixing $\boldsymbol{g}$ to $\boldsymbol{g}^*$ and $\boldsymbol{w}^*$ to $\boldsymbol{w}$ in $(P1)$:

$$\begin{aligned}
\min_{\boldsymbol{\tau}\in\mathbb{R}^{|\mathcal{R}|}} \quad & z && && \text{(P3a)}\\
\text{s.t.} \quad & \tau_{r'} \ge \tau_r + T(r,r') && \text{if } w^*_{r,r'}=1 && \text{(P3b)}\\
& t_1 g^*_s + t_2(1-g^*_s) - \Delta t_s \le \tau_r \le t_1 g^*_s + t_2(1-g^*_s) && \forall s\in\mathcal{S}, r\in\mathcal{R}_s. && \text{(P3c)}
\end{aligned}$$

We first show $(P3)$ has at least one feasible solution, obtained by setting

$$\tau_r = \begin{cases} t_1 - \Delta t_s, & \forall r\in\mathcal{R}_s, g^*_s = 1;\\ t_2, & \forall r\in\mathcal{R}_s, g^*_s = 0.\end{cases}$$

Clearly this choice of $\boldsymbol{\tau}$ verifies constraint (P3c). Moreover, constraint (P2e) for each r,r' such that $w^*_{r,r'}=1$ implies $t_2 \ge t_1 - \Delta t_s + T(r,r')$, guaranteeing that our choice of $\boldsymbol{\tau}$ verifies constraint (P3b).

Since $(P3)$ is trivially bounded, let us now denote by $\boldsymbol{\tau}^*$ the optimal solution of $(P3)$. We claim that $(\boldsymbol{g}^*, \boldsymbol{w}^*, \boldsymbol{\tau}^*)$, with associated objective z^*, is optimal for $(P1)$. Consider a feasible solution $(\boldsymbol{g}^\dagger, \boldsymbol{w}^\dagger, \boldsymbol{\tau}^\dagger)$ with objective $z^\dagger < z^*$. We first show that $(\boldsymbol{g}^\dagger, \boldsymbol{w}^\dagger)$ is feasible for $(P2)$. Constraints (P2b) and (P2c) are trivially satisfied. Further, combining constraints (P1d) and (P1e) with the two-tier property ($w^\dagger_{r,r'} = 1 \Rightarrow g_s = 1, g_{s'} = 0$), we obtain $t_2 \geq t_1 - \Delta t_s + T(r,r') - M(1 - w^\dagger_{r,r'})$, verifying feasibility of $(\boldsymbol{g}^\dagger, \boldsymbol{w}^\dagger)$ in $(P2)$.

Applying the two-tier property one more time, we know that if $w^\dagger_{r,r'} = 1$, then $\tau^\dagger_{r'} \geq t_2 - \Delta t_{s'}$ and $\tau^\dagger_r \leq t_1$, i.e. $\tau^\dagger_{r'} - \tau^\dagger_r \geq t_2 - \Delta t_{s'} - t_1$. From feasibility, we also know that $w^\dagger_{r,r'} = 1$ implies $\tau^\dagger_{r'} - \tau^\dagger \geq T(r,r')$. Thus, $z^\dagger_3 \geq \hat{z}^\dagger_3$, implying $z^\dagger \geq \hat{z}^\dagger$.

As a final step, we claim that for all pairs of routes $r, r' \in \mathcal{R}$, $w^*_{r,r'} = 1$ implies $\tau^*_{r'} - \tau^*_r = \max(t_2 - \Delta t_{s'} - t_1, T(r,r'))$. Assume there exists a pair of routes r, r' such that $\tau^*_{r'} - \tau^*_r > \max(t_2 - \Delta t_{s'} - t_1, T(r,r'))$. Consider first the case where $t_2 - \Delta t_{s'} - t_1 > T(r,r')$. Then either $\tau^*_{r'} > t_2 - \Delta t_{s'}$, or $\tau^*_r < t_1$. Let us assume the latter without loss of generality, and define $\epsilon > 0$ such that $\tau^*_r + \epsilon \leq t_1$. By construction, we see that either τ^*_r or $\tau^*_{r'}$ appear in exactly three constraints in $(P3)$: one of type (P3b), and two of type (P3c). It is thus easy to check that a new solution $\boldsymbol{\tau}^{**}$ constructed from $\boldsymbol{\tau}^*$ by incrementing only τ^*_r by ϵ is feasible in $(P3)$, and with a lower objective than $\boldsymbol{\tau}^*$, a contradiction. A similar argument can be made if $t_2 - \Delta t_{s'} - t_1 \leq T(r,r')$.

Since $\tau^*_{r'} - \tau^*_r = \max(t_2 - \Delta t_{s'} - t_1, T(r,r'))$ for all pairs of routes r, r' with $w^*_{r,r'} = 1$, we conclude that $z^*_3 = \hat{z}^*_3$, thus $z^* = \hat{z}^*$. By optimality of $(P2)$, we therefore have $z^* = \hat{z}^* \leq \hat{z}^\dagger \leq z^\dagger$, contradicting our original assumption that $z^\dagger < z^*$, and completing the proof. □

Proof of Proposition 2 We take a constructive approach, showing how from an optimal solution z^* we can build a feasible solution z' to (P') with the relevant guarantees. Consider a pair of routes r, r' such that r serves a school in the first tier, r' serves a school in the second tier, and $w^*_{r,r'} = 1$. Then $t_1 - \Delta t + T(r,r') \leq \tau^*_r + T(r,r') \leq \tau_{r'} \leq t_2$, so it is feasible to set $w'_{r,r'} = 1$. Second, consider any route r'' for a third-tier school such that $\sum_r w^*_{r',r''} = 1$ (summing over optimal second-tier routes r' only). Obviously it is feasible to set $w'_{r',r''} = 0$ for all second-tier routes. Finally, consider a route r serving a first tier school such that $\sum_{r' \in \mathcal{R}} w_{r,r'} = 0$, and consider a route r' for a third-tier school such that $\sum_r w^*_{r,r'} = 1$ and $w'_{r,r'} = 0$ for all second-tier routes r. By property (5) of a three-tier system, it is feasible to set $w'_{r,r'} = 1$. Summing up these changes, we see that $z' = z^* + N^*_{2,3} - N^*_1$. Furthermore, by proposition 1 of Zeng et al. (2022), we have that $N^*_{2,3} \leq z^*$ which yields the second inequality.